Response to Reviewer Report #2 by Janicka & Stachlewska on 14 May 2019

Dear Referee, Dear Editor,

We are grateful to Referee for comments and suggestions that allowed us to improve this manuscript.

In the following, the answers to Referee's comments and issues raised are reported directly below each related comment. All modifications of the initial version of the manuscript as well as additions are reported in color highlight in the revised version of the manuscript. We believe that we have fulfilled required changes in the final version of the manuscript.

We would like to thank Referee for her/his opinion on the initially submitted manuscript. We have carefully considered all of the comments and suggestions; below point-by-point answers are given (in blue). In general, we would like to stress that crucial points raised by the Referee have been revised, so as following:

- the methodology for interpretation of aerosol layers, as not described well enough

Now: Aerosol source/composition/advection better combined, plus referred to values reported in literature

- the uncertainties of obtained properties, as not addressed sufficiently

Now: Systematic and statistical measurement uncertainties discussed in terms of signal-to-noise ratio (averaging/smoothing), data evaluation methodology, etc.

Therefore, we are grateful to Referee for helping in substantially improving the initial version of the manuscript.

The manuscript introduces a dataset of lidar observations of aerosol backscatter,aerosol extinction, and aerosol depolarization at multiple wavelengths plus relative humidity, for a heat wave event in Warsaw comprising many identified aerosol layers over two nights. It provides aerosol type identification for these layers and gives the mean lidar properties for layers of each type.

In fact, it was never meant to focus this study on the aerosol typing in the sense of an application of an automated aerosol typing routine (so as e.g. Papagiannopoulos et al., 2018, Nicolae et al., 2018, *this issue*). This was the reason why we avoided using the wording *aerosol typing* throughout this manuscript. We performed a classical aerosol layer interpretation, this using both the aerosol optical properties (as reported in literature), which were combined with backward trajectory analyses as well as information provided by the satellite sensors and synoptic charts).

Unfortunately, most aspects of aerosol type analysis are not clear and there are many results with inadequate support or no support whatsoever.

In the revised version an effort was put on better describing the approach undertaken for the interpretation of the identified aerosol layers (and sub-layers), through adding also a methodology work-flow scheme, a better description of the aerosol layers in the Tables and Figures, and most of all – relevant references to place this research in a wider range.

The overall motivation or objective of the manuscript also seems confused.

This is true, We regret not to state this clear enough. The introduction has been modified accordingly with focus on motivation and objectives. This is explicitly addressed in the answered to specific comments of the Referee, below.

The lack of clarity about the objectives and especially the flaws in the analysis make it impossible for me to recommend the manuscript for publication in ACP.

Several aspects have been improved, all accordingly to both Referees suggestions and comments, and therefore we believe the revised version will satisfy Referees.

Before getting into the details, I will say that I agree that the dataset itself (as distinct from the aerosol type analysis) is potentially quite valuable, and I am particularly impressed by the inclusion of relative humidity profile information which adds significant potential for science value when combined with the lidar aerosol measurements.

Thank you for this comment. In fact, adding information on lidar-derived RH in the presented data set is an advantage over the EARLINET data sets, as for RH profiles being not yet reported in the EARLINET-ACTRIS Data Base. All profiles derived in this study are published in this data base. The evaluation scheme was introduced with much detail in Baars et al, 2016. All of the obtained sets of profiles are accomplished for the measurements comprising with the QA of the EARLINET-ACTRIS recommendations. Moreover also the QC tests were performed these data, whereby the Warsaw site has one of the highest scores of according to last evaluation in March 2019 for 97% of all Warsaw profiles stored fulfill the QC). Now, this all is in regard of all profiles reported in our paper but RH profiles, due to the lack of QA&QC procedures for RH in the EARLINET-ACTRIS Data Base (majority of sites are not working yet with RH). The evaluation scheme and results validation for RH are reported by Stachlewska et al 2017, where details of the retrieval, including uncertainty analyses are given.

One option might be to make a streamlined manuscript without the aerosol type analysis, but with a more in-depth uncertainty analysis for the primary lidar measurements (backscatter, extinction, intensive parameters, and RH), plus an explanation of the quality control, and submit this alternate paper as a data paper in a data journal such as Earth System Science Data.

We would like to thank Referee for this recommendation, as publishing the data sets in such a highly values journal as the Earth System Science Data of Copernicus Publications would be precious. However, we feel that it is not feasible for the current study and the discussed data set. First of all, the quality assurance and control of the obtained data set has been done accordingly to the EARLINET-ACTRIS procedures. Secondly, the data has been already published in the EARLINET-ACTRIS Data Base. The EARLINET Community (e.g. Nemuc 2019, Short Comment, *this issue*) show potential interest for using this data set. Even Referee himself seems to have interest to use it – questions on data public availability). Thirdly, we added a new section related to measurement uncertainty analyses, as suggested. This is why we reckon that submitting this research to the current EARLINET Special Issue of the ACP is appropriate and valuable for this issue/community.

I will organize my more detailed comments by manuscript section.

Thank you, below we answered them point-by-point.

Section "Introduction":

My primary criticism in the introduction is that it does not adequately state the objective and motivation.

The introduction has been revised taking to account following:

The main objective:
Derive unique set of comprehensive profiles that can be used in future for further exploration in e.g. microphysical inversion (as in Veselovskii et al., 2002; Böckmann et al., 2005; Müller et al., 2016); testing of aerosol typing algorithms (e.g. Nicolae et al., 2018; Papagiannopoulos et al., 2018); testing the aerosol separation algorithms (e.g. Mamouri and Ansmann, 2017).

The motivation for this study:
The aerosol properties reported in literature are often limited in terms of: i) existence of wavelength dependent depolarization and/or RH, ii) profiles obtained for a single measurement (1-2h average, at single location, at particular time, no temporal dependence), iii) profiles evaluated for the cases of well-defined aerosol (practically pure aerosol type/source) in well-defined layers (geometrically and optically thick).
But how to understand such reported in literature profiles?
What can be said on their representability in vertical and temporal extend?
Is the significant averaging (over time/height) necessary?
Is it meaningful to divide atmosphere to a high spatial-temporal resolution sectors and assess the aerosol properties within them?
Is it feasible to find in these many very thin sub-layers a coherence of the aerosol properties?
Are those connected to air-mass transport to unambiguously interpret/estimate their possible origin?
Finally, to confirm the results with advances microphysics inversion applied in a 2-dim height/time space, and yet be able to provide consistent microphysical parameter retrieval?

(Note that the latter is the topic of the related part 2 paper. The first results of microphysical retrieval show promising results (paper accepted for presenting at the next ILRC-2019).

The key statement of the objective in the introduction seems to beat page 2, lines 30-33, "The algorithms that deal with the inversion problem of micro-physical retrieval require an accurate aerosol layer selection and a high quality 3beta+ 2alpha optical data set as well as the depolarization information [Veselovskii et al.2002; Bockmann et al. 2005; Muller et al. 2016]. Thus the manual data evaluation which allows for an insightful analysis of the lidar signals with the individual approach to the considered case is still much-needed."
In this study, the high quality 3beta + 2alpha optical data set and the QA on that dataset is provided by the retrieval algorithm which is considered to be an input to this study and not part of the study itself.

The algorithm is published in ACP by Baars et al 2016.
The QA&QC of the EARLINET-ACTRIS are in place for all data profiles obtained.
We doubt repetitions are necessary, however we will comment on both in the revised text.

Similarly, although the layer selection is done here, the explanation of how it's done is presented as a black box.

Thank you, this is very important point, especially as the layers in our work are not defined in a classical approach (not based on gradients as marking the layers borders).
The explicit layer selection work-flow applied in our work is added in the revised version.

The statement is a non-sequitur if the "insightful analysis" is taken to mean the aerosol typing analysis, since that is not needed for microphysical retrievals.

In the fragment Referee cites above we spoke of aerosol layer selection - not interpretation, nor typing. We shall rephrase, as this seem is not clear to the reader.

I also don't understand why automated aerosol typing algorithms are acknowledged but discarded.

Well, we acknowledge the automated aerosol typing algorithms them as professional and great software tools. They are not disregarded, as they will be likely used in the future.
Indeed, we do not aim at coming up with another typing-algorithm ourselves but we are looking for collaboration! In part 2 paper, we explore microphysics inversion of this data set but we are willing to consider also an automated aerosol typing (and this, not as an option, but as a must). Thank you.

In short, I just don't understand the motivation for doing this study this way.

We apologize for not being able to convince Referee to the proposed analyses approach in the initial version of the manuscript. On the positive site, at least potentially nothing can stop Referee from performing this study in her/his way. We are keen to openly publish or provide directly all raw data files (additionally to the profiles stored in the ERLINET ACTRIS Data Base), so that the Referee can perform analyses.

A less important point is that there is a lot of information in the introduction, such as the first several paragraphs about the effect of aerosol properties on radiative forcing, whose relevance to this study is not explained at all.

Thank you, we removed from Introduction all comments/statements which are indirectly related, thus irrelevant to this paper.

On a positive note, I appreciate the introduction to the specific heat wave event being studied, and also the statement about the uniqueness of the dataset.

Thank you, in the revised version we strongly focused on including only relevant information and commenting on why we include it in the sense of supporting objectives and motivation of this study.

Section "Methodology":

The methodology describes the retrieval methodology for obtaining the measurements of e.g. backscatter and extinction but not the analysis methodology.
Most of the analysis in this paper is about aerosol type attribution, but there is no description of the methodology for this (here or anywhere in the paper), leaving me completely without a foundation for understanding the rest of the paper.

Thank you, we added the analysis methodology work-flow diagram.
We also improved the Reference list and added more descriptive information in the aerosol attribution Table.

For a few other key aspects of the analysis, the hints in the methodology are inadequate, for example "The sub-layers selection was based mainly on RH and extinction profiles". This is too vague to be useful for understanding how the layers are selected.

This sentence has been rephrased to: 'The definition of the sub-layer in this paper is not standard as it is not based on the typically used gradient of the signal, e.g. ref. The sub-layers were discriminated to fulfill the quality requirements to further microphysical retrieval (part II paper). The sub-layer selection is appropriate for such inversion if the extensive optical properties (α, β) follow the same tendency at each wavelength (i.e. constant, increasing or decreasing).
In the sub-layer discrimination, the first step was to select the constant sections in the relative humidity profile and compare them with the extinction coefficient profiles. In the next step, the correspond to the backscatter coefficients profiles as well as to the sub-layers visible in the depolarization ratio were checked and corrected if necessarily. Examples of how the sub-layers were selected are shown on the profiles in Fig. 5 and in the Appendix A.

Finally, the sub-layers were collected into groups (layers) characterized by similar properties and described in the statistical way by the mean values of the intensive optical properties and the relative humidity with the standard deviation. The layers were discriminated based on the relative humidity and depolarization plots, as the differences between the layers were the most pronounced using these properties. Then, the first choice was revised with the spread between the properties using the plots of mutual relations of intensive optical properties and relative humidity shown in Appendix B. If particular layer had too much noise (especially upper layers) it was no further analyzed.'.

Also, I think it is important to understand the uncertainties on the lidar measurements.
Page 5, line 30 says the uncertainties for the intensive parameters are propagated from the backscatter and extinction uncertainties, but does not say how the backscatter and extinction uncertainties are calculated, and anyway, it doesn't seem that the uncertainties are actually presented in the paper. Error bars in the figures and tables seem to be just the standard deviation calculated for selected layers, not measurement uncertainties. Standard deviation might be a reasonable stand-in for measurement random uncertainties, but for some of the analysis, such as understanding quite small particle depolarization values, the authors need to develop an understanding of the expected systematic uncertainty.

Indeed, in the initial version the measurement uncertainties were not discussed. Although in the Appendix B the uncertainties for all values plotted in scatter plots are given with uncertainty. Only the mean values (right hand side plots) show standard deviations.
Section on uncertainties of lidar measurements has been added, as recommended.

We do not really understand why Referee thinks it is necessary to comment on *quite small particle depolarization values*. The values of depolarization obtained in our study align in the possible ranges of depolarization values reported for different aerosol types, as listed in the extensive table based on the most recent literature review in Nicolae et al. (2018).

The statement "The columnar value of the extinction related Angstrom exponent of 1 was assumed for the extinction coefficient profiles calculations" confuses me. I don't understand why you would need this assumption and it seems that this assumption would have too much impact on the layer angstrom exponents. I think I'm probably just misunderstanding what this means. I would like to be able to read a reference about the retrieval to be sure, but I can't find one cited. Did I miss it?

We clarified this in the revised manuscript. The AE=1 is recommendable and applied in the classical Raman approach introduced by Ansmann et al. (1990). The uncertainty related to this assumption is assessed based on the sensitivity study (for AE threshold values of 0 and 3). This is not a dominating source of uncertainty, more important is SNR (for extinction coefficient and RH) and ref. height (for backscatter coefficient and depolarization ratio).
The sources of the measurement uncertainties are addressed in the revised version.

Section 4.3 "Aerosol source analysis"

On first read-through of the paper, I thought the aerosol source analysis was meant to be a major part of the analysis. There is a significant amount of analysis to calculate and collate the backtrajectories in an attempt to identify source regions, and this could potentially be very informative in the inference of aerosol type and source for the observed layers.

Thank you, this is what we actually aimed at: an independent analysis of the source regions. Backward trajectories were calculated at regular height intervals at the beginning /end of each night. Then we constrained the trajectories with possible aerosol sources (satellite data). It was necessary to proceed this way, as the aerosol sub-layers in our case are not defined based on the search for significant gradients in the profiles. Then, the individual layers were (one by one!) *connected* to the transport/source origin based on height-space apportioning.

However, the analysis stops far short of that. There is very little attempt to link the backtrajectories to specific layers and the two segments of analysis (aerosol source analysis vs. aerosol typing) seem to be separate, disjointed sections.

As said above, there was careful apportioning performed. In the revised version of the manuscript we put more effort to explicitly state the link between each layer vs aerosol source, i.e. in Fig. 6 (layers) we add legend relating them to Fig.4 (aerosol source), using the same nomenclature and in Table 1.

There are also several confusing or contradictory statements.
For example, page 8, lines 14-15 say where two-days old smoke would have come from but then lines 17-19 says that above 3 km, it's about 3 days old and below 3 km, it's older; so where is the 2-days old smoke?
Then at line 20 is a statement that aged biomass burning aerosol is possible in the lowermost levels but at line 22, it's "moderately fresh".
I'm not sure if all of this is actually contradictory or if the authors are expressing the thought that there are multiple possibilities and a lot of uncertainty.

Well, yes. We were suggesting that there are multiple possibilities for interpretation. We took wrong approach – we wanted to point out all possible explanations and then argue to eliminate the less likely. As this seem confusing to the reader we revised these sentences and left unequivocal interpretation (or none).

If there are any layers at all where the airmass analysis allows for assessing probable sources or a range of ages of the aerosol layer, it would be very helpful to see that information included explicitly in Table1 where it can be easily used for understanding the aerosol types.

Thank you for this suggestion, it has been added as suggested.

Section 5 "Optical properties of the BBA and pollution mixtures"

In this section, individual aerosol layers are labeled according to aerosol types. In a few cases, the backtrajectories are referred to, but mostly the methodology seems to be based on haphazardly comparing the aerosol intensive parameters to a few case studies in prior literature.

After adding the methodology flow it is much clearer. Also addition of the respective legends and more descriptive information in Table 1 helps out. In fact trajectories were related to each sub-layer. Thank you.

There is no discussion of how exactly this is done, no consistent thresholds are presented, and there is no quantitative comparison with prior literature that present ranges of parameters that might be expected for different types (e.g. Muller et al. 2007, Gross et al. 2012, Burton et al. 2012).

As a manual interpretation of data and not an automated aerosol typing is applied in this study, therefore there is/was no intention to *fix thresholds*. However, indeed the quantitative comparisons with prior literature can be improved, and so was done in the revised version, and special case was sent on referring to work of Müller et al. (2007), Groß et al. (2012), Burton et al. (2012).

Unfortunately, there are many, many examples of type labels being applied without adequate explanation, support, or comparison:

Page 9, line 29: "properties typical for the aged biomass burning aerosol (CRLR >1)".No support is given for this value being "typical" for aged biomass burning aerosol. Although it is not cited in reference to this statement, I believe this argument follows from the case studies presented by Nicolae

et al. (2013). However, since those are just a few case studies, it may be an overgeneralization to describe this relationship as "typical".

Thank you, references are added, sentences rephrased, wording for *typical* is no longer used.

Page 10, line 22: "The relatively low LR and high AE...indicates pollution domination. " No reference is given for LR and AE of pollution or a description of how to determine them.

Thank you, references are added, sentences rephrased, discussion is elaborated.

Page 10, line 22: The low lidar ratios that are said to indicate pollution seem very low compared to prior literature that I know of. No references are given here. How do these values compare to other published values for pollution cases?

Thank you, references are added, sentences rephrased, results are discussed against literature.

Page 10, line 24, "relatively high value of 532 nm particulate depolarization of 4.8 plus or minus 0.3 ... may reflect contamination by pollen". This is not a very high value of particulate depolarization, and is only for the earlier layer (L1b), and the larger AE for the earlier layer would suggest smaller particles rather than larger. These seem like subtle and confusing trends. Is it clear that these distinctions are actually significant compared to measurement uncertainty (random + systematic)? Also, why is the conclusion that pollen is present, and not some other depolarizing type such as dust? 4.8% is small enough that even smoke could have such a value. The Sicard reference cited here does not help me understand this, since it's not about this case, but is just a general reference for lidar measurements of pollen.
(And a minor related point: if the purpose is to credit lidar measurements of pollen, much earlier papers such as Sassen et al. 2008 should get that credit, doi: 10.1029/2008gl035085).

Thank you, references are added, sentences rephrased.
Special care was dedicate to more discussion of this dry polluted layer with only a slight contamination of the residual pollen particles, which were observed during pollination event at daytime of 10 August 2015). For pure pollen values of 6-12% over Warsaw are reported. Obtained here 4.8% is attributed to pollen contamination of polluted layer. Pollen contamination is more likely than dust/smoke, as for this layer there is evidence of local origin no long-range transport.
As suggested, the reference to Sassen et al. (2008) is added and the reference to Sicard et al. (2018) is removed.

Page 10, line 26-29 Contradictory statements that the ratio of particulate depolarization at two wavelengths "did not vary significantly" but the difference between them "confirms the hypothesis of pollen contamination". If the variation is not significant, it doesn't confirm the hypothesis.

Thank you, references are added, sentences rephrased. We apologize for misleading phrasing.
In fact, what was meant is that values are low, being lowermost for pollution layer contaminated with residual pollen particles.

Page 10, line 31. Higher lidar ratio values "indicate domination by the biomass burning aerosol." Again, based on what methodology and what references? All the papers I'm familiar with (e.g. Muller et al. 2007, Gross et al. 2012, Burton et al. 2012,Papagiannopoulos et al. 2018) have a significant overlap for the lidar ratio of biomass burning and pollution, at whichever wavelength they look at.

Thank you, suggested references are cited, sentences rephrased, comparability of results addressed.

Page 11, line 1, imaginary refractive index values of 0.12i or 0.20-0.30i seem extremely high. This is quote from a paper in preparation. It doesn't seem appropriate to present something so unexpected without any support or discussion.

Page 11, line 5, "as the LR values were relatively high of 81 plus or minus 6 sr and 60 plus or minus 6 sr at 355 nm and 532 nm, respectively, the layer can be treated as fresh biomass burning aerosol...although the angstrom exponent value of 1.34 plus or minus 0.3 is rather low for fresh BBA". How is this conclusion reached? What values are these measurements being compared to? Couldn't this be urban pollution?

Page 11, lines 15-16 "Such low depolarization ratios are characteristic for aged biomass burning". No references are given to support this statement. Aren't low depolarization ratios also characteristic of pollution and other aerosol types?

Page 11, line 21. "The AE value of about 0.6 [i.e. the same value as measured in this study] was reported by Muller et al. 2007". Actually, Mueller et al. 2007 reported a value of 1.0 plus or minus 0.5. Maybe that could be considered "about 0.6" but it's misleading way to say it.

Page 12, line 4, appears to suggests that the backscatter angstrom exponent is expected to be monotonically related to the aerosol age. I don't know of any reason to think this. Are there references?

This section also contains statements two pages apart which directly contradict each other, giving me the impression that the authors do not fully understand their subject material and this manuscript was really not ready for submission in the first place. Regarding an apparent correlation between backscatter and depolarization, on page 11,"it is probable that increased depolarization is associated with the higher participation of background aerosol in the tropical air containing some mineral dust particles." And on page 13, "the negative dependence of backscatter and particulate depolarization is unlikely related to mineral dust contamination".

Another question about the measurements.
On page 11, line 28, a lidar ratio of 114 plus or minus 33 sr at 355 nm. Isn't this really very high? Are there other published observations of 355 nm lidar ratios this high? I see from the figure that this very high value occurs where there is significant slope in the backscatter and extinction profiles, on the underside of a layer with larger backscatter. The peak of the backscatter and extinction have very

different shapes, with the extinction looking like a smoothed version of backscatter. If there is a difference in the resolutions of the backscatter and extinction profiles such that the edges of the layers are not being captured the same way, it creates spurious values in the ratio. In this case, it looks like the lidar ratio is large in consequence of the fact that the backscatter and extinction show the edge with different slopes. Are the backscatter and extinction resolutions compatible when the lidar ratio is computed?

Note that all available profiles in the EARLINET data base are contained in so called e and b files, which by definition are stored with different resolution and/or smoothing applied to the data (they contain: e- extinction and b- backscattering and depolarization). This is a standard EARLINET procedure. The EARLINET Warsaw site is one of a few that provide in the b-file not only also extinction (all in the same resolutions!).

Now, according to Referee comment, it seems essentially not possible to derive correctly lidar ratios from profiles stores in this data base. Taking to account vast literature based on the EARLINET data base (e.g. Mattis et al. (2004); Matthias et al. (2004); Papayannis et al. (2008); Papagiannopoulos et al. (2018)) using the LR obtained with different resolutions... I have difficulty to agree with this comment.

Secondly, all of the profiles we calculated are shown in the Appendix A, and looking at these We can easily find the ranges for which even for high slope (in B and A) we get low values of LR. This is why it is difficult to agree that such slope shall *in general* causing large LR (we have large values in only 5-7 layers).

Moreover, LR values of > 100 sr are not impossible, the EARLINET data base threshold for LRmax is at 150sr. In the present study only 4 values are ranging >100 sr. Also, such high LRs were discussed as possible to exist, e.g. during the ILRC-2017 Summer School (lecture course by known lidar expert Prof. A.Pappayannis. Thus, we should not exclude such values for the analyses.

Now, as for the slope itself – in our case it is not much of a slope for the 4 values of high LRs (red points in Fig.6), they also show temporal dependence (likely due to aging, as it is the same aerosol arriving at the site in this layers). If you like, there is a strong temporal change in these high LR (> 100sr) values, if one explore individual values as given on the Appendix A plots (the high LR is becoming lower with time during the night). We shall doubt that this is a coincidence or an artefact.

We regret we did not stress this well enough in the initial manuscript. Note that a sensitivity study was performed to assess if LRs are not overestimated, whereby discussion on backscatter and extinction coefficient calculated with the same resolutions vs applied smoothing to extinction and its effect on resulting LR was addressed. For the presented case the high values of LR are obtained for both cases.

The wording is quite rough. If this manuscript were to be published or resubmitted, it would be very helpful to have a round of English-language editing.

Thank you, obviously neither of us is English native speaker. We apologize for ignoring this fact, and not sending the initial manuscript for language proof. The resubmitted version of the manuscript will be sent for professional English proof.

Finally, where are the data available? Given that the paper is primarily about introducing a new dataset, I would like to see a statement of data availability.

Thank you for this comment, all of the derived data sets are already stored in the EARLINET-ACTRIS Data Base as (b and e files). The RH is currently not therein but it is available via PolandAOD (www.poland aod.pl).
We will extra explore the possibility to add the RH profiles to the EARLINET b-files category instead of volume depolarization profiles. If we obtain agreement form the AQ&AC team, we will add the full sets of profiles to the data base.