Anonymous Referee #2

We would like to thank Referee #2 for the valuable contribution to this manuscript and for the time dedicated to providing the comments and suggestions for the interactive discussion. We made all efforts to follow all recommendations and we hope that the updated version is satisfactory for publication. Below you will find a point by point description of how each comment and suggestion was addressed.

The paper surveys different procedures for calibration of lidar systems to improve the accuracy of depolarization measurements, which are of primary importance to infer aerosol particle shape, hence typology. It briefly introduces the formalism to be used in light polarization measurements and reviews the current calibration techniques, describing relative merits and drawbacks. Then present results from the application of such calibration techniques on some case studies. There are many paper dealing with the calibration of polarization diversity lidars, but the originality of the present one resides in its cut, more oriented to the description and practical implementation of the calibration systems, than in their theoretical description. This, in addition to a praiseworthy review of the theoretical assumptions of existing calibrations, assures it of the interest of the community, and for this reason I believe that the work deserves the publication.
Thank you for this positive appreciation.

Specific comments:
(2,8) typo ": : : about to: : :"

(2,20) ": : : distinguish between rather spherical particles with low depolarization ratio, and non-spherical particles with higher depolarization ratios."
I would here drop the adverb "rather", as the whole sentence tends to suggest (at least to me) a univocal relationship between polarization and a "degree of asphericity", which is misleading, as it has been proved, for instance with theoretical T-matrix computations, that particles that are "rather" but not exactly spherical (i.e. prolate or oblate spheroid with aspect ratio close to unity) may have values of depolarization higher than considerably "more aspherical" (i.e. with aspect ratio much different than one) particles.
The text was modified accordingly. Thank you for this suggestion.

(2,22) ": : :low depolarizing (e. g. local aerosol) : : :", well this claims depends on where you lidar is placed, that in turn dictates what is to be considered "local". I guess that a scientist working in Tamanrasset would have different views on what to consider "local". So you may consider to change "local", to "urban aerosol", as instance?
Indeed, this is correct. Thank you. The text was modified accordingly.

(4,19) A polarization purity of 95% is definitely a problem, and this should be stressed (actually is quite pessimistic, but even a more common 100:1 polarization purity still is a problem). Here you can quote that the residual non polarized laser light can be easily filtered out. It is said thereafter but I think the best place to pose that remark is here.
Yes indeed. Thank you for the suggestion. The text was updated accordingly:
"...commercial Nd:YAG lasers is higher than 95% and the elliptical light component of the remaining light should be even lower. To overcome this issue, the residual non polarized laser light can be easily filtered out by including additional optics in the emission block of the lidar instrument..."

(7,12) please use "responsivities" instead of "quantum efficiencies", as the latter is only a factor of the former. This has an impact in what follows.
The text was updated accordingly

Formula (19): this is basically the ratio of the overall photodetector responsivities for the two channels. What follow is my crucial remark, and I would like the authors to discuss it in some more length. The responsivity, or the "gain" of a detector, is the ratio between the power input (in our case the photon flux) and output, (the current, or photoelectron rate). One would like this gain to be constant, i.e. the idealized detector should have an output which is linearly related to the input. Unfortunately, this is seldom the case for the PMTs and APDs, as the gain may be dependent from the level of the input (this claim is straightforward in the photoncounting acquisition mode, due to "dead time" counting effects, but it is also true in current acquisition mode). This makes the ratio in (19) possibly dependent on the measurement conditions, i.e the altitude at which this ratio is computed (this is somehow implicitly – too much implicitly - addressed in fig 7 a-b) and, in some cases more important, on the level of sky background. It may well be that for "EARLINET-like" systems, which often use high power, low pulse repetition rate lasers, and very narrow interferential filters to reduce background to levels much lower than the actual signal, this effect is not apparent; but in general, and especially for systems with larger spectral bandwidth and low power, high pulse repetition rate lasers, this may be an issue. This is an issue which, to my knowledge, has never been addressed in any study putting forward the merit of calibrations others than the "0 calibration", and I think it is worthwhile to mention it.
We are aware of this issue, therefore, several quality assurance (QA) measures are commonly used:

- During the operation of the lidar instrument the indoor environmental parameters are kept constant (especially temperature).
- The EARLINET QA program recommends that the depolarization calibration procedure is performed prior to each measurement. This provides extra information on the stability of the instrument.

- For all calibration procedures all reference lidar signals are provided either in photon counting detection mode or as glued signals (analog and photon counting with photon counting as the reference signal)
- The height stability of the calibration constant is also checked during the QA process to investigate any height dependency.
- The background level is reduced to the minimum

The authors are fully aware that this is still an open issue and further work is required but according to our current experience the effects related to the responsivity instability of the detectors could be neglected for this study.

(8,18-19) The sentence is unclear and should be rephrased.
The text has been updated.
Updated text:
"In such an atmospheric region, the total volume linear depolarization ratio can be approximated by the well known value of the air molecule linear depolarization ratio (Behrendt and Nakamura, 2002). Usually this procedure can introduce additional uncertainties since for an accurate calibration at least two reference points are required. Another drawback is the presence of small amounts of highly depolarizing aerosol (e.g. ice crystals) in the assumed clean range that can easily lead to large errors in the depolarization products (Freudenthaler et al., 2009; Freudenthaler , 2016). Other calibration techniques include the use of"

(16,18-20) I am somehow uncomfortable with the whole sec. 4.4. I guess everyone is already well aware of ": : : the importance of calibrated depolarization lidar products: : :". This is not the main goal of the paper, but rather to discuss at length the different calibrations; hence it would have been much more interesting to show what is the effect of these calibration procedures, i.e. to show uncorrected vs corrected profiles, which I think is a display much more in line with the rest of the paper. Therefore, I would ask the authors to do that.
This is correct. This section has been modified accordingly. Thank you
Old text:
"In order to make a first estimate on the depolarization accuracy of the discussed lidar instruments and to emphasize the importance of the depolarization calibration for the long range transport and aerosol typing studies, several experimental results obtained using calibrated depolarization lidar instruments from different EARLINET stations are presented and discussed."
New text:
"In order to make a first estimate on the depolarization accuracy of the discussed lidar instruments, several experimental results obtained using calibrated depolarization lidar instruments from different EARLINET stations are presented and discussed."

(16,24) I understand that the presentation of particle depolarization is functional to the aim of showing ": : : the importance of: : :" (see above), but again I think

this is not the main message the paper is delivering. Moreover, as correctly stated, the computation of particle depolarization is affected by uncertainties on the aerosol backscatter coefficient and this is especially true in the case of low aerosol loading, as correctly stated in (17, 14-16). This should open a completely new and wide discussion which is clearly beyond the scope of the article, which is on calibration procedures: the effect of these may be heavily masked by other sources of inaccuracies in the computation of particle depolarization.
In synthesis, I would ask to rewrite sec. 4.4, presenting calibrated and uncalibrated profiles on selected case studies, or to drop it entirely.
The purpose of section 4.4 is to provide the support to make a first estimate on the depolarization accuracy at 532nm.
To provide a full set of uncalibrated and calibrated profiles, additional information is required: the correction of the alpha angle must be performed in the post processing (applies only to Bucharest and Granada for several case studies) and the diattenuation values for the receiving optics should be significant (does not apply to Munich and Potenza). One example of calibrated and uncalibrated profiles should be sufficient to present the accuracy of calibrated depolarization products. We would like to keep this section as it is but if the referee considers this a key factor, additional case studies could be included for Bucharest and Granada stations.

(18,24-25) This is a very nice result which I think is understated, as the comparison of the observations in regions supposedly free of aerosol, with the theoretical values of the molecular depolarization is the key factor to assess the goodness of the calibration procedures hereby described. The authors may consider to add a table reporting the values of "low aerosol height range values", vs the molecular depolarization as expected from theory. The bandwidth of the interferential filter should of course be also quoted, as it impacts that value. Incidentally, it might be quoted (18,27) that also the presence of small amount of liquid aerosol may impact the profile, in a different direction and to a smaller extent.
This type of study would require a statistical approach since the aerosol free region would be influenced by several factors. For a comprehensive approach, extensive studies must be performed for each lidar instrument:

> Freudenthaler, V., Seefeldner, M., Gross, S. and Wandinger, U.: Accuracy of linear depolarisation ratios in clean air ranges measured with POLIS-6 at 355 and 532 nm, in EPJ Web of Conferences, vol. 119, p. 25013, EDP Sciences. [online], 2016.

The effect of the interferential filters should be much lower that the first estimate on the depolarization accuracy provided in this manuscript. Also this approach would add an additional complexity level to the manuscript. Since the aim of the study is to present experimental techniques for the calibration of lidar depolarization channels in EARLINET we consider that an additional complexity level is not required. Still, if the referee considers that this topic is essential to the manuscript, we will perform the required updates in the next iteration.