Reviewer#2

This paper deals with an important and challenging issue, certainly acceptable for ACPD. The authors argue that the aerosol absorption data most widely used in climate modeling is likely biased high at low AOD, based on coincident and climatological in situ data at two rural sites. I've included some notes below; in summary, there is a lot of good data presented here, but I think the estimates of uncertainty need to be tightened up in order to reach a strong conclusion. Also, evaluating AERONET SSA at AOD below the value they state as the lower limit of quality results is a key caveat, though I agree that the AERONET results are widely applied beyond their stated validity range. Note that this is actually my full review, so it can be considered as part of the formal review process rather than just as a "quick" review for ACPD posting.

We thank the reviewer for the 'quick' review and helpful comments. We've responded below to each of them.

Lines 73-79. This essentially makes the case for selection bias in the AERONET SSA and AAOD values by itself, though I don't think it negates the value of going further and comparing with in situ observations. If the in situ data can show that in general, SSA is lower when AOD is higher, that could make a useful contribution to the argument.

The in-situ data make a useful contribution to assessing AERONET SSA and AAOD regardless of the observed relationship between SSA and AOD. That said, in general at individual sites (at least in the US) the SSA seems to be lower when AOD is lower – for both in-situ and AERONET data.

Line 112. This is supported by the AERONET data themselves. AERONET does not offer global spatial coverage, but it does provide overwhelming evidence AOD_440 is generally <0.4 via direct-sun AOD measurements, which don't suffer from the uncertainties entailed in the model estimates.

We agree with the reviewer that the AERONET data would likely also support our assertion that AOD_440 is rarely greater than 0.4. But the point of our paper is not to quantify rigorously the global coverage of the AERONET Level 2.0 AAOD/SSA retrieval products, so we don't feel that an estimate of the global coverage based on AERONET data would significantly improve the paper.

Lines 242 to 258. As you know, in addition to collocation, the big challenges for this study are probably getting the total column data from the aircraft sampling right, and accounting for the difference between the properties of the ambient particle observed by AERONET and the dessicated ones measured in situ. Assuming that absorbing aerosol is hygrophobic seems a bit risky, especially for an SSA calculation, though this would be less of an issue for cases where the ambient RH is also low.

It is risky, but it is the standard assumption that is made (i.e., in every other direct comparison paper cited in Table 3 and 4), based on limited lab and field data about absorption hygroscopicity. Nonetheless, we also performed a sensitivity test where we assume that the absorption enhancement due to RH is the same as the hygroscopicity scattering enhancement. More details are included in response to the reviewer's comments related to lines 594-596 and we've added the following sentences to section 3.1.2:

“A sensitivity test was performed assuming absorption enhancement due to RH was the same as the hygroscopicity scattering enhancement, i.e., $\sigma_{ap}(RH_{amb})/\sigma_{ap}(RH_{dry})=a*(1-(RH_{amb}/100))^{-\gamma}$. While this is likely an extreme assumption, it had little effect on the comparisons of AOD, AAOD and SSA.”

(Do Lines 332-334 raise another question about getting SSA right?)
Lines 332-334 are: “For SSA there appeared to be no correlation between AERONET retrievals and in-situ calculated values regardless of match window length (highest SSA correlation coefficient was 0.12, but most were less than 0.05 for both sites).” We’ve added the following sentence after that sentence:
“The poor correlations for SSA are not surprising given the uncertainties at low loading.”

Ok. I see that you deal with these issues in Section 2.4.1. I’m thinking that the hygroscopicity issue might need a bit more consideration; there does not seem to be a conclusion about the uncertainty in SSA from the in situ observations, and it is not clear whether the general discussion derived from the literature is applicable to the aerosols observed over the AERONET sites in the current study.
We haven’t responded specifically to this comment as it seems to summarize the previous several comments which we have responded to.

For the column AOD question, again the discussion does not seem to come to a real conclusion about the uncertainties. Having coincident lidar would help, and this might be available for at least some cases at one or both sites.
We’ve now looked at the Raman lidar best estimates of aerosol extinction profiles at SGP for the 14 flights with AERONET matches (there is no lidar data available from BND). We found three cases where there appeared to be an aerosol layer in the vicinity of the highest in-situ flight levels, but in each case the profile flight provided a hint of the presence of this layer. Looking at the actual shape of the in-situ profiles, these three flights exhibited a significant increase in measured loading at the highest flight levels. We’ve removed those flights from the comparisons reported here. There may still be aerosol above the height of the Raman lidar but we have no means for identifying it. Based on the criterion of observing a strong increase in aerosol loading at the highest flight levels, we also removed 3 flights from the set of BND profiles. We’ve added the following text:
“Although statistical profile results (e.g., Turner et al., 2001; Yu et al., 2010; Ma and Yu, 2014) suggest little contribution from high altitude aerosol layers in the region of these two sites, Schutgens et al. (2016) demonstrates the importance of considering the specifics rather than the statistical. We used the Raman lidar best estimate data product of extinction profiles at SGP to evaluate the presence of aerosol above the highest flight level at the site. For the SGP in-situ profiles that had matches with AERONET inversion retrievals, we identified three lidar profiles that exhibited aerosol layers at high altitudes, but in all three cases the presence of these layers was also hinted at by an increase in the aerosol loading at the highest flight levels of the in-situ measurement. Thus, we further screened in-situ/AERONET comparisons by removing flights at SGP and BND with significant increases in loading at the highest flight levels. There may still be aerosol layers above the level measured by the Raman lidar, but we have no

means of assessing that. The AOD comparison presented in Figure 3 suggests we are unlikely to be missing significant aerosol at high altitudes."

Section 2.4.2. There are other possible factors to consider here. For example, the AERONET retrievals report only one pair of (real, imaginary) refractive index values. If there are two or more modes in the column, this assumption will skew the result. You mention the possible surface reflectance contribution to the AERONET AOD uncertainty; there is a paper assessing this which might be worth considering (Sinyuk et al., Remote Sensing Environment 2007, doi:10.1016/j.rse.2006.07.022).
We have no particular insight or expertise concerning the AERONET retrievals, and can only rely on the published uncertainty estimates. If the retrieval experts have not assessed the uncertainty associated with a particular assumption in the retrieval, then we are unable to include that uncertainty in our paper. However, the point the reviewer makes about AERONET retrieving a column RI is a good one and we've added the following to Section 2.4.2:
"Another potential issue is that the AERONET retrievals report only one pair of (real, imaginary) refractive index values for the total size distribution (for each wavelength). If there are two or more modes in the column, this assumption may skew the resulting SSA and AAOD values, although the effect of such skewing would depend on the aerosol properties and cannot be assessed here. Potential impacts in the case of uneven mode absorption in the retrieved size distribution have been found to be minor since the retrieved size distribution is more linked to forward scattering than absorption (pers. comm., O. Dubovik)."

Then there is a question about whether the direct sun AOD measurements are used to obtain the extinction in the determination of aerosol absorption properties, or whether the scattering and extinction are both determined from the almucantar scan. In the latter case, the measurement uncertainty will be larger than 0.01 or 0.02, whereas in the former case, heterogeneity could affect the result, as the extinction and scattering data would be taken in different parts of the sky. Either way, the SSA result in most cases would be the small difference between two larger numbers, so accuracy could be an issue.
We are using the reported values of the aerosol absorption properties from the almucantar scans/inversion retrievals and we rely on the published uncertainty estimates for AERONET products. We had helpful discussions with several AERONET gurus (David Giles/Brent Holben) they provided comments to our discussion of the AERONET uncertainties (hence the mention of surface reflectance referred to in the previous comment!).

Lines 503-505. Perhaps the AOD comparisons address the total-column sampling question for the aircraft measurements, in addition to the uncertainty related to the hygroscopicity adjustment and possible large-particle under-sampling. Note that in general, a high correlation does not assure good quantitative values, as might be required for SSA assessment. So, quantitatively, how does this affect the uncertainty in subsequent SSA estimation?
We are not totally clear about what "this" refers to in the question. It could be "total-column sampling", or "hygroscopicity adjustment", or "possible undersampling". We can (and did) assess the uncertainty in the SSA derived from our in-situ measurements for all of these issues, but rely on published uncertainty estimates for AERONET products. However, we think the

reviewer is referring to the effects of quantitation vs correlation. Figure 3-5 (for 440 nm) include an indication of both the in-situ and AERONET uncertainties. For AOD we see that those uncertainty estimates cross the 1:1 line for almost all cases (red or blue) and definitely for all the blue cases at both sites. This suggests that the in-situ measurements provide a reasonable representation of the total column aerosol loading as represented by AERONET and student t-tests at the 95% level support this. In contrast, for AAOD and SSA the uncertainty bars don't cross the 1:1 line for any of the measurement comparison points at SGP and for only do so for a small subset of the comparison points at BND. Student t-tests on the AAOD and SSA data suggest the AERONET and SSA values are different at the 95% level. We've added comments about the uncertainty bars and student t-tests in the discussion of each figure.

Line 561. Again, it is not clear how much the measurement uncertainty contributes to the discrepancies between in situ and AERONET AAOD.
Uncertainty doesn't contribute to the discrepancies, but rather provides the means for assessing the significance of the discrepancy. As we note in our response to the previous comment, the uncertainty bars for AAOD and SSA suggest that even taking into account the uncertainty estimates for the measurements there are very few points (and only at BND) that overlap the 1:1 line. This suggests that there is a significant discrepancy between the in-situ and AERONET AAOD (and SSA) measurements that we don't see in the AOD comparison.

Line 567-568. Does this mean the in situ measurements are missing the extremes, either due to sampling, or to perhaps to conservative estimates of the hygroscopicity effect?
We don't think the in-situ measurements are missing the extremes. The aircraft results are very consistent with the long-term surface measurements at both sites which show much less variability in SSA than is obtained from the AERONET retrievals (e.g., Sherman et al., 2015). Figure 3 in Andrews et al. 2004 shows a comparison of the scattering at the lowest flight leg at SGP with the surface scattering measurements for a 2 year time period suggesting the aircraft is capturing the overall variability at the site…at least over the vertical range the aircraft samples at. We've also updated figure 8 in the paper to show the surface SSA data adjusted to ambient conditions for better comparisons with the ambient SSA values from the airplane and AERONET. The ambient-adjusted SSA from the continuous surface measurements (day/night, 1 min frequency, more than 15 years of data) shown in figure 8 is very similar to the SSA from the aircraft.

We've added the following sentence to the first paragraph of Section 2.1:
"Previous work has shown that the airplane measurements appear to capture the variability in aerosol properties observed by the long-term, continuous measurements at the surface (e.g., Figure 3 in Andrews et al., 2004)."

It is unclear whether assuming constant hygroscopicity fit parameters (that are used in conjunction with the variable ambient RH) will narrow or expand the variability of the calculated SSA. The discussion of the 'SSA_{hybrid}' ($SSA_{hybrid}=(AOD_{AERONET}-AAOD_{PSAP})/AOD_{AERONET}$) in the new last paragraph of Section 3.1.1 and now included on Figure 5 provides some additional thoughts on this. (Calculation of SSA_{hybrid} was proposed by Reviewer#1)

Lines 594-596. Right. But this does not address whether the underlying assumption that absorbing particles are non-hygroscopic is valid. If the absorbing species are OC rather than entirely BC, one might expect at least some hygroscopic growth is possible. And I think you concluded earlier that there must be something like OC, at least at one site.

As we've noted elsewhere in the manuscript there are VERY FEW studies (ambient or lab) investigating water uptake by absorbing aerosol and those that exist tend to suggest that water uptake is minimal. We have no data to assess the underlying assumption that the absorbing particles are non-hygroscopic. Since we expect that the aerosols at both sites are likely to be well-aged and internally-mixed, it is possible that the absorbing particles are hygroscopic, but we don't know the extent to which it would affect the absorption coefficient. At SGP, Sheridan et al (2001) showed that the aerosol hygroscopicity decreased in the presence of aerosol thought to contain dust or smoke.

There is organic at both sites – the IMPROVE measurements suggest 30%+/-13% OC at BND and 40%+/-14% OC at SGP for sub1um aerosol. Parworth et al., 2015 suggests anywhere between 25-75% organic at SGP for non-refractory portion of the1um depending on season. But the hygroscopicity and absorbing nature of that organic aerosol has not been assessed. The parameterization for hygroscopicity that we use (from Quinn et al., 2005) was derived using tandem nephelometer measurements of hygroscopicity on ambient aerosol (i.e., both scattering and absorbing aerosol) and the measurements of aerosol chemistry (specifically organic carbon (OC) and sulfate (Sulf)). The observed hygroscopicity (scattering as f(RH)/scattering_dry) was shown to decrease as the organic mass fraction (defined by Quinn as the ratio of OC/(OC+Sulf)) increased. This is a simple parameterization and does not account for all the individual chemical species which may influence water uptake nor does it account for interaction between absorbing species and water. We've added the following text to the sentence describing the parameterization to make this a little more clear:

"Climatological IMPROVE network surface aerosol chemistry measurements of sulfate and organic carbon (Malm et al., 1994) were utilized to determine a value for the hygroscopic growth parameter 'γ' for each site based on the Quinn et al. (2005) parameterization which relates aerosol hygroscopicity to organic mass fraction."

We've also done a sensitivity test to see how figures 3-5 would change if we assumed that the absorption enhancement due to RH was the same as the scattering enhancement due to RH. This assumption has little affect on the AOD comparison (in-situ absorption is only ~10% of in-situ extinction). The slopes in the AAOD comparison decrease by ~30%, but the AERONET AAOD values are still predominantly and significantly above the 1:1 line (i.e., all points at SGP and all but 3 points at BND are above the 1:1 line). At both sites the SSA values shift slightly closer to the 1:1 line; at BND 19 out of 21 AERONET SSA points are below the 1:1 line and at SGP all the SSA points are below the 1:1 line. We've added the following sentences to the text:

"A sensitivity test was performed assuming absorption enhancement due to RH was the same as the hygroscopicity scattering enhancement, i.e., $\sigma_{ap}(RH_{amb})/\sigma_{ap}(RH_{dry})=a*(1-(RH_{amb}/100))^{-\gamma}$. While this is likely an extreme assumption, it had little effect on the comparisons of AOD, AAOD and SSA."

Lines 614 to 617. Does this call into question whether the in situ measurements adequately sample the entire column observed by AERONET? I'm thinking Section 3.1.2 does not put to rest the question in the title of this section. So I'm uncertain whether you have established the conclusion stated in Lines 625-627, though I think AERONET might overestimate absorption in many cases, due to the way they relate the measured extinction and scattering in order to derive absorption.

The 'this' referred to by the reviewer is:

*"*The in-situ measurements would need to preferentially under-sample absorbing aerosol relative to scattering aerosol in order to come into line with the AERONET observations."

Section title is: "3.1.2 How might AOD discrepancies affect SSA and AAOD comparisons"

Conclusion sentences (which actually start next section) are:

"Direct comparisons at BND and SGP suggest that AERONET retrievals underestimate SSA and, consequently, that AERONET overestimates AAOD relative to in-situ measurements of AAOD for the low AOD conditions typical at these two sites."

We've tried to address the limitations of the in-situ measurements as best we can. Given that we do fairly well in the AOD comparison we don't think we are missing a significant amount of the aerosol. We can see two ways that the in-situ measurements would collect enough scattering aerosol to simulate the AERONET AOD but miss absorbing aerosol:

(1) not accounting properly for the effect of coatings (organic or water) on absorption enhancement which we've discussed in detail in the manuscript

and

(2) not sampling layers of predominantly absorbing aerosol below, between, and/or above the flight layers. These layers couldn't have much scattering associated with them or they would affect the AOD comparisons. Weigum et al (2012) do report on BC plumes over the remote Pacific although they don't comment on the aerosol scattering associated with these plumes and the levels of BC in the plumes they observed are significantly (factor of 10 or more) lower than what would be needed to bring the in-situ AAOD up to the level of the AERONET AAOD.

We've added the following text:

"In summary, we can only see two ways that the in-situ measurements can sample aerosol efficiently enough to represent AERONET AOD fairly well but significantly underestimate AAOD and overestimate SSA: (1) not accounting properly for the effect of coatings (organic or water) on absorption enhancement which we've discussed in detail (e.g., see Section 2.4.1) and (2) not sampling layers of predominantly absorbing aerosol below, between, and/or above the flight layers. We suspect that the SSA required of such layers in order to explain the AAOD and SSA discrepancies is physically impossible."

Note: We've also changed the title of section 3.1.2 to:

"*How might in-situ hygroscopicity assumptions and under-sampling of the aerosol affect SSA and AAOD comparisons?* "

Lines 689-690. This might be stated differently, as it assumes no systematic underestimation of absorption for the in situ measurements.

We've re-written the entire paragraph to be a bit more even-handed:

“In summary, the literature survey featuring measurements across the globe for many aerosol types suggests that even at higher AOD conditions, direct comparisons of AERONET with in-situ aerosol profiles find that AERONET column SSA is consistently lower than the SSA obtained from in-situ measurements (although often within the combined uncertainty of the AERONET SSA retrieval and in-situ measurements). If there was no consistent bias in the AERONET/in-situ comparison we would expect (AERONET_SSA – INSITU_SSA) to be evenly distributed around zero. Instead, Figure 6, which summarizes the literature survey, suggests either that AERONET retrievals are biased towards too much absorption, or that in-situ, filter-based measurements of aerosol absorption are biased low. We note that the results from the literature indicate that the hypothesized low-bias in in-situ absorption is not associated with a single airplane’s measurement system or the atmospheric conditions encountered in a single experiment. That leaves us with possible bias in the in-situ experimental methods (instrument issues (nephelometer, PSAP), treatment of f(RH), vertical coverage, sampling artifacts), all of which we have attempted to address above.”

We’ve also come up with a different title and edited sentences throughout the manuscript that suggest the only bias may be with AERONET retrievals.