This is an interesting, well written paper proposing a methodology for categorizing on a physical basis aerosols and clouds (the latter within certain conditions and limits) with high temporal resolution from lidar-only data.

The methodology is illustrated with practical examples and appears to be adaptable to aerosol lidars with the capabilities of that operated by the authors, provided the instrument constant of the different channels exhibits a minimum stability. The mutiwavelength, polarization-sensitive type of lidar required to apply the methodology is certainly sophisticated, but such instruments are becoming increasingly common thanks to the expertise developed and disseminated through initiatives and networks such as EARLINET, PollyNET, GALION, etc.

Lidar-only categorizations are compared to Cloudnet ones, the authors pointing out the synergy obtained by the lidar- and radar-based methodologies.

The paper should be published, although its reading prompts some questions / remarks detailed in the following that the authors are requested to address.

Main questions:

1. One of the parameters used in the proposed categorization method is the so-called quasi-particle backscatter coefficients. Profiles of this parameter are obtained assuming a typical lidar ratio and correcting the attenuated backscatter for the molecular extinction and the particle extinction, the latter estimated by multiplying the lidar ratio by a “first guess” of the quasi-particle backscatter coefficient obtained from the attenuated backscatter corrected only for the molecular extinction. Eqs. (6)-(8) describe clearly the procedure. However one wonders if the cycle could not go on – i.e. if the quasi-particle backscatter coefficient so obtained could not be used to refine the estimate of the particle extinction (Eq. (7)), which in turn would be inserted again in Eq. (8) to obtain a new estimate of the quasi-particle backscatter – until a convergence criterion is attained. It doesn’t seem that implementing this iterative cycle would be too cumbersome from the computational point of view and it would probably converge to what the authors call the “real backscatter coefficient” on lines 293, 305, and 616, and in the captions of figs. 4 and 5. Have the authors tested if this would improve the performance of the categorization?

2. I find the term “real backscatter coefficient”, as used where indicated in the above remark, somewhat misleading, because, even though the authors do not indicate explicitly how it is obtained (Klett-Fernal algorithm, iterative algorithm?), it relies probably on an assumed lidar ratio. The authors should clarify this.

3. In connection with the previous remark, the authors should also clarify what is understood by the “truth” on line 313.

4. On lines 14-16, it is stated: “By analyzing the entire HOPE campaign, almost 1 million pixel (5 min times 30 m) could be successfully classified from the two months data set with the newly developed tool”. I wonder if the claim (repeated in a somewhat different and possibly less strong and more appropriate way on line 598-599: “more than 1 million pixels of 30 m vertical and 5 min temporal resolution were successfully analysed”) is not too exaggerated. If we understand by “successfully classified” that the pixels actually contained the particle class assigned by the algorithm, can the authors be sure of that?

5. In the criteria for the categorization, clearly summarized in table 1, the Ångström exponents are not always taken into account (e.g. for the aerosol mixture, partly non-spherical, for large, non-spherical aerosols or for likely water droplets). Could the authors comment on the reason for that? Wouldn’t the consideration of the Ångström exponents provide confirmation or allow a refinement of the classification?

6. In the explanations in the text about the categorization criteria summarized on table 1, I suggest, for the convenience of the reader, a brief explanation of the Cloudnet algorithm used for some cloud categories.

7. When comparing the results of the lidar categorization to those of a Cloudnet station operating in the same location (or close enough) as the multiwavelength lidar (e.g. lines 467-479, 523-527), the authors conclude that either classification is reasonable, although they do not correspond to the same types of particle, and that the comparison results show the synergy between radar- and lidar-based classification, each system being sensitive to different types of particles. But, can a misclassification be completely ruled out, e.g. the supercooled droplets or the drizzle identified by Cloudnet being mistaken by the large aerosol particles identified by the lidar?

8. On lines 647-649 it is stated: “Ice crystals were also often classified correctly, but sometimes remained unclassified or even false classified as aerosol as a consequence of multiple reasons (a priori information aiming at aerosol, low depolarizing characteristics in certain temperature ranges, etc.)”. I think this is not sufficiently emphasize in the paper. These instances where the classification outcome is doubtful should be pointed out in section 4.2

Other issues

1. The description given in the text of the $\mathrm{Polly}^{\mathrm{XT}}_{\mathrm{IFT}}$ system does not coincide completely with that found in the reference Engelemann et al. 2016. Only a depolarization channel at 532 nm is mentioned in the text, while two, at 355 nm and 532 nm, are indicated in the reference.

2. On lines 191-192, referring to Fig.1, the authors say that “One can see that during most of the intervals of no setup change, the lidar system parameter is relatively stable and only

some of the setup changes have caused a significant change in C^{λ}". Is this really sustained by Fig. 1? There seems not to be periods between changes with many C^{λ} measurements.

3. Also related to Fig.1, I found it difficult to relate the vertical lines indicating changes in the lidar setup to the specific dates mentioned in the text. I suggest labeling those lines (at least those referred to in the text) with the precise dates.

4. On lines 194-196 the authors say: "It was found that changes in the indoor temperature of the cabinet due air conditioning malfunctioning had led to a change of the alignment and thus a change in C^{λ} during this period". Didn't this also lead to a change of the overlap function?

5. On lines 199-200 the authors say: "On three days (18 April, 25 April, and 10 May), for which multiple system setup changes were performed, more than one lidar system parameter was used to account for these setup changes". Is this shown in Fig. 1?

6. In fig.1, is it only an "optical effect", because the absolute values are higher, or it is true that the lidar system parameter has more relative variability for the 1064-nm channel than for the two other channels. If it is true, is the reason known?

7. On lines 211-212 it is stated: "For days with inappropriate weather conditions a standard value (mean of HOPE) [of the depolarization calibration constant] is used". I suppose this refers to inappropriate weather conditions for determining V^*. This should be explained.

8. On line 262 a molecular depolarization ratio equal to 0.0053 is mentioned for Polly^{XT}. A reference is in order.

9. The sentence on lines 368-371: "This threshold yields a ratio of molecular to particle backscattering at 532 (355) nm higher than 60 (180) at sea level and thus is valid for a Rayleigh calibration by means of the Raman or Klett-Fernald lidar method which might be one future application of the target categorization presented herein" is not clear (because of its ending: "which might be one future application of the target categorization presented herein") . Please check and clarify.

10. While the figures have high resolution, which allows to zoom in when reading the paper on a computer, the size of some of them is too small for a comfortable reading in print. This affects especially figs. 3, 4, 7, 8, 9, 10, and 12. In addition the left and central panels of fig. 4 have probably an excess of information, with some curves with similar colors, which makes difficult to distinguish the different parameters represented. A similar problem occurs with fig. 5. I suggest either splitting the figures to reduce the information content of each or using, in addition to colors, different line types to make easier to distinguish between them.

11. The lidar categorization figures bear the title “Lidar target classification” in an insert, while for the Cloudnet ones the titles read just “Target classification”. I suggest that these are modified to read “Cloudnet target classification”.

12. At the beginning of section 4.2 I suggest a brief rationale on the selection of three case studies selected. Why these cases instead of others? What makes them especially interesting?

13. The convective cloud observed “shortly past 12 UTC” (lines 432 and 458-459) on figs. 7 and 8 is very difficult to distinguish. I suggest some means (arrow, circle around…) to draw the reader’s attention to it.

14. Do the statement on lines 599-601 (“From these pixels, clean (i.e. molecular scattering dominating) atmosphere was observed in 29%, clouds in only 7%, aerosol in about 37% and non-typed particles/particles with low concentration in 27% of the analysed and feature-classified pixels”), and the statistics presented in fig. 13, make sense without specifying the maximum exploration height?

Minor issues:

1. The authors use in many instances the E-notation (e.g. 2e-5 to represent 2×10^{-5}). The 10 with superscript exponent notation should be used throughout.

2. There are several instances of “even so” that should probably be “even though”.

3. The paper should be revised to correct typos (e.g “Ångstöm” instead of “Ångström” on line 283, small punctuation issues (e.g. missing commas), missing spaces between words (e.g. line 269 “2013started”), etc.