GENERAL
The paper provides an in-depth investigation of CALIPSO AOD retrievals for different aerosol types in comparison to AERONET data. Statistically significant biases are found for marine and dust aerosols. In the case of dust the discrepancies can be explained with the assumption of a too small lidar ratio in the CALIPSO retrieval scheme. Therefore, the authors further investigate the lidar ratio values of dust in different regions based on AERONET inversion results. They also provide an extensive discussion of the possible influences of dust mineralogy on the findings. The paper is well written and based on a very careful statistical analysis. It is a very interesting contribution to the field. However, to my opinion, there are some aspects which have not been adequately tackled in the discussion or have not been mentioned at all. The following issues should be considered in a revised version of the paper.

SPECIFIC COMMENTS
Discussion on the assumption of the dust lidar ratio in the CALIPSO retrieval has been ongoing in the community for several years (see, e.g., Pappalardo et al. 2010, J. Geophys. Res., 115, D00H19; Wandinger et al. 2010, Geophys. Res. Lett., 37, L10801) since a clear discrepancy was found between the CALIPSO a priori value of 40 sr and direct measurements with ground-based lidars giving values mainly in the range of 50-60 sr (e.g., Tesche et al. 2009, Tellus B, 61, 144-164; Pappalardo et al. 2010, J. Geophys. Res., 115, D00H19; Tesche et al. 2011, Tellus B, 63, 677-694). This history of scientific discussion is not adequately reflected. The paper, at least in the beginning, gives the impression that it is the first thought in this direction. In general, the introductory part of the paper is rather short and does not acknowledge the variety of validation efforts for CALIPSO. The structure of the paper could be improved by providing a broader discussion on CALIPSO comparison and validation efforts in the beginning of the paper. At the moment, many arguments, which would be helpful for the reader to know at an early stage, appear only very late in the discussion part.

Thank-you for your comments. We agree that the introduction is too brief, and have expanded it accordingly.

We appreciate your interest in the dust lidar ratio, as well as your desire to have this issue placed in the beginning of the paper. This paper is not just about dust, though; the purpose of the paper is to give the reader a sense of the accuracy of the CALIPSO AODs for all aerosol subtypes. We believe that placing the dust debate at the beginning of the article will de-focus the reader from the other elements of the paper.

We do cite some previous investigations of the dust lidar ratio in Section 4, choosing not to phrase our study in the contentious form of a debate. We note that we had already included some of the citations that you mention above in the original draft (Wandinger, 2010; Tesche, 2009),

and have happily added Pappalardo (2010). We note that Tesche (2011) pertains to smoke and smoke/dust mixtures, though, so we do not include it in our dust discussion.

Finally, we note that we have included a discussion of dust lidar ratios available in the literature, but the range of values that we have found is much higher than the 50-60 sr suggested by the reviewer. The range of dust lidar ratios that we have found is 30-80 sr (Mattis et al., 2002; Mona et al., 2006; Papayannis et al., 2008; Liu et al., 2008; Tesche et al., 2009; Esselborn et al 2009), as mentioned on Page 13, line 21 of our original submission.

Section 2, CALIPSO products: Are opaque aerosol layers (no ground return) included or not, and how could this influence the AOD statistics?
Although we do not exclude opaque layers, the maximum AERONET AOD(532) in our analysis is 2.17, and the maximum CALIPOSO AOD(532) is 2.31. Thus, none of our comparisons include opaque layers, which we now mention in the revised draft. Additionally, we provide some discussion about the effect of opaque aerosol layers on the lidar ratio on Page 27, lines 19+.

Section 2, on spatial/temporal collocation: It should at least be mentioned that there are other approaches which make use of trajectories and wind profiles to get best relation between ground-based and space-borne observations. The "closest approach" seems to be well justified for cross-wind conditions, but might fail when the transport is parallel to the CALIPSO track.
Folks have successfully used back trajectories for comparisons with small sample sizes (less than about 10 or 20), but the problem rapidly becomes difficult to track as datasets get larger. How far backwards in time and space does one need to run the trajectories, what altitudes do you run the trajectories, how do you handle changes in aerosols between the source and receptor regions, etc.? All of these things are easily managed during field missions with abundant ancillary data, but attempting to do this globally at all of the AERONET sites in a sensible way is a much more difficult task. To the best of our knowledge, no-one has ever done this with a global dataset. Hence, we feel that documenting the issues associated with back trajectories and wind profiles is beyond the scope of this paper.

Section 4, pages 11658-11660, depolarization ratio >0.2: It is unclear what the depolarization ratio used in the CALIPSO typing scheme means (particle depolarization, volume depolarization, or other?), how it is obtained, and how it compares to the depolarization ratio retrieved from AERONET. What are the consequences for possible misclassifications of dust and polluted dust?

CALIPSO uses layer-integrated particulate depolarization ratios greater than 0.2 to classify aerosols as dust (see Omar et al., 2009). We use the column aerosol size distributions retrieved from AERONET to compute depolarization. Ideally, these values are identical for columns with a single aerosol layer.

Misclassification of dust is discussed on the top of page 14 in the original draft, and more generally, the variability of lidar ratios for aerosol layers containing dust is considered section 3.1.2. The range of lidar ratios for "dust" reported in the literature is 30-80 sr. The range of average values that we report in Table 4 is 39-61 sr. Clearly, a single lidar ratio for dust (or polluted dust) does not exist, and using a single lidar ratio for dust will often cause substantial error in the CALIPSO AOD retrievals. This theme is emphasized throughout the paper.

Section 4: One of the weakest points of the paper is certainly the "belief" in the AERONET retrieved lidar ratios. This issue should be discussed more critically. Retrieval of the lidar ratio requires exact knowledge of the backscatter phase function at 180°, which sensitively depends on particle shape. The lidar ratio also depends on the absorbing properties of the particles, i.e. the imaginary part of the refractive index. The latter one is not considered in the discussion at all. It has been found that AERONET retrievals could not reproduce measured lidar ratios and depolarization ratios of dust adequately. Also here, there is extensive discussion in the literature available (see, e.g., Müller et al. 2010, J. Geophys. Res., 115, D07202; Müller et al. 2010, J. Geophys. Res., 115, D11207 and citations therein).

Our original paper submission includes some discussion of this topic on Page 26, lines 23+, where we discuss the consistency of the AERONET-retrieved lidar ratios with respect to Wandinger (2010), Tesche (2009), Muller (2007), and Esselborn (2009). Thus, the AERONET-derived lidar ratio climatologies are consistent with the literature. We have also added the following paragraph to Section 4:

> *We note that it is important to retrieve the correct backscattering phase function for these lidar ratio computations, and that this is possible for size distributions with irregularly-shaped particles using the spheroid approximation (at least in some cases; see Nousiainen, 2009). Also, although it is not possible to directly measure the aerosol backscatter phase function with the AERONET instrument, the Level 2.0 AERONET products are constrained to measured radiances at scattering angles of 3 to 100–154 degrees (depending upon the solar zenith angle). Single-scatter albedo is the other important parameter for determining the lidar ratio, and the accuracy of the single-scatter albedo for dust in the Level 2.0 AERONET database is estimated to be 0.03 (Dubovik et al., 2000).*

Regarding the imaginary refractive index: indeed, the only mention of the imaginary refractive index in the original submission is on Page 19, line 18, where it is mentioned as one of the parameters that causes variability in the box plots of Fig 9. We did not mention the known variability of lidar ratio with respect to imaginary index because the variability of the lidar ratio associated with the imaginary refractive index is secondary when compared to the variability of the lidar ratio associated with the real refractive index in the AERONET retrievals. However, other readers will probably be interested in the effect of absorption on the lidar ratio as well, so we have added this paragraph to the paper:

> *Lidar ratio is also sensitive to aerosol absorption, of course, but we do not observe high correlations of lidar ratios with imaginary refractive indices for AERONET retrievals in the dust belt. Indeed, the correlation coefficient of the imaginary refractive index with lidar*

Regarding the literature suggested by the reviewer: the Muller D11207 paper indicates a decent comparison to AERONET at the 532 nm wavelength for four data points (when considering error bars; see their figure 2). However, the lidar ratio is *assumed* to be 55 sr at 355 nm and 50 sr at 1064 nm in this paper (according to the caption of their figure 2), so these two wavelengths are not suitable for lidar ratio comparisons. We could cite the D11207 article as a "good" comparison to AERONET at the 532 nm wavelength, but unfortunately the tone of the article suggests a poor comparison, and citing this article as a positive comparison would confuse many readers. The Muller D07202 paper does not discuss AERONET lidar ratios, but found poor agreement of the AERONET size distribution with in situ measurements for the single day of data that they analyzed. Muller D07202 also inferred the refractive index using a different technique, and obtained a different result than AERONET for that day. Several assumptions are required to infer the refractive index in the Muller paper, and a full discussion of the issues associated with this comparison is beyond the scope of our paper.

Ultimately, although both of these suggested Muller papers are very interesting, the sample sizes are too small to draw meaningful conclusions about the robustness of the AERONET retrievals (4 data points for Muller D11207, and a single retrieval for Muller D07202). Retrievals are ill-posed problems by definition, and are not rigorous measurements. Thus, one should not expect any retrieval to provide correct results on a case-by-case basis. Ideally, a good retrieval will produce the correct result "on average," though. As stated above and in the manuscript, the AERONET lidar ratio climatology is consistent with the lidar ratios reported in Wandinger (2010), Tesche (2009), Muller (2007), and Esselborn (2009).

Section 4: The authors investigate the presence of the fine mode fraction, which is certainly a good indication for the influence of pollution aerosol. However, although they mention the influence of marine aerosols at coastal and island AERONET stations as well, they do not provide an explanation what it means for the lidar ratio retrievals. Marine aerosol would mainly be present in the coarse mode and thus cannot be distinguished from dust in the size distribution. Because of its low lidar ratio, it can decrease the column lidar ratio drastically even in the presence of strong dust layers above the marine boundary layer.

We agree. Dust layers located above marine aerosol layers will result in an AERONET-retrieved lidar ratio intermediate of those two layers. We also agree that coarse marine aerosols can not be distinguished from dust via particle size.

We are not sure of where you want us to mention dust over coarse marine aerosols in section 4, though. Most of that section is devoted to discussion about why the AERONET-retrieved lidar ratio varies over northern Africa, and how this might relate to the variability of lidar ratios

reported in the literature. Pollution is common over Africa, but there are only three AERONET sites in Figure 8 that could possibly be contaminated by marine aerosols (Bahrain, Dakar, and Capo Verde), and the lidar ratios at these three sites are consistent with their neighboring sites (see Table 4). Additionally, one would expect a significant percentage of spheres in the AERONET retrievals when maritime aerosols have a significant presence, but the percentage of retrievals with more than 5% spheres is less than 2% at all of these sites. If we filter the data for retrievals with less than 5% spheres, we obtain identical lidar ratios to the ones presented in Table 4.

Hence, it appears as though marine aerosols are not contributing significantly to the average lidar ratio at these three sites. Additionally, the AERONET-retrieved lidar ratios are already quite high at the two African sites (54-60 sr), so if marine aerosols are significantly reducing the column lidar ratios at those two sites, then some other aerosol must be located above those sites with lidar ratios that are much higher than 54-60 sr. The two possibilities with high lidar ratios are biomass burning and pollution. Biomass burning does not play a significant role in section 4, since we have restricted our analysis to summer months (May through Sep). Pollution does not play a substantial role, either, when the fine volume fraction (fvf) is less than 0.05 (as the reviewer noted above). Thus, it seems unlikely that coarse marine aerosols are substantially affecting the analysis of section 4.

Section 4.4: The Raman lidar measurements at Cape Verde are summarized by Tesche et al. 2011, Tellus B, 63, 677-694. Wandinger et al. (2010) discussed the discrepancies between the ground-based measurements and the CALIPSO retrievals and the possible influence of multiple scattering on effective lidar ratios obtained by constrained CALIPSO retrievals. Multiple scattering influences the CALIPSO retrievals when the dust particles are large enough and can explain the low effective lidar ratios obtained especially for opaque dust layers/dust storms (see discussion in Sec. 4.4, third paragraph).
Indeed, multiple scattering is discussed at length in Liu (2011), who concluded that multiple scattering had little effect for moderately dense dust cases (optical depths of ~1), and that "...these studies together strongly suggest that the CALIOP modeled lidar ratio values of 40 sr at 532 nm and 55 sr at 1064 nm are appropriate for the thin to moderately dense African dust layers which dominate the African dust observations." We note that 97% of our lidar ratio retrievals over Africa occur at optical depths of less than 1.5, and therefore should not be significantly affected by multiple scattering. Thus, there is a large discrepancy between the Liu (2011) "thin to moderately dense African dust layers" and the values we obtain for non-Sahel Africa, even when multiple-scattering effects are small. The reviewer's point is well taken, though, and we have added an additional sentence to that paragraph acknowledging multiple scattering effects and the Wandinger work. The Tesche (2011) paper suggested here pertains only to smoke and smoke/ dust mixtures, though, so we do not cite it in this discussion.

TECHNICAL COMMENTS:
Table 4: What is the meaning of INM, TMP, IER, DNM?

These are part of the names of the AERONET sites; they mean something to the people who named the sites, but their meanings are undefined on the AERONET webpage.

Fig. 10: The caption is an interpretation rather than a description of the figure.
Thank-you for pointing this out. The caption has been modified.

Use either Cape Verde or Capo Verde throughout the paper.
Our intent was to use "Capo Verde" for the Capo Verde AERONET site on the Cape Verde islands, but there were several errors in this scheme. Hopefully we have rectified all of them -- thank-you for catching this.