We would like to first thank the anonymous reviewer for the comments to help make better our manuscript. We address the reviewer's comments below. The original comments are in **bold font** and our responses are in *italic font*.

Major comments:

Most of the comparisons deal with AOD from AERONET and in situ measurements. Therefore and for the reasons mentioned later, I think that the title of the paper is not accurate and even misleading. To meet the title the paper would have to include comparisons of particle size distributions and refractive indices which are unfortunately missing. I therefore suggest renaming the paper to something like "Closing the gap: sources of differences in aerosol optical thickness derived from sun photometer measurements and in situ aerosol profiling."

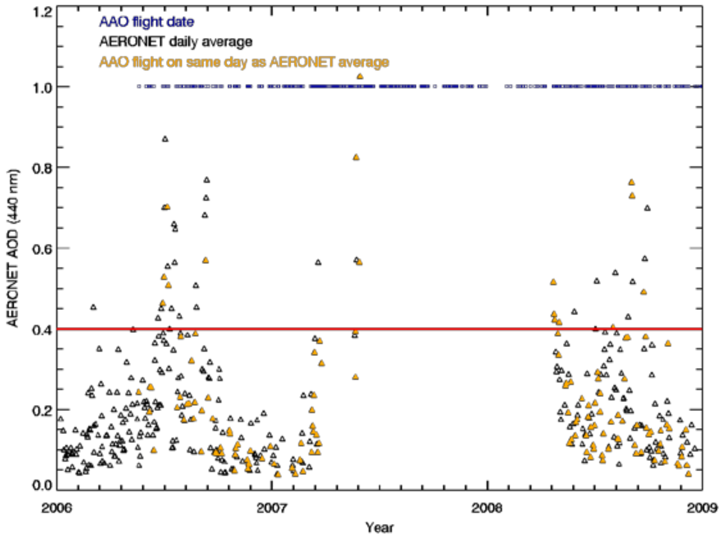
The title of the paper has been changed to "Sources of discrepancy between aerosol optical depth obtained from AERONET and in-situ aircraft profiles"

I missed an introduction to the measurements used in this study. The authors jump right into comparing the two data sets. For the reader it would be good to get an impression about the aerosol situation at Bondville in terms of histograms of AERONET measurements and mean aerosol profiles from research flights within the time period under investigation (Maybe between Sections 2 and 3). Also, it would be good to know if there is any seasonal variation of the optical properties or of the number of encountered aerosol layers.

A climatology of aerosol properties observed at Bondville was reported by Koloutsou-Vakakis et al. (2001) and Delene and Ogren (2002), and nowadays there is a paper in preparation about the AAO measurements:

Sheridan, P. J., Andrews, E., Ogren, J. A., and Hoff, R. M.: The NOAA airborne aerosol observatory (AAO): In situ aerosol optical properties, their annual and seasonal variability and comparison with satellite observations over central Illinois, Atmos. Chem. Phys. Discuss., in preparation, 2012.

A figure has been added with a summary of the AERONET AOD observations during the study period, along with indicators of the days when AAO data are available.



The comparison approach should be described in more detail. I could only find that in situ measurements are compared to AERONET measurements no later than 2 h after the measurement flight. What was used from AERONET? Was it a single measurement during the flight? Or averaged data until 2 h after landing (but starting when)? This is not clarified in Section 2.2. Some more information is given in Section 4.7. But this is also not clear (and certainly too late in the text.) Maybe it is possible to visualize the comparison approach in a sketch?

The comparison approach is now described in more detail in section 2.2.

AERONET data were matched with AAO data using the criteria that the single AERONET measurement to be used in the comparison was within +/-2 hours of the end of the corresponding AAO flight. For those cases where there was more than one AERONET measurement within that time period, the closest to the end of the AAO flight was the one used in the comparison.

The authors address a lot of possible error sources but they seem to forget the biggest source of uncertainty (which could be covered by giving a more general introduction to the measurement at Bondville): Do they encounter aerosol conditions under which column-integrated findings can be compared with height-resolved measurements (besides temporal homogeneity)? I would suggest that this is only the case if one encounters a single well-mixed aerosol layer. How many aerosol layers were usually observed? What is the influence of long-range transport? Based on these considerations, can you find sub-sets for which better

agreement is achieved? Even if all layers are covers by the in situ measurements, it could be the case that aerosol properties are very different within these layers and the average from the in situ profiles is not identical to the AERONET-derived mean of the entire aerosol column. Is there a systematic error by comparing column-integrated values to a mean over individual values for different height layers? The authors should make better use of the height-resolved measurements to address these issues. Please comment on these sources of uncertainty.

We examined the nephelometer data to test the hypothesis that the discrepancy was due to aerosols between our fixed-altitude legs, and were able to reject that hypothesis (section 4.3). If there were multiple layers between our upper and lower flight altitudes, or if the layers were not well-mixed, we would have detected them in the analysis of the nephelometer data. Likewise, we would have detected such layers whether they were due to long-range transport or not. Our comparison scheme started from the column integral of scattering, backscattering, and absorption, so we averaged the effects of layers with different properties in the same way as AERONET, i.e., there is no systematic error as suggested by the Reviewer. It would appear possible that the Reviewer hadn't read through the entire manuscript before making this comment, as his/her main question was addressed in section 4.3.

Point 3 of the Discussion (aerosol layers below, between or above the fixed flight levels) should be addresses entirely. I don't really understand why the important point of investigating the presence of aerosol layers above the flight level (or rather the use of CALIPSO measurements for such a task) is considered to be "beyond the scope of this study". Such layers would originate from distant sources and should be visible over a rather wide area. Thus, why not taking a look at the top of the first (highest) aerosol layer identified in the CALIPSO level 2 aerosol layer product for measurements in an area of 100 km (or more) around Bondville? Using the CALIPSO subset tool I found 199 overpasses between 20 June 2006 and 31 October 2008 for 100 km distance from Bondville. This could be used to estimate the influence of aerosol layers above 5 km height which I would assume to be non-negligible. Such an investigation could be presented without getting too much into details and would cover an important but yet omitted point of the study.

The possible presence of aerosol layers above the flight level is now addressed in section 4.3.

Yu et al (2010) reported seasonally-averaged profiles of aerosol extinction coefficient for the Eastern US in 2007, and showed that aerosols between 5 and 10 km asl made a negligible contribution to AOD for that year. An examination of the CALIPSO 'level 3' mean extinction profile for June 2006-October 2008 in a 2 degree latitude by 5 degree longitude grid box, roughly centered on Bondville, also reveals negligible extinction above 4.6 km (J. Tackett, personal communication).

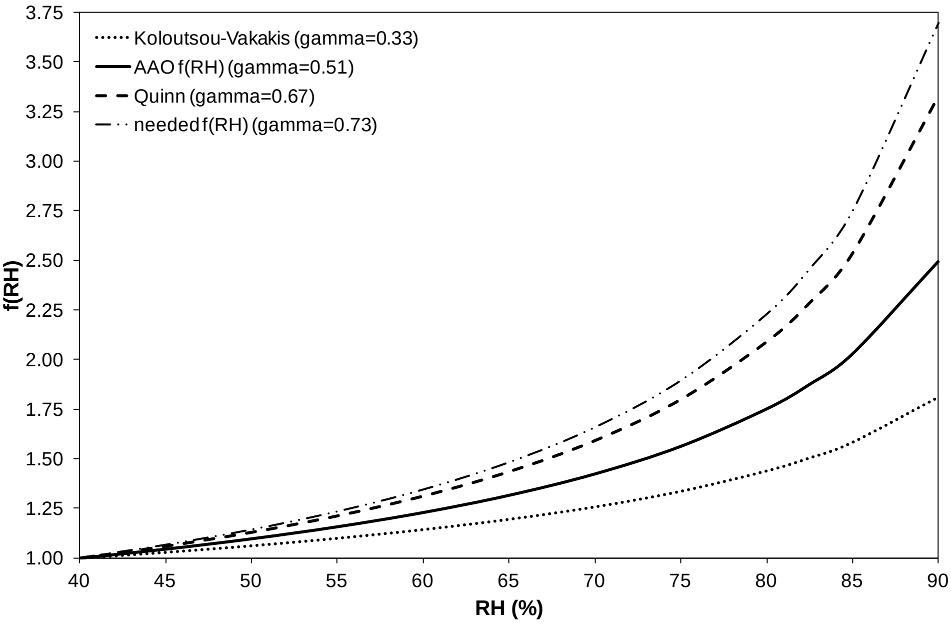
Yu, H., Chin, M., Winker, D. M., Omar, A. H., Liu, Z., Kittaka, C., and Diehl, T.: Global view of aerosol vertical distributions from CALIPSO lidar measurements and GOCART simulations: Regional and seasonal variations, J. Geophys. Res., 115, D00H30, doi:10.1029/2009JD013364, 2010.

If the authors find that the effect of humidity is dominating (minor effect of elevated aerosol layers and no systematic error in the comparison approach), it could be interesting to determine the needed growth factor for which agreement between AOD from AERONET and AAO could be achieved and weather such values are realistic.

The best comparison between the AOD from AERONET and the one calculated from the AAO measurements (y=0.99x+0.05; R^2=0.80) is obtained for a gamma value of 0.73, which yields an estimated f(RH) (at 82.5% RH) of 2.46. A sentence explaining this has been added to the text.

The figure below shows the hygroscopic growth factor f(RH) for the three parameterizations used in our paper plus the needed growth factor to achieve the best agreement between the AOD from AERONET and AAO (still with the offset of 0.05).

This value of γ = 0.73 is closer to the value of γ = 0.67 obtained at Bondville (during the limited-duration campaigns as part of the Koloutsou-Vakakis et al. (2001) study of f(RH) and the IMPROVE chemistry measurements) than the original value of γ = 0.33 that we have used in our paper. However, it is far from the value of γ = 0.51 obtained from the 3 Radiance Research nephelometers on the AAO aircraft, which should be the "most realistic" of all of them since it was obtained from in-situ measurements on the aircraft (although there were very little reliable data available).



Furthermore, it is beyond the scope of our work to determine what causes the differences in these parameterizations, and we just used three of them to show the range of possibilities. We didn't use the value of γ =0.73 because it is not justified, it's simply the one that provides the best comparison.

The discussion in the conclusion should be expanded to cover the general comparability of column-integrated findings with means of height-resolved measurements. I would also like to read a more significant conclusion or at least some kind of ranking of the influences that lead to the differences between AERONET and AAO.

We have added a sentence stating the difficulty of verifying the AERONET retrieval algorithm at a site that is not highly polluted.

We conclude that the largest portion of the observed AOD discrepancy is probably due to an incorrect adjustment of light scattering to ambient RH; improved measurements of the aerosol hygroscopic growth factor would be needed to confirm this diagnosis. Another part of the discrepancy might come from either less aerosol at the lowest flight altitude, relative to the surface, or that the aircraft inlet excludes larger aerosol particles; however, it was not possible to separate these two possibilities with this data set, making it difficult to completely rule them out. Spaceborne lidar observations indicate a minimal contribution to AOD from aerosol layers above the highest flight level.

Some of the figures (e.g., Figs. 3 and 9) could be omitted when their content can clearly be described in the text.

Figures 3 and 9 have been removed from the text, and their content is described in the text.

Specific comments:

Symbols for scattering and absorption coefficients are introduced in Sections 1 and 2.1 (with varying indices). AOD is also introduced several times. Please only introduce abbreviations if you are going to use them later in the text.

Done.

Page 29006, line 5: I suggest adding when between term and using to avoid confusion about the price of aircraft measurements.

Done.

Page 29006, line 13: You are not evaluating AERONET AOD! I think everybody agrees that you can find no better standard instrument for the measurement of AOD than an AERONET sun photometer. The inversion products on the other hand depend on the accuracy of the AERONET aerosol model and can always use independent evaluation. You should emphasize that the main objective of this paper is to investigate the sources of the observed disagreement and to assess their influence. Please rephrase.

Done, although we disagree with the Reviewer's claim that "everybody agrees". Some AOD specialists argue that precision filter radiometers that continuously point at the sun are the gold standard. The CIMEL sun photometer agrees well with these instruments, though.

Page 29007, line 20: I think (D50) should be moved behind 50% for better understanding.

Done.

Page 29016, line 18: What is meant by "dry"?

The "dry" AOD here means the AOD not adjusted to ambient relative humidity (RH) using the RH measured on the aircraft and the hygroscopic growth equation derived by Koloutsou-Vakakis et al. (2001). This is now explained in the manuscript.

Page 29018, line 24: This exercise would be easier to understand when you clearly state/repeat that size distribution measurements were not performed in the aircraft.

A sentence stating that size distributions on the aircraft were only measured up to 0.5 μm diameter has been added to the text.

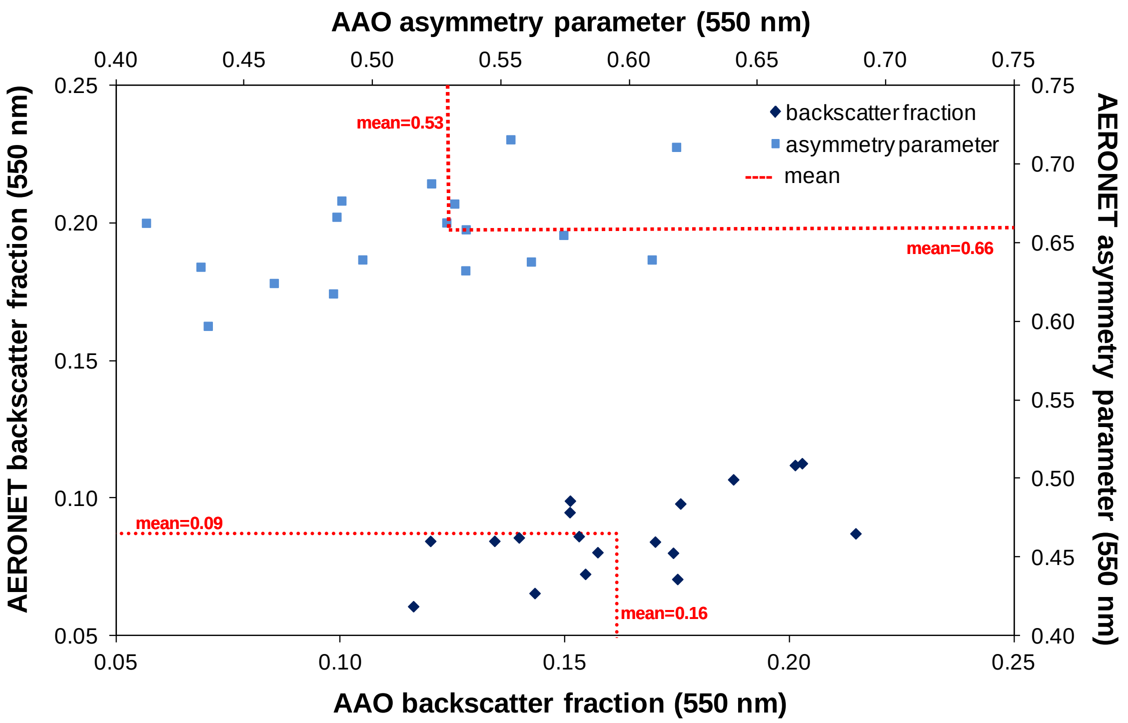
Page 29022, line 1: This section 4.5 should be shortened. I don't think there is need for Fig. 19.

Section 4.5 has been lengthened in order to address the question about cirrus contamination raised by Review #1. We have removed Figure 19 and described the result of the regression analysis in the text of section 4.5.

I suggest combining Figures 1 and 2 to one figure that gives the correlation of the optical data. Figures 4 and 5 could be combined in the same way.

We thank the reviewer for the comment, but prefer to keep separate figures for the AOD and Ångström exponent comparisons.

Figures 4 and 5 have been combined to one figure.



Please remove Figure 3. This one data point can also be discussed without the picture.

Figure 3 has been removed.

Figure 5: Is there some information in the larger variation of values obtained from the AAO measurements compared to AERONET?

A sentence has been added explaining that the high AAO values observed for the backscatter fraction ($b > 0.13$) were obtained on relative clean days ($AOD \approx 0.07$), i.e., days when the signal/noise ratio of the nephelometer data was low, and thus less reliable.

Figure 9: Since no huge difference is found, I don't see a need to show this figure.

Figure 9 has been removed.

Figure 19 shows the same as Figure 1 and is therefore not needed.

Figure 19 has been removed.