General comments.

The main objective of this paper is the improvement of daytime CALIOP retrieval that are limited by low SNR. The second objective is "to supplement column integrated measurements from AERONET sun photometers with information about the vertical distribution of aerosols". About this second objective, I don't see it as a strictly separate achievement: it seems a little tautological, to some extent, because those AERONET sites will anyhow have aerosol profiles during CALIOP overpasses. The method proposed in the paper, is based on coupling the CALIOP data with the AOD from ground stations. Because AOD data from the AERONET network are quality assured and also because the network covers a growing portion of the planet, it's of interest if those data can give an improvement on lidar profiles from space. This approach is not entirely new, although the authors shows a long record of measurements and then some statistic are also presented. They compared their results with the CALIOP L2 extinction profiles, but probably somewhere should be remembered that, strictly speaking, is the aerosol backscatter the main (vertical resolved) parameter that can be extracted from a backscatter lidar without scanning capabilities. The overall methodology adopted for data processing appears correct, although not always the results obtained are sufficiently commented and some revision should be made also in Tables and Figures. Sometimes claims on the performances of the C+A method seems questionable and not entirely convincing. Generally there is no strong evidences that the C+A results are inherently better than the L2 retrieval (see also Specific and Technical comments).

When (and if) there are the conditions that makes it meaningful, an improvement would be the comparison of the C+A results with the nighttime products from Barcelona Raman lidar. The Saharan dust case is interesting because, from the presented results and referenced literature, seems that the aerosols classification adopted by CALIOP is less accurate, at least for this particular site. Only one of those Saharan dust case is reported: I think that a more complete analysis of all cases could reinforce, if confirmed, that point.

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.

Regarding to the results of the comparison with ground-based data, the referee is right when he/she points out that the results for C+A and Level 2 data are fairly similar in the three studied cases. Claims about the better performance of C+A against Level 2 data in the studied cases will be re-written so that it does not seem such a categorical statement, although the goodness of the results obtained from Level 2 data have been already remarked in the previous version. Anyway, we have to bear in mind that in these three cases the performance of the Level 2 data is remarkably good compared to the rest of the overpasses, as it is shown in section 5.3, and thus these similar results are not expected to be found when the Level 2 AOD meaningfully differs from the AERONET AOD.

Comparing the C+A method with nighttime Raman measurements is limited by several factors. We cannot use the sun-photometer data, and although we could just use the AOD obtained from the Raman inversion, the closest CALIPSO nighttime overpasses over Barcelona lay outside the defined averaging area (the minimum distance is around 40 km), and significant deviation due to spatial variations could occur.

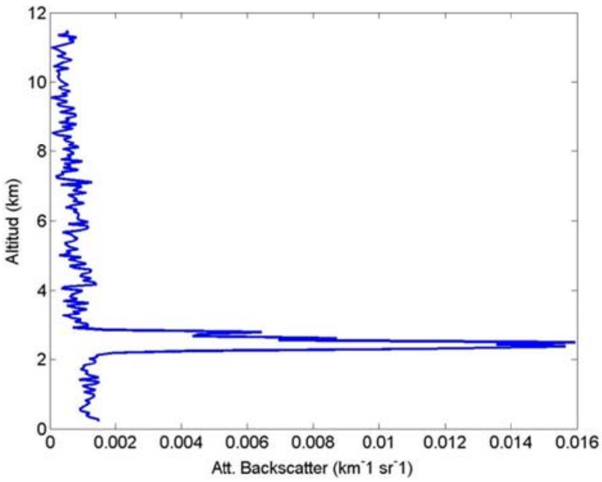
Finally, a deeper study has been carried out for the desert dust cases and it has been included in the revised manuscript.

Specific comments:

p3991, r4-8. Have you verified somehow how good is the cloud screening criterion adopted?

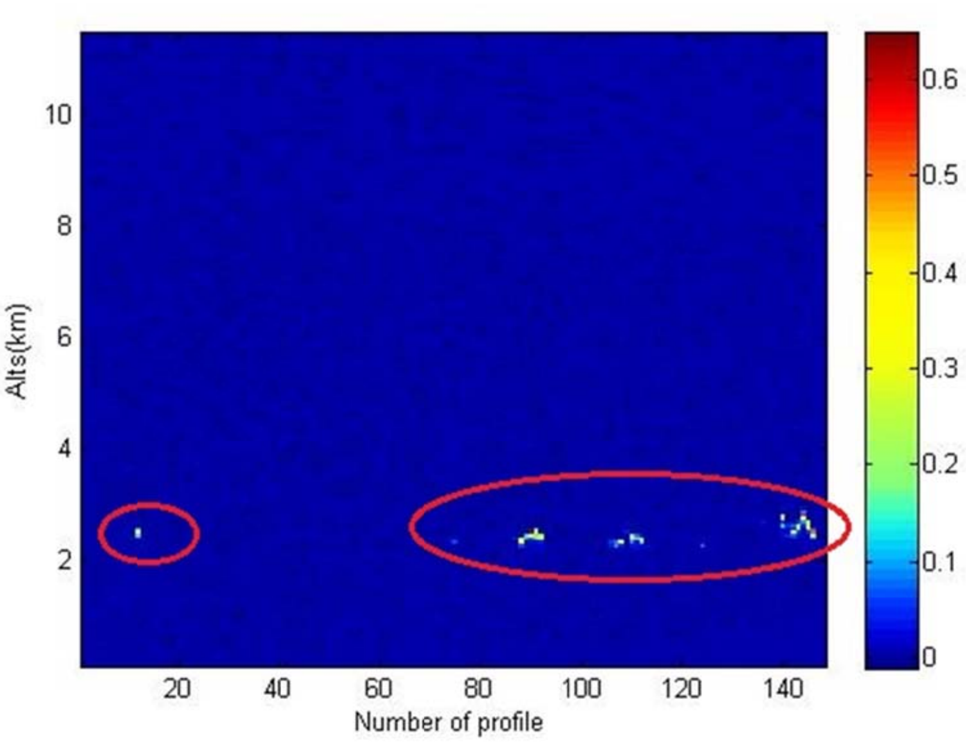
Yes, all cases were "manually" checked. We found that the automatic cloud screening worked fine for most of the days. However, in some cases it was not able detect all the clouds.

For example, for the 2008/04/18 overpass over Burjassot, the averaged Total Attenuated Backscatter profile we got after applying the cloud screening was this:

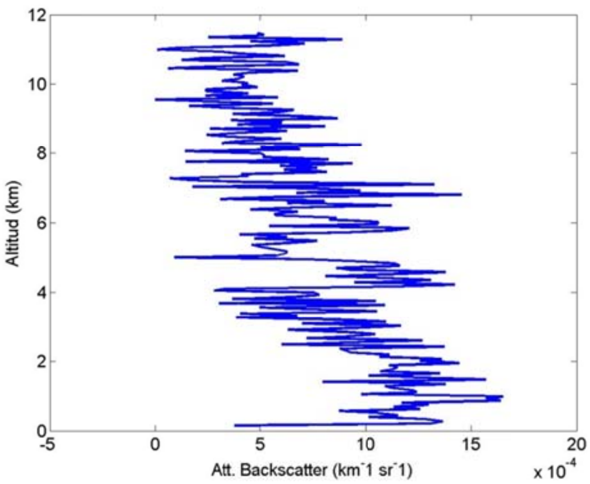


We could see a sharp variation in the TAB at 2 km, which was unlikely to be produced by aerosols. When this kind of sharp variations appeared, we checked the AQUA (also in A-Train) quick response image to see if these structures could be produced by the presence of undetected clouds. The AQUA quick response image for this case is this:

We see the presence of small clouds inside the averaging area (red circle). In this kind of cases we manually deleted the contaminated profiles. In order to do that, we represented an image of the TAB we got after applying the automatic cloud screening. For the 2008/04/18 case, we got this situation:



This way we could easily find the undetected clouds (surrounded by the red circle) and erase the profiles containing them. After clearing the cloudy profiles, the averaged TAB we got is this:

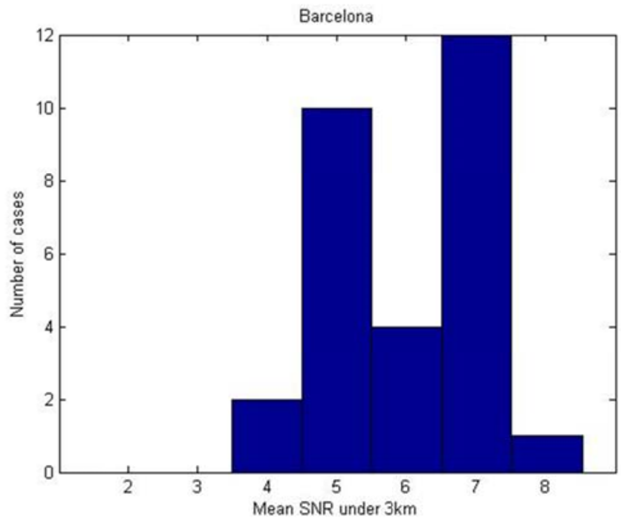


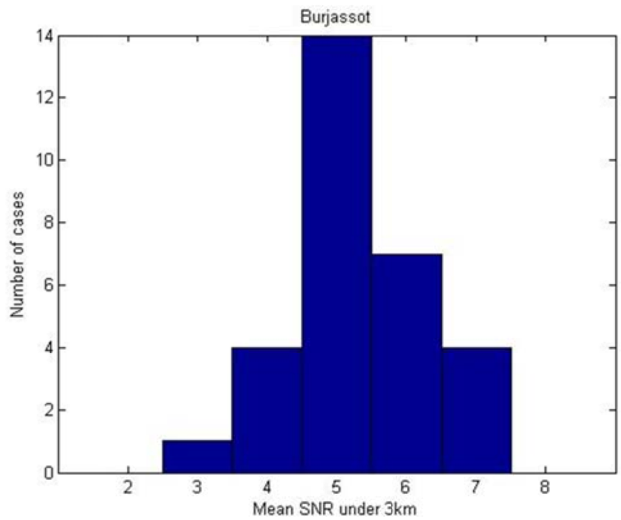
The fact that some cases had to be revised has been pointed out in the revised version of the manuscript.

p3990, r10; p3991, r11. You describe in details the criterion adopted for the averaging area, but no numbers are given about the acceptable SNR taken as reference.

We considered the number of averaged profiles (usually more than 100) enough to get a fair SNR. For example, in the work by Royer et al. (2010) they average between 150 and 300 daytime profiles to get a good enough to be combined with MODIS data. In our case, however, we have to use less profiles in order to limit aerosol spatial variability.

We have calculated the SNR at a certain altitude as the Total Attenuated Backscatter (TAB) divided by the estimated uncertainty. If we take into account all cases, we get a mean SNR of 6±1 for altitudes under 3km. These are two histograms showing the distribution of the mean SNR for altitudes under 3km for the Barcelona and Burjassot sites:





We can see that in most of the cases, a SNR greater than 4.5 is obtained. However, although lower values of SNR have been found, we can use the estimated uncertainty to see how reliable are the extinction profiles found for each case.

p3992, r1. What is impact of the uncertainty on Angstrom exponent on the overall uncertainties?

We estimated that if we take an instrumental uncertainty of 0.015, using the Angstrom interpolation would increase this uncertainty, at least, by 20%. This is not expected to affect the uncertainty of the extinction profiles in a meaningful way, since the mayor source of error is caused by the signal noise, as it has been shown in section 4. However, the effect on the uncertainty of the lidar ratio is not negligible. If we take into account the results in section 4, the uncertainty on the Angstrom exponent can increase the uncertainty of the lidar ratio up to 15%.

P3996 r19-25. Could the impact of the 70sr limit on L2 retrieval be assed? You know the LR used by you and the ones used by CALIOP in each case.

The impact of the 70sr limit can be assed comparing the results obtained when the C+A method retrieves LR values over and under that value. The results of the comparison are shown in the following table:

Cases with LR > 70 sr		
N of cases for comparison		12
	AERONET	L2
Mean AOD	0.145	0.092
Mean LR	89 sr	57 sr
Cases with LR <= 70 sr		
N of cases for comparison		33
	AERONET	L2
Mean AOD	0.199	0.183
Mean LR	57 sr	50 sr

We can see that for cases with LR > 70 sr, the AERONET mean AOD is 1.6 times bigger than the L2 mean AOD. However, for cases with LR <= 70 sr this relation is reduced to 1.1.

We also made a comparison of the AOD obtained by AERONET and L2 data (similar to the one done in the work), for cases with any LR and cases only with LR < 70sr.

	All LR	LR<70sr
$\overline{AOD}_{AERONET}$	0.18	0.21
$\overline{AOD}_{L2}$	0.16	0.20
MD	0.02	0.015
RMSD	0.09	0.09
R^2	0.5	0.5

As we can see, the only statistical parameter meaningfully affected when we discard cases with LR > 70 sr is the Mean Difference.

P3996, r25: Anyway seems that the annual LR value by Sicard (2011) concur with L2 better.

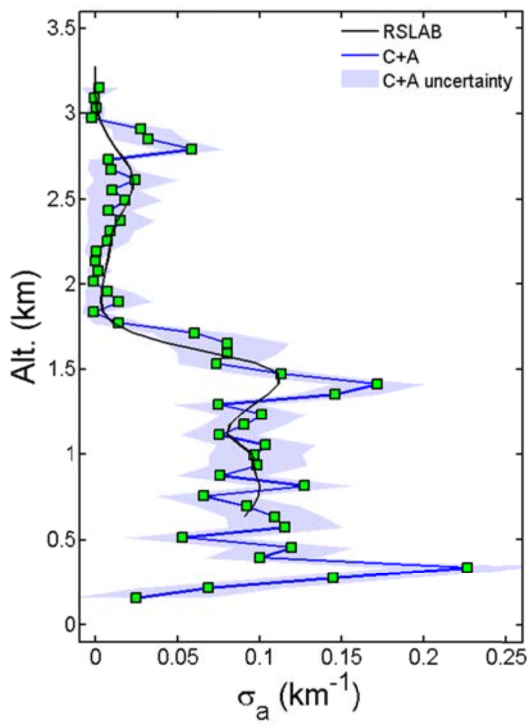
A deeper explanation has been done in the revised version of the manuscript:

"*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*."

P3997, r16-17. Especially in Fig.4 and Fig.5, seems that lidar profiles exhibits details not less than 400/500m, so apparently the resolution seems coarser than 60m.

The resolution of the level 1 products is 30m, which is changed to 60m to match the level 2 data, no further changes are performed in the resolution. The fact that the resolution of the profiles in the figure appear to be coarser might be caused by several factors: the dashed line makes the profiles look "smoother", the Total Attenuated Backscatter (TAB) given at a certain altitude has a non-zero correlation with the upper and lower TAB measurements, or the presence of actual aerosol layers. A further explanation of this fact is also given in a following answer.

Here we show a new representation of Fig. 4



Every green dot shows the location of an extinction coefficient measurement.

Fig.4 and Fig.6. Are you confident about the very first hundred meters of the C+A profiles?

We have used the “Surface Elevation” product to remove the data corresponding to non-atmospheric backscatter and we have also discarded all CALIOP data within the first 150m above the ground level, in order to avoid surface effects.

Due to the orography near the Barcelona site, the number of profiles used to average the signal in the very first hundred meters (<600m) is lower than the number used for higher altitudes. However, this is taken into account when we calculate the signal uncertainty.

It is expected that the lower altitudes are more influenced by the local conditions, and that aerosol variability within them might be higher, but we cannot check the accuracy of the C+A method under 500 m due to the RSLAB lidar system limitations.

P3997, r25. Here and also in other pages of the paper. I think that too much emphasis is posed to weak profile details that have a relative uncertainties of about 100 percent (as seems to be from vertical profile viewgraph). The CALIOP L2 algorithm could have cut those details also because the too high error.

We emphasized these facts because could see that these layers actually existed, thanks to the RSLab lidar. Moreover, these results match those obtained by Sheridan et al. (2012) who found that that for daytime conditions CALIOP algorithms are unable to detect aerosols with σ^a <0.02 km^{-1} in a 50% of the cases.

In our work we show that those weak aerosols layers can be detected using level 1 CALIOP signal (although an accurate estimation of their extinction cannot be done) and that the detection limit of the level 2 data is partly conditioned by the CALIOP algorithm, as the referee points out.

In the revised version of the manuscript we have indicated the large uncertainties found for the weak layers:

“*C+A has also been able to detect weak aerosol layers (σ^a < 0.025km-1, as measured by RSLab), although accuracy in the determination of σ^a within them is limited due to relative uncertainties around 100%*”

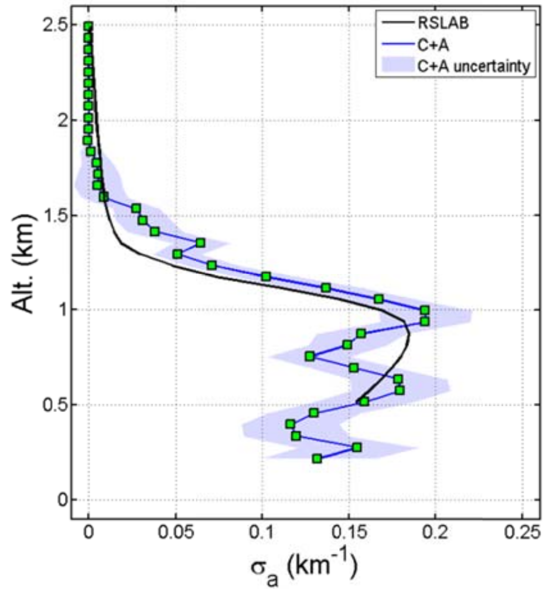
Sect 5.2.1. Significant differences on parameters are simply stated and not critically commented. Why linear regression results are not inserted for RSLAB vs. CALIOP L2?

In section 5.2.1 a description of each case has been done. However, as the referee remarks, critical comments should be added to this descriptions.

For example, for all cases we see that the C+A method estimates a bigger lidar ratio (between 10 and 15 sr). This, roughly speaking, means that CALIOP “sees” fewer aerosols than RSLab, and thus more extinctive properties have to be given to them in order to reach the same AOD. That can be explained by the space-time aerosol variability: the actual time-averaged aerosol load over the RSLab (and photometer) might be bigger than the actual space-averaged aerosol load under CALIPSO overpass. We can also see that the lidar ratios obtained from RSLab lay outside the calculated margin of uncertainty by C+A, indicating an underestimation of the aerosol variability.

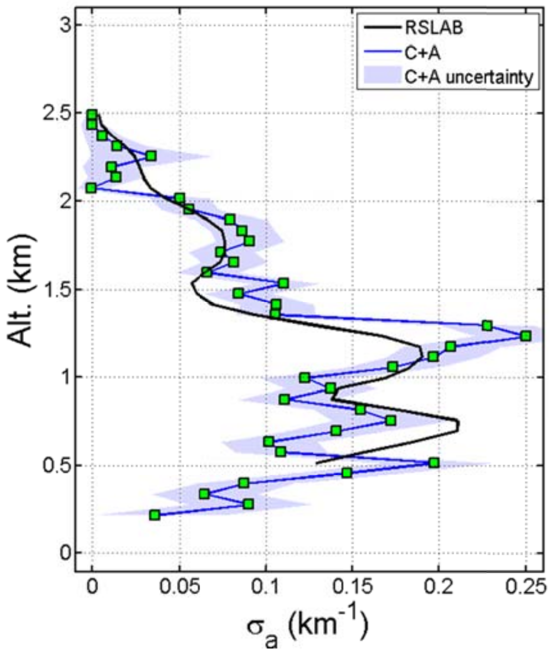
Regarding to the profiles, for case *a* we can see oscillations in the extinction coefficients under 1.5 km, which have amplitude between 0.02 and 0.05 km^{-1}, and a spatial scale limited by the profile resolution. This can be explained by the presence of signal random noise, which does not allow us to detect small variations in the extinction profiles that can be seen by RSLab.

If we look to case b:



we can also see oscillations within the boundary layer. However, these oscillations have bigger amplitude than in case a, and also a larger spatial scale than the resolution of the profile. This means that in addition to noise, the C+A results at these altitudes are also affected by the spatial aerosol variability, which does not correspond to aerosol time variability over the RSLab, at least between 0.5 and 1 km where a smooth and continuous decrease in σ^a is seen.

Finally, in case c we can see a more complex shape in the extinction profile.



This time, the presence of different layers, although with different values of the extinction coefficients, has been detected by both methods. We can see, from the study of these three cases, that the validity of the averaging criterion can be limited, especially for the lower layers, when the time variability of aerosols does not correspond to an equivalent special variability.

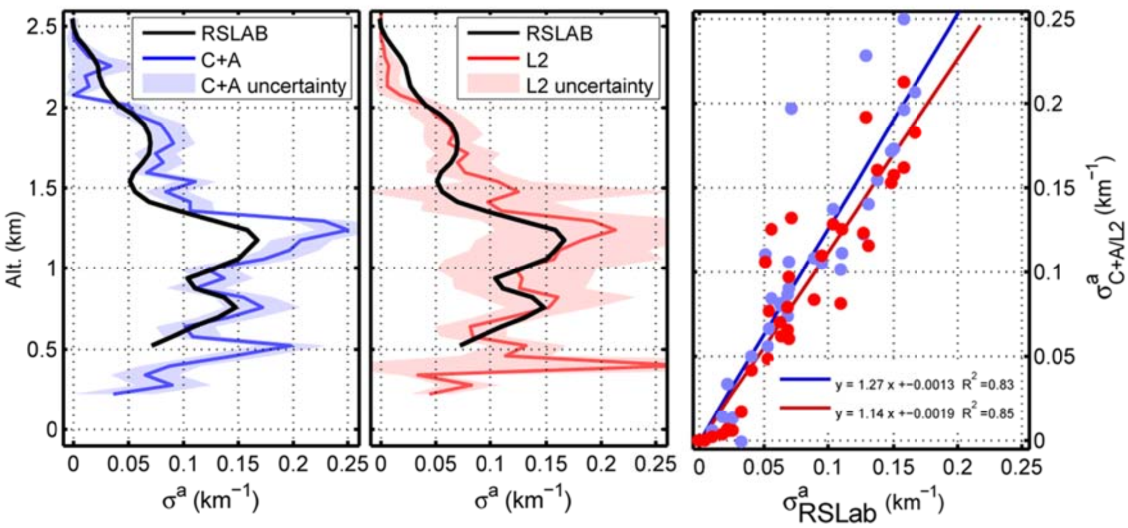
The importance that aerosol variability may have in the values of the lidar ratio and the shape of the profiles, as well as the limitation of seeing smooth variations in the profile caused by the noise was not pointed out in the previous version. However, these facts can help us to understand the results in a better way and have been included in the revised version of the manuscript.

Finally, linear regression parameters between RSLab and L2 data have been added in the new tables and figures, as it is shown in the figure included in the following answer.

P3999, r14. Fig.8 should be Fig.7. The uncertainties on CALIOP L2 data should be inserted in Fig.7 as it was done for C+A vertical profiles in Fig 4-6. The lower part of the L2 profiles needs to be cut at some meaningful altitude. Could deep blue color on the aerosol classification figures gives hints on the lower cut for CALIOP L2 profiles?

Yes, the deep blue stands for altitudes below the surface level and they can be used as a reference to “cut” the profiles.

Fig. 7 has been changed and the comparisons of C+A and level 2 data have been joined in a single subsection (as it is suggested in one of the following remarks). An example of the new figures for case b is can be seen here:



Note: Some changes might be done between this figure and the one in the revised version

Sect 5.2.3. The better agreement of C+A in case B seems quite marginal: AOD is within 10 percent, LR is quite similar and I already said my opinion about weak details in the profile. Is not clear to me how strong AOD differences could directly explain the better R^2 in Barcelona respect to Kampur.

As the referee says, differences between C+A and level 2 data cannot be considered meaningful. This can be explained because of the very similar estimated values of the lidar ratio: 53 and 55sr respectively. As it has been said in the answer to the general comments, claims about the better performance of C+A against Level 2 data in the studied cases will be re-written so that it does not seem so categorical, although the goodness of the results obtained from Level 2 data have been already remarked in the previous version.

The strong differences in the AOD between Barcelona and Kampur would explain the discrepancies found in the root mean squared deviations (RMSD), not the correlation coefficient, as the referee says. The discrepancies in R^2 might be caused by the good performance of CALIOP L2 algorithms in the three studied cases, compared to the usual performance that have been seen when comparing AERONET to CALIOP.

Some parts of this section have been re-written for the new version in order to make them clearer:

"*For the extinction profiles, the results obtained can be compared to those present in the work by Misra et al. (2012), where a similar comparison is made over Kanpur, India. For four cases where CALIPSO overpass distance is lower than 25 km, Misra et al. (2012) find a mean RMSD of 0.18 km-1 and a mean R2 of 0.63 between level 2 data and a ground-based lidar plus a photometer. This is a poor performance compared to our results. The disagreement in the RMSD can be explained by the difference of the AOD values over Kanpur (over 0.5) and Barcelona (under 0.25) and); or by the particularly good performance of level 2 data in the three cases studied over Barcelona, compared to the average results, that could also explain the differences in R2.*"

Sect 5.3. In Fig.12 several (5 or 6 over 23 cases) CALIOP L2 AOD values at BUR site are quite low (<0.02) compared to corresponding AERONET BUR values (about 5-10 times higher). Can you give more insight on that? What about CALIOP L2 profiles in those cases? Those CALIOP L2 low AOD data have a non-marginal impact on some numbers in Tables for BUR site.

We have updated the figure in the revised version of the manuscript, so that all the points can be seen more clearly. In the old figure, some points were "duplicated" and they looked like two points, when they only represented one case.

As the referee says, there are two points where CALIOP AOD is quite low (<0.02) compared to the AERONET measurements (AOD > 0.11). These measurements correspond to 2008/01/29 and 2010/05/26.

These disagreements can be explained in terms of the applied methodology. In order to estimate the mean AOD given in a single overpass by CALIOP, we averaged the AOD given by the "Column_Optical_Depth_Aerosols_532" product in the level 2 profiles where no clouds were found. However, in some of those profiles CALIOP algorithms were unable to find aerosols, giving an AOD value of 0 that was included in the estimation of the mean.

We made a new calculation of the AOD values obtained when level 2 profiles with no aerosols were excluded. These are the results.

	Burjassot		Barcelona		Total	
$\overline{AOD}_{AERONET}$	0.19 (0.12)		0.17 (0.09)		0.18 (0.10)	
	Aerosol profiles	All profiles	Aerosol profiles	All profiles	Aerosol profiles	All profiles
$\overline{AOD}_{L2}$	0.14(0.10)	0.12 (0.09)	0.18(0.14)	0.17 (0.13)	0.16(0.12)	0.14 (0.11)
MD	0.05	0.06	-0.003	0.006	0.02	0.04
RMSD	0.09	0.11	0.08	0.07	0.09	0.09
R^2	0.5	0.5	0.7	0.7	0.5	0.5
Regression Slope	0.62	0.53	1.32	1.14	0.85	0.74

We can see that the AOD estimated when we only take into account profiles where aerosols were detected offers slightly better results than the previous estimations for the Burjassot site and when we take the overall data. However, no noticeable improvement has been seen for Barcelona.

For the Burjassot station, we get the following scatter plot:

We can see that the effect of the L2 low AOD values is partly corrected when we use the new method.

Since the mean level 2 aerosol extinction profiles are estimated using only profiles where aerosols were detected by CALIOP algorithms, this new method has been used in the corrected version of the manuscript, in order to unify the methodology.

P4004 r13. The relative vertical distribution differences are not that strong.

Although different values where estimated, a 70% against a 66%, the referee is right when he/she points out that these differences are not strong, especially if take into account the uncertainties of both methods. This fact has been remarked in the corrected version of the manuscript.

Technical comments:

The captions associated to Tables need to be more clear. Es. in Table 4, it's not clear what parameters are referred both RMSD and R2.

In table 4 the analyzed parameter is the aerosol extinction coefficient (σ^a). Based on this parameter, the root mean squared deviation (RMSD) 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 R^2 parameter is the square of the Pearson product-moment correlation coefficient estimated within the same range:

$$R^2 = \frac{\left[\sum_{z=overlap}^{rayleigh}\left(\sigma^a_{RSLab}(z)-\overline{\sigma^a_{RSLab}}\right)\left(\sigma^a_{C+A}(z)-\overline{\sigma^a_{C+A}}\right)\right]^2}{\sum_{z=overlap}^{rayleigh}\left(\sigma^a_{RSLab}(z)-\overline{\sigma^a_{RSLab}}\right)^2\sum_{z=overlap}^{rayleigh}\left(\sigma^a_{C+A}(z)-\overline{\sigma^a_{C+A}}\right)^2}$$

A brief description of how these terms have been calculated has been made at the beginning of the results section, similar to the one presented here.

Captions will be changed in the new version order to clarify the content of the tables.

Typo: The subscript N2 in all tables, probably should be L2.

Done.

For better reading the results, it would be helpful to put few grid lines in figures with vertical profiles.

Grid lines will be added to the plots in the new version as it has been shown in the figure for the new comparison.

p3992, r8. A typo, I think: 0.15 should be 0.015.

Done.

P3995, r2: in Sect. 3.2 you determine only the AOD (Did you meant Sect. 4?).

Done.

Sect 5.2. In my opinion is better to have a single paragraph for each case study vs. RSLAB, without splitting into C+A and CALIOP L2.

In the new version, the comparisons between RLab, C+A and L2 have been included in just one section.

P3998, r4. “between both cases”. It’s probably better write instead something like: “: : :between the aerosol extinction profiles: : :”.

Done.

P4000, r6. If I understand well, Fig.6 should be Fig.10 and so "obtained with RSLab and level 2 data" should be changed in "obtained with RSLab, level 2 and C+A data".

Done.