Reply to the comments by the anonymous Referee #2

The authors gratefully acknowledge the reviewers' effort in improving the quality of the manuscript. Below, all the major, general and specific reviewers' comments are addressed (in bold). The remaining minor comments have been all of them fixed in the new version of the manuscript.

1 General comments:

The paper describes an assessment of the performance of a miniMPL and two ceilometers using collocated Raman lidar measurements during the INTERACT II field campaign. This reviewer greatly appreciates the paper's clear organization and good writing, which made it relatively easy to read and review. Another strength is that the discussion is realistic and straightforward about the findings although the findings are not all positive. I think it is an important part of good research to straightforwardly describe both positive and negative findings and I commend the authors for not over- reaching in their motivational discussion or "spinning" their conclusions to sound more positive than what's justified.

On the negative side, some of the graphs don't seem well designed to answer the questions being asked, and consequently some of the interpretations of the results appear somewhat off-base. There is at least one remaining error in data labeling, and some details need to be explained better. These should be addressed in revision.

The introduction gives a bullet list of the objectives of the campaign but not the objectives of the paper. It would be good to be explicit whether the intention is to address all of these objectives in the paper or only some of them.

Then, please be sure to address those specific objectives in the paper's conclusions also.

Although I was somewhat uncertain about the intended objectives for this paper, some of the objectives listed in the bullet list are addressed rather superficially or even incorrectly. For example, "assess the signal to noise ratio and dynamic range". I don't see analysis specifically addressing these, although some of the analysis of the figures includes some confusion about SNR (see specific comments below). Similarly, "assess the ceilometers' calibration stability and accuracy". There is some discussion about the relatively poor accuracy and about the stability of the lidar instrument itself, but there appears to be confusion in this paper about how to assess stability of the calibration.

2 Specific comments:

44. I'm not sure this link is an adequate reference. It links to a pdf of documentation of a piece of software for getting data but doesn't say where to get it. Maybe instead use the link that allows users to get the data– https://www.dwd.de/EN/research/projects/ceilomap/ceilomap_node.html

The authors have been not able to find a reference, therefore they simply modified the link according to the reviewer's suggestion.

47. The authors criticize the lack of global lidar coverage and lack of homogeneity within current lidar networks as if this is a motivator of the current work, but it isn't clear how the current work advances the goal of homogeneous lidar coverage. I do understand (from the next paragraph in the manuscript) and agree that vetting cheaper lidars might enable more global coverage, but it doesn't follow that such a network with a wide range of lidar capabilities will be more homogeneous than existing networks. Better to delete the sentence starting "Even when federated" or put in more discussion making the motivation and link to this work more clear.

The sentence indicated by the reviewer #1 states:" Even when federated networks have been set-up by international stakeholders (e.g. GALION – GAW Lidar Observation Network), the different practices adopted within each of the federated networks (e.g. EARLINET, MPLNET, ADNET, LALINET) significantly affect the homogeneity of the collected measurements; at present only one example of a coordinated monitoring of a global scale event (Nabro volcanic eruption) has been provided in literature (Sawamura et al., 2011)."
The authors used this sentence to acknowledge that GALION could improve our understanding of aerosol at the global scale but not without a harmonization effort to spend across the federated network. Clearly this is the same for commercial instruments, though a smaller number of "degrees of freedom" should be involved to achieve a global harmonization of the provided data. Actually there 7-8 commercial instruments (including the old models) which covers the majority of stations available worldwide and equipped with an automatic lidar or a ceilometer.
In contrast, a network like GALION, though based on a smaller number of station should deal with many home-made and commercial instruments together, which surely led to an increased level heterogeneity within the network. EARLINET and MPLnet already spent a lot of effort to increase the data harmonization level at their stations but, for example, they have never tried to perform a joint harmonization of the respective products.

Nevertheless, the authors smoothed the tone of the mentioned paragraph as follows: "*Federated networks set-up by international stakeholders (e.g. GALION – GAW Lidar Observation Network) are slowly evolving towards the harmonization of the different practices adopted within each of the federated networks (e.g. EARLINET, MPLNET, ADNET, LALINET), and, therefore, towards the homogeneity of the respective measurements and products; at present only one example of a coordinated monitoring of a global scale event (Nabro volcanic eruption) has been provided in literature (Sawamura et al., 2011).*
***It is useful for the scientific community to understand to what extent automatic lidars and ceilometers (ALCs) are able to provide an estimation of the aerosol geometric and optical properties and fill in the geographical gaps of the existing advanced lidar networks, like EARLINET,…..*".**
The second paragraph is indeed saying that using commercial lidars/ceilometers, some effort must be spent to learn how these systems can fill in the observation gaps, also in terms of providing a support to the global lidar data harmonization.

58. "have been already investigated". That's fine, but in the next sentence or soon thereafter, you need to explain what the new contribution of this paper is.

At lines 88-89, the authors added the following paragraph: "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 networks for the aerosol observation and the monitoring of pollution. Behind this motivation, the INTERACT campaign was arranged and took place at CIAO……".

62. “retrieval . . . can be performed using the molecular backscattering profile”. Please be more specific. You mean the calibration can be performed using the molecular backscattering profile in a region where there is negligible aerosol. Without these additions, the phrase “the retrieval can be performed using the molecular backscatter profile” sounds like the Raman or HSRL retrieval.

The text has been modified accordingly.

215. Each lidar instrument has its own version of the quantity being considered, all with different names: attenuated backscatter, range-corrected signal (RCS), and normalized relative backscatter (NRB). It’s a little confusing, but with some effort I see why you made these choices. It would be helpful to have a paragraph (earlier than this) where all three quantities are described in one place and the reasons for using different quantities for each instrument are provided.

Given that the difference between the RCS and the NRB is in a constant term, and given that a normalization is operated to compare the MiniMPL and CIAO lidars, the authors will make use only the values of the RCS and of the attenuated backscatter, which have been defined in the text of the new manuscript version.

275. Please explain the normalization further. Is the MiniMPL normalized to match PEARL in the normalization region on a profile-by-profile basis for every profile?

Each single MiniMPL profile is normalized to match CIAO lidars in the normalization region on a profile-by-profile basis for every profile. This has been further clarified in the text of the new manuscript version.

322. If you mention after-pulse correction as a possibility, I think it needs to be supported. Otherwise it just sounds random and speculative.

Indeed, this is a speculative discussion given that the the MiniMPL data processing has been performed by the manufacturer. Tests to assess the effect of a wrong after pulses correction have been performed by the authors though we agreed that the manufacturer shall investigate this hypothesis.

325. Are the 12 cases all the measurements available from the whole six months deployment period, or have these been selected from a larger dataset? (Were the lidars operating continuously?)

The MiniMPL, the CL51 and the CS135 were operated on a continuous basis from the respective deployment dates at CIAO, except for a few weeks when the MiniMPL had thermalizing problems and for a few days when the CS135 had communication issues. The CIAO EARLINET lidars, PEARL and MUSA, have been operated according to the EARLINET measurements schedule (3 night time measurements per week only with clear sky). This is the main reason why the number of cases is restricted though the automatic lidars and ceilometers were operated on a continuous basis.

325. RCS or normalized relative backscatter? I thought that RCS meant non- normalized signals, so they could not be compared between two instruments? If there’s no useful distinction in the terminology, it would be better to just use one name instead of three.

To the authors knowledge, the term normalized-relative-backscatter (NRB) signal is in use within MPLnet (Micro-Pulse Nidar NETwork) which is defined by the equation:

$$P_{\mathrm{NRB}}(r) = C[\beta_M(r) + \beta_P(r)]T_M^2(r)T_P^2(r).$$

which indeed is equivalent to the RCS. According to the information provided by the manufacturer, it appears to the authors that there was a difference between RCS and NRB in the definition of the constant "C" of the lidar equation. However, as already mentioned above, to avoid misunderstandings, given that MiniMPL vertical profiles were normalized with respect to CIAO lidars profiles, in the new version of the manuscript the authors make use only of the values of RCS.

338. "The good stability of the MiniMPL calibration ... is shown by the small variability (10%) of differences in the normalization region". I have a major problem with this statement. This one must be addressed. Aren't the profiles for all 12 cases normalized to the Raman lidar in the normalization region? So, for each profile, the normalization constant is divided out and each profile is independently set to have zero average difference in the normalization region. That means that to assess the profile-to-profile variability in the normalization constant, you'd have to explicitly look at the 12 normalization constants. That information is not present in the data shown in Figure 5. Variability in the normalization region in Fig 5 is only representative of high-frequency noise within the normalization region, so it informs you about the precision, but not about the stability over time.

The authors agree with the reviewer. The numbers reported in the new version of the manuscript have been calculated using the real variability of the values for the normalization constant. The authors also remarked that these values have been calculated on small datasets, due to the number of available cases. The text of the new version of the manuscript has been refined accordingly.

Figure 7. The differences in the scatter plots are very hard to make out given the data being compared are in two different plots. A figure showing both in the same figure would be better. For example, consider making a scatter plot of CL135 vs. MUSA/PEARL attenuated backscatter directly, and color code by extinction or stratify different ranges of extinction into multiple sub-plots. Accompanying them with another set color coded or stratified by altitude would also be helpful, I think, given that the interpretations of this figure in the text are related to specific altitude regions (the overlap region and the free troposphere). (Same comment for Figure 13.)

Figure 7 and 13 have been modified according to the reviewer's suggestions.

348. You say that the choice of aerosol extinction for the y-axis of Figure 7 (and 13) is to reveal differences in sensitivity to different aerosol types. In fact, you have not mentioned different aerosol types in your interpretation at all. Indeed, this task would be quite difficult with the information given in the figures since the relationship between attenuated backscatter and aerosol extinction is related not just to lidar ratio (indicator of aerosol type) but also to the amount of attenuation, and the attenuation may be a more dominant effect in this data set. If you really wanted to distinguish different aerosol types, you might consider including lidar ratio (from the Raman lidar) in the analysis. If you don't care about aerosol type, then probably just delete the statement about them.

The term "aerosol types" has been deleted.

354. "The most evident differences between the two lidars can be identified for values of extinction larger than about $5.0x10{-}5m{-}1$ where miniMPL shows a broader scatter." It's very difficult to see this. I see that miniMPL has a bit more data at the low end of the x-axis and MUSA has a bit more data at the high end of the x-axis, but it is by no means obvious.

The two panels in Figure 7 have been replaced using a plot simultaneously showing CIAO lidars and MiniMPL using different colors. This clarifies where are the differences between the RCS values measured by the two lidars.

355. "Described above" What does this refer to, the unexplained statement about after- pulse correction?

Yes, now this is clarified in the text.

371. SNR. There seems to be some confusion between signal level and signal-to- noise ratio. The text says the CS135 SNR decreases above 3500 and the CS 51 SNR is higher. To me, it looks like the signal of CS135 decreases and the signal of CS51 is higher, but the noise in the CS51 signal is also quite high and it clearly does not agree with the more reliable Raman lidar, so it's likely this higher signal is an artifact. Your graph doesn't show SNR explicitly enough to aid in analyzing the SNR. I think you would have to look at both signal and noise and analyze noise levels explicitly to be able to make quantitative statements about how the SNR for the two instruments compare. From figure 8 I think you can say "The CS135 signal strongly decreases . . .. The CL51 signal is higher but the noise suggests that it is not reliable to detect the residual aerosol. . ."

The paragraph has been modified accordingly.

Figure 10. Please check the units. Is the exponent -6? Or -5? Compare to Figure 7 which I think should be the same PEARL profile.

The exponent is -6. This mistake has been fixed in the new version of the manuscript.

422. Similar to my comment at line 338, I don't agree with this. Is each case normalized separately? If so, then the variability of the normalization constant is not represented in this plot. The error bars are related to the amount of variability over the few hundred meters of normalization range, but not to the stability of the normalization constant over time.

Please see previous comments for CIAO lidars and MiniMPL.

425. "The standard deviation of the normalization constant." Is this calculated separately by keeping track of the individual profile normalization constants? That's the correct way to do it.

Please see previous comments for CIAO lidars and MiniMPL.

439. "better performance of the CL51 when the values of extinction are larger for corresponding small values of backscatter and therefore indicates its improved SNR in the FT". Is it really true that these large values of extinction with corresponding small values of backscatter are in the free

troposphere? Wouldn't small backscatter values in the free troposphere more likely be accompanied by small values of extinction? It seems more logical that if there are small backscatter values when the extinction is large, that means there is significant attenuation, so the points are more likely low in the atmosphere below significant aerosol layers. This is important to check and clarify since you seem to be drawing a major conclusion (better performance of CL51 in the FT) almost wholly from this subtle and hard-to-interpret pattern.

The manuscript refers only to night time data when the boundary layer height is very low and it is very often difficult to identify it with the MUSA-PEARL data. Therefore, most of the data and the reported value effectively corresponds to value in the FT. This is also consistent with the CIAO boundary layer climatology and modelling estimation of the nocturnal boundary layer. However, to avoid misinterpretations, the authors replaced FT with the term "aerosol residual layer" which improves the reader understanding.
The paragraph is now as follows: "*these threshold values reveal the slightly better performance of the CL51 when the values of α are larger for corresponding small values of β' and, therefore, indicates CL51 improved SNR in the night time aerosol residual layer, in particular below 2.0 km asl where the profiles measured by both the ceilometer may be still affected by the correction for the incomplete overlap*"

443. "The overall stability of ceilometers' calibration constant . . .has been addressed in a statistical sense." I don't see any analysis of the overall stability of the calibration constant, see comments at line 338 and 422.

Please see previous comments for CIAO lidars and MiniMPL.

464. "in general is embedded" – please be more specific. Do you mean "is directly proportional to"? If so, please say that. If the relationship is more complicated than that, please include the equation.

The sentence has been modified as follows: "*The number of lasers pulses is included as a multiplying factor in the CHM15k data processing and it is one of the factors contributing to the so-called lidar constant (i.e the constant depending only on the lidar system experimental setup) within the lidar equation.*".

470. "the calculated embedded constant". Does this mean lidar constant? Please say that. I think it would be good to put in some clarification that the calibration constant is an operational assessment of the lidar constant (which may have some noise or error). So, I think what you're saying is that if the true lidar constant has seasonal variability but a calibration constant is only assessed infrequently, then there will be a systematic error in the calibration constant.

The authors thank the reviewer for noting this inconsistency. The text of the manuscript has been modified as follows: "*This indicates that, across a fixed calibration range (i.e an aerosol free range to perform the molecular calibration), the normalization constant will range with a behaviour similar to that shown by the laser pulses in order to correct for the change in transmitted energy. As a consequence, given that the normalization constant is an operational assessment of the lidar constant plus a residual uncertainty due to the noise, also the true lidar constant will have seasonal variability. The reported …….*".

471. “what was reported during INTERACT”. What was reported? Be more specific.

With regard with the last two comments above, at line 470-471 of the old manuscript version, the authors have modified the paragraph as follows: “***<u>This indicates that, across a fixed calibration range (i.e. an aerosol free range to perform the molecular calibration), the normalization constant will range a behaviour similar to that shown by the laser pulses in order to correct for the change in transmitted energy. As a consequence, given that the normalization constant is an operational assessment of the lidar constant plus a residual uncertainty due to the noise, the true lidar constant will have the same seasonal variability as the normalization constant. The reported laser pulses variability can contribute to explain the large variability of the calibration constant (about 58 %) calculated during the six-month period of INTERACT-I (Madonna et al., 2015) which was only partly due to the variability of MUSA reference lidar (19%). During INTERACT-I, a direct correlation between the variability of the calibration constant and the seasonal temperature changes was found to be limited (R^2=0.6). Nevertheless, the seasonal change in the absolute value of the calibration constant was quite evident and addressed to the coupling of two simultaneous effects (temperature change and decrease in the aerosol loading). The reported seasonal variability of laser pulses also confirms that a calibration constant assessed infrequently will increase the systematic uncertainty contribution. It is possible to estimate over a period longer than 6 months a systematic uncertainty in the calibration constant of 10-20 %; over a period of three months the additional uncertainty may reduce to 5-10%.”.</u>***

471. “This partly explains”. To me this finding of a temperature dependence suggests a hypothesis, but I don’t see any testing or exploration of the hypothesis. Is there any indication that the variability during INTERACT was correlated with temperature? (I see in the earlier paper it was believed that there was, but there was no quantification of the correlation, and that information is missing entirely from this paper).

To clarify this point, the authors has modified the corresponding paragraph as reported in the reply to the previous comment.

471. As your continuing discussion points out, it doesn’t seem that the size of the effect matches well at all. If the lidar constant is linearly related to the number of laser pulses, then the variabilities are also linearly related, and so 10% variability in pulse count can hardly explain 58% variability in the calibration constant. I think maybe it would be best to change the wording to remove or further deemphasize the “This partly explains” clause. While I agree that you have demonstrated that operators must be aware of temperature as a source of variability, as an investigation of the cause of the observed variability in the INTERACT observations, this is inconclusive at best.

Please see previous comment.

486. “most of the difference could be reduced after a reevaluation of the overlap correction”. This statement in the conclusions is quite a bit stronger than the statement in the body of the text. In the text you demonstrated that reduction of the error was possible for a single case when the Raman lidar is available to show the true shape of the overlap region, but that it couldn’t be corrected in most cases.

To properly evaluate the overlap correction, an observation scenario with a low aerosol content is required which was not very common during the period of INTERACT-II. Nevertheless, the results of this test, along with the experience gained in the overall data analysis, allowed us to be optimistic on the possibility to improve the MiniMPL performances if a more robust evaluation of the overlap corrections function is carried out.
The sentence commented by the reviewer has been modified as follows: "*The RCS values measured with MiniMPL and CIAO lidars agree within 10-15 % and a re-evaluation of the overlap correction applied in the data processing could further reduced the discrepancies.*"

492. "The CL51 is able to detect the molecular signal in the free troposphere". I'm not convinced this was demonstrated.

Also on the basis of the results reported in the manuscript, the sentence has been smoothed as follows: "*The CL51 appears to have the capability to detect the molecular signal in the free troposphere and, therefore, in order to retrieve the aerosol backscattering coefficient, the calibration of the attenuated backscatter using a molecular profile as a reference can be attempted over integration times longer than 1-2 hours.*"

500. Since the introduction suggested a main motivation was "to understand to what extent automatic lidars and ceilometers are able to provide an estimation of the aerosol geometric and optical properties and fill in the geographical gaps of the existing advanced lidar network", it would be good to see some conclusion about this question here. You have said earlier "the only possible CL51 normalization to provide a reliable estimate of attenuated backscatter profile must be performed over a profile of attenuated backscatter from a reference lidar (like MUSA or PEARL)." These seems to argue against the usefulness of ceilometers for filling in existing gaps. Whether or not I am correctly guessing your conclusion, some discussion belongs in the conclusion section.

The following paragraph has been added to the conclusions: "*The experience gained during INTERACT-I and INTERACT-II confirms the ceilometers' good performances in the qualitatively monitoring of aerosols in the boundary layer with enhanced profiling capabilities in the free troposphere only for the most advanced models. Nevertheless, the retrieval of aerosol attenuated backscatter (and of any related optical properties) appears to be affected by the instrumental issues which must be improved by the manufacturers in cooperation with the scientific community. It is possible therefore to argue that, compared to automatic (backscatter) lidars, more expensive but more powerful, the capability of ceilometers of filling in the existing observational gaps in lidar networks is continuous improving but it is still limited.".*"

3 Technical & grammatical:
143. Is it 16 optical channels? The description in the following sentences seems to say 16, not 17. Is something left out or is there a typo, maybe?
231. Probably "temperature" rather than "thermostat". A thermostat regulates temperature.
235. Instead of using "beta", spell out attenuated backscatter or use the symbol β′ that was already introduced.
344, 353, 354, elsewhere? Fix formatting of numbers in scientific notation
367. Possible missing word "between" 2.5 and 3.5 km asl
416. Delete the word "average"? I think you probably are reporting the standard deviations of the fractional differences, not the standard error of the mean. If you are reporting the standard error

of the mean, please use that terminology rather than “standard deviation of the average”.
448. Replace indifferently with interchangeably
451. “over the time”, delete “the”
504. “INTERACT-II”. Should this be “INTERACT-I”?
Figure 1. A log scale might be more informative for this quantity.
Figures 3, 4, 6, 8 . The label “LIDAR” should be “MUSA”, “PEARL” or “MUSA/PEARL”
Figure 7. the axis labels are really small and it’s not possible to zoom them in enough to make them clear. It would be good to remake these with bigger axis labels. (But see above: I also have a suggestion for a different plot style altogether.)
Figure 7 caption. Please state the time & date of the comparisons.
Figure 8. “Using t[w]o normalization ranges (below 3 km and above 8 km)”. It appears that this is incorrectly pasted from another figure. Figure 8 doesn’t seem to have two normalization regions.
Figure 11 caption, line 841. “standard deviations of the fractional differences” not “average”, I think (see above)

All the technical corrections have been fixed in the new version of the manuscript according to the reviewers’ suggestions.