General comments

The paper shows comparison between the standard automatic CALIPSO extinction retrieval (level 2) and a different retrieval based on the combination of CALIPSO and AERONET data. The main objective is demonstrating that the C+A retrieval gives more accurate results during daytime respect to the standard ones. The method is not original and new, but if assessed, this would be of great interest for the community because AERONET is widely distributed world-wide and moreover the same retrieval could also use other AOD data as MODIS ones.

For demonstrating this method advantages, the authors compare the two retrievals against ground-based lidar measurements. The ground based lidar used in this work is a Raman lidar, but only its elastic capability is used here because limited to daytime measurements. So that measurements that are considered as reference in this evaluation (RSLAB) are actually affected by a large uncertainty because of the lidar ratio assumption. The RSLAB extinction profiles are obtained using the same method applied for C+A: forcing the lidar profile to result in a columnar optical depth as provided by AERONET using a lidar ratio constant with the altitude. RSLAB and C+A extinction profiles are therefore based on the same assumptions: S constant with the altitude and AOD equal to the AERONET one. Because of this, I would expect better agreement between these 2 retrievals rather than between CALIPSO lev 2 and RSLAB. Actually RSLAB and C+A profiles are not completely independent.

However I do not see significant improvement of the agreement against RSLAB data: Table 6 shows similar results for the 2 retrievals when compared to RSLAB. More interesting is the analysis of dust cases even if, it should be supported by all the info available from the Raman lidar measurements available for Barcelona About dust cases, the authors state that they "noticed relevant differences between cases where air masses came from Northern of Africa and the rest of the cases". This, I think, is related to the altitude-independent S assumption in C+A retrieval: this assumption can be suitable if there are not intense lofted layers, but it is probably far from the truth when a significant aerosol load is present above the PBL. More suitable is the CALIPSO lev 2 approach in these cases: different S for different aerosol types, even if the assumed values could be improved.

In synthesis, the idea of combining CALIPSO and distributed AERONET data is great, but its performances should be better analyzed and major revisions are needed. Authors have the opportunity to use their nighttime measurements as reference for more reasonable assumptions on S profile (no constant values) for the daytime observations. These improved daytime RSLAB measurements would be independent from the C+A

As the referee points out, the assumption of a constant S in the whole profile is the main limitation of the C+A method, and it unavoidably introduces some error in the extinction profiles. Regarding to this, CALIOP algorithms do not make such assumption, which is an advantage.

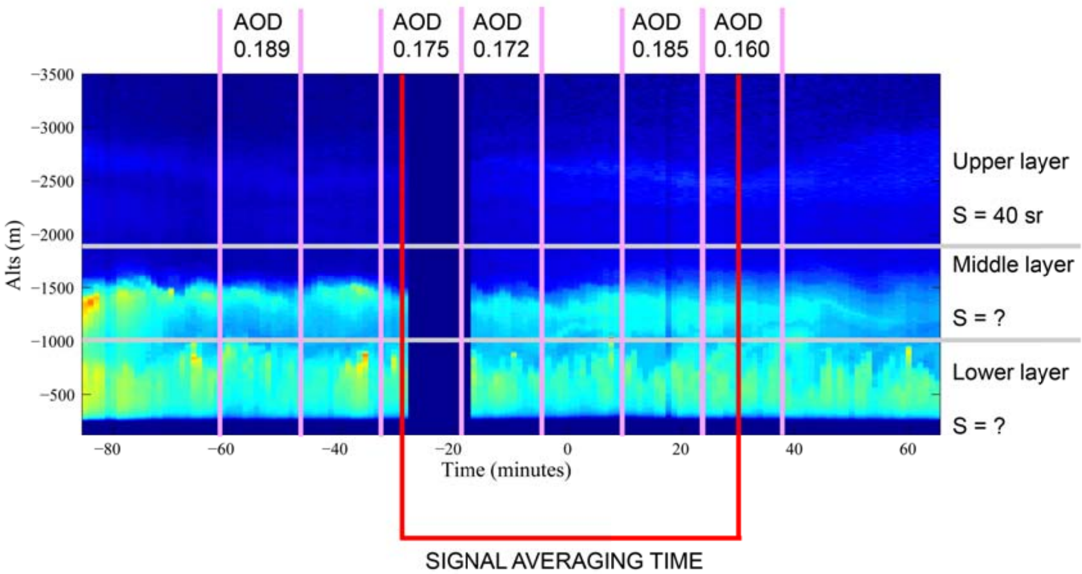
In order to retrieve aerosol extinction profiles, CALIOP algorithms need to classify the aerosols in the atmosphere, and then associate a certain S value to these aerosols. Thus, the accuracy of the level 2 profiles is limited by the CALIOP capability of classifying aerosols, which is highly affected by background noise in daytime cases, and by the choice of S, which can fail for some kind of aerosols. In section 5.3 we showed that the AOD obtained with this method meaningfully differs from ground-based photometer measurements.

One of the objectives of this work is to see if, despite the assumption of a constant S, the C+A method could offer better results than the level 2 data. In order to do this, we decided to make a comparison with a ground-based lidar. However, as the referee says, in the inversion of the ground-based data we had to make the same assumptions that we made in the C+A method. In this way, we could study the differences between RSLab and C+A caused by space-time aerosol variability, but we could not study the error that we introduced when assuming a constant S. That is why we restricted the comparison with Level 2 to the case b, where a constant S is also found by CALIOP algorithms.

The referee makes an interesting suggestion that could make the ground-based data independent from C+A: "*use their nighttime measurements as reference for more reasonable assumptions on S profile*". The use of the Raman results was discarded in the previous version because the possibility of using nighttime measurements of the lidar ratio for daytime cases is limited by the different aerosol conditions, especially in the planetary boundary layer, since we would expect more homogenous conditions during daytime due to a higher mixing. Moreover, changes in the relative humidity might also affect the values of S, especially during winter. In the work by Gross et al. (2011)[1], the nighttime lidar ratio is used for the inversion of daytime measurements of the PBL, but only in cases where atmospheric stability is guaranteed by continuous lidar measurements. In our case, however, these measurements were not available.

Nevertheless, since the "uncoupling" of the C+A and RSLab results would improve our comparison, we searched for new ways of inverting ground-based data. If we look at the three cases, we can see that most of the aerosols lay within the PBL, but that some layers can be found in the free troposphere, especially in cases a and c, which correspond to summer. These kinds of layers are commonly found in summer between the PBL and 3000m, and are mainly caused by recirculation processes. Studies about these layers have been carried out by Sicard et al. (2011) and Millán et al. (1997)[2], both finding mean lidar ratio values between 40 and 43 sr within them. Since these layers lay above the PBL, they are expected not to be so affected by daytime-nighttime changes, and thus using the nighttime values of their lidar ratio in daytime measurements would be associated to a smaller uncertainty.

For the PBL we studied the lidar continuous measurements close to the CALIPSO overpass to see the aerosol layer structure. For example, for case a, the RSLab range corrected signal (PR^2) shows this situation:



We classified the atmosphere into three layers, according to the PR^2 values: an upper layer above 1800m, corresponding to a recirculation layer with an associated S of 40 sr, a middle layer between 1000 and 1800m, and a lower layer below 1000m, both with unknown S. The time in this figure is given relative to the CALIPSO overpass, and the red lines show the 1-hour averaging period.

During these continuous lidar observation, there are single AOD measurements available from the sunphotometer (pink lines), which we consider to be representative of the 15-minutes (the sunphotometer measuring frequency) average RSLab signal surrounding them. We choose the S values for the lower and middle layer so that we obtain, taking into account all 15-minutes periods with available sunphotometer measurements, a mean absolute difference between the RSLab-AOD and the sunphotometer-AOD lower than 0.015. Finally, the pairs of values of S that fulfill that condition are averaged in order to get the final S values. Using this methodology we obtained a S of 29±7 sr for the lower layer and a S of 56±14 sr for the middle layer. In order to make the inversion for all altitudes, we considered a well-mixed atmosphere for the first hundreds of meters, which enable us to assume a constant aerosol backscatter equal to the first aerosol backscatter coefficient above the overlap range for altitudes below 500m.

Calculating the RSLab extinction profiles this way enable us to extend the comparison with level 2 data for all cases. This method has also been applied to the other two cases in the revised version of the manuscript in a similar way.

Regarding to the desert dust cases, we agree with the referee when he/she estates that the assumption of the constant S is associated to larger errors when aerosol layers with different properties coexist in the atmosphere. In order to estimate this effect we already included at the end of section 5.4 of the previous version a brief study of the differences between the extinction profiles of C+A and level 2 for a dust case where two clearly different aerosol layers were found in the atmosphere. In this case big differences were noticed in the AOD given by each method, but not in the shape of the extinction profiles.

It is also clear that the differences seen in Fig. 14 are largely influenced by the constant S assumption. However, the fact that aerosols can be found above 4.5km for desert dust outbreaks but not for the rest of the cases does not depend on S, but merely on CALIOP signal. In the revised version of the manuscript we have made a deeper study of the desert dust cases, in order to get a better understanding of the differences between C+A and level 2 data and the effect of the choice of S. Raman results for desert dust outbreaks over Barcelona by Reba (2010) has also been added.

1: GROß, S., GASTEIGER, J., FREUDENTHALER, V., WIEGNER, M., GEIß, A., SCHLADITZ, A., TOLEDANO, C., KANDLER, K., TESCHE, M., ANSMANN, A. and WIEDENSOHLER, A. (2011), Characterization of the planetary boundary layer during SAMUM-2 by means of lidar measurements. Tellus B, 63: 695–705. doi: 10.1111/j.1600-0889.2011.00557.x

2: Millán, M. M., R. Salvador, E. Mantilla, and G. Kallos (1997), Photooxidant dynamics in the Mediterranean basin in summer: Results from European research projects, J. Geophys. Res., 102(D7), 8811–8823, doi:10.1029/96JD03610.

Pp 3984: as referee 1 noted, info about vertical distribution of aerosol is anyhow available from CALIPSO even if not combined with AERONET data.

As the referee says, CALIOP measurements are still available over AERONET sites whether the C+A method is applied or not. However, since the described method is intended to reduce the uncertainty in the extinction profiles, the synergetic combination of data would lead to more accurate results that, in principle, are not available only with CALIOP data.

Pp 3986lines 21-pp 3987line 11: it is not clear which kind of CALIPSO data are used here.

The data used by Kim et al. (2008) is level 1 data (versions 1.10 and 1.11); while the data used by Chiang et al. (2011) is also level 1 data (attenuated backscatter). It has been pointed out in the revised version of the manuscript.

pp3990 line 19: this quantity depends on the place, period of the year and conditions. It should be stated clearly that this is a typical / average value and for what region/period is obtained.

A deeper explanation of the averaging criterion is made in the revised version, taking into account these remarks:

"The criterion we used to choose the averaging area is based on the one described by Ichoku et al. (2002), who set that MODIS measurements averaged in a 50×50 km window can be compared to the hourly averaged measurements of a sun photometer centered in that window. To get to this window size, Ichoku et al. (2002) found that, for a certain number of reference points, the mean AOD from MODIS did not show significant dependence on the averaging area between 30×30 km and 90×90 km windows. Moreover, they also estimated that mean speed of an aerosol front is about 50 km/h, according to measurements of the aerosol fronts travelling from Africa to America for July to September 1988, as measured by the TOMS (Total Ozone Mapping Spectrometer) sensor. This averaging scale is similar to the one used by the CALIOP algorithms to detect aerosol layers."

Pp 3992: equ. 4: the error 3 is a maximum error and cannot be combined with a statistical error in this way. The final maximum error should be 3 times the instrumental error plus the error of equ 3

As the referee points out, equation 3 can be used to define the maximum error for a group of measurements. However, in our case it has been used as an estimation of the uncertainty in the AOD caused by the aerosol variability during the average period.

As it is already said in the previous version, we used this expression because the maximum number of AOD measurements we can find in an hour is only four, and thus using the standard deviation would not be appropriated.

Strictly speaking, if we consider that the uncertainty represents the 68.2% confidence interval (integrated probability within 1 standard deviation range from the mean in a normal distribution) for a limited number of measurements, we should use the Student's t formula:

$$\varepsilon = \mathrm{A}\frac{S_n}{\sqrt{n}}$$

Where Sn is the square root of the variance of the sample, n is the number of measurements and A is the Student t-value. For a sample with 4 measurements and an 68.2% confidence interval we get a Student t-value of 1.19.

The Student-t distribution instead of the normal distribution has been used in the revised version of the manuscript in the estimation of the uncertainties caused by the aerosol variability in order to take into account the limited number of measurements of AOD for each case.

Pp3993 and here after: S instead of LR would be preferred. It is more widely used in the community.

It is changed in the revised version of the manuscript.

Pp 3996line 24: actually the agreement is better with lev2.

A deeper explanation has been done in the revised version of the manuscript:

"*Finally, the mean S values obtained in Barcelona, both for C+A (66 ± 12 sr) and level 2 data (55 ± 13 sr), concur, within the margin of error, with the mean annual value of 55.5 sr retrieved by Sicard et al. (2011). However, we can see that the level 2 values are closer to those obtained by the C+A method, which are around 10 sr higher. These higher values have also been seen in the comparison against RSLab, showing that the C+A method might tend to overestimate the S values*."

Pp3997, lines 12-15: here you are using results from C+A for the RSLAB inversion. So that they are based on 2 common assumptions (see general comments) and RSLAB is dependent in some way on C+A

This has been changed in the revised version of the manuscript as it is explained in the answers to the general comments.

Section 5.2: not clear, within the text, RMSD on which parameter is evaluated.

The root-mean-square deviation (RMSD) is evaluated on the aerosol extinction coefficients (σ), and is calculated as:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{z=overlap}^{rayleigh}\left(\sigma^{a}_{RSLab}(z)-\sigma^{a}_{C+A}(z)\right)^2}{N}}$$

where N is the number of extinction coefficients estimated between the overlap height and the Rayleigh reference height used for the RSLab inversion.

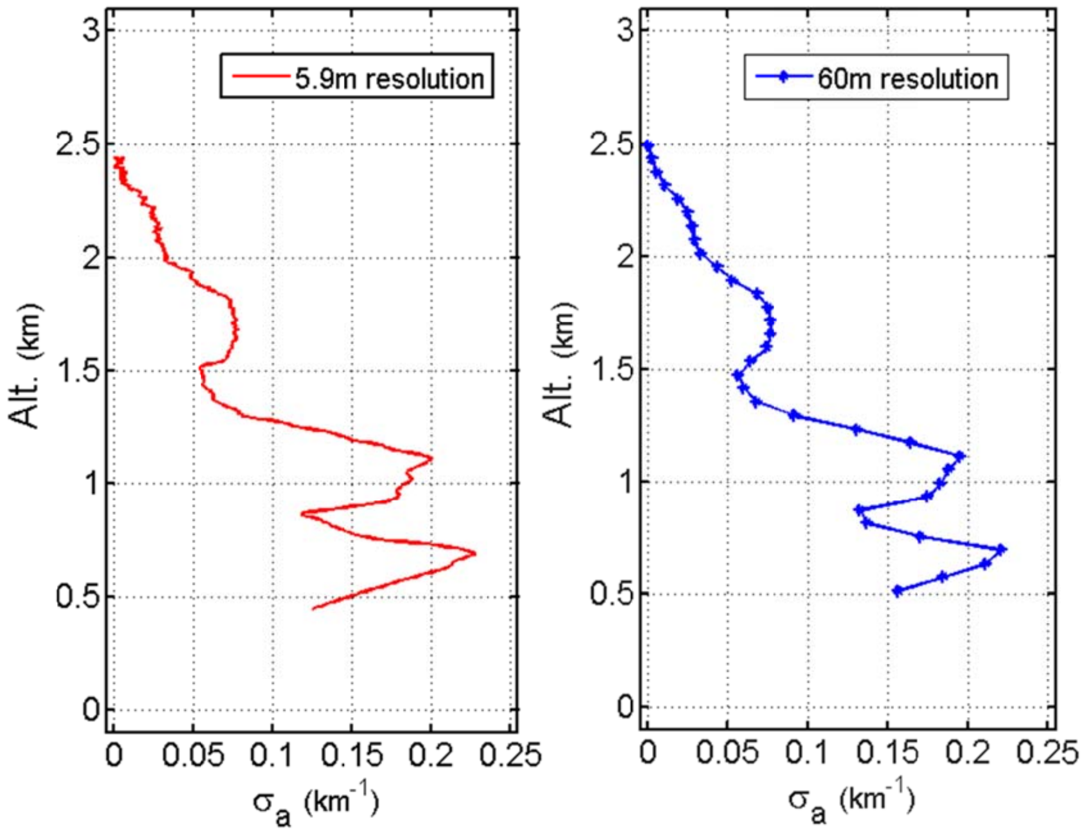
The results should be discussed respect to the Raman Barcelona measurements available for the same period, if they are, or from the values collected at that station so far.

Data obtained from Raman measurements by Sicard et al. (2011) and Reba (2010) at Barcelona have been compared in the revised version of the manuscript to our results. As it has been said in the answer to the general comments, Raman measurements of the tropospheric layers have been also used in the new inversion of the daytime RSLab data.

All RSLAB profiles have clearly a larger than 60m effective resolution. This should be discussed comparing small structure and values

As it is said in the previous version, the raw effective vertical resolution of the RSLAB data is 5.9 meters. The signal is averaged during one hour centered in the CALIPSO overpass and then the lidar inversion is made. Finally, the resolution of the inversion profile is reduced to 60 meters based on the altitude array given by the CALIOP level 2 metadata.

Here we show the effect of the change in the vertical resolution for case c:



We can see that although we lose small details in the profile, the main structures are not meaningfully affected by the change in the resolution.

Values of linear fit should be reported in text and tables (intercept and slope). Maybe weighted fit it would be more suitable

These remarks are taken into account in the revised version.

pp 4000: Figure 10 instead of 6.

It is corrected in the revised version.

Pp 4001: lines1-2: this better agreement is on the base of R^2? Actually in this case the slope is different from 1, so the agreement/disagreement should be better discussed.

The parameters used to analyze the agreement between level 2 data and AERONET measurements are the mean difference (MD), root-mean-square deviation (RMSD) and correlation coefficient (R^2). As the referee points out, the slope of the linear fit, as well as the intercept, also offers valuable information about the goodness of the comparison. In the revised version of the manuscript these parameters are also included in the discussion.

Pp 4002, 2nd paragraph: the correlation analysis between S profiles case-by-case sounds really strange: they are constant with the altitude (C+A) and constant within the layers (lev 2) profiles.

The correlation study of the lidar ratio cannot be made within the profiles, because we cannot obtain lidar ratio profiles with the C+A method, as the referee remarks.

In this study we compared the values of S obtained by C+A for each overpass with the mean values of S estimated from level 2 data as:

$$\overline{S}_{L2} = \frac{\sum_{i=1}^{N} \sigma_i^a}{\sum_{i=1}^{N} \beta_i^a}$$

as it is pointed out in the previous version. Thus, this is a comparison between the values of S estimated with C+A versus the mean values of S obtained from level 2 data for each case, similar to the comparison made for the AOD.

We clarify this point in the corrected version of the manuscript.

Pp4003: figure 16 shows data for a different day. Check it.

Checked. It is corrected in the revised version.

Pp 4003-4004: see general comments about dust cases

Answer can be found in the general comments.

Pp4004 line 23: three cases are not sufficient for proposing a methodology

In this case, we used the word “proposed” as a synonym of “presented”. As the referee points, the number of available cases for the comparison with ground-based data is limited in order to state that the C+A method has a better performance than the Level 2 data. Moreover, in the three studied cases the results given by level 2 data are not significantly different to those obtained from C+A, which is can be explained by the fact that similar values of AOD were found by AERONET and CALIOP due to a good choice of the lidar ratios by the Level 2 algorithms. However, this is not expected to happen when values between AERONET and CALIOP differ in a meaningful way, which has been shown to be the typical situation in section 5.3.

Finally, we do think that the new inversion methodology described in the answer to the general comments will allow us to assess the limitations of the C+A method, as well as to find out the conditions in which its best performance is expected. That is why we have included a deeper comparison study of C+A and Level 2 data against RSLab results in the revised version of the manuscript.

Conclusion section: all this section should be rewritten on the light of considerations reported in the general comments.

The conclusions have been rewritten taking into account the remarks of the referee. Special emphasis has been made in the limitations of the C+A method due to the choice of a constant S, and the effect that this choice has in the differences with level 2 data.

Pp 4005, line 5: a correlation for r2 of 0.65 is very low.

It is, indeed. However, if we compare this value with the ones obtained in the comparison for Burjassot cases - 0.5 - or the results by Bréon et al. (2011) – 0.4 – the value of 0.65 can be considered to show a better correlation. Anyway, other statistical parameters are also used in the revised version to state the better performance of L2 data over Barcelona.

Figure 12: this figure is too briefly discussed within the text. AOD is one of the key parameter for aerosol studies. Differences between the CALIOP retrieval and the AERONET measurements should be discussed in more details.

The results of the comparisons between AERONET and CALIOP (both ours and the ones obtained by other authors) are presented as one of the main motivations for this work. However, as the referee says, results of section 5.3 regarding the AOD can be studied in a deeper way. A critical discussion of the causes of the great differences between AERONET and CALIOP has been added to the revised version of the manuscript.