AAOD Response to Reviewers and other Comments

Nick Schutgens' comment:

A very useful paper, and it will be interesting to see how to reconcile various measurements. Here I just want to mention the issue of temporal sampling, see also: http://www.atmoschem-phys.net/16/1065/2016/

While this temporal sampling issue is important for model evaluation, it is equally important in comparing different observational datasets. I see two issues relevant to the current paper:

Page 18, Line 732-735: Fig 7 was apparently made with different samplings of the in-situ and AERONET measurements. Is that the case for other figures as well? How might that affect results?

Figures 3-5 were made for matched samplings of the in-situ and AERONET measurements where matched means that the measurements were made within 3 h and 15 km of each other. Given the high correlation in AOD we are reasonably confident that with these sampling constraints the in-situ and remote sensing instruments were measuring in the same atmospheric column. The literature studies cited in Table 4 and included in Figure 6 used similar temporal and spatial matching criteria (see comments column in Table 4), with the exception of the DABEX dust/biomass burning flights (Osborne et al 2008; Johnson et al 2009) which were matched in time but less so in distance (flight profiles were within 100 km of AERONET retrievals).

In contrast, Figures 7 and 8 utilized the multi-year climatological data sets available for each measurement which have different samplings. Your work (e.g., Schutgens et al., 2016) shows there can be large differences when comparing values obtained with different samplings (more than 100% for AOD), particularly when there are high levels of variability in the data. In our manuscript the different temporal samplings are likely one contributor to the relatively small differences observed between the in-situ (red line) and AERONET 1.5 AOD (black line) although other things (e.g., assumptions about aerosol hygroscopicity, missed aerosol (i.e., due to size cut or flight limitations)) will also contribute. The relatively small differences between the in-situ and AERONET 1.5 AOD suggest there may not be much year-to-year variability at these two sites. The long term surface measurements at the site also suggest there is not much year-to year variability. The effects of different sampling are definitely the primary reason for the difference between the AERONET level 2 almucantar values (AOD and AAOD) and the in-situ measurements.

Page 22, Line 925 - 934: The authors suggest better estimates of AAOD may be obtained by using SSA measured at high AOD and applying it to low AOD cases. They mention possible sampling impacts but seem to feel those may not be that important. I'd like to caution against that.

I attach a figure of the difference in yearly SSA, when that SSA is taken at high AOD or at any AOD, for three different models. At least two models allow differences of more than 0.05. (In general, the MIROC-SPRINTARS model agrees best with AERONET Lev 2 SSA while HadGEM-UKCA is often too high and ECHAM-HAM too low.)

We agree that this is an approach to be cautioned against, particularly as systematic variability between loading and SSA has been observed by both in-situ and AERONET measurements at BND, SGP and many other sites (Delene and Ogren, 2002; Andrews et al., 2011b; Schaefer et al., 2014 and our Figure 8). Current work by our group shows this systematic variability is also simulated by many global models. We've re-written the abstract, discussion of Fig 8 and the conclusions to highlight the importance of the systematic variability we've observed and to note that such systematic variability cautions against the use of applying SSA obtained from high loading to obtain AAOD at low loading conditions via the relationship AAOD=SSA*AOD. That said – for the specific case of these two sites, we note that using the monthly median SSA from the high loading retrievals would result in a reasonable monthly median AAOD if the high loading SSA was applied to all AOD values.

Finally, it would be useful if the authors made a suggestion under what conditions AERONET SSA conditions may be used. Is AOD > 0.4 sufficient?
I don't think we can say definitively. Our Figure 6 comparing many field campaign measurements suggests that AOD440>0.25 or 0.3 may be reasonable. Oleg Dubovik (pers. comm. with co-author Stefan Kinne) thinks AOD440>0.4 may be too restrictive but did not suggest a lower alternative. We've added the following text to the discussion of Figure 6: "Figure 6 suggests that AERONET retrievals of SSA could perhaps be used at AOD_{440}<0.4, perhaps down to AOD_{440}~0.25 or ~0.3 – even at those low AOD values the differences in SSA between AERONET and in-situ still tend to be within the AERONET uncertainty. However, as Figure 6 shows, there are not a lot of direct comparisons to support such a choice."

Reviewer #1

This work compares AERONET column retrievals of aerosol single-scatter albedo to column single-scatter albedos obtained from in situ flight profiles. This is an important topic, because modeling groups regularly use the AERONET absorption products to infer or constrain atmospheric black carbon concentrations. This work focuses especially on assessing the capability of the Level 1.5 AERONET retrievals that are obtained at low aerosol optical depth (AOD(440) < 0.4) common to many locations throughout the world, which has received little or no attention thus far. The paper is clear and well written, scientifically solid, and thoroughly covers nearly all issues associated with this topic. I have a few topical comments and a laundry list of small details that will make the paper stronger, in my opinion.

We'd like to thank reviewer#1 for the thoughtful and well-organized review. Below we describe how we've addressed specific comments and suggestions.

Column Single-scatter Albedos

Once concern is the hygroscopic growth corrections for the nephelometer and PSAP measurements. Since particles are dried prior to entering the aircraft instruments, the authors employ an empirical correction to the scattering measurements and no correction to the absorption measurements. The scattering correction is an exponential function of RH (Eq. 1), and the exponents that they choose for the corrections are based upon aerosol chemistry measurements at nearby IMPROVE sites. This is all very standard in the business, and the pitfalls of empirical scattering corrections at high RH are well documented. The authors argue that they are mostly operating at low RH, so that their column AODs are not very sensitive to the hygroscopic growth corrections.

This is all fine, except that Figure 3 shows a significant amount of scatter in the author's AOD comparisons that undoubtedly works its way into the SSA comparisons. This is unavoidable when applying climatological hygroscopic growth corrections to in situ scattering profiles, but why not use AERONET AODs to compute column SSAs? Unlike AERONET's absorption and size distribution retrievals, AERONET AOD is a robust measurement that is widely accepted by the community. Thus, the authors could compute a "hybrid" SSA, whereby
SSAhybrid = (AODaeronet – AAODpsap)/AODaeronet

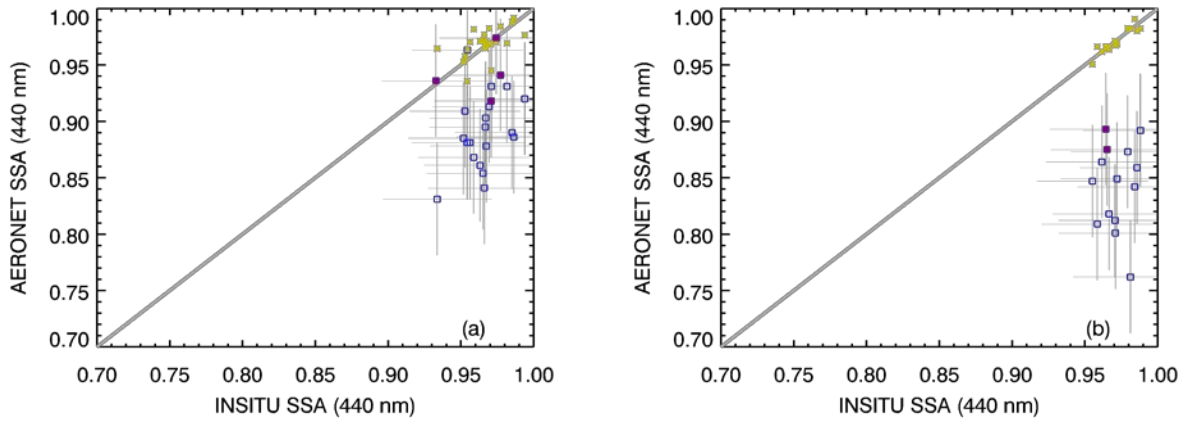
This will isolate the differences in SSA caused by absorption, which is of primary interest to the community. This approach will also remove the uncertainty associated with the empirical hygroscopic growth factors applied to the scattering measurements, and it will remove the scattering uncertainty associated with undersampling the coarse mode. Since AAOD is very well-correlated at SGP (per Fig 4), I suspect that this approach will tighten up the SSA scatter at that site in Figure 5.

Now, some readers may find issues with this approach as well, so I suggest including this approach as an additional alternative to the material already presented in the paper. For some of us, this approach will remove doubts about hygroscopic growth corrections and inlet size cutoffs for the scattering measurements.

This is an interesting suggestion. We've done this analysis and found that, as the reviewer suggested, it tightens up the SSA scatter at SGP and actually even at BND.
We've added this paragraph in section 3.1.1:
"Figure 5 also includes a set of 'hybrid SSA' (SSA_{hybrid}) points in yellow. These points have been calculated using the AERONET AOD and the in-situ AAOD:

$$SSA_{hybrid} = (AOD_{AERONET} - AAOD_{PSAP})/AOD_{AERONET} \quad (3)$$

This hybrid approach to SSA eliminates the uncertainty associated with the empirical hygroscopic growth factors applied to the in-situ scattering measurements, and also removes the scattering uncertainty associated with undersampling the coarse mode. It does not, however, eliminate the uncertainties associated with assuming the absorbing aerosol is hygrophobic, that there is little absorption in the potentially undersampled coarse mode, or the unknown contribution from absorption enhancement. SSA_{hybrid} is very similar to the SSA derived from in-situ measurements, suggesting the primary discrepancy between the AERONET SSA and the in-situ SSA is due to the determination of the absorbing nature of the aerosol, either due to issues with the limitations of the filter-based measurements or to the interpretation of the relative contribution of aerosol absorption from the AERONET inversion retrieval products."



Merging of in situ and AERONET data

Figure 3-6: The authors use multiple AERONET AODs and AAODs for each flight profile (i.e., the number of AERONET AODs at BND is 662, but there are only 72 flight profiles). This is not appropriate, in my opinion. The flight profiles occur over a 2-hour period, so I argue that they are comparing quasi 2-hour averages to AERONET's quasi-instantaneous retrievals. The atmosphere can easily change during this timeframe, especially since the authors allow an additional 1 hour before the flight and 3 hours after the end of the flight for including additional AERONET retrievals. Apparently the spread of AOD for a given flight profile can be quite large over this 6 hour period, too, as seen in the upper left panel of Figure 3 (for instance the AERONET AOD(440) ranged from 0.45 to 0.55 when in situ AOD(440) = 0.3). Additionally, the authors have argued that aerosols sampled within a 3-4 hour period have high autocorrelations, per their Figure 2. Since the premise of their comparison is that they are essentially sampling homogeneous air masses on average, why not use averaged AERONET values for comparison to the flight profiles? (After all, there is a heck of a lot of averaging going on for each in situ flight leg.) This will reduce the number of points in figures 3,4,5, and 6, but I believe that this will

provide a more accurate presentation of the data. Having said all that, I am OK with multiple AODs per in situ measurement (red points in Fig 3), as this provides an indication of the variability in aerosol optical properties that occur in the atmosphere during the flight profiles. A comment from the authors pointing this out would be helpful, though (if I've got this correct). Multiple AERONET AAODs and SSAs per flight in Figures 4 and 5 serve no purpose, though.
We've remade the plots as suggested:

- keep all the red points in Fig 3, to show variability in AOD during the flights
- change the blue points in Fig 3 to average AERONET AOD during the flights
- change Figs 4&5 to use average AERONET AAOD and SSA during the flights
- Figure 5 now includes the 'hybrid SSA as well'
- The plots reflect the new uncertainty values as described below.

In terms of new text, we've rewritten the first paragraph of section 3.1.1 to reflect the new plots.
"The red points represent all direct sun AERONET Level-2 AOD retrievals during the +/-3 hours window around the end of each profile– this provides an indication of the variability in AOD during the in-situ profiling flight.**"**
We also refer to the blue points in the discussion of Figures 3-4-5 as 'flight-averaged'

Autocorrelation
The authors present autocorrelations in Figure 2, which is a robust way to identify acceptable collocation lag times. They conclude that a 3-hour lag is acceptable for their analysis. They mention that the auto-correlations of scattering is greater than 0.8 for up to a 4-5 hr time lag (lines 316-317), but they do not specify that the auto-correlation of absorption is only 0.75 at BND and 0.55 at SGP. This should be explicitly stated here, because it is important for putting Figure 4 into context.
DONE

If the auto-correlations in figure 2 are robust values computed with single instruments, then Figure 2 represents the maximum correlation that we can realistically expect to achieve in a comparison of two different instruments. Thus, R2 = 0:77 at SGP in Figure 4 (or, R = 0:87) is actually an excellent result for AERONET, as the correlation is substantially greater than the auto-correlation in Figure 2. It is also an excellent result for the authors, as it demonstrates that they have put great care into producing a clean analysis. This is something worth mentioning in the article. Surprisingly, the BND site has higher 3-hour autocorrelations for absorption than SGP (r = 0.75 for BND and r = 0.55 for SGP, per figure 2), but the results for BND in Fig 4 are not as clean as SGP. Nonetheless, the correlation coefficients for BND in Figure 4 (R = 0.53-0.58) are not that far from the auto-correlation of 0.75 for absorption at BND in Fig 2.
We've added the following sentence to the end of the first paragraph of Section 2.3 (thank you reviewer for some lovely text!):
"Additionally, Figure 2 represents the maximum correlation that we can realistically expect to achieve in a comparison of two different instruments with temporally offset measurements and provides context for the AERONET/in-situ comparisons presented in Section 3."
We've also added these sentences to the discussion of Figure 4 in Section 3.1.1:
"Surprisingly, while the BND site has higher 3-hour autocorrelations for absorption than SGP (R = 0.75 for BND and R = 0.55 for SGP, per Figure 2), the results for BND in Figure 4 indicate

less correlation than at SGP for absorption. Nonetheless, the correlation coefficients for BND in Figure 4 (R^2=0.49 (blue) and 0.37 (red) correspond to R = 0.70 (blue) and 0.61 (red)) are not that far from the 3 h auto-correlation of R=0.75 for absorption at BND in Figure 2."

This brings up another confusing point: The authors define R2 as the correlation coefficient on lines 518, but R2 is usually reserved for the coefficient of determination, and R is the correlation coefficient. The authors also use "r" for the auto-correlation in Figure 2, which adds to the confusion. Thus, I am unsure if the R2 values in Figures 3-5 indicate correlation coefficients or coefficients of determinations.
Our mistake! R2 values in Figures 3-5 represent coefficient of determination –we've fixed this and the sentence now reads:
"The coefficients of determination (R^2) are within the …"

Additionally, how does one determine either parameter (correlation coefficient or coefficient of determination) when the number of points on the y-axis is different than the number of points on the x-axis (e.g., R2 = 0.84 for 56 AERONET retrievals and only 24 flights at BND for the upper left panel of Figure 3, etc.)? The authors need to explain how they were able to do this (mathematically) in both the text and in the figure captions.
In the original submission we determined the R2 values by matching the single in-situ value with each AERONET retrieval that fell within the +/-3h time window, i.e., if there were 3 AERONET retrievals for a given flight they would all correspond to the same x-value to create 3 xy pairs. Since we're now using flight-averaged AERONET retrievals for the comparisons (blue points in Figures 3-5), this is no longer an issue. We still provide a fit for the red points in figure 3.

Mixing State
It is easy to demonstrate (with publicly available Mie codes) that an internally-mixed absorbing aerosol particle has a much higher absorption coefficient than an identical externally-mixed counterpart. The amount of absorption enhancement varies with the particle size and coating thickness, but factors of two or more absorption enhancements are reasonable. Recent measurements utilizing the SP2 instrument indicate that roughly half of biomass burning and background aerosols are internally mixed (e.g., Schwarz, GRL, 2008).
It is not clear to me that a filter-based instrument can differentiate between internally and externally mixed particles because the EM field on the particle changes when it touches the substrate. Therefore, any enhanced absorption associated with internal mixing of atmospheric carbon particles might not be detectable with a filter measurement.
We agree that filter-based measurements may have issues with coated particles and may not report the absorption properly and we mention this in the text (original submission, lines 376-379). Coatings appear to increase the absorption measured by the PSAP, but we don't know by how much – i.e., whether it is more or less than the actual absorption enhancement observed in the atmosphere. The problem is that the effect of the interaction of coated-particles with filters on the measured absorption is unknown – the literature we've seen suggests that filter-based measurements of absorption tend to be *higher* than those made by non-filter based methods. As we already discuss in the text, Lack et al. (2008; 2012) show absorption enhancements for PSAP measurements relative to photo-acoustic (PAS) measurements which are not filter-based.

In the Lack studies, the fact that the PAS observed less absorption than the PSAP is unsettling as the PAS absorption measurements should also include the effects of coatings. Lack et al., (2008) also showed that the discrepancy between the PSAP and PAS increased as the organic aerosol concentration increased. We should also note that, thermally-denuded, filter-based measurements of absorption are lower than non-denuded absorption measurements suggesting the filter-based measurements are capturing at least some of the contribution of the coating. Comparisons of filter-based absorption measurements for denuded and un-denuded particles (e.g., Kanaya et al., 2013; Sinha et al., in revisions, 2017) suggest the un-denuded particles have absorption enhancements of 5-25% relative to those that have been through a denuder. These comparisons show that stripping off coatings and evaporating the non-absorbing particles reduces the measured absorption, i.e., that the effects of coatings is not completely lost in filter-based measurements.
We're left with considerable uncertainty and a need for more research to really understand the effects of coatings on absorption measured with filter-based instruments. We've added the information about the denuded vs. non-denuded filter-based measurements to the text in section 2.4.1 and have doubled the uncertainty of the PSAP measurements to account for this (see our response to your comment about the coating uncertainty below).

The altering of aerosol absorption is further aggravated by the drying process, which can modify aerosols by removing coatings that are composed of semi-volatile compounds, inorganics, and certainly water.
We also agree that the drying process we use (gentle heating of 40 C or less) may remove some volatile components which could affect both the scattering and absorption measurement. We've added the following text:
"One aspect of the in-situ system that will affect both the scattering and absorption measurement is the gentle heating used to dry the particle to RH<40%. The drying process we use (heating of 40 C or less) may remove some volatile components but we believe the removal to be minimal (<10-20%) based on lab and ambient volatility studies in the literature. Thermal denuder studies suggest little removal of volatile components (<10%) at 40 C (e.g., Mendes et al., 2016; Hakkinen et al., 2012; Huffman et al., 2009, Bergin et al., 1997) although thermal denuders results may be limited by short residence times (<20s). However, smog chamber evaporation studies on ambient aerosol over longer time periods (minutes-hours) at ambient temperature also suggest ambient aerosol may be less volatile than previously thought – Vaden et al. (2011) showed that ambient SOA lost just ~20% of its volume after ~4h."

The authors state on line 406 that "measured absorption Ångström exponents are quite low (close to 1) suggesting little influence of coatings." This is a weak argument, however, because the absorption Ångström exponent is measured after the drying process, which could have easily removed semi-volatile coatings.
Additionally, coated carbonaceous particles of about 0.1 um diameter and a wide range of coating thicknesses can easily have absorption Ångström exponents close to 1, per Gyawali (ACP, 2009, Figure 8).

Figure 8 in Gyawali shows that for a given core size, a wide range of non-absorbing coating thicknesses (from no coating to 4x the size of the core) can have the same AAE (within a factor of 0.25 which is the contour thickness). This suggests if some (even a lot) of the non-absorbing coating is removed the AAE won't change significantly. Even in the case of a slightly absorbing coating (Gyawali's Figure 9), 50% of the coating mass would have to be removed to see a significant change in AAE (we estimate that removing 50% of the mass of coating would change the AAE by ~0.25 which is one contour line in the Gyawali's plots). But we agree with the reviewer that the statement is weak since we know nothing about the coating (thickness, composition, geometry, level of mixing, etc). With the limited measurements available to us we can't say much more than we already have about the effects of coatings. We've deleted the offending phrase!

Finally, the authors conclude on line 434 that ". . . it is not possible to estimate the actual uncertainty in the in-situ light absorption measurements reported here due to coating effects." Since accurately estimating this uncertainty is not possible, the authors seem to be choosing an uncertainty of zero; I believe that many readers will not be satisfied with this approach. Hence, I suggest that the authors include a more conservative uncertainty of 50% for the PSAP absorption coefficient (rather than 25%) to account for all possible PSAP errors (you've already included the 50% value in much of your discussion on pages 10-11). This number still comes out of thin air (just like zero), but you can point to the 50% enhancement factor recommended by Bond and Bergstrom (AST, 2006). Thus, I recommend including a comment on line 445 that if the PSAP absorption uncertainty is 50%, the SSA uncertainty at SSA=0.9 is approximately 0.06.

We've redone the uncertainty calculation as suggested. We've also added this sentence to the discussion of the PSAP uncertainty due to coating in Section 2.4.1:

"To address this, we double the assumed PSAP uncertainty of ~25% to 50% in the calculations of uncertainty."

Optical and Aerodynamic sizes

The authors state that their inlet samples particles with aerodynamic diameters less than 7 μm, and that the particle diameter for 50% sampling efficiency is 5 μm (lines 196-198). They go on to address this issue on lines 601-617 by using the AERONET size distributions to estimate the fraction of aerosol extinction that occurs at AERONET diameters of less than 5 μm. By doing this, the authors are assuming that aerodynamic diameter is equal to volume equivalent diameter, which is not generally the case (unless the aerosols are spherical and the aerosol density equals 1). Since: $D_{vol} \propto D_{aer}/p^{1/2}$, the authors should use a much smaller cutoff diameter than 5 m in their equation on line 606. For instance, if one assumes = 2, then Dvol = 3.53 for spherical aerosols; if = 2:7, then Dvol = 3.04. These smaller cutoff values will decrease their estimated extinction fractions of the particles entering their inlets. However, the authors correctly argue on lines 612-617 that under sampling large particles probably lowers the in situ SSA, and that correcting this artifact will probably make the discrepancy between AERONET and in situ worse. Nonetheless, some recognition of the optical/aerodynamic size difference will indicate to readers that you considered this issue.

This is a good point and we've updated the paragraph in section 3.1.2 (lines 601-617 in original submission) to reflect this. We suggest that the inlet cutoff would be closer to 3 or 4 um. A quick survey of the literature (e.g., Kannosto et al., 2008; Topping et al., 2011; Zhang et al., 2016) suggests a value of 1.5 g/cm3 is probably a reasonable value for ambient aerosol density which would lead to a volume equivalent size cut of 4.1 um (density of ammonium sulfate is 1.77, density of secondary organics is often assumed to be 1.4, water is 1). If we assumed all of the aerosol was AmSulf, the size cut would be 3.75 um. The values we reported in the initial submission were a rough calculation and actually correspond to diameter<4um due to the rather coarse AERONET size bins. We've changed the numbers in the paragraph to reflect a cut size of 3um as a worst case scenario for this effect. The mean extinction fraction shifts down by 1-2% to 0.88+/-0.09 for SGP and 0.93+/-0.07 for BND.

Tables and Figures

I like Table 1, but the authors need to include the number of AERONET retrievals and the number of in situ flights that are used to compute these statistics. Standard deviations of the means (SDOM) would be nice, too, so that readers can quickly see that the difference between the averages are statistically significant.

Standard deviations are already included in Table 1. The first line in each row is the median, the second line in each row is the mean and standard deviation of the mean. We've added a final row with the number of retrievals and flights. The number of retrievals and flights is the same for all 3 parameters as we wanted the values to be for the same data points - the values in Table 1 represent the blue points in figures 3-5.

We've also added a Table 1b now which includes the values for the purple points (i.e., AERONET AOD440 > 0.2)

I am having difficulty understanding Table 2. You list 56 retrievals with Level 2 AOD + almucantar at BND, but only 6 retrievals with Level 2 + almucantar + AOD440 > 0:2. The 2nd number should be greater than the first number, right?

(The first number only includes AOD > 0.4, but the 2nd number includes all retrievals with AOD > 0.2). Unless... do you mean "size distribution retrieval" instead of "almucantar retrieval?" The almucantar products include size, AAOD, SSA, and complex refractive index; size distributions are the only Level 2 almucantar products available at AOD(440) < 0.4.

You've kind of answered your question – but clearly we need to be more clear. There are three sections to the table: (1) number of profile flights (one row); (2) AERONET statistics for AOD for various constraints (three rows); and (3) AERONET statistics for AAOD for various constraints (three rows). For the specific values you ask about, the '56' includes observations with AOD<0.4, i.e., it includes V2 Level 2 values where there was a successful almucantar retrieval (of any property) – for these, no AOD threshold is applied. In contrast, the second number '6' represents how many of the 56 points correspond to AOD values >0.2.

To address this we've add further demarcations in the table splitting the 3 sections. We've also added a footnote stating that an almucantar retrieval does not necessarily imply an AAOD retrieval. Finally we've updated the numbers in Table 2 both to reflect the fact that we removed 3 flights at each location due to the potential for aerosol aloft and also so they represent the

number of flight matches (rather than overall number of retrievals as there are variable numbers of retrievals/flight)

The legend in Figure 1 is confusing. The legend contains two solid black lines (Direct RF BCFF and Global), but there is only one solid black line in the figure. The legend also contains two variations of dashed black lines (Land and Sea), but there are no black dashed lines in the figure. The figure contains two red dashed lines that are not shown in the legend.
We're sorry if this was confusing – the idea was that the colors indicate the variable ("AOD", "direct RF, all comp" and "direct RF, BCFF", while the line style represented whether it was global, or just over land or sea. We've remade the legend to make this more clear.

The authors need to explain that "r" represents auto-correlation in Figure 2, as it is not obvious from the caption.
We've augmented the caption as suggested. There is now a caption sentence that says:
"The value r(k) on the y-axis represents the autocorrelation at lag time 'k'."

Also, presumably the correlograms correspond to scattering and absorption coefficients measured with the airborne nephelometer and PSAP; that should also be stated in the caption.
Figure 2 – as stated in the caption - represents the continuous surface measurements at each site not the airborne measurements – it's difficult to look at autocorrelations with non-continuous data – AERONET has lots of gaps due clouds and nighttime and the in-situ profiles are even more 'gappy' than AERONET.

Line-by-line Details
The authors lament on line 27 that the terminology for absorbing aerosols are imprecise, but Andreae and Galencsér (2006) provide precise definitions for soot, soot carbon, brown carbon, light absorbing carbon, elemental carbon, apparent elemental carbon, black carbon, and equivalent black carbon. It would be nice if the community embraced this paper as the "go-to" article for carbonaceous aerosol definitions.
The article cited (Petzold et al 2013) expands upon the important work of Andreae and Gelencsér (2006) and makes specific recommendations for terminology for carbonaceous aerosol as a function of measurement technique. We've added the Andreae and Gelencser reference.

Line 77-78: the authors state ". . . by invalidating low AOD cases, the AAOD values that are retained in the AERONET Level-2 data may be biased high." This is not quite correct. It would be more accurate to state that ". . . the averaged AAODs" or ". . . the climatological AAODs" are biased high.
Changed the sentence to read: "…by excluding low AOD cases, the climatological statistics of AAOD derived from the AERONET Level-2 data may be biased high."
We've also changed a similar sentence that occurs later in the paper (lines 817-818 original submission)

Lines 82, 93: Authors mention several Aerocom models, but do not tell us which ones or which Aerocom experiments they are pulling the data from.
The model information was in the caption of Figure 1 but we've now added the model names to the text and additionally noted that the model simulations were from the AeroCom Pha II control experiments.

Line 185: How much horizontal distance is covered in these 5- and 10-minute flight legs? This is important for readers to understand the auto-correlation within the profiles.
Good point – we've added this to the text. Airplane speed was ~50 m/s resulting in the 10 min upper level legs being approximately 30 km long and the 5 min lower level legs approximately half that (15km) length.

Lines 192-194 states that there are 253 complete profiles at BND and 132 complete profiles at SGP. However, only 24 and 14 flights were used in the analysis of Figures 3-6. I think that it is important to mention this here, because this description sounds like you have a much larger dataset available than you really do. This might be a good opportunity to mention that the stringent AERONET cloud screening procedure drastically reduces the number of possible comparisons, and this must also be considered by modelers when they utilize AERONET retrievals.
The intention was to first separately describe the in-situ and AERONET measurements and then in Section 2.3 we describe the merging of the data sets which leads to reduced numbers of comparisons. We've added a sentence after lines 192-194 (original submission line numbers):
"The number of flights that could be compared with AERONET measurements is significantly less than this, as discussed in Section 2.3 where the merging of the AERONET and in-situ data sets is described"
We've also added these sentences to the first paragraph in section 3.1.1:
"The low number of points on Figures 3-5 and in Table 1 indicate both the effects of AERONET stringent cloud screening routine and the constraints imposed by the almucantar retrievals. In addition to limiting the number of comparisons available in this study this limited data availability also has implications for modellers utilizing AERONET data – for example, Schutgens et al. (2016) has shown the importance of temporal collocation in measurement/model comparisons."

Lines 322-325: The authors state:
"Because the profiles are "stair-step" descents from 4600 m asl down to 450 m asl (e.g., see Figure 4 in Sheridan et al., 2012), matching with AERONET retrievals at the end of the profile means that the matches are more closely aligned with when the airplane is in the boundary layer and thus, typically, sampling the highest aerosol concentrations." This is inconsistent with other statements in the paper. The flight takes 2 hours and the authors are matching AERONET retrievals within +/- 3 hours of the end of the profile. Hence, the actual retrieval could occur up to 1 hour before the flight profile begins, and anytime during the flight. Thus, the AERONET retrievals are not necessarily "more closely aligned when the airplane is in the boundary layer.
We've added the following text:
"This way the maximum time difference between the boundary layer portion of the flight and the AERONET retrieval is 3 h; if we'd chosen to match based on the start of the flight the maximum

time difference between the boundary layer measurements and the AERONET retrieval could be as large as 5 h."

Line 359: I don't understand how the authors can test γ +/-1 standard deviation when they do not measure γ.
We've added the following text: "As described above, γ was calculated from the climatological chemistry measurements made by the IMPROVE network (14 years of data, ~1700 data points at BND; 10 years of data, ~1000 data points at SGP) using the Quinn et al. (2005) parameterization. We calculated the mean and standard deviation of γ based on those climatological chemistry measurements."

Lines 360-361: Authors estimate the uncertainty in AOD by varying by up to 2 S.D., but there are hundreds of sunphotometer AODs available during their flights (per figure 3). Why not just compare their in situ AODs to AERONET AODs?
Here we were trying estimate the uncertainty in the AOD specifically due to the relative humidity adjustment. We can't do this by comparing to the AERONET AODs as other factors may also contribute to discrepancies between insitu and AERONET AOD. We have not changed the text.

Line 379: The Lack 2008 citation is not included in the references.
Oops! Added.

Lines 434-435: Authors state that it is not possible to estimate the uncertainty in measured coatings due to coating effects. Consequently, they seem to be choosing zero uncertainty associated with coatings. Many readers will not agree with this choice.
As noted above we've redone the uncertainty calculation including an additional uncertainty to account for the unknown effect of the coating.

Line 516: Authors state that AERONET AOD tends to be higher than in situ AOD in Figure 3; however, the slope is only greater than 1 in one of those figures, although the offset is always positive. So AERONET > in situ AOD is not obvious, especially since many of the regression lines in Figure 3 fall below the 1:1 line. Additionally, AERONET median and average AODs are almost always less than the corresponding in situ AODs in Table 1, indicating that line 516 is not correct for the medians or the averages.
Table 1 has new numbers based on some other analysis and Figure 3-5 are different as well. We've let the sentence stand as it definitely relect the new analysis. Here's a table describing the AOD flight info for the revised paper (it's not included in the paper):

	BND	SGP
RED	#total points=629 #points above 1:1=441 median aod ratio=1.99	#total points=347 #points above 1:1=202 median aod ratio=1.03
BLUE	#total points=21 #points above 1:1 = 16 median aod ratio=1.22	#total points=11 #points above 1:1=6 median aod ratio=1.01

Also, please note that the values in Table 1 in the manuscript have changed for two reasons: (1) we are now using the flight averaged AERONET values (2) we eliminated 3 flights from each

comparison due to the potential for layers aloft identified using Raman lidar and/or shape of profiles (see details in other responses to reviewer comments).

Line 519: Should be changed to “The correlations improve when subsetted for the more restrictive Level-2 almucantar retrievals." Otherwise, it sounds like the almucantar scans are incorporated into the AOD measurements, which is not the case.
We changed the sentence to read:
The R^2 values increase when sub-setted for the more restrictive Level-2 almucantar retrievals.

Lines 541-543:
Authors state: “although the scatter in the relationships (particularly at BND) suggests that a multiplicative factor doesn’t represent the relationship very well." I agree that the BND data looks quite scattered, but the SGP values do not look scattered. The R2 values are 0.76-0.77, which is excellent compared to the autocorrelation values for absorption provided in Fig 2 (rabs = 0.45-0.55). Additionally, the scatter shown for AERONET AAOD is deceiving. There are approximately 2 AERONET retrievals for each in situ flight. The in situ flight represents a 2-hour average, but each AERONET retrieval occurs over the period of a 10-minute almucantar scan. The AERONET scans could occur up to 6 hours apart, and up to 3 hours after the flight is complete. Finally, R2 = 0.3 implies that R = 0.55. Thus, the correlation in Figs 4 and Fig 2 at BND are not very different.
The SGP AAOD linear fits have a large y-intercept, which is why we stated that a multiplicative factor doesn’t represent the relationship very well. The BND data in addition to looking less linear than the SGP data also have a large y-intercept.
The reviewer is right that some of the scatter in Fig 4 could be due to our matching criteria and that we are comparing averages derived from 10min almucantar scans with 2h averaged flight profiles. Both of these issues could contribute to the observed scatter in Figs 3-5 if there are changes in the airmass on short timescales (<1h). But we thank the reviewer for pointing out that the correlations are not that dissimilar after all – we’ve now added the following text to the discussion of figure 4:
“Surprisingly, while the BND site has higher 3-hour autocorrelations for absorption than SGP (R = 0.75 for BND and R = 0.55 for SGP, per Figure 2), the results for BND in Figure 4 indicate less correlation than at SGP for absorption. Nonetheless, the correlation coefficients for BND in Figure 4 (R^2=0.49 (blue) and 0.37 (red) correspond to R = 0.70 (blue) and 0.61 (red)) are not that far from the 3 h auto-correlation of R=0.75 for absorption at BND in Figure 2.”

Also just to comment on the timing of retrievals - at both sites the AERONET retrievals tended to occur during the flights. At BND the retrievals generally occurred between 1-2 h before the end of the flight (i.e., at the start or during the flight), while at SGP the retrievals generally occurred 40 min before the end of the flight (i.e., around the time the plane was entering the boundary layer). For flights with multiple retrievals the difference between first and last retrieval was <2h at SGP and typically 3h or less at BND (two flights had a ~5h range). We haven’t added this information to the text though.

Lines 574-576: Well... you don't really have enough data to draw this conclusion, in my opinion. Why not include the averages for the purple points in Table 1, though?

This comment was directed at this statement: "As with Figure 4, the purple points on Figure 5 indicate when the AOD440>0.2; there does not appear to be an improvement in the relationship between in-situ and AERONET SSA when only these purple points are considered."

We've changed the sentence to read: although there aren't enough points to draw a robust conclusion, there does not appear to be an improvement in the relationship between in-situ and AERONET SSA when only these purple points are considered."

We've also added Table 1b which includes the values for the purple points.

Line 579: "Figure 3 shows that the AERONET AOD may be slightly larger than the in-situ AOD." Again, this is inconsistent with Table 1.

We've rewritten the sentence to read:

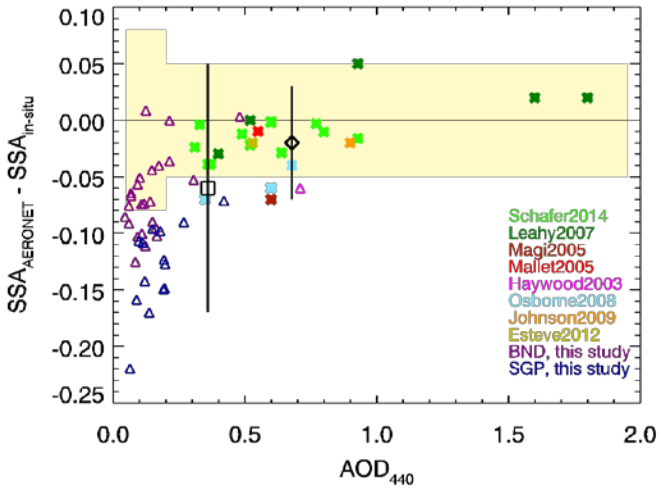
"Figure 3 shows that the AERONET AOD is similar (SGP) to or be slightly larger (BND) than the in-situ AOD,"

Line 644: It should be noted that 100 km is a long ways away! What kind of autocorrelations did Anderson (1998) find at those distances? That value should be noted here.

We've added the following text:

For non-plume data sets, Anderson et al. (2003) found autocorrelations $\geq$ 0.8 at 100 km (their figure 6). For plume-influenced data sets they found autocorrelations ~0.6**.**

Line 673: Authors state that SSA differences are greater than would be expected from random error, even when AOD > 0.4. Why not quantify that? That is, compute the mean and 2xSDOM of all independent measurements in that figure to see if the null hypothesis is valid; add this point with the resulting errorbar to the figure. It is important to isolate the independent measurements, though – multiple values that are obtained within 4-5 hours are not independent measurements (as you argued in Fig 2).



This is the plot the reviewer suggests. We've put two points with bars representing 2*SD on the plot. The plot was remade with flight average values for the BND and SGP flights so it doesn't include multiple points for a single flight. The black diamond represents just the literature

studies, while the square represents all the studies (lit+BND+SGP). If we assume that ΔSSA is normally distributed, we can use the characteristics of the normal distribution to say where ΔSSA is likely to fall. In the case of random error we would expect the values of ΔSSA to be evenly distributed above and below ΔSSA=0. However, based on where the standard deviation lines cross the ΔSSA=0 line, for both cases (lit and lit+BND+SGP) we ~80% of the ΔSSA points will be negative (while random error should lead to only ~50% of the points being negative). We can also calculate the confidence interval that a value will fall within based on t-statistics. The t-statistics suggest with 99.9% confidence that the ΔSSA literature values will fall in the interval 0.0 and -0.04 (i.e., the literature mean -0.02+/-0.02.) Similarly for the all data confidence interval of 0 to -0.12 there's a greater than 99.9% confidence that the data will fall in that range. We've added the following to the text:
"Figure 6 also shows the mean and 2*standard deviation of all of the points (black square and vertical lines) and just the literature value points (black diamond and vertical lines). Based on the characteristics of a normal distribution the standard deviation lines suggest ~80% of the points will be negative – random error would suggest only 50% of the points should be negative."

Page 10+: Since the authors are discussing non-synchronized data here, I think that they should note that the models and the in situ flights include cloudy periods, whereas the AERONET data are stringently cloud-screened.
We've added the following text in the second paragraph of section 3.3 before the statistical data comparisons are discussed:
"It should be reiterated here that we are comparing asynchronous data and that there are some additional differences amongst the data sets that need to be kept in mind: the AERONET data are rigorously cloud-screened and only obtained during daytime; the in-situ measurements are also daytime-only and the airplane did not fly in-cloud due to FAA flight restrictions, but may have flown near clouds; and the model data include day and night with clouds and also represent values over a 1x1 degree grid."

Lines 830-837: This is a nice approach in some regions, but it is virtually impossible to capture seasonal variability over North America with this approach. For instance, from 4/1994 through 10/2012 the Cart Site had 3 lev2 retrievals in DJF, 22 Lev2 retrievals in MAM, 122 lev2 retrievals in JJA, and 51 retrievals in SON. Thus, results would be skewed to the high humidity summer periods.
We agree that using AERONET to describe seasonality over North America (and specifically over BND and SGP) is limited by frequent cloudiness and/or the cleaner, dryer conditions prevalent over the US Midwest, particularly in the cooler months. We did note in lines 764-765 of the original submitted manuscript that 'During the cleanest months of the year (December-February) there are none to few Level-2 almucantar retrievals of SSA and AAOD at either BND or SGP.' We've expanded on that statement to mention that lower humidity during the winter also plays a role and pointed to the gray lines in Figure 7ab showing the lack of level2 almucantar retrievals in Jan, Feb, Dec (at BND) and Jan, Dec (SGP).

Lines 840-844: This passage needs some massaging, as it makes no sense to me.
This passage has been re-written and is (hopefully) clearer:

A similar, though statistical, approach was used in Bond et al.'s (2013) bounding BC paper in order to reduce uncertainty and better represent AERONET SSA and AAOD retrievals at low AOD. Bond et al. (2013) worked with AERONET monthly local statistics for the time period 2000-2010. Monthly values of AAOD and SSA at 550 nm were calculated from size distributions and refractive index when there were at least 10 valid inversion retrievals for that month at that site in the 2000-2010 period (most sites had more than 10 retrievals in a given month over the 11 year period). It was assumed in Bond et al. (2013), based on AERONET reported uncertainties, that the retrieved absorption-related values were more reliable at larger AOD and so they made some adjustments to account for this. For each site, AAOD and SSA values were binned as a function of AOD (there were five AOD bins, with each bin corresponding to 20% of the AOD probability distribution). For lower AOD conditions, the calculated AAOD and SSA values were replaced by values obtained during larger AOD conditions for the same month as follows: (i) the SSA and AAOD values corresponding to AOD_{550} of 0.25 were prescribed for all SSA and AAOD observations at lower AOD and (ii) for locations where all $AOD_{550}<0.25$, the average SSA and AAOD of the upper 20^{th} percentile of AOD observations at the site was prescribed for all lower AOD bins. Finally, the average of all five bins was used to determine the overall monthly average. In the case of AAOD the bin averages were simply averaged to get the monthly value while for SSA the AOD-weighted bin averages were averaged to get the monthly value. Note: the $AOD_{550}=0.25$ cutoff point corresponds (approximately) to $AOD_{440}=0.35$ for smaller particles and $AOD_{440}=0.25$ when large particles are present. This is less strict than the AERONET recommended constraint of $AOD_{440}>0.4$, but it had been suggested $AOD_{440}>0.4$ might be too restrictive (pers. comm., O. Dubovik).

Line 872: Emphasize that this comment only pertains to AOD < 0.2.
We've added the phrase "below $AOD_{440}=0.2$" to the end of the sentence

Line 1125: Müller (2012) is not included in the text. This is an important citation that needs to be presented in the text, as it discusses the difficulties associated with using in situ instruments to measure absorption.
Müller et al. (2012) is included in the text (line 127, original submission); however, Müller et al. (2011) was included in the citations but was not included in the text and should have been.
We've added the following sentence to the third paragraph of section 2.4.1:
"Müller et al. (2011) describe detailed experiments to characterize filter-based absorption instruments and describe some additional limitations of the instruments."

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.

Reviewer#3

We appreciate the reviewers detailed reading and commenting on the manuscript and hope we have address the concerns raised.

General comments: The authors present data of situ measurements from aircraft profile flights from which calculations of AOD, single scattering albedo (SSA), and Absorption Aerosol Optical Depth (AAOD) are compared to remote sensing measurements (of AOD) and retrievals of SSA and AAOD from AERONET sun-sky radiometers. These comparisons are made for two sites in the USA and for primarily low AOD levels, mostly less than 0.25 at 440 nm. This is well below the AERONET recommendations for use of absorption parameters from their retrievals (>0.4 at 440 nm is recommended), and although the authors discuss this in the text this recommended low AOD threshold is conspicuously absent from both the Abstract and Conclusions sections (and this needs to be remedied).

We've re-written the abstract and conclusions to reflect these points.

The authors state that in prior publications "... the in-situ derived AOD values tend to be slightly lower than the AOD retrieved from remote sensing measurements." They fail to point out that the in situ measurements rarely if ever measure the total column AOD, which includes both mid- to upper-tropospheric aerosol plus stratospheric AOD. The authors should include some references and discussion on the AOD that is not measured by in situ instruments in the upper troposphere and stratosphere since the aircraft do not fly complete profiles from the surface on upward into the stratosphere.

Not covering the entire column is indeed a limitation of all aircraft measurements. We've added some additional discussion of this issue in section 2.4.1. Additionally, we've also now used SGP Raman lidar data and also assessed the shapes of the profiles to better account for aerosol above the highest flight level of the aircraft. We've also added altitude ranges and information on how each campaign dealt with aerosol below (and above in the case of Magi et al. 2005) their flight profiles if that information was provided. Please see our responses to the specific comments related to this issue below.

Discussion of the fact that the aircraft profiles presented (with 4.2 km above ground level as the maximum in situ sampling altitude) do not actually measure the total atmospheric column AOD needs to be included in this manuscript. Therefore differences in in situ versus AERONET AOD are indeed expected and the AOD would be expected to be somewhat higher for sunphotometer total column AOD than for in situ in most aircraft sampling strategies. Moreover lidar measurements sometimes show mid to upper altitude aerosol layers that this aircraft sampling strategy (max at 4.2 km agl) would not measure.

See our response to the previous comment.

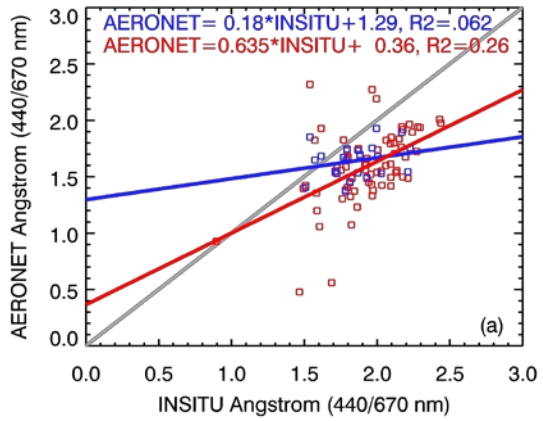
We should note that the AERONET and in-situ AOD are in fair agreement, whereas the AAOD comparisons look much, much different. Suppose the AOD discrepancy were entirely due to particles above 4.2 km agl – what SSA would those particles need in order to eliminate the AAOD discrepancy? We suspect that the required SSA is physically impossible, which means that missing particles can't explain the AAOD discrepancy.

Additionally, Sunphotometers in general and AERONET instruments in particular measure AOD more directly than any other technique and as such these data are considered by the scientific community to be the gold standard of accurate AOD measurement for the total atmospheric column. AERONET measured AOD represent the ambient aerosol optical properties and do not have to be corrected for RH humidification growth effects, loss of large particle sampling, etc. as is required and/or discussed for in situ data utilized in this paper. Nyeki et al. (2012) found that AERONET measured AOD agrees very well with other well-calibrated sunphotometers.
We apologize if any part of this paper came across as questioning AERONET AOD measurements. We recognize them as the gold standard for AOD and indeed our NOAA colleagues making solar radiation measurements have discussed this (e.g., Augustine et al., 2008)

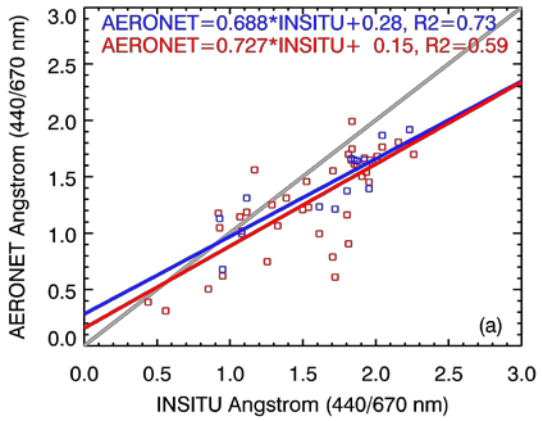
At the Davos, Switzerland site the comparison of the time co-located and matched 500 nm AOD differences between AERONET and GAW-PFR from 2007 through 2010 resulted in a mean AOD difference of -0.0024 and a root-mean square error of 0.0071. These issues should be included in the discussion on AERONET data, and in the section on comparison of AOD from AERONET measurements to in situ estimates. Accuracy of AOD is very important in this paper as AAOD is derived from AOD values and the AOD values derived from aircraft profiles (after corrections to make ambient estimations) and also from models (such as within AEROCOM) can be either biased or have significant uncertainties.
First, we'd like to correct a possible misunderstanding by the reviewer. The in-situ AAOD values are NOT derived using the in-situ AOD values. The in-situ measurements include a separate measurement of aerosol absorption and that absorption is what we integrate over the vertical range to calculate the in-situ AAOD value. Figure 3 was included to show that we can use the in-situ measurements to estimate AOD reasonably well.
As we say above, we recognize that AERONET is a gold standard for AOD measurements and have already noted the standard reference for AERONET AOD uncertainties in the text (e.g., Eck et al., 1999) as advised by our communications with the NASA AERONET scientists. The uncertainty of 0.01 for AERONET AOD is the same as the uncertainty Nyeki et al. (2012) report for the PFRs: "The combined uncertainty related to instruments and retrieval algorithms is estimated to result in an AOD uncertainty <0.010 at λ = 500 nm.". The AERONET AOD uncertainties are certainly less than those for in-situ AOD. The uncertainties in other variables (e.g., SSA and AAOD from both AERONET and in-situ measurements) are the important ones to consider because they are much larger than the AERONET AOD uncertainties.
In case the title of section 3.1.2 was confusing to the reviewer we've changed it to:
"3.1.2. *How might in-situ hygroscopicity assumptions and under-sampling of the aerosol affect SSA and AAOD comparisons?"*

Furthermore, in order to better understand the comparisons of aircraft profiles to AERONET measurements a scatterplot of computed Extinction Angstrom Exponent (EAE; 440 - 675 nm) estimated from the aircraft data versus AERONET measured EAE needs to be added to Figure 3. This comparison of EAE is pertinent to the discussion in the current text of aircraft inlet sampling issues regarding possible large-sized particle losses.

Esteve et al. (2012) presents a plot of airplane column Angstrom exponent vs AERONET Angstrom exponent for BND which shows the AERONET Angstrom exponents to be consistently lower than the airplane column Angstrom exponents (med_aeronet=1.53, med_airplane=1.82. Andrews et al., (2011) provides a statistical comparison of Angstrom exponent (their figure 3) from the airplane and AERONET at SGP and there's a similar offset of ~0.3 between AERONET and in-situ Angstrom exponent with AERONET being lower. Delene&Ogren, 2002 (their Fig 9b) shows that a difference of 0.3 in SAE corresponds to a difference of about 0.05 in submicrometer scattering fraction. For BND, this means that the supermicrometer scattering fraction might drop from 0.26 to 0.20, i.e., a 25% loss of supermicron-mode scattering. But since supermicron scattering is only about 20-25% of the total, losing 25% of the supermicron-mode means only a 5% loss in total scattering. This indicates that possible losses of supermicrometer particles has a minor effect on the in-situ AOD. We've also used the AERONET size distributions (lines 604-617 original manuscript) to evaluate super micron particle undersampling – the AERONET size distribution analysis suggests a 5-10% loss of total extinction.



BND angstrom comparison



SGP angstrom comparison

We've included the Angstrom exponent plots here for the reviewer, but as versions of them appear in other papers we have not added another figure to this manuscript.

There is a lack of discussion in the paper of how the uncertainties of the in situ measurements change as aerosol concentrations decrease. All measurement methodologies suffer from issues related to a decrease in signal at low concentrations (relative to potential instrumental noise and offsets), therefore I think that discussion of how the in situ measurement uncertainty changes with aerosol concentration is a very important aspect that needs to be included in the manuscript. Since the paper focuses primarily on low AOD cases, this is a critical issue that is surprisingly neglected in the current manuscript.

While BND and SGP are termed low loading sites in terms of their AERONET AOD climatology, the boundary layer aerosol loading is not typically low enough to significantly impact the uncertainty in the in-situ measurements. The uncertainty as a function of loading and averaging time for the in-situ measurements has been discussed in detail in many previous publications

(e.g., Table 2 in Sheridan et al 2002; Table 2 in Andrews et al., 2011; supplemental materials of Sherman et al., 2015). We already provided those references in the manuscript and have used their methodology to determine the uncertainty values reported here. For example, we state in the first paragraph of the in-situ uncertainty discussion:
"Sheridan et al. (2002) calculated uncertainties in aerosol light scattering for the TSI nephelometer to be 7-13% for 10 min legs depending on amount of aerosol present – the higher uncertainty value applies to very low aerosol loadings (scattering < 1 Mm^{-1})."
We've now added the following text to the in-situ uncertainty discussion:
"For the higher altitude flight segments the loading does tend to be quite a bit lower and thus has higher uncertainty but those upper-level segments contribute little to the overall AOD or AAOD. Because the flight column SSA is calculated using extinction-weighted SSA flight segments, segments with very low aerosol concentrations will have little impact on the column SSA derived from the flight measurements."

Additionally it is necessary to summarize in the text a description of the methodology used for computation of profile weighting of the in situ SSA estimates during each aircraft flight. Are these SSA values at each altitude weighted by the extinction coefficient at that altitude, thereby effectively giving higher weighting to the measurements at altitudes that had the highest aerosol concentrations? The AERONET retrievals of SSA are effective optical extinction weighted values for the total atmospheric column, therefore extinction weighting of the in situ data would be the most rigorous way to compare similar quantities**.**
We've updated the description of how flight column SSA was calculated:
"As described in Andrews et al. (2004), the in-situ column SSA (which is compared to the AERONET SSA value in section 3.1) was calculated for each flight level and then extinction-weighted and integrated to determine column SSA. This results in SSA values which are virtually identical to SSA values calculated using: $SSA_{col,in\text{-}situ} = (AOD_{in\text{-}situ} - AAOD_{in\text{-}situ})/AOD_{in\text{-}situ})$ and effectively gives higher weighting to the SSA values at altitudes that had the highest aerosol concentrations."

The authors also need to show plots of the in situ aircraft measured/computed SSA altitude profiles to show how SSA varies as a function of altitude for several days of varying AOD magnitude. This is important as it can provide some needed information on how the in situ SSA measurement data look at very low concentrations, especially higher than 3 km above ground level on very low AOD days and also on some moderately high AOD days.
Examples of profiles of multiple variables including SSA are presented in Figure 2 of Andrews et al. (2004). The AODs aren't noted in the text of Andrews et al (2004), but Figure 2a corresponds to an AOD_440 ~ 0.15 , Figure 2b corresponds to an AERONET AOD_440 of ~1.0 and Figure 2c corresponds to an AERONET AOD_440 of ~0.3. Box-whisker statistics for various parameters (including SSA) for the flight profiles can be found Andrews et al. 2004, Andrews et al., 2011a and Sheridan et al 2012.
Additionally, profile plots of various parameters (including SSA) for each individual SGP flight can be found here: http://www.esrl.noaa.gov/gmd/aero/net/iap/iap_profiles.html
And for the first two years of the BND flights can be found here:
https://www.esrl.noaa.gov/gmd/aero/net/aao/aao_prof2007.html

As these individual profile plots and statistics on the profiles are available in other locations, we have not included them here. We've added a sentence mentioning the availability of these plots in other locations:
"Profile statistics for various parameters including SSA are provided in Andrews et al. (2004, 2011a) and Sheridan et al. (2012). Individual flight profiles for various parameters are available online at: http://www.esrl.noaa.gov/gmd/aero/net/iap/iap_profiles.html (for SGP) and https://www.esrl.noaa.gov/gmd/aero/net/aao/aao_prof2007.html (for BND)."

In the abstract you state: "The tendency of AERONET inversions to overestimate absorption at low AOD values is generally consistent with other published comparisons." However the published comparisons between AERONET retrievals and in situ measured SSA shown in Figure 6 are not for low AOD (the AOD are moderate to high in the Figure) and also the SSA differences are generally within the combined uncertainty estimates of the two different techniques (see numerous additional comments on Figure 6 data below in 'Specific Comments'). Since it has never been established that the in situ measurements of SSA have no bias of their own then it is not possible to say that the AERONET measurements of SSA at moderate to high AOD are biased since there is no absolute benchmark for comparison purposes.
In order to more accurately reflect Figure 6, we've rephrased the sentence in the abstract to read: "The tendency of AERONET inversions to overestimate absorption at low AOD values relative to the in-situ measurements is generally consistent with other published comparisons across a range of locations, atmospheric conditions and AOD values." We've now noted in the abstract that the comparisons tend to fall within the reported uncertainty range. We've also rephrased the comments about bias to note that the in-situ measurements could be biased low. We feel it's important to note here that we do have absolute benchmarks for the accuracy of in-situ measured scattering (CO2) and absorption (various, PAS uses molecular absorption or scattering, EXT-SCA uses physical length). AERONET's absorption products lack such absolute benchmarks. However, we do not have characterization of bias vs random error in those benchmarks and our instruments that are referenced to those benchmarks. So our end conclusions are (a) that either AERONET overestimates absorption or INSITU underestimates it, and (b) there is bias in one or the other or both, because the comparisons in Fig 6 are not symmetrical about the "no-error" line.

Additionally since you state that the science is unclear on absorption enhancement due to coated absorbing particles (Section 2.4.1, line 404-407) you need to give a detailed explanation as to how this unknown factor was incorporated into the uncertainty estimates you made for the in situ measured single scattering albedo (it seems to have been ignored in your estimates). The aircraft sampled aerosols are dried first and therefore true atmospheric ambient state aerosol optical properties are actually not measured directly during the profiles. This is important regarding your claim of relative bias of single scattering albedo from one measurement type versus another since you cannot rigorously state (or indirectly suggest) that the aircraft measurements of single scattering albedo are unbiased given that the ambient state optical properties of the aerosols were not directly measured by your in-situ instruments.

There are studies (Lack, Cappa) that suggest coatings cause the PSAP to overestimate absorption, and numerous studies that suggest that coatings enhance absorption of suspended particles. If those coatings are lost or evaporate in our sampling system, then we would expect PSAP to underestimate absorption. As a result, we cannot treat the effects of coatings as a clear bias, as they could enhance or reduce the absorption measured by the PSAP. We should also point out that the particles are not completely dried or dessicated by the sampling system on the airplane. The heater only supplies enough heat to reduce the RH to 40%.

Based on the recommendation of another reviewer we've doubled the PSAP uncertainty to account for the effect of coatings, since the coating enhancement is unknown. We've added the following sentence in Section 2.4.1 when discussing the absorption enhancement:
"To address this, we double the assumed PSAP uncertainty of ~25% to 50% in the calculations of uncertainty."

The climatological comparison of in situ and AERONET values in Figure 7 is a very important figure in this paper. This figure suggests that if AERONET data are wisely utilized (as done by Bond et al. (2013), for example) then the SSA differences between the two methodologies can be relatively small. The large differences in the time matched data at Bondville and SGP sites from in situ flights and the AERONET retrievals shown in previous figures (Figs. 4-6) are not nearly as evident in Figure 7, especially for the SGP site. There is a surprising lack of discussion of this apparent discrepancy between the matched aircraft profile/AERONET data and the 'climatological' comparisons and the reasons for it. There is also a surprising lack of emphasis on the SSA comparison results shown in Figure 7 in the Abstract and Conclusions given the importance of this result.
We agree with the Reviewer that Figure 7 shows good agreement of monthly medians of SSA between AERONET Level 2.0 SSA and INSITU measurements and that is already stated in the text. This comparison is subject to considerable sampling bias, however, as we note in the discussion of the figure that the AERONET Level 2.0 almucantar data are restricted to more polluted cases with AOD440>0.4. Directly comparing the climatological AERONET Level 2.0 SSA with INSITU measurements requires an implicit assumption that SSA does not show a systematic co-variance with AOD, which does not seem to be valid (e.g., for in situ data sets: Delene and Ogren, 2002; Andrews et al., 2013; Pandolfi et al., 2014; Sherman et al., 2015 and for North American AERONET data sets: Schafer et al. (2014; their figure 6) and our own analysis as described in the text (Figure 8 and lines 855-861 of original submitted manuscript)). As a consequence, the combined results of Figure 7 and Figure 8 do not suggest that a "wise" utilization of AERONET data can minimize the differences between the two methodologies if "wise" implies the Bond et al., 2013 methodology of using SSA from high loading events and applying it to low loading conditions. Our Figure 8 suggests that a global climatology based on SSA measured at high AOD will lead to an underestimate of the global average AAOD. We've added the following text to the discussion of Figure 8:
"This relationship implies that a global climatology based on SSA measured at high AOD will lead to an underestimate of the global average AAOD."

Additionally it is very interesting that in Figure 8 the in situ surface measurements of SSA agree quite well with AERONET retrieved SSA for both sites, with excellent agreement at SGP site and within uncertainty bounds for the BND site except for extremely low AOD of less than 0.05.
We've replaced the dry surface measurements previously shown in figure 8 with those same surface measurements adjusted to ambient humidity using the hygroscopic growth parameterizations that were applied to the aircraft measurements. The surface ambient RH measurements used in the adjustment came from DOE/ARM at SGP (2m ambRH) and from NOAA/GMD at BND (10m ambRH). We did this so that shape of the three curves and the SSA values are more directly comparable. We've updated the figure caption and the paragraph describing figure 8 to reflect this change (lines 862-870 in original submission). We've also adding the following text:
"The AERONET SSA values are also lower than the surface in-situ SSA values – the surface in-situ SSA values adjusted to ambient conditions are quite similar to those obtained from the in-situ vertical profiles."

The authors have stated that the RH for the surface measurements are all <40%, although this is somewhat surprising given that the surface RH is typically >40% at this location, or perhaps measurements are never made when RH exceeds 40%? Or maybe the surface data that are shown are for the dried aerosol only? If so, you should apply the same humidification factors to the surface in situ data that you have applied to the aircraft profile data in this analysis to make the comparisons in Figure 8 consistent. A look at climatological data for Enid, Oklahoma and Ponca City, Oklahoma (same region as the SGP site) show daily average surface minimum RH of >40% and average Maximum RH of >75-80% for almost all days of the year.
The surface in-situ measurements are made at RH<40% for consistency with the GAW program protocols. As we note in the manuscript (original submission, lines 867-870) "…adjustment of the surface measurements to ambient conditions would tend to shift the SSA values upward (assuming absorption is not affected) and the scattering values to the right but would not significantly change the shape of the curve". We've now provided the surface data adjusted to ambient conditions so the shapes of the three curves and the SSA values are more directly comparable (see response to previous comment).

The surface in situ measured SSA to AERONET retrieval comparison result may be particularly interesting since the aerosol concentrations are often highest near the surface and therefore the in situ measurements made at the surface should have less uncertainty than those made at high altitudes where the concentrations may be very low. The authors should also present a comparison of the SSA and Extinction Angstrom Exponent measurements made at the lowest flight altitudes during profiles to those made at the surface by similar in situ instrumentation to show how good the agreement is between these measurements and to prove that the aircraft inlet sampling issues mentioned in the manuscript do not result in significant measurement uncertainties.
These comparisons are shown in Andrews et al (2004) and Sheridan et al (2012). Further, many of the sampling issues (RH adjustment, size cut, discrete flight levels) are discussed in detail in Esteve et al., (2012) as mentioned in the text (see for example lines 447-468 in the original submitted manuscript). For example, Sheridan et al (2012) shows plots of surface

measurements versus lowest level flight leg at 157 m agl. Their plots represent 5-min AAO low-level flight segment averages over the BND site vs. two-hour BND surface data centered on the flyby time. They show a slope of 0.87 for sub10um surface data vs the aircraft and a slope of 0.97 for sub1um surface data vs the aircraft. This suggests that the airplane measurements are capturing virtually all of the submicron aerosol but could be missing 10-15% of the super micron aerosol For scattering Angstrom exponent and SSA the slopes are 0.92 and 0.99 respectively. The airplane scattering Angstrom exponents are actually slightly smaller than the surface scattering Angstrom exponents which is the opposite of what might be expected. Andrews et al. (2004) show that SGP for an earlier version of the inlet with a 1um size cut the surface vs lowest level flight leg slopes were 1.02, 1.04 and 1.00 for sub-1um scattering, scattering Angstrom exponent and SSA.

Specific Comments:
Abstract: You also state: "We conclude that scaling modeled black carbon concentrations upwards to match AERONET retrievals of AAOD may lead to aerosol absorption overestimates in regions of low AOD." This statement is somewhat simplistic and mis-leading since it does not reflect the much better comparisons shown in Figure 7 for 'climatological' analyses. It also ignores the well thought out application of the use of AERONET retrieved SSA values as weighted by higher AOD observations and then applied to highly accurate AOD measurements at all AOD levels from AERONET, similar to the approach of Bond et al. (2013).
We've re-written the abstract significantly.

Introduction (lines 77-79): You state: "Moreover, by invalidating low AOD cases, the AAOD values that are retained in AERONET Level-2 data may be biased high." Again, it is misleading and simplistic to suggest that careful investigators would take the AAOD values from only Level 2 data and assume that they can be utilized as is. Many researchers have already utilized a much more intelligent approach: first estimate SSA at higher AOD from AERONET, and then apply those values to ALL levels of AOD (see Bond et al., 2013). I suggest that you remove or modify this sentence.
This sentence has been modified as suggested by Reviewer#1.
Changed the sentence to read: "…by excluding low AOD cases, the climatological statistics of AAOD derived from the AERONET Level-2 data may be biased high."
We've also changed a similar sentence that occurs later in the paper (lines 817-818, original submission).

Introduction (lines 142-143): I assume you mean Dubovik et al. (2000). Dubovik et al. (2002) is not in the reference list.
Correct. Fixed.

Section 2.1 (lines 185-187): Please elaborate what you mean by improving measurement statistics here. It would seem that the aircraft instruments 10-minute sampling rate at higher altitudes is an attempt to overcome issues associated with low aerosol concentrations and associated limits of instrumental sensitivity. Therefore on very low AOD days it would seem that an even longer time interval than 10 minutes would be justified. Please elaborate on the

sampling strategy and state whether it was modified for very low aerosol concentrations (very low AOD days).
The sampling strategy was the same regardless of loading. It is described in Andrews et al (2004), Andrews et al (2011) and Sheridan et al (2012). We updated the sentence about improving statistics to read:
"…in order to improve measurement statistics at the typically cleaner higher altitude flight levels."
We've also added the following sentence in the first paragraph of section 2.1 (actually now the second paragraph – we split the first paragraph into two):
"The pilot flew within the constraints provided (specifically-defined stairstep profile, vary the time of day, cross wind, over the instrumented field site, during daylight and not within clouds) but without day-to-day scheduling input from scientists."

Section 2.1 (line 194): "Only complete profiles were used in this analysis." Please state here that complete profiles as made by the aircraft do not equate to complete atmospheric profiles. None of the aerosol from 4.2 km agl through the stratosphere is sampled in the flights. Especially for very some low AOD days and (and also for some moderate-high AOD days in summer with strong convection) it is expected that a significant amount of the AOD actually occurs above 4.2 km agl. These upper aerosol layers that are often seen in lidar data may have different optical properties than lower altitude aerosols.
We've clarified that sentence and added a second sentence:
"Only complete profiles (all 10 flight levels) were used in this analysis. As is obvious from the vertical range of the flight levels, complete in-situ profiles do not equate to complete atmospheric profiles – this is discussed more in the in-situ uncertainties discussion (Section 2.4.1).

In the in-situ uncertainties section (section 2.4.1) we now discuss in greater detail the fact that the aircraft does not cover all the way up to the stratosphere. There is no lidar data available for BND, but we did retrieve the Raman lidar best estimate data product for SGP for the direct flights and compared the lidar extinction to the extinction obtained from the in-situ profiles. There were three SGP flights we removed from the comparison based on that analysis. We also took a harder look at the BND profiles and removed profiles that appeared to have increasing extinction at the highest flight levels as this was the smoking gun in the lidar comparisons for SGP. This is now discussed in more detail in the text of section 2.4.1.

Section 2.1 (line 216-219): Discuss how assuming a constant hygroscopic growth parameter would cause uncertainties when seasonal variation in aerosol type exits. Especially in spring, aerosol type may include biomass burning (crop waste or grass burning) and also dust from the Great Plains region (see Ginoux et al. (2010)), plus pollen from grass and trees.
It turns out that we'd been exploring this concept – we'd forgotten that we'd turned off the hygroscopicity adjustment for one BND flight because the hygroscopicity adjustment resulted in the flight's ambient in-situ AOD being ~2 times higher than the AERONET AOD (the value w/o hygroscopicity correction was within 0.01 of the AERONET AOD). This is the very high AOD point for BND (Blue point on fig 3, now labeled BB, with AERONET AOD440~0.5) and was

associated with smoke from wildfires in Canada being transported to the US Midwest (Flight date: June 28, 2006).
We now use the same hygroscopicity adjustment for that flight as we do for all the other flights, but we've labeled the point BB for biomass burning in Figs 3-5. We've added the following text to the manuscript about the issues with assuming a constant hygroscopic growth factor:
"While Equation 1 takes into account differences in hygroscopic growth due to RH for each segment of each flight, it does not account for compositional changes that might affect the scattering enhancement due to hygroscopicity. For aerosol events such as biomass burning and dust episodes with significantly different composition than the 'normal' aerosol we would expect to over-predict the aerosol hygroscopicity relative to the normal aerosol. Sheridan et al., (2001) showed that the SGP surface aerosol had lower hygroscopicity when it was influenced by dust or smoke."
We've also added this to the discussion of Figure 3:
"One thing to note on Figure 3a is the blue point marked BB (the BB stands for biomass burning). This measurement occurred on June 28, 2006 and appears to have been strongly affected by forest fire smoke transported from Canada. We applied the same hygroscopicity adjustment to the measurements of this flight as we did to all of the BND flights and, in this BB case, the hygroscopicity correction was the primary reason the in-situ AOD value is significantly higher than the AERONET AOD value. This point would lie much closer to the 1:1 line if the in-situ BB data were assumed to be hygrophobic. Previous work at the surface site at SGP has shown that dust and smoke aerosol types tend to exhibit lower hygroscopicity than the background aerosol normally observed at the site (Sheridan et al., 2001). This BB point provides an extreme example of the downside of using a constant hygroscopic growth parameter as a function of RH, although without additional information about the aerosol for each profile it is difficult to do otherwise. The light blue dotted line on Figure 3 represents the relationship between AERONET and in-situ data if the BB point is excluded."

Section 2.1 (line 267): Please change "column average properties" here to "flight profile average properties" to accurately reflect the fact that the aircraft does not measure the total atmospheric column, as AERONET does.
Done

Section 2.2 (line 290-292): Please add 'calibrations' before 'corrections' as the consistent high accuracy calibration of AOD and sky radiances are the basis for what makes AERONET data so valuable.
Done

Section 2.2 (line 294): Please add that the AERONET data is Version 2 data, since the Version 3 database will be available in the near future.
Done – we've also mentioned that the Version 3 data are coming in the same paragraph.

Section 2.4.1 (line 356-358): Please discuss whether you accounted for soil dust and biomass burning aerosols in the 'aerosol chemistry' mentioned here.

Our hygroscopicity relationship accounts for the hygroscopic growth based on the 'typical' aerosol chemistry - we did not specifically account for soil dust and biomass. We see little indication that the comparison flights were influenced by BB or dust (with the exception of one flight at BND which we now note in the text). We utilized the parameterization by Quinn et al. (2005) which uses the organic mass fraction (defined in Quinn et al as OC/(OC+sulfate) where OC= organic carbon concentration and sulfate = sulfate concentration) to estimate hygrosocopicity. She developed this parameterization based on chemistry and hygroscopicity measurements at several sites, including sites impacted by dust and biomass burning.
We've add some more details about this, in response to the reviewer's previous comment on this topic in both Section 2.1 and in the discussion of Figure 3. (See our response to previous related comment.)

Section 2.4.1 (line 361-364): In your discussion of RH levels during the profile flights please include some mention of the higher RH (RH halos) that typically exist in the vicinity of non-precipitating cumulus clouds that are imbedded within the aerosol layer < 4 km. Higher RH near cumulus clouds and higher AOD in the near Cu cloud environment (likely due to combined humidification, cloud processing of aerosols and rapid gas-to-particle conversions) were observed by Jeong and Li (2010) at the SGP site utilizing both AERONET data and in situ aircraft data. If you only flew aircraft profiles on cloudless time periods or avoided flying near clouds then this needs to be documented in the manuscript, as the sampling could possibly be skewed to specific meteorological and/or cloudiness conditions.
Thanks for bringing the Jeong and Li (2010) paper to our attention. I've also passed it on to the DOE Arm Aerial Facility manager as they try to keep track of papers using data from the IAP aircraft (e.g., Schmid et al., BAMS, 2014). We would not have expected to have AERONET retrievals available for comparison with the aircraft data under such conditions due to the rigorous cloud screening the AERONET Level 2.0 data undergoes. Jeong and Li (2010) made use of earlier measurements made by the same SGP aircraft flying the same profiles (albeit with a 1um inlet and max altitude of 3659 asl). Both the BND and SGP aircraft were operated under visual flight regulations and could not fly in clouds – they would skip a flight level if there was a cloud on that level and we did not use any flights that had missing flight levels in this analysis. We've now specifically mentioned this by adding the following sentence to the first paragraph of section 2:
"The flights at both sites were subject to 'visual flight regulations' which means they took place during daylight hours and the plane did not fly in-cloud."

We've also added another paragraph to the discussion of in-situ uncertainties and cited the Jeong and Li (2010) paper in there. Here is the text we've added:
"Jeong and Li (2010) have noted that the presence of nearby clouds may influence AOD values. They've investigated the effect of high RH-halos embedded in aerosol layers that typically exist in the vicinity of non-precipitating cumulus clouds. If the AERONET retrieval went through such a halo it could result in an increased AOD due to the combined effects of hygroscopic growth, cloud processing of aerosols and rapid gas-to-particle conversions. If the aircraft also flew through this RH-halo then the effect would also be accounted for in the RH-corrected in-situ measurements. However, if the high RH layer was between two flight levels then the aircraft

measurements would not account for it. Addressing this effect is outside the scope of this paper."

Section 2.4.1 (line 401-402): Although biomass burning does not have a consistent influence at these sites it is episodic, therefore did you exclude these biomass burning aerosol episodes from your data analysis? If so how did you identify the biomass burning episodes?
No data were excluded due to type of aerosol (biomass burning or otherwise). We did try to identify points that were affected by biomass burning and we note in section 3.2 line 705-706 that the BND point with AOD~0.4 represents a day we believe was affected by biomass burning.

Section 2.4.1 (line 443-445): Please state here that the uncertainty estimate for the in situ SSA of 0.04 is a lower bound since it does not take into account the effect of particle coatings on aerosols since the aerosols are modified (dried) before the measurements are taken, plus some fraction of the coarse mode particles are not sampled.
Based on the recommendation of another reviewer we've doubled the PSAP uncertainty to account for the effect of coatings, since the coating enhancement is unknown. We've added the following sentence in Section 2.4.1 when discussing the absorption enhancement:
"To address this, we double the assumed PSAP uncertainty of ~25% to 50% in the calculations of uncertainty."

Section 2.4.1 (line 451-454): You state here that 15% of the aerosol in the column is not sampled below the lowest flight altitude (150 m agl) of the aircraft for in situ measurements, and it can also be inferred that possibly another 15% or more is not sampled above the highest flight altitude on very low AOD days or high AOD days with layering from convective vertical aerosol transport. Therefore it is likely that 30% of the aerosol in the total atmospheric column is not sampled by your aircraft vertical profiles. This issue needs some discussion in the text and also should be factored into your uncertainties of in situ measured SSA (or clearly state that it has been ignored).
We've significantly augmented, rearranged, and rewritten this discussion of potentially missed aerosol below and above the aircraft as described below.

We're not sure that the reviewer's suggestion that the aircraft is likely missing 30% of the aerosol is reasonable. We based our comment that the aircraft could be missing 15% below the lowest flight level on previously published results for the comparing the lowest level leg (LLL) with the surface (S) measurements (e.g., Sheridan et al., 2012; Andrews et al., 2004; 2011; Esteve et al., 2012). We've now gone back and looked at the lowest level leg/surface comparison for just the flights included in the direct comparisons reported on here. At BND the relationship is LLL=1.0*S-0.99, R2=0.99 suggesting the lowest level leg and surface are seeing virtually identical aerosol. At SGP the relationship is LLL=1.17*S+0.43, R2=0.96, so for these particular SGP flights the airplane is actually seeing ~17% higher aerosol than is observed at the surface. The implication is that at SGP we may be over-estimating by applying the lowest level leg value down to the surface in order to obtain the column values. We've added the following text:

“We’ve looked at the surface/lowest flight leg relationship specifically for the flights with matching AERONET retrievals studied here. We found that at BND the surface and lowest level flight aerosol measurements were virtually identical. At SGP the lowest level leg actually measured slightly higher aerosol loading than was observed at the surface, which could lead to an overestimate of the aerosol optical depth in that layer, depending on the shape of the profile.”

Obviously if there are layers above the highest aircraft flight level they wouldn’t be sampled and that will negatively impact the AOD and AAOD comparisons. Depending on the layer loading that impact could be significant. It is however unclear to us how the reviewer can infer that the aircraft might be missing 15% or more above the highest flight on very low AOD days. Turner et al. (2001) segregated lidar aerosol extinction profiles at SGP by season and loading. Their results (their Figure 1) suggest that, for the vast majority of cases observed at SGP, 5% or less of the extinction will be found above 4 km. For low AOD cases (AOD_{355}<0.3) their mean extinction profiles suggest little to no aerosol extinction between 4-7km.
A 30% upward adjustment of the in-situ measurements would worsen the AOD comparisons shown in Figure 3 but not greatly improve the AAOD comparisons shown in Figure 4.

However, to further address this concern, 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.”

We’ve also added the following three sentences to the section.
“Missing aerosol above and below an aircraft profile is a potential issue in all aircraft/column comparisons.”

And
"Turner et al. (2001) segregated lidar aerosol extinction profiles at SGP by season and loading. Turner et al.'s results (their Figure 1) suggest that for the vast majority of cases observed at SGP, 5% or less of the extinction will be found above 4 km. For low AOD cases (AOD_{355}<0.3) their mean extinction profiles suggest no aerosol extinction between 4-7km."
And
"Regionally, seasonal average profiles from CALIPSO also suggest there is minimal aerosol above the flight's highest level (Ma and Yu, 2014; Yu et al., 2010)."
And
"Andrews et al. (2004) also assumed assumed an AOD contribution of 0.005 from stratospheric aerosol which was not done here."

Section 2.4.1 (line 464-467): Please elaborate here on whether the estimates of the percentage of aerosols above the highest flight altitudes as analyzed by Andrews et al. (2004) were comparisons made for all AOD levels and seasons. It would not be surprising for a greater percentage of AOD above flight altitudes to occur in summer when convection is stronger (transporting boundary layer aerosols upwards), or also in all seasons when AOD is very low since there is always some background midtropospheric to stratospheric AOD present which constitutes a greater percentage of total AOD when AOD magnitudes are very low.
The estimates in Andrews et al., 2004 were made by matching Raman lidar observations with each individual flight. The flights discussed in Andrews et al., 2004 covered all seasons and loadings. We've now utilized the Raman lidar data at SGP to further evaluate the potential high altitude contribution of aerosol as described in our response to the previous comment.

Figure 3: In Figure 3, please explain how you can have 3 observation points of AERONET measured AOD at 675 nm ranging from 0.5 to 1.0 at INSITU AOD of 0.15 when there does not seem to be any corresponding data at 440 nm in the plot above it. This does not seem possible, and should be explained in the text.
This is in reference to the BND plots. These 3 points match up with the flight on DOY 187, 2006 (July 6, 2006) for a flight ending at 187.80191. The AERONET values at 675 nm are: 0.493, 0.824, and 0.993 and the corresponding values at 440 nm are: 0.888, 1.464, 1.754. The two points greater than 1 are off the scale of the 440nm plot (Figure 3a) as it only goes up to 1. The 3rd 440 nm point (AERONET AOD value 0.888, insitu AOD value 0.32) is (barely) visible under the blue linear fit equation. I've added the following text to the caption:
"Note: two BND direct sun AOD440 points corresponding to the two highest AOD675 points in the figure below are off the scale of the plot and not shown. The third high AOD440 point is partly obscured by the legend."

Section 3 (line 528-530): Please also add to this paragraph that fact that the in situ under sampling of the total atmospheric column AOD is due to the restricted altitudes of the flight profile measurements (150 meters to 4200 meters).
We've added the following text in this section:
"Some of the discrepancy between the in-situ and the AERONET values may also be due to the limited vertical range covered by the airplane (150 – 4200 m asl)."

We've also included the reported altitude ranges and additional altitude information for all flights in Tables 3 and 4 (flight ranges are in column 2, additional altitude related info is in the comments column):
Schafer 2014: 250-5000 m (doesn't say if agl or asl) for column comparison flights average altitude range is 367-3339 m. They required flights to be less than 500 m and greater than 1500 m to obtain adequate representation of column.
Magi 2005: 170-1500 m agl
Mallet 2005: 100-2900 m (doesn't say asl or agl)
Leahy 2007: 100-5320 m asl (that's min and max over 5 flights – no flights covered that entire range). They used AATS to account for aerosol above plane and extrapolated down to acct for aerosol below plane. (Altitude range obtained from flight info in Magi et al, 2003)
Haywood 2003: 330-3420 m agl, extrapolated down to ground acct for aerosol below plane
Osborne 2008: 100-5000 m (doesn't say agl or asl) (that's min and max over 4 flights – no flights covered that entire range).
Johnson 2009: 150-3000 m (doesn't say agl or asl)
Corrigan 2008: 0-3200 m asl

Section 3.1.2 (line 604-607): Please note that the in situ instrument known cutoff of 5 micron for particle diameter for the aircraft sampling would also contribute to an under sampling of total column AOD, in addition to the incomplete altitudinal atmospheric profile for the total column AOD.
The first sentence of this paragraph has been adjusted to read:
'The other likely candidate to explain the in-situ AOD being slightly lower than the AERONET AOD is aircraft under-sampling of super-micron aerosol due to the 5 μm inlet cutoff'.

Figure 6 (numerous comments follow regarding some of the referenced data sets plotted in the Figure, especially note the issues regarding aircraft sampling and also the fact that some papers published Version 1 data that were biased due to inaccurate surface albedo assumptions, versus current Version 2 data that became available in 2006):
We utilized Version 2 data for all studies that used/reported Version 1 data. That is noted in the comments column of Table 4 for the relevant papers – that's what the note 'Used AERONET 2.0' was supposed to indicate. I imagine that could be confused with level 2.0 data so we've changed 'Used AERONET 2.0' to 'Used V2 AERONET Level 2.0'. We've also added the following sentence to the end of the first paragraph in Section 3.2:
"Please note that some of the earlier studies shown in Figure 6 and described in Table 4 used values from Version 1 AERONET data. Where that was the case, we retrieved Version 2 AERONET data from the AERONET website and those Version 2 data are what is depicted in Figure 6. The comments section of Table 4 mentions the cases where this was done."

Osborne et al. [2008] compared three cases of aircraft flights (on three different days) over the same site during the same experiment with the same instruments and aircraft but found that the aircraft in situ measured SSA values ranged from 0.04 to 0.07 higher than the AERONET version 2 retrievals. However, for all three of these cases the aircraft measured Angstrom exponents were found to be about 0.40 lower than the AERONET measured values. This

discrepancy in AE suggests that the aircraft may have sampled a different fine and coarse mode fraction mixture than the column integrated value measured by AERONET, and the higher SSA in conjunction with lower AE measured by the aircraft is consistent with this possibility. In fact, for the linear fit of SSA versus AE for all aircraft data from DABEX, reported in the work of Johnson et al. [2008], a difference of 0.40 in AE corresponds to a difference in SSA of about 0.06, almost the same value of the bias reported in Osborne et al. [2008].
It's already noted in Table 4 that there was a large discrepancy between the aircraft and AERONET AOD comparison and that the aircraft may have over-sampled large particles (or over corrected for large particles).

Johnson et al. [2009] compared in situ measured aerosol optical properties from an aircraft vertical profile flight over the Banizoumbou (Niger) AERONET site on 19 January 2006. This was a mixed aerosol case with Angstrom exponent (450–700 nm) of approx. 0.8–0.9 and high 550 nm AOD of approx. 0.75, where a shallow dust layer up to 1 km altitude was overlain by a layer of predominantly fine mode smoke. Both aircraft and AERONET measurements of column integrated AOD at 550 nm and of AE were in good agreement for this case, with dAOD = 0.08 (INSITU was 7% higher) and dAE = 0.06, suggesting that both were sampling the same aerosol mixture. The aircraft measured column mean SSA at 550 nm (from PSAP and nephelometer) was 0.87, in good agreement with the AERONET retrieval of 0.85 (interpolated to 550 nm).
These are the values reflected in Figure 6.

Magi et al. (2005; JAS) Note: Version 1 retrievals were 0.015 lower than V2 retrievals on this day at 1310 UTC at COVE site: From the paper: "Ground-based retrievals of SSA were obtained by the Aerosol Robotic Network (AERONET) sun photometers (e.g., Dubovik et al. 2000) during the CLAMS field campaign from a site known as the Clouds and the Earth's Radiant Energy System (CERES) Ocean Validation Experiment (COVE; 36.98 N, 75.78 W). The vertical profiles were often spatially located close to COVE. The mean value of SSA at 550 nm from AERONET retrieval data (processed to remove clouds and manually quality assured) is 0.94 +- 0.03. Therefore, the mean value of SSA retrieved from AERONET agrees with mean value of SSA derived from our in situ airborne measurements (0.96 +- 0.03) to within one standard deviation. On 17 July 2001, measurements were made from the UW aircraft and the COVE site that were both temporally (the aircraft vertical profile was from 1304–1337 UTC and the AERONET retrieval was at 1310 UTC) and spatially (the aircraft was ; 2.5 km from COVE) collocated. The mean value of SSA calculated from the airborne in situ measurements made in polluted layers during this vertical profile was 0.97 +-0.02; the corresponding column-averaged value of SSA for accumulation mode particles retrieved from the AERONET data was 0.90 +- 0.03 (VERSION 1 data). Particle losses in the sampling system for the in situ instruments could have contributed to an underestimate of the absorbing component of the aerosol. Spatial variability may have played a role as well."
As stated above and in the comments section of Table 4, we retrieved the Version 2 AERONET AOD440 values from the AERONET website.

Mallet et al. (2005): V2 almost same as V1 at 0.932 at 550 nm at 6 UTC: AERONET retrieval (at 0600 UTC ; on June 25, 2001) of the single scattering albedo at Avignon indicated a coherent

value (SSA 0.93 at 550 nm) compared to the one obtained from optical measurements for flight 41 (0515–0537 UTC, SSA 0.94 in the PBL).
As stated above and in the comments section of Table 4, we retrieved the Version 2 AERONET AOD440 values from the AERONET website.

Haywood et al., (2003): Comparison of aerosol size distributions, radiative properties, and optical depths determined by aircraft observations and Sun photometers during SAFARI 2000 V2 SSA = 0.84, 0.83, 0.81, 0.80 for 440, 675, 870 and 1020 nm V1 SSA = 0.88, 0.87, 0.84, 0.82 for 440, 675, 870 and 1020 nm The corresponding SSA for the mean size distribution used in the calculations derived from the PCASP distributions and from the nephelometer and PSAP on the C-130 is 0.90, 0.87, 0.85, and 0.82 (Table 1).
We thank the reviewer for pointing out our error – in Figure 6 we used the the V1 SSA value for AERONET of 0.88 when we should have used the V2 SSA value to be consistent with the other comparisons. The pink triangle for this campaign shifts to -0.06 (instead of -0.02). We have updated Figure 6 and the comments column of Table 4 accordingly.

Figure 7 caption: One detail regarding the caption of Figure 7 that is misleading is the Level 2 AOD shown in the plot. Please note that Level 2 AOD exists for all AOD levels not just for AERONET almucantar retrievals for which AOD is >0.4 at 440 nm. Level 2 almucantar retrievals of size distributions are made for all AOD levels, but refractive indices are only given for AOD>0.4 at 440 nm. Therefore the authors need to clearly describe and accurately label in the figure caption that this data as only associated with AERONET almucantar retrievals for which AOD>0.4 and therefore have error bars on SSA of 0.03.
The caption as originally submitted says “AERONET AOD medians are for observations with Level-2 almucantar retrievals, with corresponding AAOD and SSA retrievals at Level-1.5 (black) or Level-2 (gray). AERONET 2.0 values are biased high by definition, because of the AOD_{440}>0.4 constraint.”
We’ve added in the words ‘Level-2 almucantar’ and ‘AOD and AAOD’ in the following caption sentence to further clarify:
AERONET Level-2 almucantar AOD and AAOD values are biased high by definition, because of the AOD_{440}>0.4 constraint.

The actual complete Level 2 AOD data set (for all AOD levels) shows monthly means that are significantly higher in summer with many more days of data sampled and many more partly cloudy to mostly cloudy days sampled also (see Jeong et al., 2010; JGR for a discussion of higher AOD in the near cloud environments at the SGP site).
Figure 7 shows monthly medians not means. Below, in the other comment related to this point, I show a version of the AOD plots from Figure 7 that also includes the direct sky AOD medians – they lie directly on top of the 1.5* median values.

Section 3.2 (line 692-695): Your statements here assume that the in situ determinations of SSA are un-biased (despite the fact that ambient aerosol properties are not actually measured). This has not been proven in the paper, especially since the in situ data have to be corrected for

humidification effects, the total column aerosol is not sampled, and the effects of aerosol coatings are not accounted for (therefore blindly assumed have no effect**).** Please revise or eliminate these sentences.

We've rewritten the paragraph as follows:

"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 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."

Section 3.3 (line 754-768): Please note that AOD sampled by AERONET in the Level 2 dataset (not just for the subset that have L2 retrievals) includes many more days of data than the in situ flights, and is therefore a much more statistically robust data sample. Please note that the Level 2 AOD climatology for the SGP site (average of 13-19 years per month) shows significantly higher AOD (440 nm) than shown for L1.5* in Figure 7, For example for the SGP site the August monthly mean AOD is 0.272 and the September monthly mean is 0.215 at 440 nm. Similarly for the BND site the L2 monthly means of AOD(440 nm) for June, July, August, and September are 0.282, 0.329, 0.343 and 0.283 respectively (computed from 15-17 years of data per month). These monthly means are significantly higher than the AERONET values shown in Figure 7, since the data in Fig 7 are only AOD associated with the Dubovik and King algorithm retrievals.

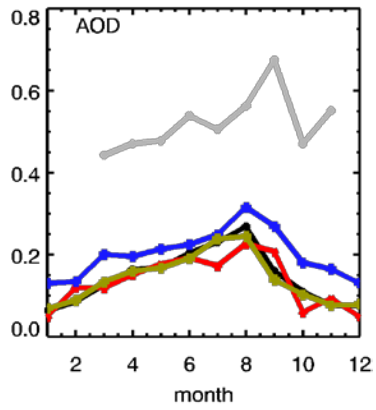
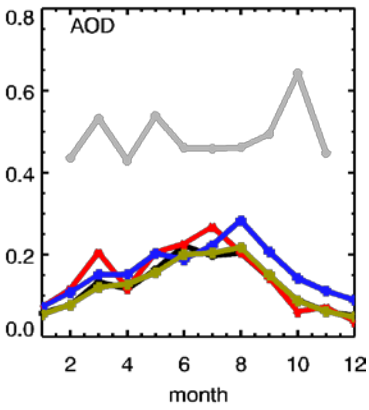
The plots in the manuscript show medians, not means. Below I've pasted the AOD portion of Figure 7 that also includes the medians for the version 2 Level 2 direct sky AOD measurements (in mustard). The direct sky medians lie pretty much directly on top of the 1.5* median AOD values (black lines). We have not added the direct sky AOD line to the plot in the manuscript.

This sentence (lines 739-740 of original manuscript) still stands:

"The AERONET Level-1.5* AOD monthly medians are representative of the direct sun AERONET Level-2 AOD climatology at the two sites."

We've clarified this sentence in the following paragraph by adding the phrase 'direct sky':

"…AERONET Level-1.5* retrievals (recall that the AERONET 1.5* AOD is representative of the overall direct-sky AERONET AOD climatology at each site)…"

		LEGEND AERONET 1.5* AOD Version 2, level 2 AOD for successful Almucantar scans with AAOD and SSA retrievals (i.e., has the AOD440<0.4 restriction) INSITU AOD AEROCOM (phase II) AOD Version 2, level 2 direct sky AOD
BND monthly median AOD values	SGP monthly median AOD values	

Section 3.3 (line 781-783): You state: “The AERONET 1.5* SSA values tend to be quite a bit lower than the other data sets at both sites, which is why the AERONET 1.5* AAOD values tend to be higher (recall that for AERONET data AAOD is calculated using AAOD=(1-SSA)*AOD).” No, this is not really accurate, since as shown on Figure 7, at the SGP site the agreement between the AERONET L1.5* data and in situ measurements of SSA are well within the uncertainty of the measurements for all months (and you have not proved that the in situ is not biased). Please revise this sentence to reflect this fact as presented by the data shown in Figure 7b.

We’ve revised this sentence to read: “The AERONET 1.5* SSA values tend to be quite a bit lower at BND, and somewhat lower at SGP which is why the AERONET 1.5* AAOD values tend to be higher (recall that for AERONET data AAOD is calculated using AAOD=(1-SSA)*AOD).

Section 3.4 (line 875-879): Again, you have omitted the fact that for the SGP site the agreement between the in situ estimations of SSA and the AERONET retrievals of SSA are within the uncertainty levels of these data sets over the entire range of AOD shown in Figure 8.

Note: we’ve remade Figure 8 so that the surface in-situ SSA values are also now at ambient conditions. We’ve added the following sentences to the discussion of Figure 8:

“It should however be noted that despite the discrepancy between in-situ and AERONET SSA values, Figure 8 shows that the SSA values for all three sets of measurements at SGP are within the reported AERONET SSA uncertainty range of 0.05-0.07 for AOD_{440}<0.2 across the narrow and low AOD range shown in the figure. At BND the SSA values are within the AERONET SSA uncertainty range down to AOD_{440}~0.1.”

The way the paper is written there seems to be a consistent attempt to steer the reader to the conclusion that the AERONET retrievals are biased low despite significant uncertainties in the in situ determinations and despite the fact that the in situ instruments do not measure ambient aerosol properties directly without corrections. Therefore it is not proven that the in situ determinations of SSA are unbiased themselves**,** so the text and title require rewriting to acknowledge this.

New title: “Comparison of AOD, AAOD and column single scattering albedo from AERONET retrievals and in-situ profiling measurements”

We’ve gone rewritten the abstract, text and conclusions to emphasize that (a) there is a systematic difference in the comparisons that would suggest a bias in one or both of the methods, (b) the majority of SSA comparisons for AOD440>0.2 are within the uncertainty bounds, and (c) there is a systematic relationship between SSA and aerosol amount (AOD or scattering) that should be considered in analyses of global-averaged AAOD.

~~Is there a bias in AERONET retrievals of aerosol light absorption at low AOD conditions?~~

Comparison of AOD, AAOD and column single scattering albedo from AERONET retrievals and in-situ profiling measurements

Elisabeth Andrews[1], John A. Ogren[2], Stefan Kinne[3], Bjorn Samset[4]

[1]CIRES, University of Colorado, Boulder, CO 80309, USA
[2]NOAA/ESRL/GMD, Boulder, CO 80305, USA
[3]Max Planck Institute for Meteorology, 20146 Hamburg, Germany
[4]Center for International Climate and Environmental Research – Oslo (CICERO), 0349 Oslo, Norway

Abstract

Here we present new results comparing aerosol optical depth (AOD), aerosol absorption optical depth (AAOD) and column single scattering albedo (SSA) obtained from in-situ vertical profile measurements with AERONET ground-based remote sensing from two rural, continental sites in the US. The profiles are closely matched in time (within +/-3 h) and space (within 15 km) with the AERONET retrievals. W~~-~~e have used Level 1.5 inversion retrievals when there was a valid Level 2 almucantar retrieval in order to be able to compare AAOD and column SSA below AERONET's recommended loading constraint (AOD>0.4 at 440 nm). While there is reasonable agreement for the AOD comparisons, ~~T~~t~~he~~se direct comparisons of in-situ-derived to AERONET-retrieved AAOD (or SSA) reveal that AERONET retrievals yield higher aerosol absorption than obtained from the in-situ profiles for the low aerosol optical depth ~~(AOD)~~ conditions prevalent at the two study sites. However, it should be noted that ~~–~~the majority of SSA comparisons for AOD_{440}>0.2 are, nonetheless, within the reported SSA uncertainty bounds. The ~~tendency of~~observation that, relative to in-situ measurements, AERONET inversions ~~to overestimate~~ exhibit increased absorption potential at low AOD values ~~-~~is generally consistent with other published AERONET/in-situ comparisons across a range of locations, atmospheric conditions and AOD values.~~-~~ This systematic difference in the comparisons suggests a bias in one or both of the methods, but we can not assess whether the AERONET retrievals are biased towards high absorption or the in-situ measurements are biased low. Based on the discrepancy between the AERONET and in-situ values, we~~W~~e conclude that scaling modelled black carbon concentrations upwards to match AERONET retrievals of AAOD should be approached with caution as it may lead to aerosol absorption overestimates in regions of low AOD. Both AERONET retrievals and in-situ measurements suggest there is a systematic relationship between SSA and aerosol amount (AOD or aerosol light scattering) – specifically that SSA decreases at lower aerosol loading. This implies that the fairly common assumption that AERONET SSA values retrieved at high AOD conditions can be used to obtain AAOD at low AOD conditions may not be valid.

1. Introduction

The amount and location of absorbing aerosol in the atmosphere is critical for understanding climate change (e.g., Hansen et al., 1997; Ramanathan and Carmichael, 2008; Bond et al., 2013; Samset et al., 2013). Ramanathan and Carmichael (2008) note the effects of absorbing aerosol (which they termed black carbon (BC)) on atmospheric heating rates, precipitation and weather patterns. (Note: The terminology used to refer to absorbing aerosol is imprecise (Petzold et al., 2013, Andreae and Gelencsér, 2006) and encompasses the terms describing chemistry, e.g., 'black carbon' (BC) and terms describing optical effects, e.g., absorption. The measurements reported herein all refer to light absorption.) The vertical distribution of BC can also influence its effect on climate (e.g., Haywood and Ramaswamy, 1998; Samset et al., 2013; Ramanathan and Carmichael, 2008). Single scattering albedo (SSA) is an indicator of the absorbing nature of the aerosol; higher SSA values indicate a more reflective (whiter) aerosol while a more absorbing aerosol will have lower SSA values. SSA is a primary determinant of whether the aerosol will have a warming or cooling effect (e.g., Haywood and Shine, 1995; Hansen et al., 1997; Reid et al., 1998). Uncertainty in the value of SSA due to uncertainties in the amount of absorbing aerosol can even prevent determination of the sign of aerosol forcing on local to regional scales. Bond et al. (2013) assessed BC as the second most important global-average warming species (top-of-atmosphere forcing +1.1 W m^{-2}, 90% bounds: +0.17 to +2.1 W m^{-2}) after CO_2 (in Bond et al. (2013) the direct effect of BC is 0.71, 90% bounds: +0.09 to 1.26 W m^{-2}).

Currently, the only way vertical profiles of aerosol absorption can be obtained is via airborne in-situ measurements. Such flights are expensive and tend to primarily occur during intensive field campaigns, which are usually aimed at studying specific aerosol types (e.g., biomass burning, African dust, urban/industrial pollution). This reliance on short-term campaigns results in profile data sets that are sporadic in both space and time, and not necessarily representative of typical conditions. Additional issues with airborne in-situ measurements include adjustment of measurements to ambient conditions, particle losses in sample lines, and instrument uncertainties. Nonetheless, in-situ vertical profiling of absorbing aerosols has provided useful information to modelers trying to understand climate effects, transport, and lifetimes of these important atmospheric constituents (e.g., Koch et al., 2009; Schwarz et al., 2010; Skeie et al., 2011).

The limited availability of in-situ vertical profile measurements means modelers must rely on globally sparse and/or temporally sporadic airborne measurements to evaluate BC vertical distributions in their models. Alternatively, the column properties retrieved from AERONET measurements and inversions have been widely used to provide a first constraint on modeled vertical aerosol properties (e.g., Sato et al., 2003; Koch et al., 2009; Bond et al., 2013; He et al., 2014; Wang et al., 2014). Use of the AERONET data as an absorption constraint has suggested upscaling of modeled AAOD values by a factor of 2-6 depending on location (e.g., Bond et al., 2013), although Wang et al. (2016) has shown that better spatial resolution of models and emission inventories can reduce some of the previously observed model/AERONET discrepancies.

Ground-based remote sensing of both direct attenuation and sky radiances permit inversions of atmospheric column averaged absorption. By retrieving the complex refractive indices at different solar wavelengths as well as the average aerosol size-distribution, absorption related properties can be determined (e.g., aerosol absorption optical depth (AAOD), single scattering albedo (SSA) and, absorption Ångström exponent (AAE)). The AERONET network has a fairly wide spatial coverage on land, with long data records at many sites (Holben et al., 1998; Dubovik et al., 2000; Dubovik and King, 2000). One obvious limitation of the AERONET inversion retrievals is that the uncertainty of the derived single scattering albedo (SSA) becomes very large at low values of AOD (Dubovik et al., 2000). To minimize the effects of this uncertainty, the AERONET Level-2 data invalidates all absorption-related values if the AOD at wavelength 440 nm (AOD_{440}) is below 0.4 (Dubovik et al., 2000; Dubovik et al., 2002; Holben et al., 2006). Unfortunately, this restriction greatly reduces the spatial and temporal coverage of absorption-related data that can be obtained from AERONET. Moreover, by ~~invalidating~~ excluding low AOD cases, the climatological statistics of AAOD ~~values that are retained in the AERONET Level-2 data~~ derived from the AERONET Level-2 data may be biased high.

Model analysis of global AOD values suggest that 95% of global AOD_{440} values are below 0.4 (Figure 1), while 89% of the AOD_{440} values over land are below the 0.4 threshold. Five models in the AeroCom suite (GMI-MERRA-v3, GOCART-v4, LMDZ-INCA, OsloCTM2, and SPRINTARS-v385) have reported daily-average values of AOD_{440} (for AeroCom Phase II control experiment), which can be used to develop a cumulative frequency distribution of the percent of the Earth's surface and days where a Level-2 AERONET retrieval of AAOD might be possible (ignoring the presence of clouds and absence of sunlight). Figure 1 indicates that, at best, Level-2 AERONET AAOD retrievals might represent 5% of the days, globally, and less than 11% of the days over land. In other words, the AOD constraint on Level-2 AERONET almucantar inversion retrievals means these retrievals represent only a small fraction of the Earth's surface and are biased to conditions of high aerosol loading.

The other information that Figure 1 provides is the fractional contribution of regions with different AOD_{440} amounts to the total aerosol and the fossil fuel black carbon (BCFF) radiative budget. These values were derived from monthly data from 4 models in the AeroCom suite. The fractional contribution to the radiative budget can be mathematically described as follows: for each model grid box there are three quantities: (i) the radiative forcing (W m^{-2}), (ii) the horizontal area of the box (m^2), and (iii) the AOD_{440}. The product of the radiative forcing term and area is the perturbation to Earth's radiative budget due to total aerosol (or BCFF) in the box. The sum of this product over all the boxes is the total perturbation. Figure 1 shows the fraction of the radiative budget perturbation as a function of AOD_{440}. It suggests that approximately 75% of the total aerosol forcing and 83% of BCFF forcing is due to regions of the globe where AOD_{440} <0.4. This highlights the significant contribution of aerosol in these cleaner areas to the total global radiation budget.

It should be noted that there is significant inter-model variation in the AeroCom cumulative AOD_{440} and radiative forcing plots shown in Figure 1. In particular the BCFF cumulative forcing fraction varies with the lifetime of BC predicted by the models. A long BC lifetime results in more

dilute AOD and BCFF radiative forcing distributions. Other issues include the fact that global models have limited spatial and temporal resolution, and generally simulate less variability in aerosol properties than is observed in measurements. However, all models used to generate Figure 1 follow the same general trend as is shown in Figure 1 with the take-away point being that AOD_{440} values >0.4 are a relatively rare occurrence.

Because of the potential of the AERONET absorption-related retrievals (e.g., AAOD and SSA) for understanding global distributions of absorbing aerosol, there have been many studies comparing AERONET retrieval values with those obtained from in-situ measurements in order to assess the AERONET retrieval validity. Such comparisons have taken several different forms. There have been direct comparisons where column SSA or AAOD values calculated from individual in-situ vertical profiles have been compared with AERONET retrieved values for retrievals close in time and space (Haywood et al., 2003; Magi et al., 2005; Mallet et al., 2005; Leahy et al., 2007; Corrigan et al., 2008; Osborne et al., 2008; Johnson et al., 2009; Esteve et al., 2012; Schafer et al., 2014). In addition to direct comparisons there have been general, statistical assessments between AERONET and in-situ measurements for both SSA and AAOD including: (a) comparing surface in-situ measurements with AERONET retrievals (e.g., Dubovik et al., 2002; Doran et al., 2007; Mallet et al., 2008; Corr et al., 2009); (b) comparing in-situ SSA (or AAOD) from a few flight segments to the corresponding column SSA (or AAOD) from AERONET (e.g., Kelektsoglou et al., 2012; Müller et al., 2012) and (c) comparison of statistical distributions or averages of AERONET retrievals for a given time period with airborne in-situ measurements (e.g., Ramanathan et al., 2001; Leahy et al., 2007; Andrews et al., 2011a; Ferrero et al., 2011; Johnson et al., 2011). Many of these statistical comparisons have shown good agreement between the AERONET and in-situ values. This ~~increased~~ increases general confidence in the AERONET retrievals. However, such statistical comparisons are not appropriate for the evaluation of the accuracy of individual retrievals.

The primary scientific question to be addressed in this paper is: ~~*Are the AERONET estimates for SSA biased low (or, alternatively, are estimated AERONET AAOD values biased high) under low AOD (AOD_{440}<0.4) conditions, or are they just highly uncertain?*~~*Is there a consistent bias observed between AAOD and column SSA obtained from in-situ profiling flights and AERONET retrievals?* The answer to this question may help determine the validity of adjusting model estimates of AAOD to agree with AERONET retrievals (e.g., Sato et al., 2003; Bond et al., 2013). It should be noted that AERONET does not recommend the use of absorption-related parameters (e.g., single scattering albedo, absorption aerosol optical depth, and complex index of refraction) at AOD_{440} below 0.4. Dubovik et al. (~~2002~~2000) suggests the uncertainty of AERONET SSA values more than doubles for AOD_{440} less than 0.2.

In what follows, we first evaluate how direct AERONET AAOD retrievals compare with those derived from multi-year, in-situ measurements obtained from vertical profiles over two rural continental AERONET sites in the U.S. Second, we create a summary of all direct AAOD or SSA comparisons between in-situ vs. AERONET data previously presented in the literature in order to place our results about AERONET aerosol absorption-related retrievals in a wider context. Finally, we look at the seasonality of in-situ, AERONET, and modelled (AeroCom) SSA

and AAOD values to see if the annual cycles can provide any insight into observed discrepancies in the direct comparisons. Because this study focuses on only two low AOD sites in the continental US which are unlikely to be generally representative of other low loading sites around the globe, and because other factors (e.g., Wang et al., 2016) may contribute to reported differences between modelled and AERONET AAOD we do not attempt to suggest implications for global BC forcing.

2. Methods

This study utilizes data from two sites with collocated AERONET measurements and multi-year, in-situ aerosol profiling measurements. The two sites are Bondville (BND, 40.05ºN 88.37ºW, 230 m asl) and Southern Great Plains (SGP, 36.61ºN 97.49ºW, 315 m asl). Surface in-situ measurements and AERONET column measurements have been made at both locations since the mid-1990s (e.g., Delene and Ogren, 2002; Sheridan et al., 2001: Holben et al., 1998). Weekly to twice-weekly flights measuring in-situ vertical profiles of aerosol optical properties over these two sites were made for a subset of the years of ground-based observations. At SGP the in-situ profile flights were centered over the site's central facility where the AERONET sunphotometer is deployed. Due to FAA flight restrictions, the BND in-situ profiling flights took place approximately 15 km to the WNW of the AERONET sunphotometer location at the BND surface site (Sheridan et al., 2012). Additionally, for BND, a low level flight leg (200 m agl) was flown directly over the instrumented BND surface site. The flights at both sites were subject to 'visual flight regulations' which means they took place during daylight hours and the plane did not fly in-cloud.

At BND and SGP, the median AOD_{440} values are 0.14 and 0.11, respectively (based on all AERONET Level-2 data from the start of AERONET measurements at each site). These median values fall right around the 50% mark on the AOD cumulative distribution plot (Figure 1), indicating BND and SGP may be appropriate sites to explore potential ~~biases~~ discrepancies between~~in~~ AERONET and in-situ AAOD and SSA retrievals at lower AOD conditions.

2.1 IN-SITU

The in-situ aerosol profiles were obtained with dedicated Cessna 206 airplanes flying stair-step profiles one to two times per week over the two sites. Between 2006 and 2009, 365 flights were flown over BND (out of a total of 401 flown in the region (Sheridan et al., 2012)), while 171 aerosol profile flights were flown over SGP in the 2005-2007 time period (Andrews et al., 2011a). The profiles consisted of 10 (at BND) or 12 (at SGP) level flight legs between approximately 450 and 4600 m asl (corresponding to approximately 150 and 4200 m agl). The profiles, which were 'stair-step' descents, took approximately 2 hours to complete as the airplane spent set amounts of time at each level (10 min/flight level for flight legs above ~1600 m asl and 5 min/flight level for flight legs below that altitude) in order to improve measurement statistics at the typically cleaner higher altitude flight levels. Airplane speed was approximately 50 m/s, resulting in the 10 min upper level legs being approximately 30 km long and the 5 min lower level legs approximately half that (15 km) length. This flight pattern means the last 30 min

of the profile were typically in the boundary layer for these two sites and encompassed the majority of the aerosol contribution to column aerosol loading. 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)

Descriptions of the flight profiles and aircraft package have been described in detail in other papers (Andrews et al., 2011a; Sheridan et al., 2012) so only a brief description is provided here. The pilot flew within the constraints provided (specifically-defined stairstep profile, vary the time of day, cross wind, over the instrumented field site, during daylight and not within clouds) but without day-to-day scheduling input from scientists. Here, we utilize the same 10 flight levels for both profiling sites: 457, 609, 915, 1219, 1829, 2439, 3050, 3659 and 4575 m asl. Of the 365 flights at BND, 253 flights had complete profiles (all flight levels) with valid scattering, absorption and relative humidity data; at SGP, 132 flights out of 171 were complete. Only complete profiles (all 10 flight levels) were used in this analysis. As is obvious from the vertical range of the flight levels, complete in-situ profiles do not equate to complete atmospheric profiles – this is discussed more in the in-situ uncertainties discussion (Section 2.4.1). -The number of flights that could be compared with AERONET measurements is significantly less than this, as discussed in Section 2.3 where the merging of the AERONET and in-situ data sets is described.

The aircraft were equipped with an inlet that sampled particles with aerodynamic diameter D_p<7 μm, and losses in downstream sample lines were estimated to reduce the particle diameter for 50% sampling efficiency to 5 μm (Sheridan et al., 2012). Aerosol light absorption (σ_{ap}) was measured at three wavelengths (467, 530, 660 nm) using a Radiance Research Particle-Soot Absorption Photometer (PSAP) and aerosol light scattering (σ_{sp}) was measured at three similar wavelengths (450, 550, 700 nm) using an integrating nephelometer (TSI model 3563). The measurements of absorption and scattering were made at low relative humidity (RH<40%). Absorption data were corrected for scattering artifacts, flow and spot size calibrations, etc., using the Bond et al. (1999) algorithm, with appropriate modifications for wavelength (Ogren, 2010). The Anderson and Ogren (1998) correction for instrument non-idealities was applied to the nephelometer data.

Ambient temperature (T_{amb}) and RH (RH_{amb}) were measured by a sensor (Vaisala Inc, Model Humicap 50Y) mounted on the aircraft fuselage inside a counterflow inlet shroud, and the nephelometer sample pressure was used as a surrogate for ambient pressure. These measurements of ambient meteorological parameters were used to adjust the in-situ optical data to ambient conditions in order to compare with the AERONET measurements and retrievals, which are made at ambient conditions. 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. For BND γ=0.71±0.08, while for SGP γ=0.65±0.08. At BND the IMPROVE chemistry measurements are co-located at the profile location, while for SGP the measurements at the

IMPROVE Cherokee Nation site (approximately 56 km southwest of the profile location) were used. This γ value was then used in conjunction with the airborne RH_{amb} measurements to adjust the in-situ scattering profiles for both SGP and BND.

The equation used to adjust the dry, in-situ scattering to ambient relative humidity (RH_{amb}) is a commonly used aerosol hygroscopic growth parameterization (e.g., Kasten, 1969; Hanel, 1976; Kotchenruther et al., 1999; Carrico et al., 2003; Crumreyrolle et al., 2014):

$$\sigma_{sp}(RH_{amb})/\sigma_{sp}(RH_{dry})=a*(1-(RH_{amb}/100))^{-\gamma}. \quad (1)$$

where $\sigma_{sp}(RH_{amb})$ is the aerosol scattering at ambient RH, $\sigma_{sp}(RH_{dry})$ is the measured scattering at low RH, and γ is the hygroscopic growth parameter derived from the IMPROVE aerosol chemistry. The value of 'a' can be determined using: $a = (1/(1-RH_{dry}/100))^{-\gamma}$ (e.g., Crumreyrolle et al., 2014; Quinn et al., 2005). Here we assume a=0.9 based on the typical RH values measured inside the nephelometer for both profile locations (BND RH_{dry}=12+/-11%; SGP RH_{dry}=14+/-10%). RH_{amb} at BND and SGP averaged 47.4% and 38.6%, respectively, over all flight levels and seasons (56% (BND) and 43% (SGP) below 1500 m asl). The 95th percentile RH_{amb} values (calculated over all flights and flight levels) were 79.3% and 76.6% at BND and SGP, respectively. (Note: scattering-weighted column average RH values were 54% at BND and 43% at SGP). Applying eq. 1 to the observed RH_{amb} and $\sigma_{sp}(RH_{dry})$ profiles, the average enhancement of column-average σ_{sp} due to hygroscopic growth was 1.52 and 1.36 at BND and SGP, respectively. The corresponding 95th percentiles of column average enhancement of scattering were 2.06 and 2.10. While Equation 1 takes into account differences in hygroscopic growth due to RH for each segment of each flight, it does not account for compositional changes that might affect the scattering enhancement due to hygroscopicity. For aerosol events such as biomass burning and dust episodes with significantly different composition than the 'normal' aerosol we would expect to over-predict the aerosol hygroscopicity relative to the normal aerosol. Sheridan et al., (2001) showed that the SGP surface aerosol had lower hygroscopicity when it was influenced by dust or smoke.

The absorption measurements were adjusted to ambient temperature and pressure, but not to ambient RH because the parameterization of the correction and its magnitude are unknown. It is typically assumed that absorbing aerosol is hydrophobic (e.g., Schmid et al., 2003; Reid et al., 2005; Schaefer et al., 2014), i.e., does not take up water. The uncertainties associated with this assumption are discussed in section 2.4.

Both the scattering and absorption in-situ measurements were adjusted to the two nominal Level-2 AERONET wavelengths in the mid-visible spectrum (440 nm and 675 nm). The 440 nm wavelength is of interest as that is the wavelength for which the AOD constraint for retrieving SSA and hence, AAOD, is given; the 675 nm wavelength is also presented because it is less sensitive to NO_2, organics, and dust which could potentially bias the in-situ/AERONET comparison. Also, evaluating data at both wavelengths helps in attributing aerosol absorption to BC versus dust, since at 675 nm absorption is almost entirely caused by BC. The measured

scattering Ångström exponent was used to adjust the in-situ scattering measurements to the AERONET wavelengths. For the in-situ aerosol absorption wavelength adjustments we used a constant absorption Ångström exponent of 1.2 to minimize the effects of noise in the measurement. Previous studies have shown that for both BND and SGP the absorption Ångström exponent is ~1.0 in the BL and 1.5 at higher altitudes (Andrews et al., 2011; Sheridan et al., 2012). Using the incorrect absorption Ångström exponent will have a negligible effect on the resulting absorption value because of the small difference between the measured and target wavelengths; using an absorption Ångström exponent of 1.2 instead of 1.0 will result in a 1% difference in adjusted wavelength while using an Ångström exponent of 1.2 instead of 1.5 will result in a 2% difference in adjusted absorption.

Finally, using these in-situ values adjusted to AERONET wavelengths and ambient conditions the ~~column~~ flight profile average properties can be determined. Aerosol extinction (σ_{ep}= σ_{sp} + σ_{ap}) was calculated and integrated vertically for the profile to obtain the in-situ AOD. The aerosol absorption for each profile was integrated vertically to obtain the in-situ AAOD. As described in Andrews et al. (2004), ~~T~~the in-situ column SSA (which is compared to the AERONET SSA value in section 3.1) was calculated for each flight level and then extinction-weighted and integrated to determine column SSA. This results in SSA values which are virtually identical to SSA values calculated using: $SSA_{col,in\text{-}situ}$ = ($AOD_{in\text{-}situ}$ – $AAOD_{in\text{-}situ}$)/$AOD_{in\text{-}situ}$) and effectively gives higher weighting to the SSA values at altitudes that had the highest aerosol concentrations. ~~determined using the AOD and AAOD calculated from the in-situ flights, e.g., $SSA_{col,in\text{-}situ}$ = ($AOD_{in\text{-}situ}$ – $AAOD_{in\text{-}situ}$)/$AOD_{in\text{-}situ}$)~~. Details of the procedure for calculating the vertical integral are given in Andrews et al. (2004), although, in this study, the in-situ profiles contained two additional high altitude flight levels (at 3659 and 4575 m asl) and the layer at the highest altitude was assumed to extend 457 m above the measurement altitude. As described in Andrews et al. (2004), the calculated column values for SSA were calculated from extinction-weighted SSA values at each flight level, effectively giving higher weighting to the measurements at altitudes that had the highest aerosol concentrations. Profile statistics for various parameters including SSA are provided in Andrews et al. (2004,2011a) and Sheridan et al. (2012). Individual flight profiles for various parameters are available online at: http://www.esrl.noaa.gov/gmd/aero/net/iap/iap_profiles.html (for SGP) and https://www.esrl.noaa.gov/gmd/aero/net/aao/aao_prof2007.html (for BND).

2.2 AERONET

AERONET measurements have been made at BND since mid-1995 and at SGP since mid-1994. The AERONET network makes spectral measurements of aerosol optical depth (AOD) using CIMEL sun/sky radiometers (Holben et al., 1998). The measurements are typically made at seven wavelengths, with an eighth wavelength used for water vapor measurements. The AERONET website (http://aeronet.gsfc.nasa.gov) provides links to data from more than 500 sites across the globe. The column extinction Ångström exponent (å) can be directly calculated from the wavelength-dependent AOD measurements (Eck et al., 1999). In addition to AOD and

å, algorithms have been developed utilizing both the spectral AOD and the spectral angular distribution of the sky radiances obtained from almucantar scans, which enable retrieval of other column aerosol properties including AAOD, SSA, size distribution, complex refractive index, and fine mode fraction of extinction (FMF_e) (Dubovik and King, 2000; Dubovik et al., 2000; O'Neill et al., 2003; Dubovik et al. 2006). The nominal wavelengths of the almucantar inversion retrievals are 440, 675, 870 and 1020 nm. An additional advantage of the AERONET database is that the retrieval values are obtained consistently – the calibrations, corrections, QC and algorithms are applied identically for each AERONET location.

For Version 2 AERONET data, ~~T~~there are different levels of AERONET data available for download from the AERONET website. Level 1.0 is unscreened data while Level-1.5 undergoes automated cloud-screening (Smirnov et al., 2000). Level-2 represents data with pre-field and post-field calibrations applied, manual inspection, and quality assurance (Smirnov et al., 2000). In addition to the Level-1.5 screening, the criteria for Level-2 almucantar inversion products include a check of the sky residual error as a function of AOD_{440}, solar zenith angle must be greater than or equal 50 degrees, and almucantars must have a minimum number of measurements in each of the four designated scattering angle bins. Further, for Level-2 absorption-related products (including SSA, AAOD, AAE, and the complex refractive index) the AOD_{440} must be greater than 0.4 to exclude more uncertain aerosol absorption estimates (Holben et al., 2006). Version 3 AOD products are now available but the Version 3 inversion products are not.

The AAOD values reported in the AERONET almucantar inversion files are obtained using the relationship: AAOD=(1-SSA)*AOD. Schafer et al. (2014) has a nice description of how SSA is obtained from the AERONET measurements. In the present study, in order to maximize the number of AERONET data points available for comparison with the in-situ measurements, Level-1.5 retrievals of AAOD and SSA were included in the analysis if there was a corresponding valid Level-2 AOD value (i.e., the same primary criterion as was used in Bond et al. (2013)). We will refer to these AAOD and SSA values as 1.5* data.

2.3 Merging the IN-SITU and AERONET data sets

Merging of collocated (within 15 km), but temporally disparate data sets can induce discrepancies in the combined data set. Lag-autocorrelation analysis (e.g., Anderson et al., 2003) is used to determine an appropriate time window for comparison of the AERONET and in-situ profile measurements. Figure 2 shows that, at the surface, at both BND and SGP, scattering is well correlated (r(k)>0.8) out to 4-5 hr lag, while absorption is less correlated than scattering (r(k) for absorption is 0.75 at BND and 0.55 at SGP). Based on the correlograms, AERONET retrievals were merged with the in-situ profile data when the retrievals were within +/-3 h of the end of the in-situ profile. This is the same time range constraint used to compare AERONET and PARASOL SSA values (Lacagnina et al., 2015). Additionally, Figure 2 represents the maximum correlation that we can realistically expect to achieve in a comparison of two different instruments with temporally offset measurements and provides context for the AERONET/in-situ comparisons presented in Section 3.

Because the profiles are "stair-step" descents from ~4600 m asl down to ~450 m asl (e.g., see Figure 4 in Sheridan et al., 2012), matching with AERONET retrievals at the end of the profile means that the matches are more closely aligned with when the airplane is in the boundary layer and thus, typically, sampling the highest aerosol concentrations. This way the maximum time difference between the boundary layer portion of the flight and the AERONET retrieval is 3 h; if we'd chosen to match based on the start of the flight the maximum time difference between the boundary layer measurements and the AERONET retrieval could be as large as 5 h. The boundary layer portion (<1800 m asl) of the ~2 h profile takes approximately 30 min. While the +/- 3 h match window was chosen based on the surface in-situ aerosol lag-autocorrelation statistics (Figure 2), other time windows were also examined. For time windows less than +/-3 h (e.g., 1 h and 2 h) the fit coefficients (slope, intercept) did not change significantly although the AOD and AAOD correlation coefficients did improve for those smaller time windows. For time windows longer than +/- 3 h (e.g., 6 h and 12h) there were changes in AOD and AAOD fit parameters and the correlation coefficients decreased significantly. 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). The poor correlations for SSA are not surprising given the uncertainties at low loading. The AERONET/in-situ comparisons for the +/-3 h window are discussed in section 3.1 below.

2.4 Uncertainties in IN-SITU and AERONET data

In any study comparing parameters obtained from different instruments and/or methods, an understanding of the uncertainties in each of the parameters being compared is critical. Below we discuss the uncertainties inherent in both the in-situ and AERONET datasets.

2.4.1 IN-SITU uncertainties

Uncertainties for measurements by the in-situ instruments have been described previously (e.g., Sheridan et al., 2002; Formenti et al., 2002; Shinozuka et al., 2011; Sherman et al., 2015) so only an overview is provided here. Sheridan et al. (2002) calculated uncertainties in aerosol light scattering for the TSI nephelometer to be 7-13% for 10 min legs depending on amount of aerosol present – the higher uncertainty value applies to very low aerosol loadings (scattering < 1 Mm^{-1}). We assume that uncertainty in the profile scattering measurements is 13%. 13% is appropriate for the higher altitude flight legs (10 min duration with, typically, low aerosol loading) and is also reasonable for the lower altitude flight legs which are only 5 min in duration but have significantly higher loading. At both BND and SGP the median boundary layer scattering is typically >10 Mm^{-1} while median scattering for the upper altitude flight legs is typically between 1-10 Mm^{-1} (Andrews et al., 2011; Sheridan et al., 2012).

Unfortunately, because profile-specific aerosol hygroscopicity measurements were not available for the in-situ aircraft measurements described here, a single hygroscopic growth parameterization was applied for all profiles at each site as described in Section 2.1 and equation 1. To determine the uncertainty in AOD induced by the uncertainty in the scattering adjustment to ambient RH, AOD values were calculated using different γ values representing

the range of hygroscopic growth factors suggested by the aerosol chemistry. Specifically, AOD_{440} was calculated for $\gamma \pm 1$ standard deviation and $\gamma \pm 2$ standard deviations. As described above, γ was calculated from the climatological chemistry measurements made by the IMPROVE network (14 years of data, ~1700 data points at BND; 10 years of data, ~1000 data points at SGP) using the Quinn et al. (2005) parameterization. We calculated the mean and standard deviation of γ based on those climatological chemistry measurements. Using this approach, the uncertainty in AOD due to adjustment to ambient RH was determined to be between 9% and 16%. This uncertainty might seem to be low, but recall that the 95th percentiles of ambient RH values observed throughout the profiles were ~80% but that more typically ambient RH in the boundary layer was less than 70% at BND and less than 60% at SGP. Sum of squares uncertainty analysis suggests the overall uncertainty in the in-situ AOD is approximately 30% for higher ambient humidities (RH_{amb}>70%) and approximately half that at RH_{amb}<50%.

Jeong and Li (2010) have noted that the presence of nearby clouds may influence AOD values. They've investigated the effect of high RH-halos embedded in aerosol layers that typically exist in the vicinity of non-precipitating cumulus clouds. If the AERONET retrieval went through such a halo it could result in an increased AOD due to the combined effects of hygroscopic growth, cloud processing of aerosols and rapid gas-to-particle conversions. If the aircraft also flew through this RH-halo then the effect would also be accounted for in the RH-corrected in-situ measurements. However, if the high RH layer was between two flight levels then the aircraft measurements would not account for it. Addressing this effect is outside the scope of this paper.

The PSAP measurement of aerosol absorption is more uncertain than the aerosol scattering measurements – PSAP uncertainty is reported to be in the 20-30% range (e.g., Bond et al., 1999; Sheridan et al., 2002; Sherman et al., 2015). It should be noted that the PSAP absorption measurement represents all absorbing aerosol collected on its filter, as opposed to being specific to 'black carbon' absorption. That is actually helpful for this particular study as the AERONET retrieval of AAOD also represents all flavors of absorption (e.g., 'black carbon', 'brown carbon' and dust). Müller et al. (2011) describe detailed experiments to characterize filter-based absorption instruments and describe some additional limitations of the instruments.

There is, however, some question of whether the PSAP (or any filter-based measurement) is able to accurately represent absorption by particles coated with semi-volatile or liquid organics, due to the possibility of such coatings changing the characteristics of the filter substrate (oozing!) after impaction (e.g., Subramanian et el., 2007; Lack et al., 2008). Comparisons of filter-based absorption measurements for denuded and un-denuded particles (e.g., Kanaya et al., 2013; Sinha et al., in revisions, 2017) suggest that the un-denuded particles have absorption enhancements of 5-25% relative to those that have been through a denuder. These comparisons show that stripping off coatings and evaporating the non-absorbing particles reduces the measured absorption, i.e., that the effect of coatings is not completely lost in filter-based measurements. The effect~~s~~ of ~~such oozing~~coatings appears ~~are~~ to increase the

absorption value reported by the PSAP relative to that reported by a non-filter-based instrument (Lack et al., 2008); in other words the aerosol absorption values obtained from PSAP measurements may have a positive bias. It is worthwhile to explore the potential magnitude of such a bias. The mean mass concentrations of organic aerosol determined from the IMPROVE measurements near BND and SGP (the OCf value in the IMPROVE data set; Malm et al., 1994) are similar for both sites and less than 2 $\mu g/m^3$, putting them firmly in the rural/remote category identified by Lack et al (2008; their figure 4). Depending on whether figure 3 or figure 4 in Lack et al. (2008) is used, Lack et al.'s (2008) results suggest that the PSAP might be overestimating absorption by a factor of 1.1 to 1.5 due to artifacts caused by organic aerosols. However, in a subsequent study, Lack et al. (2012) reported a PSAP overestimate by factors of 1.02-1.06 over Los Angeles, considerably lower then the Lack et al. (2008) results.

The positive bias in absorption related to filter-based measurements is the same order of magnitude and direction of the absorption enhancement factor found by some lab and theoretical studies for coated absorbing particles suspended in the atmosphere. Absorption enhancement values of 1.3-3 have been predicted for coated particles (e.g., Bond et al., 2006; Lack et al., 2009; Cappa et al., 2012) although enhancements larger than a factor of 2 have not been measured for ambient aerosol (e.g., Lack et al., 2008; Cappa et al., 2012; McMeeking et al., 2014). Wang et al. (2014) suggested that an absorption enhancement factor of 1.1 was appropriate for fossil fuel influenced aerosol and that 1.5 was a more reasonable enhancement factor for biomass burning affected aerosol. Biomass burning does not have a consistent influence on either BND or SGP. Cappa et al. (2012) suggested that the discrepancies between ambient and modelled and/or laboratory results, could be a result of differences in particle morphology and/or chemistry. We have not made any adjustments for the absorption effects of coatings or the potential positive bias in PSAP measurements as ~~(a)~~ the science is still unclear ~~and (b) additionally, measured absorption Ångström exponents are quite low (close to 1) suggesting little influence of coatings~~.

In addition to the potential absorption enhancement due to organic coatings, it has been suggested that aerosol water on absorbing particles may also enhance absorption. There have been very few studies where the hygroscopic growth enhancement of absorption was explicitly considered. Redemann et al. (2001) modeled absorption enhancement as a function of RH based on characteristic atmospheric particles and found absorption enhancement values of up to 1.35 at 95% RH; for the 95th percentile RH_{amb} values encountered at BND (78.9%) and SGP (76.6%), the Redemann et al. (2001, their figure 2) study would predict absorption enhancements of ~1.1. Nessler et al. (2005) and Adam et al. (2012) utilized both ambient aerosol measurements and Mie theory to calculate absorption enhancement values due to hygroscopic water uptake. Nessler et al. (2005) does not provide absorption enhancements as a function of RH, but Adam et al. (2012) suggest absorption enhancements due to hygroscopic growth of less than 1.1 at 80% humidity. Brem et al. (2012) report on laboratory studies that show that aerosol absorption was enhanced by a factor of 2.2 to 2.7 at 95% relative humidity relative to absorption at 32% relative humidity, although for RH less than ~80% (i.e., the RH values observed in this study) they show no absorption enhancement (their figure 9). Lewis et

al. (2009) actually observe a decrease in absorption with increasing RH for some biomass fuels, but hypothesize the decrease might have been due to their measurement technique and/or a change in the morphology of the particles.

In summary, the positive bias in the PSAP measurements of aerosol light absorption might be as high as a factor of 1.1 to 1.5 due to oozing (e.g., the overestimate of absorption reported by Lack et al., (2008) for filter-based measurements). Atmospheric absorption may be underestimated by PSAP measurements by up to a factor of 1.5 due to not accounting for coating (organic or water) effects. Without additional laboratory and field measurements to quantify the net effect of the possible positive and negative biases in PSAP measurements of aerosol light absorption, it is not possible to estimate the actual uncertainty in the in-situ light absorption measurements reported here due to coating effects. To address this, we double the assumed PSAP uncertainty of ~25% to 50% in the calculations of uncertainty.

One aspect of the in-situ system that will affect both the scattering and absorption measurement is the gentle heating used to dry the particle to RH<40%. The drying process we use (heating of 40 C or less) may remove some volatile components but we believe the removal to be minimal (<10-20%) based on lab and ambient volatility studies in the literature. Thermal denuder studies suggest little removal of volatile components (<10%) at 40 C (e.g., Mendes et al., 2016; Huffman et al., 2009, Bergin et al., 1997) although thermal denuders results may be limited by short residence times (<20s). However, smog chamber evaporation studies on ambient aerosol over longer time periods (minutes-hours) at ambient temperature also suggest ambient aerosol may be less volatile than previously thought – Vaden et al. (2011) showed that ambient SOA lost just ~20% of its volume after ~4h.

Once the uncertainties in the in-situ aerosol scattering and absorption are known, the uncertainty in SSA (SSA= $\sigma_{sp}/(\sigma_{sp}+\sigma_{ap})$ can also be calculated. Formenti et al. (2002, their equation 5) suggests the uncertainty in single scattering albedo (δSSA/SSA) can be calculated:

$$\delta SSA/SSA = (1-SSA)*[(\delta \sigma_{sp} / \sigma_{sp})^2+(\delta \sigma_{ap} / \sigma_{ap})^2)]^{1/2} \quad (2)$$

For scattering uncertainties of 30%, (combined nephelometer and f(RH) induced uncertainty), PSAP absorption uncertainties of ~~25~~50%, and SSA values of 0.9, equation 2 results in an in-situ SSA uncertainty of ~6% or approximately 0.06. For the higher altitude flight segments the loading does tend to be quite a bit lower and thus has higher uncertainty but those upper-level segments contribute little to the overall AOD or AAOD. Because the flight column SSA is calculated using extinction-weighted SSA flight segments, segments with very low aerosol concentrations will have little impact on the column SSA derived from the flight measurements.

In addition to instrumental uncertainties there are also uncertainties associated with the aircraft flight patterns, i.e., the presence of aerosols below, between and above the discrete flight levels. Missing aerosol above and below an aircraft profile is a potential issue in all aircraft/column comparisons. Different approaches have been used to assess whether aerosol loading contributions above the highest flight level (4.6 km asl) are important. Andrews et al. (2004)

utilized Raman lidar measurements to determine that 80-90% of the aerosol was below 3.7 km asl at SGP (3.7 km was the maximum altitude flown by the original SGP airplane, although all the profile flights utilized here occurred after the maximum flight level was increased to 4.6 km asl). Andrews et al. (2004) also assumed assumed an AOD contribution of 0.005 from stratospheric aerosol which was not done here. At SGP, Turner et al. (2001) segregated lidar aerosol extinction profiles by season and AOD. Their results (their Figure 1) suggest that for the vast majority of cases observed at SGP, 5% or less of the extinction will be found above 4 km. For low AOD cases (AOD_{355}<0.3) their mean extinction profiles suggest little to no aerosol extinction between 4-7 km. At BND, Esteve et al. (2012) noted that CALIPSO data indicated negligible extinction above 4.6 km asl. Regionally, seasonal average profiles from CALIPSO also suggest there is minimal aerosol above the flight's highest level (Ma and Yu, 2014; Yu et al., 2010).

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.

Several papers (Andrews et al., 2004; Esteve et al., 2012; Sheridan et al., 2012) have shown that there is a high correlation (R^2 > 0.8) between scattering measured at the surface site (SGP or BND) with scattering measured at the corresponding lowest flight leg, although the slopes of the relationships indicated that the airplane measurements might be missing a fraction (10-20%) of the aerosol below about 150 m agl. Additionally, Esteve et al. (2012) found high correlation (slope=1.01, R^2~0.7) between scattering AOD calculated by assuming the lowest leg represented scattering in the entire layer between surface and that flight leg with scattering AOD calculated from 1-sec data obtained during descent from the lowest flight leg to landing. This result suggested that no consistent bias would result from assuming the lowest flight leg was representative of the aerosol between surface and that altitude. We've looked at the surface/lowest flight leg relationship specifically for the flights with matching AERONET retrievals studied here. We found that at BND the surface and lowest level flight aerosol measurements were virtually identical. At SGP the lowest level leg actually measured slightly higher aerosol loading than was observed at the surface, which could lead to an overestimate of the aerosol optical depth in that layer, depending on the shape of the profile.

Similarly, Esteve et al. (2012) investigated differences in aerosol scattering between and at flight levels by comparing scattering AOD from the airplane descent between layers with that calculated from the individual level legs in the profile. Again they were able to confirm that measurements made during the fixed flight altitudes are representative of the aerosol near those altitudes.

~~Different approaches have been used to assess whether aerosol loading contributions above the highest flight level (4.6 km asl) are important. Andrews et al. (2004) utilized Raman lidar measurements to determine that 80-90% of the aerosol was below 3.7 km asl at SGP (3.7 km was the maximum altitude flown by the original SGP airplane, although all the profile flights utilized here occurred after the maximum flight level was increased to 4.6 km asl). For BND, Esteve et al. (2012) noted that CALIPSO data indicated negligible extinction above 4.6 km asl.~~

2.4.2 AERONET uncertainties

Uncertainties in AERONET retrievals have been reported in several papers. Eck et al. (1999) indicate that the uncertainty in AOD is approximately 0.01 for a field-deployed AERONET sunphotometer at solar zenith angle = 0 (i.e., sun directly overhead). For the almucantar retrievals (solar zenith angle > 50) used here, the AOD uncertainty will be smaller as the uncertainty in AOD decreases inversely with air mass (Hamonou et al., 1999; their equation 1).

Dubovik et al. (2000) report AERONET retrieved SSA uncertainties in their Table 4. For water soluble aerosol (the predominant aerosol type at both BND and SGP) they report that SSA values are reliable to within ±0.03 when $AOD_{440} > 0.2$, while the uncertainty in SSA increases to (±0.05-0.07) for $AOD_{440} \leq 0.2$. The almucantar retrieval of SSA may be biased by errors in the surface reflectance when the AOD is very low. 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 aerosol 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).

Mallet et al. (2013) reports an AAOD uncertainty of 0.01 but does not indicate whether or how the AAOD uncertainty would change with AOD_{440}. Using the sum of squares propagation of errors to calculate the uncertainty in AAOD for both high and low AAOD cases results in an AAOD uncertainty of approximately ±0.015 for both high and low AOD cases (high AOD_{440}=0.5, δAOD=0.01, SSA=0.95, δSSA=0.03, AAOD=0.026; low AOD_{440}=0.2, δAOD=0.01, SSA=0.95, δSSA=0.07, AAOD=0.011). An AAOD uncertainty value of ±0.015 suggests an uncertainty of about 60% in AAOD for AOD_{440}=0.5 and more than 140% uncertainty in AAOD for AOD_{440}<0.2.

3. Results

In this section we first present comparisons of AOD, AAOD and SSA from the in-situ measurements at BND and SGP with AERONET retrievals. This includes (1) direct comparisons of each in-situ profile with contemporaneous AERONET retrievals; the BND and SGP comparisons are then put in the wider context of a literature review of similar direct comparisons of in-situ and AERONET AAOD and SSA; (2) seasonal comparisons of AOD, AAOD and SSA from Phase II AeroCom model results, AERONET retrievals and in-situ measurements for BND and SGP; and finally, (3) we discuss these results in the context of biases in determination of AAOD.

3.1.1 BND and SGP: in-situ vs AERONET – Direct Comparisons

Figures 3, 4 and 5 show the direct comparisons of AOD, AAOD and SSA at both 440 nm and 675 nm. On all 3 plots, the blue points represent the same data set – each point indicates a flight for which there was one or more successful AERONET Level-2 almucantar retrievals within +/-3 hours of the end of the flight profile (if there was more than one retrieval corresponding to a flight, the retrievals were averaged). The thin gray lines on the 440 nm plots indicate the reported (AERONET) or calculated (in-situ) uncertainties in the data. Table 1 provides a comparison of the statistical values (median, mean and standard deviation) at 440 nm for each of the parameters at both of the sites for these direct comparisons (blue points in Figures 3, 4, and 5). The low number of flights for which there are comparisons available (~10% of total number of flights) indicate both the effects of AERONET stringent cloud screening routine and the constraints imposed by the almucantar retrievals. In addition to limiting the number of comparisons available for this study, this limited data availability also has implications for modellers utilizing AERONET data – for example, Schutgens et al. (2016) has shown the importance of temporal collocation in measurement-model comparisons. Figure 3 also contains red points – the red data points represent all direct sun AERONET Level-2 AOD retrievals during the +/-3 hours window around the end of each profile. Depending on atmospheric conditions, there may be more than one AERONET retrieval within +/-3 hours of the end of each profile, which is why in Figure 3 there are more red data points plotted than there are flights. The red points have not been averaged in order to provide an indication of the variability in AOD during the in-situ profiling flight.

The comparison between in-situ and AERONET AOD is important because it can be used to evaluate how well the in-situ and AERONET retrievals can be expected to agree and, thus, set the context for the AAOD and SSA comparisons. Many studies have investigated the relationship between in-situ and remotely sensed AOD (e.g., Crumreyrolle et al., 2014; Schmid et al., 2009, and references therein). As noted in these studies, the in-situ derived AOD values tend to be slightly lower than the AOD retrieved from remote sensing measurements. Figure 3 presents the comparison of Level-2 AOD for AERONET and in-situ measurements at 440 nm and 675 nm for two sets of AERONET AOD data. The first comparison (red points on plots) is for all direct sun AERONET Level-2 AOD measurements. The second comparison (blue points on plots) is for flight-averaged AERONET Level-2 AOD measurements where all the criteria required for almucantar retrievals are satisfied. Table 2 summarizes how many points make up each of these data sets.

In general, Figure 3 shows that AERONET AOD tends to be higher than the in-situ AOD, although there is good correlation between AERONET and in-situ AOD. The uncertainty bars tend to overlap the 1:1 line suggesting that in-situ measurements provide a reasonable proxy of the total column aerosol loading as represented by AERONET AOD. Student t-test evaluation suggests that the AERONET and in-situ AODs are the same at the 95% confidence level. The ~~correlation~~ coefficients of determination (R^2) are within the range we would expect based on the lag-autocorrelation of scattering at these two sites (Figure 2) and the +/-3 h time window. The R^2 values increase when sub-setted for the more restrictive Level-2 almucantar retrievals.~~The correlations improve for the more restrictive Level-2 almucantar retrievals.~~ The lower in-situ AOD values observed at both sites, compared to AERONET, may be due to the hygroscopicity adjustment from dry in-situ to ambient RH conditions being too low or undersampling of larger particles (e.g., Esteve et al., 2012). Esteve et al. (2012) found slopes closer to 1 when they restricted AERONET/in-situ AOD comparison to low ambient RH (<60%) conditions, although the AERONET AOD values were still larger than the in-situ AOD. The effect of undersampling larger particles or underestimating aerosol hygroscopicity on the AAOD and SSA comparisons are discussed in section 3.1.2. Some of the discrepancy between the in-situ and the AERONET values may also be due to the limited vertical range covered by the airplane (150 – 4200 m asl). We've excluded flights that might have had aerosol above the highest flight level, based on Raman lidar comparisons (at SGP) and profile shapes (at BND). The relationships observed between AERONET and in-situ AOD for both sites are very similar to those observed for the recent DISCOVER-AQ campaign (e.g., Crumreyrolle et al., 2014, their figure 3).

One thing to note on Figure 3a is the blue point marked BB (the BB stands for biomass burning). This measurement occurred on June 28, 2006 and appears to have been strongly affected by forest fire smoke transported from Canada. We applied the same hygroscopicity adjustment to the measurements of this flight as we did to all of the BND flights and, in this BB case, the hygroscopicity correction was the primary reason the in-situ AOD value is significantly higher than the AERONET AOD value. This point would lie much closer to the 1:1 line if the in-situ BB data were assumed to be hygrophobic. Previous work at the surface site at SGP has shown that dust and smoke aerosol types tend to exhibit lower hygroscopicity than the background aerosol normally observed at the site (Sheridan et al., 2001). This BB point provides an extreme example of the downside of using a constant hygroscopic growth parameter as a function of RH, although without additional information about the aerosol for each profile it is difficult to do otherwise. The light blue dotted line on Figure 3 represents the relationship between AERONET and in-situ data if the BB point is excluded.

Figure 4 presents the comparison of AAOD for flight-averaged AERONET and in-situ measurements. As described above, the AERONET AAOD values shown in Figure 4 are what we have termed Level-1.5* data – i.e., they are from Level-1.5 almucantar retrievals when there was a valid Level-2 almucantar retrieval, but the AOD_{440}>0.4 constraint was not applied. In contrast to the AOD comparison depicted in Figure 3, the AERONET Level-1.5* AAOD values are significantly higher than the in-situ AAOD values. Figure 4 also shows that the correlation between the AERONET and in-situ AAOD is poorer than it was for AOD, particularly at BND (R^2 is 0.49~~34~~ at BND and 0.68~~77~~ at SGP for the 440 nm comparison). The lower~~poor~~ correlation at

BND is somewhat surprising given the lag-autocorrelation results for aerosol absorption (Figure 2a) at the BND surface site. Surprisingly, while the BND site has higher 3-hour autocorrelations for absorption than SGP (R = 0.75 for BND and R = 0.55 for SGP, per Figure 2), the results for BND in Figure 4 indicate less correlation than at SGP for absorption. Nonetheless, the correlation coefficients for BND in Figure 4 (R^2=0.49 (blue) and 0.37 (red) correspond to R = 0.70 (blue) and 0.61 (red)) are not that far from the 3 h auto-correlation of r(k=3h)=0.75 for absorption at BND in Figure 2. For AAOD the uncertainty bars, while wider, exhibit significantly less overlap with the 1:1 line (indeed no overlap at SGP) and indeed the student t-test suggests the AERONET and in-situ AAOD values are different at the 95% level at both sites.

Both Figure 4 and the median values provided in Table 1 indicate that AERONET Level-1.5* AAOD tends to be larger than the in-situ AAOD, although the scatter in the relationships (particularly at BND) suggests that a multiplicative factor doesn't represent the relationship very well. The purple points in Figure 4 indicate AAOD retrievals where the flight-averaged AOD_{440}>0.2. There is no obvious improvement of the relationship between in-situ and AERONET AAOD when these points are considered (although there are only 1-46~~-7~~ comparison points above AOD_{440}>0.2 for each site).

The AAOD comparisons at 675 nm at BND (Figure 4c) are quite similar to those at 440 nm, suggesting that there is little contribution to absorbing aerosol from dust, organic carbon and/or NO_2. In contrast, at SGP, there is a change in the relationship between AERONET and in-situ AAOD from 440 to 675 nm indicating that one or more of these components may affect the 440 nm comparisons at that site (Figure 4d). Ångström exponent values from the matched AERONET and in-situ profile data do not support the presence of dust, while the rural nature of the site suggests significant levels of NO_2 are unlikely. Thus the most likely explanation is the presence of organic carbon, although the IMPROVE sulfate and organic data used to estimate aerosol hygroscopcity do not support this. The IMPROVE measurements tend to suggest a relatively small contribution of organics to the aerosol mass with the average mass concentration of organics only 40 to 60% that of sulfate aerosol mass concentration for BND and SGP, respectively. In contrast, the Aerosol Chemical Speciation Monitor (ACSM) measurements by Parworth et al. (2015) indicate that, depending on the month, organic aerosol can contribute up to 70% of the total aerosol mass at SGP.

Figure 5 presents the comparison of column SSA retrieved from flight-averaged AERONET inversions (Level-1.5* data) with the column SSA calculated from in-situ profile measurements of aerosol scattering and absorption at BND and SGP. Consistent with the AOD and AAOD comparisons (Figures 3 and 4) the SSA retrieved from AERONET tends to be much lower than the SSA calculated from the in-situ profile measurements. . As with AAOD, the SSA uncertainty bars exhibit little overlap with the 1:1 line and a student t-test suggests the AERONET and in-situ SSA values are different at the 95% level for both BND and SGP. At both sites the range in AERONET-retrieved SSA is much wider than the range in column SSA obtained from the in-situ profiles. Long term, in-situ measurements at the BND and SGP surface sites yield mean SSA values of 0.92 and 0.95 respectively (Delene and Ogren, 2002, based on monthly-averaged data). Delene and Ogren's (2002) surface SSA values are reported at low RH

(RH<40%) and 550 nm; adjusting them to ambient conditions and 440 nm would likely cause them to increase making them more comparable to the in-situ column SSA depicted in Figure 5 but even less like the AERONET Level-1.5* SSA values. As with Figure 4, the purple points on Figure 5 indicate when the flight-averaged AOD_{440}>0.2; ~~there does not appear to be an improvement in the relationship between in-situ and AERONET SSA when only these purple points are considered~~although there aren't enough points to draw a robust conclusion, there does not appear to be an improvement in the relationship between in-situ and AERONET SSA when only these purple points are considered.

Figure 5 also includes a set of 'hybrid SSA' (SSA_{hybrid}) points in yellow. These points have been calculated using the AERONET AOD and the in-situ AAOD:

$$SSA_{hybrid} = (AOD_{AERONET} - AAOD_{PSAP})/AOD_{AERONET} \quad (3)$$

This hybrid approach to SSA eliminates the uncertainty associated with the empirical hygroscopic growth factors applied to the in-situ scattering measurements, and also removes the scattering uncertainty associated with undersampling the coarse mode. It does not, however, eliminate the uncertainties associated with assuming the absorbing aerosol is hygrophobic, that there is little absorption in the potentially undersampled coarse mode, or the unknown contribution from absorption enhancement. SSA_{hybrid} is very similar to the SSA derived from in-situ measurements, suggesting the primary discrepancy between the AERONET SSA and the in-situ SSA is due to the determination of the absorbing nature of the aerosol, either due to issues with the limitations of the filter-based measurements or to the interpretation of the relative contribution of aerosol absorption from the AERONET inversion retrieval products.

3.1.2 How might ~~AOD discrepancies~~in-situ- hygroscopicity assumptions and under-sampling of the aerosol affect SSA and AAOD comparisons?

Figure 3 shows that the AERONET AOD may be slightly larger than the in-situ AOD, while Figures 4 and 5 suggest that the AERONET retrievals significantly overestimate the amount of absorbing aerosol (low SSA, high AAOD) relative to the in-situ measurements. The slight deviation between in-situ and AERONET AOD may lead to questions about whether directly comparing other AERONET and in-situ parameters (e.g., SSA, AAOD) is a reasonable thing to do and whether the AAOD and SSA comparisons shown in Figures 4 and 5 are related to issues with the AOD comparison. As mentioned above, Esteve et al. (2012) suggested the AOD difference was most likely due to either underestimating the hygroscopic growth correction and/or undersampling of supermicron particles by the aircraft inlet. In this section we evaluate how these two possible causes of the AOD discrepancy might affect the SSA and AAOD comparisons.

Increasing the hygroscopic growth adjustment of the in-situ measurements would enhance the in-situ scattering values used to calculate the in-situ AOD, but would not change the in-situ AAOD because the absorbing particles are assumed to be non-hygroscopic. Consequently, the comparison depicted in Figure 4 would not change with a different adjustment for hygroscopic growth. Increasing the in-situ AOD, without affecting the in-situ AAOD, would result in higher in-situ SSA values and an even greater discrepancy between AERONET and in-situ SSA values

than shown in Figure 5. To evaluate the effect of assuming absorbing particles were non-hygroscopic, a sensitivity test was performed assuming the 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 minimal effect on the comparisons of AOD, AAOD and SSA.

The other likely candidate to explain the in-situ AOD being slightly lower than the AERONET AOD is aircraft under-sampling of super-micron aerosol particles due to the 5 μm inlet cutoff'. Esteve et al.'s (2012) comparison of column in-situ and AERONET scattering Ångström exponents at BND suggested that the airplane measurements might be under-sampling larger particles. Sheridan et al. (2012) estimated that the aircraft inlet 50% cut-off aerodynamic diameter is approximately 5 μm, so particles larger than that are unlikely to be sampled by the in-situ measurements but will be sensed by the AERONET sunphotometer. If we take into account that atmospheric particles are likely to have a density greater than 1 g cm^{-3}-, the actual cut size would be closer to 3 or 4 μm. The AERONET volume size distributions were used to estimate the fraction of column extinction due to particles less than 3~~5~~ μm. At BND the mean and standard deviation of the 3~~5~~ μm extinction fraction (extinction(D<3~~5~~μm)/extinction(D<30μm)) was 0.93~~4~~ ±0.07~~6~~, while at SGP the extinction fraction value was 0.88~~90~~ ± 0.09~~7~~. At the BND and SGP surface sites, most (80-90%) of the observed sub-10 μm scattering and absorption is also attributed to sub-micron aerosol, with absorption more likely to be in the sub-micron size range than scattering (Delene and Ogren, 2002; Sherman et al., 2015). This is consistent with the observation that absorbing aerosol tends to be concentrated in sub-micron particles for typical aged continental air masses (e.g., Hinds, 1982). Based on these observations, larger and primarily scattering particles are more likely to be under-sampled by the in-situ measurements than absorbing particles. This is the opposite of what is needed to explain the discrepancies between AERONET and in-situ AOD, AAOD, and SSA shown in Figures 3-5. 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.

Additionally, Sheridan et al. (2012) calculated particle transmission losses from behind the sample inlet on the airplane to both the nephelometer and PSAP to be similar and to be less than 10% in the particle diameter range 0.01<D<1 μm. This suggests that preferential losses of absorbing aerosol are also unlikely to occur downstream of the aerosol inlet. 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 in the manuscript 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.

3.2 Literature survey: in-situ vs AERONET – Direct Comparisons

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. The next question to address is whether this ~~bias~~discrepancy, found for two rural, continental sites in the central US with relatively low aerosol loading, is more widely observed for direct in-situ/AERONET comparisons at a variety of sites/conditions. As in section 3.1, the focus in this section is on direct comparisons of column-averaged SSA (or AAOD) derived from in-situ measurements made during aerosol profiling flights that were flown in close proximity (temporal and spatial) to an AERONET retrieval. Tables 3 and 4 summarize literature results describing the direct comparisons of AERONET retrievals with in-situ aerosol profile measurements for AAOD and column SSA. Figure 6 provides a graphical overview of the SSA comparisons described in Table 4. Tables 3 and 4 and Figure 6 also include the BND and SGP comparisons described in this study. With the possible exception some of the profiles reported by Corrigan et al. (2008), the literature comparisons cited in Tables 3 and 4 and shown in Figure 6 have been made at higher AOD conditions (AOD_{440}>0.3) to reduce retrieval uncertainty. In contrast, the SGP and BND comparisons are more representative of global AOD (Figure 1) with the majority of the comparisons at BND and SGP occurring for AOD_{440}<0.2. - Please note that some of the earlier studies shown in Figure 6 and described in Table 4 used values from Version 1 AERONET Level-2.0 data. Where that was the case, we retrieved Version 2 AERONET Level-2.0 data from the AERONET website and those Version 2 data are what is reported in Table 4 and depicted in Figure 6. The comments section of Table 4 mentions the cases where this was done. For some of these references we also retrieved the AOD_{440} values from the AERONET website as the AOD_{440} values weren't reported in all papers.

Tables 3 and 4 have been restricted to studies with direct comparisons of column-averaged AAOD or SSA retrieved from full in-situ vertical profiles flown near (within ~100 km) AERONET sites within a few hours of the AERONET retrieval, i.e., studies that are comparable to the BND and SGP studies described in Section 3.1. For non-plume data sets, Anderson et al. (2003) found autocorrelations > 0.8 at 100 km (their figure 6). For plume-influenced data sets they found autocorrelations ~0.6**.** Included in the tables are the field campaign name (if applicable), number of AAOD or SSA comparisons, the primary type of aerosol studied, summary of AOD comparisons (if available), altitude range covered by the airplane, instruments and data processing (e.g., instrument corrections, treatment of hygroscopicity, wavelength adjustment) and a summary of the results of the AAOD comparison. The last column in Tables 3 and 4 includes information on the spatial and temporal differences between the in-situ measurements and AERONET retrievals and comments on treatment of the AERONET and in-situ data. The last column also notes how each campaign dealt with aerosol below and above the in-situ profile if reported. It should be noted that the number of SGP and BND comparisons of AAOD and SSA in Tables 3 and 4 are only possible because we've utilized AERONET retrievals below the recommended threshold of AOD_{440}>0.4. The uncertainty for the BND and SGP comparisons is much higher than for some of the other direct comparisons due to the low AOD conditions observed at these sites.

For the three AAOD closure studies listed in Table 3 (the BND and SGP results presented here, plus results from a field campaign over the Indian Ocean) the AERONET retrievals indicate more absorbing aerosol in the column than is suggested by the corresponding in-situ measurements. The Corrigan et al. (2008) paper mentioned in Table 3 is the sole AERONET/in-situ AAOD comparison cited by Bond et al. (2013), as it was the only published direct AAOD comparison available. Corrigan et al. (2008) present no AOD comparisons that could provide an indication of their sampling system efficiency, and information about the wavelength of the comparisons and profiles specifics are lacking. To our knowledge, no other direct comparisons of in-situ and AERONET AAOD are available in the literature.

The SSA comparison studies listed in Table 4 and visually summarized in Figure 6 indicate that, even at higher AOD, AERONET ~~has a bias towards~~retrievals tend to indicate more-absorbing aerosol (lower SSA) relative to in-situ measurements. While much of the observed difference between $SSA_{in\text{-}situ}$ and $SSA_{AERONET}$ may fall within the uncertainty of the SSA values, as noted in Schafer et al. (2014), the fact that the difference ($SSA_{AERONET} - SSA_{in\text{-}situ}$) is predominately negative across all the direct comparisons found in the literature is not what would be expected from random error. Figure 6 also shows the mean and 2*standard deviation of all of the points (black square and vertical lines) and just the literature value points (black diamond and vertical lines). Based on the characteristics of a normal distribution the standard deviation lines suggest ~80% of the points will be negative – random error would suggest only 50% of the points should be negative. Figure 6 suggests that AERONET retrievals of SSA could perhaps be used at AOD_{440}<0.4, perhaps down to AOD_{440}~0.25 or ~0.3 – even at those low AOD values the differences in SSA between AERONET and in-situ still tend to be within the AERONET uncertainty. However, as Figure 6 shows, there are not a lot of direct comparisons to support such a choice.

Most of the SSA comparisons in Table 4 found fairly good agreement between AERONET and in-situ AOD, suggesting that the issue is an over-estimation of the absorption contribution to AOD rather than an underestimation of the AOD scattering contribution. This is consistent with the AERONET AAOD values being greater than those obtained from in-situ measurements presented in Table 3. Out of the 63 profiles compared in Table 4, there are ~~eight~~ four exceptions, (three from Leahy et al. (2007) and ~~five~~ one from this study for the BND site) where $SSA_{AERONET}$ is larger than the corresponding $SSA_{in\text{-}situ}$. Interestingly, the three exceptions from Leahy et al. (2007) were for their high AOD (AOD_{550}>0.6) cases; for their two low AOD (AOD_{550}<0.3) cases the opposite was found, i.e., $SSA_{AERONET}$<$SSA_{in\text{-}situ}$.

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 uncertainty of the AERONET SSA retrieval and in-situ measurements). If there was no consistent bias in the ~~AERONET retrieval~~AERONET/in-situ comparison we would expect (AERONET_SSA – INSITU_SSA) to be evenly distributed around zero. Instead, ~~the results from the literature~~Figure 6, which

summarizes the literature survey, suggests either that AERONET ~~is~~ retrievals are biased towards too much absorption, or that in-situ, filter-based measurements of aerosol absorption ~~undersample aerosol absorption~~are biased low. We note that the results from the literature (e.g., Figure 6) 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.

An alternative explanation is that the AERONET SSA uncertainties are non-symmetric. Dubovik et al. (2000) suggest that simulated retrievals of SSA for 'water soluble aerosol' are asymmetric when different 'instrumental offsets' are assumed, particularly at lower AOD values (0.05 and 0.2). Their figure 4 shows a much larger decrease in SSA for some instrumental offsets relative to the increase in SSA observed for an instrumental offset of the same magnitude but opposite sign. Asymmetry is also indicated for 'biomass burning' aerosol (their figure 7) although the asymmetry is in the opposite direction, i.e., the increase in SSA is larger than the decrease for a given pair of instrumental offset values. It is not obvious from their figure 7 whether the retrievals are asymmetric for simulated dust aerosol. Interestingly, at least three of the four ~~of the~~ points in Figure 6 with AERONET_SSA>INSITU_SSA ~~(three of the Leahy et al. (2007) points and the BND point with AOD ~0.4)~~ represent retrievals of biomass burning aerosol.

3.3 BND and SGP: in-situ vs AERONET and AeroCom model output – Statistical Comparisons

Most of the statistical comparisons between AERONET and in-situ profiles (e.g., Ramanathan et al., 2001; Leahy et al., 2007; Ferrero et al., 2011; Johnson et al., 2011) were for short-term field campaigns with a limited number of in-situ profiles. The advantage of the multi-year, in-situ vertical profiling programs at BND (401 flights) and SGP (302 flights) is that we can compare the statistics for both in-situ and AERONET values as opposed to comparing individual in-situ values to remote retrieval statistics. Figure 1 in Andrews et al. (2011) and Figure 9 in Sheridan et al. (2012) demonstrate that the BND and SGP flight programs captured the multi-year seasonality in aerosol properties at these two sites. Because of the large number of flights over an extended period of time, Skeie et al. (2011) was able to compare the seasonally averaged, in-situ absorbing aerosol profiles from BND and SGP with seasonal vertical profiles of black carbon generated by the Oslo-CTM2 model. Skeie et al. (2011) found that the model underestimated absorbing aerosol relative to the BND and SGP in-situ profiles for most seasons and altitudes, although agreement between the model and measurements tended to be better at higher altitudes.

As mentioned in the introduction, AERONET retrievals of AAOD have been used to suggest upscaling factors for modelled values of absorbing aerosol (e.g., Sato et al., 2003; Bond et al., 2013). These model/AERONET comparison studies are typically based on model and measurement statistics (i.e., properties are averaged over time and region) rather than direct comparisons due to both computational constraints and the discrete nature of the AERONET measurements. Given the statistical nature of some historical AERONET/in-situ comparisons as well as the typical model/AERONET comparison constraints, in this section we compare

monthly statistics for in-situ measurements, AERONET retrievals and AeroCom model output. It should be reiterated here that we are comparing asynchronous data and that there are some additional differences amongst the data sets that need to be kept in mind: the AERONET data are rigorously cloud-screened (although cloud halo effects may persist (e.g., Jeong and Li, 2010) and only obtained during daytime; the in-situ measurements are also daytime-only and the airplane did not fly in-cloud due to FAA flight restrictions, but may have flown near clouds; and the model data include day and night with clouds and also represent values over a 1x1 degree grid.

Figure 7 shows the 440 nm monthly medians of AOD, AAOD and SSA at BND and SGP based on the in-situ profile measurements, and two versions of AERONET retrievals as described below. For the in-situ properties, all profiles were used, regardless of whether there was an AERONET retrieval corresponding to the flight. The AERONET monthly medians in Figure 7 use the long-term (1996-2013) AERONET data record for each site. As described previously, the lines labeled AERONET 1.5* were calculated from Level-1.5 inversion data with matching Level-2 almucantar retrievals. The lines labeled AERONET 2.0 utilized only Level-2 almucantar retrieval data. In both cases the median AERONET AOD values represent those Level-2 AOD measurements for which there was also an AAOD and SSA retrieval, ensuring that the AERONET AOD medians represent the same set of retrievals as the corresponding AAOD and SSA medians in the figure. The AERONET Level-1.5* AOD monthly medians are representative of the direct sun AERONET Level-2 AOD climatology at the two sites. Figure 7 also includes the AeroCom Phase II model monthly medians for BND and SGP (Kinne et al., 2006, Myhre et al., 2013) with model emissions, meteorology and other details briefly described in Myhre et al. (2013). The AeroCom values, which were provided at 550 nm, have been adjusted to 440 nm using the reported AeroCom monthly scattering Ångström exponent to adjust AOD wavelength and assuming an absorption Ångström exponent of 1 for the AAOD wavelength adjustment. It should be noted that the three monthly data sets (AERONET, AeroCom, and in-situ) plotted in Figure 7 are derived from measurements for overlapping, but not identical time periods, i.e., these plots represent climatological comparisons rather than direct comparisons of the data sets.

At both sites, the climatological seasonal patterns for AOD (i.e., high in summer, low in winter) are similar for the three data sets: in-situ measurements, AERONET Level-1.5* retrievals (recall that the AERONET 1.5* AOD is representative of the overall AERONET AOD climatology at each site) and AeroCom model output. At BND the AeroCom model AOD tends to be larger than the in-situ and AERONET 1.5* AOD values by up to a factor of two. AERONET 1.5* AOD is larger than the in-situ AOD in the summer (by up to 50%) but quite close the rest of the year (typically within 20%). At SGP the AOD monthly medians from in-situ measurements and AERONET Level-1.5* are almost identical for August-December, with slightly more discrepancy among the AOD values in summer and early part of the year. In contrast, AeroCom model median AOD values tend to agree better with AERONET 1.5* and in-situ AOD values from January-July but are noticeably higher (up to a factor of 2) in the later half of the year. At both sites, the median AERONET Level-2 AOD values (corresponding to AAOD and SSA retrievals) are much higher (by a factor of 2 or more) than the Level-1.5* and in-situ climatologies due to

the AOD_{440}>0.4 constraint. During the cleanest, lowest humidity, and often cloudiest months of the year (December-February) there are none to few Level-2 almucantar retrievals of SSA and AAOD at either BND or SGP – the gray lines in Figures 7ab are lacking data points for Jan., Feb. and Dec. at BND and Jan. and Dec. at SGP.

For AAOD at BND, the AeroCom model output falls between the AERONET 1.5* and in-situ values, with AERONET 1.5* AAOD being higher than the in-situ data by up to a factor of 8. As with AOD, the AERONET AAOD Level-2 values are much higher than the in-situ or modelled AOD values due to the constraint that they are only retrieved at high loading conditions (AOD_{440}>0.4). The three data sets (AeroCom, in-situ and AERONET 1.5*) agree best in the month of May when the median values of AAOD are within 30%. At SGP there is fairly good agreement between AeroCom model and in-situ AAOD for the first 7 months of the year, while the AERONET 1.5* monthly AAOD values are considerably higher for that same time period. For the latter part of the year the in-situ AAOD values tend to be lower than both AERONET and AeroCom AAOD values.

The AERONET 1.5* SSA values tend to be quite a bit lower at BND, and somewhat lower at SGP~~The AERONET 1.5* SSA values tend to be quite a bit lower than the other data sets at both sites~~, which is why the AERONET 1.5* AAOD values tend to be higher (recall that for AERONET data AAOD is calculated using AAOD=(1-SSA)*AOD). Figure 7 also shows that the AERONET Level-2 SSA values are similar to the monthly in-situ and AeroCom SSA medians between April and November. There are no AERONET Level-2 almucantar retrievals of SSA in January or December at either site. For the ~~remaining months~~February and March, median Level-2 almucantar retrievals of SSA are based on very few data points resulting in bigger discrepancies between AERONET Level-2 almucantar retrievals of SSA and the in-situ and AeroCom SSA values.

Aside from differences in magnitude, there are also differences in the seasonal patterns of AOD, AAOD and SSA for the three data sets (in-situ, AERONET 1.5* and AeroCom). For example, at BND, the AERONET and in-situ AAOD both have a bi-modal annual distribution with peaks in late spring and early fall, which is not captured by the AeroCom AAOD and which is not seen in the AOD seasonality. The observed seasonal differences may be a result of (a) the different climatology time ranges for each method and/or (b) very little overlap in the measurement times for AERONET and in-situ measurements or (c) in the case of the models, not capturing local emissions near the sites. This highlights the importance of direct (i.e., near in time and space) comparisons in order to understand these seasonal differences. The seasonal cycle plots in Figure 7 also direct attention to the fact that AOD and AAOD vary independently rather than exhibiting the same seasonal pattern. This suggests that different emission sources and/or atmospheric processes control the variability of absorption and scattering aerosol over the course of the year.

3.4 Discussion

Because AERONET data are readily available and are being widely used as a benchmark data set for evaluating model output of AAOD (e.g., Chung et al., 2012; Bond et al., 2013; He et al., 2014; Wang et al., 2014) as well as for comparison with satellite retrievals and development of AAOD climatologies, we document and discuss some of the previous methods for utilizing existing AERONET retrievals that have been used to ~~minimize the bias in retrieved~~estimate AAOD at low AOD (AOD$_{440}$<0.4) where Level-2 retrievals do not exist. These approaches fall into several categories (1) use only Level-2 data; (2) use Level-2 and Level-1.5 data with acknowledgement of greater uncertainty in the retrievals and potentially additional measurement constraints for the Level-1.5 data; (3) make climatological assumptions about the representativeness of Level-2 SSA for low AOD conditions to obtain AAOD.

Clearly the simplest approach to minimize ~~bias~~ uncertainty in retrieved AERONET AAOD and SSA is to only use AERONET Level-2 retrievals which include the AOD$_{440}$>0.4 constraint ~~as those have the lowest uncertainty~~. This approach has been and continues to be used (e.g., Koch et al., 2009; Bahadur et al., 2010; Chung et al., 2012; Buchard et al., 2015; Pan et al., 2015; Li et al., 2015). However, as shown in Figure 1 the vast majority of the globe has AOD$_{440}$<0.4, meaning few if any AERONET Level-2 AAOD or SSA retrievals will be available for most locations. This approach is quite useful in regions (or for case studies) with high aerosol loading (high AOD). However, ~~removing~~ excluding low ~~absorption events~~loading conditions is likely to ~~will~~ cause AERONET AAOD statistics to be biased high. This is particularly important when evaluating models in clean locations such as the Arctic. The AOD$_{440}$>0.4 constraint may also affect the SSA statistics.

Some studies have utilized AERONET Level-1.5 retrievals of absorption-related aerosol properties in order to avoid being limited to the high AOD levels required by Level-2 data (e.g., Lacagnina et al., 2015; Mallet et al., 2013). These studies note that Level-1.5 data include more relevant AOD values but that there are accompanying higher uncertainties in the retrievals for absorption related properties. Mallet et al. (2013) use Level-1.5 data to evaluate the spectral dependence of aerosol absorption. Lacagnina et al. (2015) utilize both Level-2 and Level-1.5 AERONET data in their comparison with PARASOL satellite retrievals of SSA and AAOD. For the Level-1.5 data they apply the additional requirement that the solar zenith angle must be ≥50°. Lacagnina et al. (2015) find quite good agreement (within +/- 0.03) for AAOD and note that larger differences between PARASOL and AERONET retrieval occur at higher AOD conditions, possibly due to less homogenous aerosol (i.e., plumes).

A more sophisticated approach to deal with SSA (and hence AAOD uncertainties) at low AOD is implemented by Wang et al. (2014). They make the assumption that SSA is independent of AOD (at least as a function of season) and utilize climatological Level-2 SSA values for each season with the measured AOD in order to obtain AAOD. The seasonal climatologies of SSA are based on 12 years of Level-2 AERONET data. For the two US continental sites studied in this paper, the approach of Wang et al. (2014) would likely minimize the potential AERONET ~~bias~~ tendency towards high AAOD at low AOD conditions as the Level-2 monthly climatological SSA values are quite similar to SSA values obtained by in-situ measurements (Fig. 7).

A similar, though statistical, approach was used in Bond et al.'s (2013) bounding BC paper in order to reduce the uncertainty and better represent AERONET SSA and AAOD retrievals at low AOD. Bond et al. (2013) worked with AERONET monthly local statistics for the time period 2000-2010. Monthly values of AAOD and SSA at 550 nm were calculated from size distributions and refractive index when there were at least 10 valid inversion retrievals for that month at that site in the 2000-2010 period (most sites had more than 10 retrievals in a given month over the 11 year period). It was assumed in Bond et al. (2013), based on AERONET reported uncertainties, that the retrieved absorption-related values were more reliable at larger AOD and so they made some adjustments to account for this. For each site, AAOD and SSA values were binned as a function of AOD (there were five AOD bins, with each bin corresponding to 20% of the AOD probability distribution). For lower AOD conditions, the calculated AAOD and SSA values were replaced by values obtained during larger AOD conditions for the same month as follows: (i) the SSA and AAOD values corresponding to AOD_{550} of 0.25 were prescribed for all SSA and AAOD observations at lower AOD and (ii) for locations where all $AOD_{550}<0.25$, the average SSA and AAOD of the upper 20th percentile of AOD observations at the site was prescribed for all lower AOD bins. Finally, the average of all five bins was used to determine the overall monthly average. In the case of AAOD the bin averages were simply averaged to get the monthly value while for SSA the AOD-weighted bin averages were averaged to get the monthly value. Note: the $AOD_{550}=0.25$ cutoff point used in Bond et al. (2013) corresponds (approximately) to $AOD_{440}=0.35$ for smaller particles and $AOD_{440}=0.25$ when large particles are present. Thus it is less strict than the AERONET recommended constraint of $AOD_{440}>0.4$, but it had been suggested that the recommended constraint might be too restrictive (pers. comm., O. Dubovik).

~~A similar, though statistical, approach was used in Bond et al.'s (2013) bounding BC paper in order to reduce the uncertainty in AERONET SSA retrievals at low AOD. Bond et al. (2013) worked with monthly local statistics of at least 10 inversion samples. After recovering SSA data from the lower quality Level-1.5 into the Level-2 data (AERONET Level-1.5*), the SSA value corresponding to AOD_{550} of 0.25 was prescribed for all observations at lower AOD, while for locations with all $AOD_{550}<0.25$, the average SSA of the upper 20th percentile of AOD observations was prescribed for all lower AOD.~~

One drawback affecting approaches using climatological values of SSA (e.g., Wang et al., 2014; Bond et al., 2013) is that they may not account for the systematic variability that has been observed between SSA and loading at many sites, although AOD is usually more variable than the composition (or SSA). Still some studies with in-situ data (e.g., Delene and Ogren, 2002; Andrews et al., 2013; Pandolfi et al., 2014; Sherman et al., 2015) indicate that SSA systematically decreases with decreasing aerosol loading. A similar SSA/AOD systematic variability relationship is also observed at some North American AERONET sites. Schafer et al. (2014; their figure 6) shows SSA decreasing at lower loading for the GSFC site near Washington D.C. during the period of their field campaign; they also show similar relationships between SSA and AOD based on the long-term data for three mid-Atlantic AERONET sites. Additionally, a quick survey (not shown) of other long-term North American AERONET sites with good statistics (i.e., lots of points) for Level-1.5 SSA retrievals (e.g., Billerica (Massachusetts), Bratts Lake (Saskatchewan, Canada), COVE (Virginia), Egbert (Ontario, Canada), Fresno

(California), Konza (Kansas), SERC (Maryland), and University of Houston (Texas)) indicates this systematic relationship may be observed at a wide range of locations in North America. Such climatological analyses may mask short-lived and/or infrequent aerosol events (e.g., dust or smoke incursions) that may have significantly different optical properties.

Figure 8 shows the systematic relationships between SSA_{440} and AOD_{440} for BND and SGP for both the AERONET retrievals and in-situ profile measurements. Consistent with previous figures, we have utilized SSA values for AOD_{440}<0.4 when there was a valid Level-2 AOD inversion retrieval, i.e., what we call AERONET Level-1.5*. Also included on the figure is a line showing the SSA_{550} versus scattering ($\sigma_{sp,550}$) relationships for the surface measurements at BND and SGP. The surface measurements are ~~for low~~made at low RH conditions (RH<40%) and adjusted to ambient RH using the available meteorological measurements at the site (ambient RH at 2 m at SGP and ambient RH at 10 m at BND); adjustment of the surface measurements from dry to ambient conditions ~~would tend to~~ shifts the SSA_{550} values upward (assuming absorption is not affected) and the scattering values to the right~~, but would not significantly change the shape of the curve~~.

Figure 8 suggests that for all three sets of measurements at both sites, there is a consistent decrease in SSA as aerosol loading decreases below AOD_{440}=0.2. This relationship implies that a climatology based on SSA values measured at high AOD may underestimate the AAOD climatology. The AERONET SSA values are lower than the in-situ profile values as would be expected from the results presented in sections 3.1 and 3.3. The AERONET SSA values are also lower than the surface in-situ SSA values – the surface in-situ SSA values adjusted to ambient conditions are quite similar to those obtained from the in-situ vertical profiles. It should however be noted that despite the discrepancy between in-situ and AERONET SSA values, Figure 8 shows that the SSA values for all three sets of measurements at SGP are within the reported AERONET SSA uncertainty range of 0.05-0.07 for AOD_{440}<0.2 across the narrow and low AOD range shown in the figure. At BND the SSA values are within the AERONET SSA uncertainty range down to AOD_{440}~0.1. ~~.~~ At the lowest AOD values (AOD_{440}<~0.05) the AERONET SSA values diverge, consistent with very large uncertainties expected in the AERONET SSA retrievals in the cleanest conditions. Uncertainty in the AERONET AOD retrieval may begin to affect the AERONET SSA retrieval where +/- 0.01 AOD uncertainty is equivalent to a 20% change in AOD for AOD of 0.05. In addition, at such low AOD values, the surface reflectance uncertainties may influence AERONET's retrieval of SSA. Figure 8 suggests that, in terms of the shape of the systematic variability plot, there are no obvious retrieval issues for AERONET SSA retrievals in the range 0.05<AOD_{440}<0.2, although this is in the AOD range where high uncertainty in the SSA retrieval is expected (Dubovik et al., 2000).

There are large differences (orders of magnitude) in the number of data points in each of the data sets; the number of points in each bin is indicated by the color-coded histograms shown on Figure 8. The mean standard error (MSE) in SSA (MSE=(standard deviation)/(number of points)$^{1/2}$) is indicated by the shading surrounding the solid colored lines. The MSE is quite similar for the AERONET 1.5* and in-situ profile measurements across the AOD range plotted in Figure 8, suggesting the observed systematic variability is not merely due to small numbers of

data points in each bin, particularly at lower loading. However, the fact that the AERONET MSE is approximately the same as the in-situ profile MSE, despite having approximately an order of magnitude larger number of points/bin, indicates that variability in the retrieved AERONET SSA is larger than the variability in SSA derived from in-situ profile measurements.

This study has utilized a valuable but spatially limited (i.e., two rural continental North American sites) climatological vertical profile dataset to explore AERONET retrievals of AAOD and SSA. Clearly, one way to address the observed discrepancy~~bias~~ between in-situ and AERONET~~in~~ AAOD is to pursue a focused measurement program designed to acquire statistically robust in-situ vertical profiles over AERONET sites representing a wide range of conditions and aerosol types. This type of measurement program ~~is being considered at NASA~~has been proposed to evaluate satellite retrievals and better characterize atmospheric aerosol (R. Kahn, SAM-CAAM, pers. comm.). Further evaluation and development of in-situ instrumentation for measuring aerosol absorption is also necessary, particularly in assessing the effects of coatings and hygroscopity on the resulting absorption values. Additional evaluation of the AERONET retrieval algorithm may provide insight into a potential SSA and, thus, AAOD bias (e.g., Hashimoto et al., 2012). ~~While~~ The discrepancies reported here between in-situ and AERONET values of AAOD and SSA suggest that caution should be used in upscaling model results to match AERONET retrievals of absorbing aerosol as this will have a significant impact on ~~an AERONET bias towards high AAOD will affect~~ global radiative forcing estimates. ~~if models upscale their simulated absorption to match the AERONET retrievals, t~~The work of Wang et al. (2016) has shown that other factors (e.g., the spatial resolution of models and emissions) may also contribute to the differences observed between model and AERONET retrievals of AAOD. Thus, to really be able to understand and simulate the influence of absorbing aerosol on radiative forcing will require expanded effort on both the measurement and modeling fronts.

4. Conclusion

~~We cannot say how to estimate SSA or AAOD from AERONET retrievals for the low AOD conditions prevalent around much of the globe.~~ AERONET retrievals of SSA at low AOD conditions (below the recommended AOD_{440}<0.4 constraint) ~~appear to be biased low~~ are consistently lower than coincident and co-located in-situ vertical profile observations of SSA (based on detailed comparisons at two rural sites in the US). Correspondingly, AERONET retrievals of AAOD at low AOD are consistently higher than those obtained from in-situ profiles. A survey of the literature suggests that even at higher loading (AOD_{440}>0.4) AERONET SSA retrievals tend to be lower than SSA values obtained from vertical profiling flights, although discrepancies are within the reported uncertainty bounds down to ~ AOD_{440}>0.3. The tendency of AERONET SSA to be lower suggests either that AERONET retrievals ~~an~~over-estimate~~d to over-predict~~ absorbing aerosol or that the in-situ measurements under-estimate aerosol absorption~~relative to coincident and co-located in-situ vertical profile measurements~~. ~~This suggests~~Since the observed discrepancy in SSA can not definitively be attributed to either technique, the idea of ~~that~~ scaling modelled black carbon concentrations upwards to match AERONET retrievals of AAOD should be approached with caution. If the AERONET SSA and AAOD retrievals are indeed biased towards higher absorption, such an upscaling may lead to aerosol absorption overestimates, particularly in regions of low AOD. ~~The magnitude of this~~

~~bias is unknown and may be close to or within the uncertainty estimates for the retrievals.~~ If the discrepancy between the in-situ and AERONET AAOD is due to issues with the in-situ measurements of absorption, the only way we see to increase the in-situ absorption values is a significant enhancement (on the order of a factor of 2 or more) in absorption due to a coating effect. While that level of absorption enhancement factor is within the range suggested by modelling studies, it is significantly higher than many observations of absorption enhancement for ambient aerosol reported in the literature.

The AERONET retrievals of SSA and AAOD have been used as a primary constraint on global model simulations of aerosol absorption. Using only Level-2 retrievals of AAOD (i.e., for AOD_{440}>0.4) on a global scale (e.g., Koch et al., 2009; Bahadur et al., 2010; Chung et al., 2012; Buchard et al., 2015; Pan et al., 2015; Li et al., 2015) is likely to lead to significant over-estimates of absorption in cleaner regions although it may be appropriate for conditions of high loading. Several different approaches of varying complexity have been developed to better represent absorbing aerosol for cleaner conditions. Some of these approaches utilize SSA at high AOD to estimate AAOD at lower AOD conditions (e.g., Bond et al., 2013; Wang et al., 2014), while others utilize Level-1.5 retrievals with the added uncertainty that entails (e.g., Lacagnina et al., 2015; Mallet et al., 2013). Based on the analysis presented here, we cannot say how to best estimate SSA or AAOD from AERONET retrievals for the low AOD conditions prevalent around much of the globe.

Some in-situ measurements suggest that a systematic relationship exists between SSA and AOD, but these measurements are spatially sparse and typically not made at ambient conditions. Nonetheless, systematic relationships between SSA and AOD, similar to those seen in the in-situ data at the two sites, are also observed for multiple North American AERONET sites. The existence of such a systematic relationship may limit the accuracy of AAOD estimates when climatological values for SSA from high AOD retrievals are assumed to apply at~~for~~ low loading conditions. However, for the two mid-continental rural sites studied here, the ~~use of~~ statistically-based ~~values~~ monthly medians of SSA from Level-2.0 inversions (i.e., ~~local or regional~~ SSA values derived for AOD_{440}>0.4 ~~for higher AOD conditions~~) appear to be quite consistent with monthly SSA values obtained from in-situ measurements and AEROCOM model simulations. This suggests that, at these two sites, using the Level-2.0 inversion SSA to retrieve monthly AAOD at lower AOD conditions (e.g., AAOD=AOD*SSA) ~~will at least avoid a high bias for~~would not bias the resulting monthly AAOD high, as would occur if only AAOD values for high AOD cases are included in the AAOD statistics. ~~when absorption at low AOD (AOD_{440}<0.4) is not considered (as is the case for AERONET Level-2.0 data)~~. This may not be true for other locations or averaging times. Further, for these two sites, a more complex approach to retrieve monthly AAOD is needed for very clean months when no Level-2.0 inversions are available.

This study points to several areas where additional research would be useful in resolving the observed AERONET/in-situ absorption-related discrepancies. First, continued laboratory, field and modelling efforts are needed to elucidate and unify the current inconsistencies in the literature on the effects of coatings on absorption enhancement reported for field and lab measurements and for model simulations. Second, a more extensive evaluation of the hygroscopicity of ambient (not lab-generated!) absorbing particles would be helpful. Third,

better characterization of how filter-based measurements of absorption respond to coated particles would be useful, not just in the context of this study, but also for improving our understanding of the in-situ absorption data acquired by long-term, surface aerosol monitoring networks (e.g., GAW). Finally, the development of a focused measurement program designed to acquire statistically robust in-situ vertical profiles over AERONET sites representing a wide range of conditions and aerosol types could be used to explore the relationships between retrievals of column properties and variable aerosol profiles and to provide further validation of the inversion retrieval data products.

Acknowledgements

Funding for the SGP airplane measurements was provided by DOE/ARM, while the BND airplane measurements were funded by NOAA's Climate Program Office. Greenwood Aviation and FlightSTAR provided fabulous pilots and maintenance of the aircraft. Derek Hageman (University of Colorado) is owed much for his coding genius and Patrick Sheridan (NOAA/ESRL/GMD) for making instruments on both airplanes work and ensuring the quality of the BND aircraft data set. We thank the modelling groups for providing AeroCom Phase II results. We greatly appreciate the ease of access to and use of IMPROVE aerosol chemistry data for the Bondville and Cherokee Nation IMPROVE sites. IMPROVE data were downloaded from: http://views.cira.colostate.edu/web/DataWizard/. This study was supported by NOAA Climate Program Office's Atmospheric Chemistry, Carbon Cycle and Climate (AC4) program. BHS acknowledges funding by the Research Council of Norway through the grants AC/BC (240372) and NetBC (244141). We also gratefully acknowledge the very helpful discussions from David Giles and Brent Holben from AERONET.

References

Adam M., Putaud, J.P., Martins dos Santos, S., Dell'Acqua, A., Gruening, C., "Aerosol hygroscopicity at a regional background site (Ispra) in Northern Italy," Atmos. Chem. Phys., 12, 5703–5717, 2012.

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K., "Mesoscale Variations of Tropospheric Aerosols," J. Atmos. Sci., 60, 119–136, 2003.

Anderson, T. L. and Ogren, J. A., "Determining aerosol radiative properties using the TSI 3563 integrating nephelometer," Aerosol Sci. Technol., 29, 57–69, 1998.

Andreae, M. O., Gelencsér, A., "Black carbon or brown carbon? The nature of light-absorbing carbonaceous aerosols," Atmos. Chem. Phys., 6, 3131-3148, 2006.

Andrews, E., Sheridan, P.J., Ogren, J.A., "Seasonal Differences in the Vertical Profiles of Aerosol Optical Properties over Rural Oklahoma," *Atmos. Chem. Phys.,* 11, 10661–10676, 2011a.

Andrews, E., et al. "Climatology of aerosol radiative properties in the free troposphere," Atmos. Res., 102, 365-393, 2011b.

Arnott, W.P., Hamasha, K., Moosmuller, H., Sheridan, P.J., Ogren, J.A., "Towards aerosol light-absorption measurements with a 7-wavelength aethalometer: evaluation with a photoacoustic instrument and 3-wavelength Nephelometer," Aerosol Sci. Technol, 39, 17-29, 2005.

Bahadur, R., Praveen, P. S., Xu, Y., and Ramanathan, V.,"Solar absorption by elemental and brown carbon determined from spectral observations," P. Natl. Acad. Sci. USA, 109, 43, 17366–17371, doi:10.1073/pnas.1205910109, 2012.

Bergin, M.H., Ogren, J.A., Schwarz, S.E., McInnes, L.M., "Evaporation of Ammonium Nitrate Aerosol in a Heated Nephelometer: Implications for field measurements," Environ. Sci. Technol., 31, 2878-2883, 1997.
Bond, T. C., Anderson, T. L., and Campbell, D., "Calibration and intercomparison of filter-based measurements of visible light absorption by aerosols," Aerosol Sci. Technol., 30, 582–600, doi:10.1080/027868299304435, 1999.
Bond, T. C., et al., "Bounding the role of black carbon in the climate system: A scientific assessment," J. Geophys. Res., 118, 5380–5552, doi:10.1002/jgrd.50171, 2013.
Bond, T. C., Habib, G., Bergstrom, R.W., " Limitations in the enhancement of visible light absorption due tomixing state," J. Geophys. Res., 111, D20211, doi:10.1029/2006JD007315, 2006.
Brem, B.T., Mena Gonzalez, F.C., Meyers, S.R., Bond, T.C., Rood, M.J., "Laboratory-Measured Optical Properties of Inorganic and Organic Aerosols at Relative Humidities up to 95%," Aerosol Sci. Technol., 46:2, 178-190, 2012.
Buchard, V. da Silva, A.M., Colarco, P.R., Darmenov, A., Randles, C.A., Govindaraju, R., Torres, O., Campbell, J., Spurr, R. "Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis," Atmos. Chem. Phys., 15, 5743–5760, 2015.
Burrows, J. P., Dehn, A., Deters, B., Himmelmann, S., Richter, A., Voigt, S. and Orphal, J., "Atmospheric Remote-Sensing Reference Data from GOME: Part 1. Temperature-Dependent Absorption Cross-sections of NO2in the 231-794 nm Range," JQSRT,60, 1025-1031, 1998.
Carrico, C.M., Kus, P., Rood, M.J., Quinn, P.K., Bates, T.S., "Mixtures of pollution, dust, sea salt and volcanic aerosol during ACE-Asia: Radiative properties as a function of relative humidity," J. Geophys. Res., 108, doi:10.1029/2003JD003405, 2003.
Cappa, C.D., et al., "Radiative absorption enhancements due to the mixing state of atmospheric black carbon," Science, 337, 1078-1081, 2012.
Chin, M., T. Diehl, T., Dubovik, O., Eck, T.F., Holben, B.N., Sinyuk, A., Streets, D.G., "Light absorption by pollution, dust, and biomass burning aerosols: a global model study and evaluation with AERONET measurements," Ann. Geophys., 27, 3439-3464, 2009.
Chung, C., Ramanathan, V, Decremer, D., "Observationally constrained estimates of carbonaceous aerosol radiative forcing," PNAS, 109, 11624-11629, 2012.
Corr, C.A., Krotkov, N., Madronich, S., Slusser, J.R., Holben, B., Gao, W., Flynn, J., Lefer, B., Kreidenweis, S.M., "Retrieval of aerosol single scattering albedo at ultraviolet wavelengths at the T1 site during MILAGRO," Atmos. Chem. Phys, 9, 5813-5827, 2009.
Corrigan, C.E., Roberts, G.C., Ramana, M.V., Kim, D., Ramanathan, V., "Capturing vertical profiles of aerosols and black carbon over the Indian Ocean using autonomous unmanned vehicles," Atmos. Chem Phys, 8, 737-747, 2009.
Crumreyrolle, S., Chen, G., Ziemba, L., Beyersdorf, A., Thornhill, L., Winstead, E., Moore, R., Shook, M.A., Anderson, B., "Factors that influence surface PM2.5 values inferred from satellite observations: perspective gained for the Baltimore-Washington Area during DISCOVER-AQ," Atmos. Chem. Phys., 14, 2139-2153, 2014.
Delene, D. J. and Ogren, J. A., "Variability of aerosol optical properties at four North American surface monitoring sites," J. Atmos. Sci., 59, 1135–1150, 2002.
Doran, J.C., Barnard, J.C., Arnott, W.P., Cary, R., Coulter, R., Fast, J.D., Kassianov, E.I., Kleinman, L., Laulainen,, N.S., Martin, T., Paredes-Miranda, G., Pekour, M.S., Shaw, W.J., Smith, D.F., Springston, S.R., Yu, X.-Y., "The T1-T2 study: evolution of aerosol properties downwind of Mexico City," Atmos. Chem. Phys., 7, 1585-1598, 2007.
Dubovik, O. and King, M.D., "A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements," J. Geophys. Res., 105, 20673-20696, 2000.

Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Munoz, O., Veihelmann, B., van der Zande, W.J., Leon, J.-F. ,Sorokin, M., Slutsker, I., "Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust," J. Geophys. Res., 111, doi:10.1029/2005JD006619, 2006.

Dubovik, O., A. Smirnov, B.N. Holben, M.D. King, Y. J. Kaufman, T.F. Eck and I. Slutsker, "Accuracy assessment of aerosol optical properties retrieval from AERONET sun and sky radiance measurements," J. Geophys. Res. , 105, 9791-9806, 2000.

Eck, T.F., Holben, B.N.,.Reid, J.S., Dubovik, O., Smirnov, A., O'Neill, N.T., Slutsker, I. Kinne, S., "Wavelength dependence of the optical depth of biomass burning, urban and desert dust aerosols," J. Geophys. Res., 104, 31 333-31 350, 1999.

Esteve, A.R., Ogren, J.A., Sheridan, P.J., Andrews, E., Holben, B.N., and Utrillas, M.P., "Statistical evaluation of aerosol retrievals from AERONET using in-situ aircraft measurements," Atmos. Chem. Phys., 12, 2987-3003, 2012.

Ferrero, L., Mocnik, G., Ferrini, B.S., Perrone, M.G., Sangiorgi, G., Bolzacchini, E., "Vertical profiles of aerosol absorption coefficient from micro-aethalometer data and Mie calculation over Milan," Sci. Tot. Environ, 409, 2824-2837, 2011.

Formenti, P., et al., "STAAARTE-MED 1998 summer airborne measurements over the Aegean Sea, 2, Aerosol scattering and absorption, and radiative calculations," J. Geophys. Res., 107(D21), 4451, doi:10.1029/2001JD001536, 2002.

Giles, D.M., Holben, B.N., Eck, T.F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickerson, R.R., Thompson, A.M., Schafer, J.S.," An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source region," J. Geophys. Res., 117, doi: 10.1029/2012JD018127, 2012.

Hamonou, E.,. Chazette, P., Balis, D., Schneider, X., Galani, E., Ancellet, G., Papayannis, A., "Characterization of the vertical structure of Saharan dust export to the Mediterranean basin," J. Geophys. Res., 104, 22257-22270, 1999.

Hänel, G., "The properties of atmospheric aerosol particles as functions of the relative humidity at thermodynamic equilibrium with surrounding moist air," Adv. Geophys., 73-188, 1976.

Hansen, J., Sato, M., Ruedy, R., "Radiative forcing and climate response," J. Geophys. Res., 102, 6831-6864, 1997.

Hartley, W.S., Hobbs, P.V., Ross, J.L., Russell, P.B., Livingston, J.M., "Properties of aerosols aloft relevant to direct aerosol radiative forcing off the mid-Atlantic coast of the United States," J. Geophys. Res., 105, 9859-9885, 2000.

Haywood, J.M., P. Francis, O. Dubovik, M. Glew, and B. Holben, Comparison of aerosol size distributions, radiative properties, and optical depths determined by aircraft observations and Sun photometers during SAFARI 2000, J. Geophys. Res.,108(D13), 8471, doi:10.1029/2002JD002250, 2003.

Haywood, J.M. and Ramaswamy, V, "Global sensitivity studies of the direct radiative forcing due to anthropogenic sulfate and black carbon aerosols," J. Geophys. Res., 103, 6043-6058, 1998.

Haywood, J.M. and Shine, K.P., "The effect of anthropogenic sulfate and soot aerosol on the clear sky planetary radiation budget," Geophys. Res. Lett., 22, 603-606, 1995.

Hinds, W.C., Aerosol Technology: Properties, behavior and measurement of airborne particles, John Wiley and Sons, New York, 1982.

Holben, B. N., et al., "AERONET—A federated instrument network and data archive for aerosol characterization," Remote Sens. Environ., 66, 1–16, 1998.

Holben, B.N., Eck, T.F., Slutsker, I., Smirnov, A., Sinyuk, A., Schafer, J., Giles, D., Dubovik O., "AERONET's Version 2.0 quality assurance criteria," http://aeronet.gsfc.nasa.gov/new_web/Documents/AERONETcriteria_final1.pdf, 2006.

Huffman, J.A., Docherty, K.S., Mohr, C., Cubison, M.J., Ulbrich, I.M., Ziemann, P.J., Onasch, T.B., Jiminez, J.L., "Chemically-resolved volatility measurements of organic aerosol from different sources," Environ. Sci. Technol., 43, 5351-5357, 2009

Jeong, M.-J. and Li, Z., "Separating real and apparent effects of cloud, humidity, and dynamics on aerosol optical thickness near cloud edges," J. Geophys. Res., 115, doi:10.1029/2009JD013547, 2010.

Johnson, B.T., Christopher, S., Haywood, J.M., Osborne, S.R., McFarlane, S., Hsu, C., Salustro, C., Kahn, R., "Measurements of aerosol properties from aircraft, satellite and ground-based remote sensing: A case-study from the Dust and Biomass-burning Experiment (DABEX)," Q. J. R. Meteorol. Soc., 135, 922–934, 2009.

Johnson, B.T., Osborne, S.R., "Physical and optical properties of mineral dust aerosol measured by aircraft during the GERBILS campaign," Q.J.R.Meteorol. Soc., 137, 1117-1130, 2011.

Kanaya,Y., Taketani, F., Komazaki, Y., Liu, X., Kondo, Y., Sahu, L., Irie, H., Takashima, H., "Comparison of Black Carbon Mass Concentrations Observed by Multi-Angle Absorption Photometer (MAAP) and Continuous Soot-Monitoring System (COSMOS) on Fukue Island and in Tokyo, Japan," Aerosol Sci. Technol., 7:1, 1-10, DOI: 10.1080/02786826.2012.716551, 2013.

Kasten, F., "Visibility forecast in the phase of pre-condensation," Tellus, 21, 631-635, 1969.

Kelektsoglou, K., Rapsomanikis, S., Karageorgos, E.T., Kosmadakis, I., Optical properties of aerosol over a Southern European urban environment," Int.J. Remote Sens., 33, 1214-1233, 2012.

Kinne et al., "An AeroCom initial assessment – optical properties in aerosol component modules of global models," Atmos. Chem. Phys., 6, 1–20, 2006.

Koch, D., et al., "Evaluation of black carbon estimations in global aerosol models," Atmos.Chem. Phys., 9, 9001–9026, doi:10.5194/acp-9-9001-2009, 2009.

Kotchenruther, R. A., Hobbs, P.V., Hegg, D.A., "Humidification factors for atmospheric aerosols off the mid-Atlantic coast of the United States," J. Geophys. Res., 104, 2239– 2251, 1999.

Lacagnina, C., Hasekamp, O.P., Bian, H., Curci, G., Myhre, G., van Noije, T, Schulz, M., Skeie, R.B., Takemura, T., Zhang, K., "Aerosol single-scattering albedo over the global oceans: Comparing PARASOL retrievals with AERONET, OMI, and AeroCom models estimates," J. Geophys. Res. Atmos., 120, 9814–9836, doi:10.1002/2015JD023501, 2015.

Lack, D.A., et al., "Relative humidity dependence of light absorption by mineral dust after long-range atmospheric transport from the Sahara," Geophys. Res. Lett., 36, L24805, doi:10.1029/2009GL041002, 2009.

Lack, D.A., Cappa, C.D., Covert, D.S., Baynard, T., Massoli, P., Sierau, B., Bates, T.S., Quinn, P.K., Lovejoy, E.R., Ravishankara, A.R., "Bias in Filter-Based Aerosol Light Absorption Measurements Due to Organic Aerosol Loading: Evidence from Ambient Measurement," Aerosol Sci. and Technol., 42, 1033–1041, 2008.

Lack, D.A., Richardson, M.S., Law, D., Langridge, J.M., Cappa, C.D., McLaughlin, R.J., and Murphy, D.M.: Aircraft Instrument for Comprehensive Characterization of Aerosol Optical Properties, Part 2: Black and Brown Carbon Absorption and Absorption Enhancement Measured with Photo Acoustic Spectroscopy, Aerosol Sci. Tech., 46:5, 555-568, DOI: 10.1080/02786826.2011.645955, 2012.

Leahy, L.V., Anderson, T.L., Eck, T.F., Bergstrom, R.W., "A synthesis of single scattering albedo of biomass burning aerosol over southern Africa during SAFARI 2000," Geophys. Res. Lett., 34, doi:10.1029/2007GL029697, 2007.

Li, S., Kahn, R., Chin, M., Garay, M.J., Liu, Y., "Improving satellite-retrieved aerosol microphysical properties using GOCART data," Atmos. Meas. Tech., 8, 1157–1171, 2015.

Ma, X. and Yu, F., "Seasonal variability of aerosol vertical profiles over east US and west Europe: GEOS-Chem/APM simulation and comparison with CALIPSO observations," Atmos. Res., 140-141, 28-37, ttp://dx.doi.org/10.1016/j.atmosres.2014.01.001, 2014.

Magi, B.I., Hobbs, P.V., Kirchstetter, T.W., Novakov, T., Hegg, D.A., Gao, S., Redemann, J., Schmid, B., "Aerosol Properties and Chemical Apportionment of Aerosol Optical Depth at Locations off the U.S. East Coast in July and August 2001," J. Atmos. Sci., 62, 919-933, 2005.

Magi, B.I., Hobbs, P.V., Schmid, B., Redemann, J., "Vertical profiles of light scattering, light absorption and single scattering albedo during the dry biomass burning season in southern Africa and comparisons of in-situ and remote sensing measurements of aerosol optical depths," J. Geophys. Res., 108, doi:10.1029/2002JD00236, 2003

Mallet, M., Dubovik, O., Nabat, P., Dulac, F., Kahn, R., Sciare, J., Paronis, D., and Léon, J. F., "Absorption properties of Mediterranean aerosols obtained from multi-year, ground-based remote sensing observations," Atmos. Chem. Phys., 13, 9195-9210, 2013.

Mallet, M., Van Dingenen, R., Roger, J. C., Despiau, S., Cachier, H., "In situ airborne measurements of aerosol optical properties during photochemical pollution events," J. Geophys. Res., 110, D03205, doi:10.1029/2004JD005139, 2005.

Mallet, M., Pont, V., Liousse, C., Gomes, L., Pelon, J., Osborne, S., Haywood, J., Roger, J.C., Dubuisson, P., Mariscal, A., Thouret, V., Gouloub, P., "Aerosol direct radiative forcing of Djougou (northern Benin) during the African Monsoon Multidisciplinary, Analysis dry season experiment (special observation period – 0)," J. Geophys. Res., 113, doi:1029/2007JD009419, 2008.

Malm, W.C., Sisler, J.F., Huffman,D., Eldred, R.A., Cahill, T.A., "Spatial and seasonal trends in particle concentration and optical extinction in the United States," J. Geophys. Res. 99, 1347–1370, 1994.

McConnell, C.L., Highwood, E.J., Coe, H., Formenti, P., Anderson, B., Osborne, S., Nava, S. Desboeufs, K., Chen, G., and Harrison, M.A.J., "Seasonal variations of the physical and optical characteristics of Saharan dust: Results from the Dust Outflow and Deposition to the Ocean (DODO) experiment," J. Geophys. Res., 113, doi:10.1029/2007JD009606, 2008.

McMeeking, G. R., Fortner, E., Onasch, T.B., Taylor, J.W., Flynn, M., Coe, H., Kreidenweis, S.M. "Impacts of nonrefractory material on light absorption by aerosols emitted from biomass burning," J. Geophys. Res. Atmos.,119, 12,272–12,286, doi:10.1002/2014JD021750, 2014.

Mendes, L., Eleftheriadis, K., Biskos, G., "Performance comparison of two thermal denuders in volatility tandem DMA measurements," J. Aerosol Sci., 92, 38-52, 2016.

Müller, D., Lee, K.-H., Gasteiger, J., Tesche, M., Weinzierl, B., Kandler, K., Muller, T., Toledano, C., Otto, S., Althausen, D., Ansmann, A., "Comparison of optical and microphysical properties of pure Saharan mineral dust observed with AERONET Sun photometer, Raman lidar, and in situ instruments during SAMUM 2006," J. Geophys. Res., 117, doi: 10.1029/2011JD016825, 2012.

Müller, T., et al. "Characterization and intercomparison of aerosol absorption photometers: result of two intercomparison workshops" Atmos. Meas. Tech., 4, 245–268, 2011.

Myhre G., Samset, B.H., Schulz, M. et al., "Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations," Atmos. Chem. Phys., 13, 1853–1877, 2013

Nessler, R., Weingartner, E., Baltensperger, U., "Effect of humidity on aerosol light absorption and its implications for extinction and the single scattering albedo illustrated for a site in the lower free troposphere," J. Aerosol Sci., 36, 958–972, 2005.

Ogren, J. A., "Comment on "Calibration and Intercomparison of Filter-Based Measurements of Visible Light Absorption by Aerosols", Aerosol Sci. Technol., 44, 589–591, doi:10.1080/02786826.2010.482111, 2010.

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., Thulasiraman, S., "Spectral discrimination of coarse and fine mode optical depth," J. Geophys. Res., 108, 4559-4573, 2003.
Osborne, S. R., Johnson, B. T., Haywood, J. M., Baran, A. J., Harrison, M. A. J., McConnell, C.L., "Physical and optical properties of mineral dust aerosol during the Dust and Biomass-burning Experiment, J. Geophys. Res., 113, D00C03,doi:10.1029/2007JD009551, 2008.
Pan, X., Chin, M., Gautam, R., Bian, H., Kim, D., Colarco, P.R., Diehl, T.L.,Takemura, T., Pozzoli, L., Tsigaridis, K., Bauer, S., Bellouin, N., "A multi-model evaluation of aerosols over South Asia: common problems and possible causes," Atmos. Chem. Phys., 15, 5903–5928, 2015.
Parworth, C., Fast, J., Meib, F., Shipper, T>, Sivaraman, C., Tilp, A., Watson, T., Zhang, Q., "Long-term measurements of submicrometer aerosol chemistry at the Southern Great Plains (SGP) using an Aerosol Chemical Speciation Monitor (ACSM)," Atmos. Environ., 106, 43–55, 2015.
Petzold, A., Ogren, J.A., Fiebig, M., Laj, P., Li, S.-M., Baltensperger, U., Holzer-Popp, T., Kinne, S., Pappalardo, G., Sugimoto, N., Wehrli, C., Wiedensohler, A., Zhang, X.-Y., "Recommendations for reporting "black carbon" measurements," Atmos. Chem. Phys., 13, 8365-8379, 2013.
Quinn, P.K., Bates, T.S., Baynard, T., Clarke, A.D., Onasch, T.B., Wang, W., Rood, M.J., Andrews, E., Allan, J., Carrico, C.M., Coffman, D., Worsnop, D., "Impact of particulate organic matter on the relative humidity dependence of light scattering: A simplified parameterization," Geophys. Res. Lett., 32, doi:10.1029/2005GL024322, 2005.
Ramanathan, V., Carmichael, G., "Global and regional climate changes due to black carbon," Nature – Geosci., 1(4), 221-227, 2008.
Reddy, M. S., Boucher, O. Bellouin, N., Schulz, M., Balkanski, Y., Dufresne, J.L., Pham, M., "Estimates of global multicomponent aerosol optical depth and direct radiative perturbation in the Laboratoire de Meteorologie Dynamique general circulation model," J. Geophys. Res., 110, doi:10.1029/2004jd004757, 2005.
Redemann, J., Russell, P.B., Hamill, P., "., "Dependence of aerosol light absorption and single scattering albedo on ambient relative humidity for sulfate aerosols with black carbon cores," J. Geophys. Res., 106, 27485-27495, 2001.
Reid, J.S., Hobbs, P.V., Liousse, C., Vanderlei Martins, J., Weiss, R.E., Eck, T.F., "Comparisons of techniques for measuring shortwave absorption and black carbon content of aerosols from biomass burning in Brazil," J. Geophys. Res., 103, 32031-32040, 1998.
Samset, B.H. et al., "Black carbon vertical profiles strongly affect its radiative forcing uncertainty," Atmos. Chem. Phys., 13, 2423–2434, 2013.
Sato, M., Hansen, J., Kock, D., Lacis, A., Ruedy, R., Dubovik, O., Holben, B., Chin, M., Novakov, T., "Global atmospheric black carbon inferred from AERONET," PNAS, 100, 6319-6324, 2003.
Schafer, J.S., Eck, T.F., Holben, B.N., Thornhill, K.L., Anderson, B.E., Sinyuk, A., Giles, D.M., Winstead, E.L., Ziemba, L.D., Beyersdorf, A.J., Kenny, P.R., Smirnov, A., Slutsker, I., "Intercomparison of aerosol single-scattering albedo derived from AERONET surface radiometers and LARGE in situ aircraft profiles during the 2011 DRAGON-MD and DISCOVER-AQ experiments," J. Geophys. Res., 119, 7439–7452, doi:10.1002/2013JD021166, 2014.
Schmid, B., Flynn, C.J., Newsom, R.K., Turner, D.D., Ferrare, R.A., Clayton, M.F., Andrews, E., Ogren, J.A., Johnson, R.R., Russell, P.B., Gore, W.J., Dominguez, R., "Validation of aerosol extinction and water vapor profiles from routine Atmospheric Radiation Measurement Program Climate Research Facility measurements," J. Geophys. Res., 114, doi:10.1029/2009JD012682, 2009.
Schutgens, N.A.J., Partridge, D.G., Stier, P., "The importance of temporal collocation for the evaluation of aerosol models with observations," Atmos. Chem. Phys., 16, 1065–1079, 2016.

Schwarz, J.P., Spackman, J.R., Fahey, D.W., Gao, R.W., Lohmann, U., Stier, P., Watts, L.A., Thomson, D.S., Lack, D.A., Pfister, L., Mahoney, M.J., Baumgardner, D., Wilson, J.C., Reeves, J.M., "Coatings and their enhancement of black carbon light absorption in the tropical atmosphere," J. Geophys. Res., 113, D03203, doi:10.1029/2007JD009042, 2008.

Schwarz, J.P., Spackman, J.R., Gao, R.S., Watts, L.A., Stier, P., Schulz, M., Davis, S.M., Wofsy, S.C., Fahey, D.W., "Global-scale black carbon profiles observed in the remote atmosphere and compared to models," Geophys. Res. Lett., 27, doi:10.1029/2010GL044372, 2010.

Sharma, S., Ishizawa, M., Chan, D., Lavoue, D., Andrews, E., Eleftheriadis, K, Maksyutov, S., "16-year simulation of Arctic black carbon: transport, source contribution, and sensitivity analysis on deposition," J. Geophys. Res. 118, 1–22, doi:10.1029/2012JD017774, 2013.

Sheridan, P. J., Andrews, E., Ogren, J. A., Tackett, J. L., and Winker, D. M.: Vertical profiles of aerosol optical properties over Central Illinois and comparison with surface and satellite measurements, Atmos. Chem. Phys., 12, 11695-11721, 2012.

Sheridan, P. J., Arnott, W. P., Ogren, J. A., Andrews, E., Atkinson, D. B., Covert, D. S., Moosm¨uller, H., Petzold, A., Schmid, B. Strawa, A. W., Varma, R., and Virkkula, A.: The Reno Aerosol Optics Study: An Evaluation of Aerosol Absorption Measurement Methods, Aerosol Sci. Tech., 39, 1–16, 2005.

Sheridan, P. J., Delene, D. J., Ogren, J.A., "Four years of continuous surface aerosol measurements from the Department of Energy's Atmospheric Radiation Measurement Program Southern Great Plains Cloud and Radiation Testbed site," J. Geophys. Res., 106, 20735–20747, 2001.

Sheridan, P.J., Jefferson, A., Ogren, J.A., "Spatial variability of submicrometer aerosol radiative properties over the Indian Ocean during INDOEX," J. Geophys. Res., 107, D19, 8011, 10.1029/2000JD000166, 2002.

Sherman, J. P., Sheridan, P. J., Ogren, J. A., Andrews, E., Hageman, D., Schmeisser, L., Jefferson, A., and Sharma, S.: A multi-year study of lower tropospheric aerosol variability and systematic relationships from four North American regions, Atmos. Chem. Phys., 15, 12487-12517, doi:10.5194/acp-15-12487-2015, 2015.

Shinozuka, Y., Redemann, J., Livingston, J. M., Russell, P. B., Clarke, A. D., Howell, S. G., Freitag, S., O'Neill, N. T., Reid, E. A., Johnson, R., Ramachandran, S., McNaughton, C. S., Kapustin, V. N., Brekhovskikh, V., Holben, B. N., and McArthur, L. J. B.: Airborne observation of aerosol optical depth during ARCTAS: vertical profiles, inter-comparison and fine-mode fraction, Atmos. Chem. Phys., 11, 3673–3688, 2011.

Sinha, P., Hobbs, P.V., Yokelson, R.J., Bertschi, I.T., Blake, D.R., Simpson, I., Gao, S., Kirchstetter, T.W., Novakov, T., "Emissions of trace gases and particles from savanna fires in southern Africa," J. Geophys. Res., 108, 8487, doi:10.1029/2002JD002325, 2003.

Sinha, P.R., Kondo, Y., Koike, M., Ogren, J.A., Jefferson, A., Barrett, T.E., Sheesley, R.J., Ohata, S., Moteki, N., Coe, H., Liu, D., Irwin, M., Tunved, P., Quinn, P.K., Zhao, Y., "Evaluation of ground-based black carbon measurements by filter-based photometers at two Arctic sites," in revisions for J. Geophys. Res., 2017.

Skeie, R. B., T. Berntsen, G. Myhre, C. A. Pedersen, J. Ström, S. Gerland, and J. A. Ogren, "Black carbon in the atmosphere and snow, from pre-industrial times until present," Atmos. Chem. Phys., 11, 6809-6836, doi:10.5194/acp-11-6809-2011, 2011.

Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker, "Cloud screening and quality control algorithms for the AERONET database," Remote Sens. Environ., 73, 337– 349, 2000.

Subramanian, R., Roden, C. A., Boparai, P., and Bond, T. C.: Yellow beads and missing particles: Trouble ahead for filter-based absorption measurements, Aerosol Sci. Technol., 41(6), 630–637, DOI: 10.1080/02786820701344589, 2007.

Taubman, B.F., Hains, J.C., Thompson, A.M., Marufu, L.T., Doddridge, B.G., Stehr, J.W., Piety, C.A., and Dickerson, R.R., "Aircraft vertical profiles of trace gas and aerosol pollution over

the mid-Atlantic United States: Statistics and meteorological cluster analysis," J. Geophys. Res., 111, D10S07, doi:10.1029/2005JD006196, 2006.
Turner, D.D., Ferrare, R.A., Brasseur, L.A., "Average Aerosol Extinction and Water Vapor Profiles over the Southern Great Plains," Geophys. Res. Lett., 28, 4441-4444,
Vaden, T.D., Imre, D., Baranek, J., Shrivastava, M., Zelenyuk, A., Finlayson-Pitts, B.J., "Evaporation kinetics and phase of laboratory and ambient secondary organic aerosol," PNAS, 108, 2190-2195, 2011.
Virkkula, A.: Correction of the Calibration of the 3-wavelength Particle Soot Absorption Photometer (3 PSAP), Aerosol Sci. Technol., 44, 706–712, 2010.
Wang, R., et al., "Estimation of global black carbon direct radiative forcing and its uncertainty constrained by observations," J. Geophys. Res. Atmos., 121, 5948–5971, doi:10.1002/2015JD024326, 2016.
Wang, X., Heald, C.L, Ridley, D.A., Schwarz, J.P., Spackman, J.R., Perring, A.E., Coe, H., Liu, D., Clarke, A.D., "Exploiting simultaneous observational constraints on mass and absorption to estimate the global direct radiative forcing of black carbon and brown carbon," Atmos. Chem. Phys., 14, 10989–11010, 2014.
Yu, F., Chin, M., Winker, D.M., Omar, A. H., Liu, Z., Kittaka, C., 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.

Tables

Table 1a Statistical values (medians, means and standard deviations) of AERONET versus in-situ comparison where there was an AERONET retrieval within +/-3 h of the end of a 2 h flight profile. AERONET values are for Level-1.5 data when there was a Level-2 AOD value and an almucantar retrieval. (First value in each cell is median; second set of values in each cell are mean± Std.Dev, third row is number of AERONET retrievals corresponding to flights (in AERONET columns) or number of flights (In-situ columns)). These numbers represent the blue points in Figures 3-5.

	BND		SGP	
	AERONET	In-situ	AERONET	In-situ
AOD	0.1182; 0.14631±0.09987	0.11425; 0.13527±0.139095	0.13875; 0.14688±0.099109	0.13798; 0.14795±0.077106
AAOD	0.0130; 0.0132±0.007	0.003; 0.005±0.006	0.01925; 0.0236±0.00811	0.0045; 0.0046±0.0034
SSA	0.895906; 0.898904±0.03446	0.9617; 0.9647±0.02016	0.84754; 0.83948±0.03841	0.971; 0.9732±0.011
#	51 retrievals[1]	21 flights	23 retrievals[1]	11 flights

[1]retrievals are flight-averaged prior to calculating statistics.

Table 1b Statistical values (medians, means and standard deviations) of AERONET versus in-situ comparison where there was an AERONET retrieval within +/-3 h of the end of a 2 h flight profile and AERONET AOD_{440}>0.2. AERONET values are for Level-1.5 data when there was a Level-2 AOD value and an almucantar retrieval. (First value in each cell is median; second set of values in each cell are mean± Std.Dev, third row is number of AERONET retrievals corresponding to flights (in AERONET columns) or number of flights (In-situ columns)). These numbers represent the purple points in Figures 4-5.

	BND		SGP	
	AERONET	In-situ	AERONET	In-situ
AOD	0.306; 0.304±0.125	0.299; 0.331±0.230	0.269	0.238
AAOD	0.025; 0.019±0.012	0.010; 0.013±0.012	0.034	0.009
SSA	0.941; 0.942±0.023	0.971; 0.966±0.010	0.875	0.964
#	6 retrievals[1]	4 flights	2 retrievals	1 flights

[1]retrievals are flight-averaged prior to calculating statistics.

Table 2 Number of AERONET/IN-SITU AOD and AAOD flight matches as a function of various AERONET constraints~~AERONET match criteria~~ and the +/- 3h time window.

	BND (2006-2009)	SGP(2005-2007)
Total profile flights	402	171
Level-2 AOD	~~72~~73	~~40~~37
Level-2 AOD+almucantar retrieval[1]	~~5~~216~~6~~	~~2~~119~~9~~
Level-2 AOD+almucantar retrieval+AOD_{440}>0.20	2~~6~~	1~~7~~
Level-1.5* AAOD	~~24~~21	~~14~~11
Level-1.5* AAOD + AOD_{440}>0.20	4~~6~~	1~~7~~
Level-2 AAOD	1~~2~~	0~~4~~

[1]an almucantar retrieval does not necessarily imply an AAOD retrieval

Table 3 Direct AAOD comparisons – AERONET ("RS") vs In-Situ ("IS")

Study, # profiles Citation(s)	**Location, aerosol type AOD comments Alt. range**	**Instruments corrections size cut**	**AAOD comparison information**	**Comments**
BND 24 profiles This study	Central US Rural, continental AOD_{440} range: 0.04-0.55 AOD comparison See Fig. 3a 150-4200 m agl	PSAP-3wave TSI neph-3wave B1999, O2010, AO1998 f(RH) adjust Dp<5-7 μm	Wavelength=440, 670 nm Ångström interpolation RS AAOD>IS AAOD $AAOD_{440}$ range: 0.001-0.042	Profiles matched within 3 hours of AERONET measurement. Profiles within 15 km of AERONET measurement. Used V2 AERONET Level 1.5 AAOD values for cases with valid V2 AERONET Level 2.0 AOD value. Extrapolated from lowest altitude range to ground to account for aerosol below plane
SGP 14 profiles This study	Central US Rural, continental AOD_{440} range: 0.06-0.43 AOD comparison See Fig. 3b 150-4200 m agl	PSAP-3wave TSI neph-3wave B1999, O2010, AO1998 f(RH) adjust Dp<5-7 μm	Wavelength=440, 670 nm Ångström interpolation RS AAOD>IS AAOD $AAOD_{440}$ range: 0.012-0.052	Profiles matched within 3 hours of AERONET measurement. Profiles within 1 km of AERONET measurement Used V2 AERONET Level 1.5 AAOD values for cases with valid V2 AERONET Level 2.0 AOD value. Extrapolated from lowest altitude range to ground to account for aerosol below plane
MAC 13 profiles Corrigan et al., 2008	Indian Ocean Pollution AOD_{440} range: 0.1-0.6	Aethalometer 3-wave OPC +Mie for scattering A2005	Wavelength not provided RS AAOD>IS AAOD $AAOD_{440}$ range: 0.005-0.033	No details on how profiles matched with retrievals in terms of time or distance. No details on version of AERONET data used; this is relevant, given low AODs in first half of study – not sure if there were comparisons for low AODs.

	No AOD comparison 0-3200 m asl	Dp<5 μm		Note: this study is the one cited by Bond et al. (2013) to support the use of AERONET to scale modeled BC values

In-situ instrument corrections: B1999=Bond et al., 1999; O2010=Ogren, 2010, AO1998=Anderson and Ogren, 1998; A2005=Arnott et al., 2005; Ångström interpolation – indicates in-situ wavelength adjusted to AERONET wavelength using Ångström interpolation; f(RH) adjust – indicates the in-situ measurements were adjusted to ambient humidity conditions for the AOD comparison. IS=In-situ measurements, RS=Remote sensing (AERONET) measurements

Table 4 SSA comparisons – AERONET vs In-situ

Study, # profiles Citation(s)	**Location, aerosol type AOD comments Alt. range**	**Instruments, Corrections, Inlet size cut**	**SSA comparison information**	**Comments**
BND 24 profiles This study	Central US Rural, continental AOD_{440} range: 0.04-0.55 AOD comparison: See Fig. 2a 150-4200 m agl	PSAP-3wave TSI neph-3wave B1999, O2010, AO1998 f(RH) adjust Dp<5-7 μm	Wavelength=440, 670 nm Ångström interpolation RS SSA<IS SSA	Profiles matched within 3 hours of AERONET measurement. Profiles within 15 km of AERONET measurement. Used V2 AERONET Level 1.5 AAOD values for cases with valid V2 AERONET Level 2.0 AOD value. Extrapolated from lowest altitude range to ground to account for aerosol below plane
SGP 14 profiles This study	Central US Rural, continental AOD_{440} range: 0.06-0.43 AOD comparison:	PSAP-3wave TSI neph-3wave B1999, O2010, AO1998 f(RH) adjust	Wavelength=440, 670 nm Ångström interpolation RS SSA<IS SSA	Profiles matched within 3 hours of AERONET measurement. Profiles within 1 km of AERONET measurement. Used V2 AERONET Level 1.5 AAOD values for cases with valid V2 AERONET Level 2.0 AOD value.

	See Fig. 2b 150-4200 m agl	Dp<5-7 μm		Extrapolated from lowest altitude range to ground to account for aerosol below plane
AAO (BND) 1 profile Esteve et al., 2012	Central US Rural, continental AOD_{550} = 0.65 AOD comparison: RS AOD>IS AOD 150-4200 m agll	PSAP-3wave TSI neph-3wave B1999, O2010, AO1998 f(RH) adjust Dp<5-7 μm	Wavelength=550 nm Power law interpolation RS SSA < IS SSA	Profiles matched within 2 hours of AERONET measurement. Profiles within 15 km of AERONET measurement. Used V2 AERONET Level 2.0 AOD value. Extrapolated from lowest altitude range to ground to account for aerosol below plane
DISCOVER-AQ 12 profiles Schafer et al., 2014	East Coast US Polluted air AOD_{440}>0.2 AOD compare: RS AOD > IS AOD (by 23%)* 367-3339 m	PSAP-3wave TSI neph-3wave V2010, AO1998* f(RH) adjust Dp<4 μm*	Wavelength=550 nm AERONET "interpolated" to 550 (no detail provided) In-situ absorption interpolated to 550 using Ångström interpolation RS SSA < IS SSA	Profile matched within 45 min of AERONET measurement. Profile within 1 km of AERONET measurement. Used V2 AERONET Level 2.0 values in paper Altitude range: at least <500 m and >1500 m for column comparisons, min and max altitudes: 367 m and 3339 m Did not specify agl or asl but those are similar for the location.
CLAMS 1 profile Magi et al., 2005	East Coast US Polluted air AOD_{440}=0.60 AOD comparison: RS AOD > IS AOD (by 15%)	PSAP-1wave MSE neph-3wave B1999, AO1998 f(RH) adjust Inlet size cut not reported, Sinha,	Wavelength=550 nm Wave_adj =quadratic polynomial interpolation RS SSA < IS SSA	Profile matched within 1 hour of AERONET measurement. Profile within 3 km of AERONET measurement. ~~Used~~ Retrieved V2 AERONET Level 2.0 AOD_{440} from http://aeronet.gsfc.nasa.gov/~~,~~ ~~not stated in paper.~~

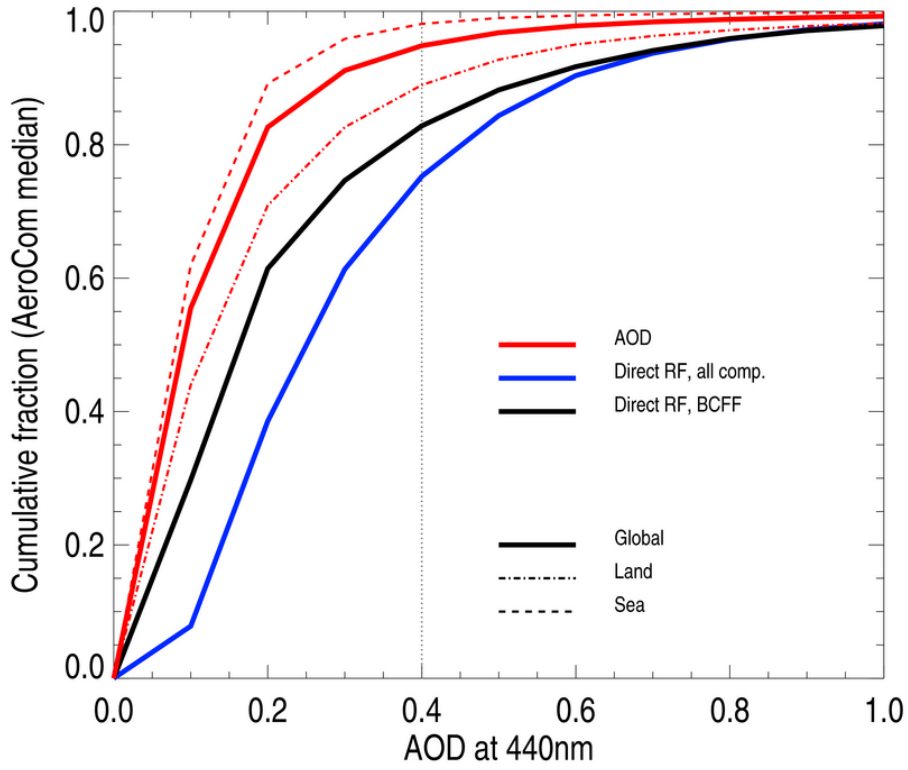
	170-1500 m agl	2003 suggests Dp<4 μm		Also compared campaign AERONET average with profile average: SSA's much closer, but profiles weren't necessarily close in time or space to AERONET site
ESCOMPTE 1 profile Mallet et al., 2005	Avignon, France Pollution AOD_{440}>0.55 No AOD comparison 100-2900 m	PSAP-1wave TSI neph-3wave B1999, A1999 No f(RH) adj Inlet Dp not given	Wavelength=550 nm Wave_adj = estimated from visual inspection (spectral dependence is relatively flat) RS SSA < IS SSA	Profile matched within 1 hour of AERONET measurement. Profile within 10 km of AERONET measurement. Used V2 AERONET Level 2.0 AOD_{440} from http://aeronet.gsfc.nasa.gov/, not stated in paper. Did not adjust in-situ measurements for f(RH), so presumably IS SSA would increase so it was even larger than RS SSA. Did not specify agl or asl
SAFARI 5 profiles Leahy et al., 2007 UW plane	Southern Africa Biomass burning AOD_{550}>0.28-1.12 AOD comparison: RS AOD > IS AOD RS=1.12*IS-0.05 R^2=0.99 100-5320 m asl	PSAP-1wave MSE neph-3wave B1999;H2000 f(RH) adjust Dp<4 um	Wavelength=550 nm Wave_adj= 2nd order polynomial For AOD_{550}>0.6 (3 profiles) RS SSA > IS SSA For AOD_{550}<0.3 (2 profiles) RS SSA ≤ IS SSA	Profiles matched within 1-4 hours of AERONET measurement. Profiles within 20 km of AERONET measurement. Used V2 AERONETLevel 2.0 Also found: AEROCOM model>insitu Altitude range is min and max over 5 flights – no flights covered that entire range). They used AATS to account for aerosol above plane and extrapolated down to acct for aerosol below plane. (Altitude range from flight info in Magi et al., 2003)

SAFARI 1 profile Haywood et al, 2003 C-130	Southern Africa Biomass burning AOD_{440}=0.71 AOD comparison: RS AOD < IS AOD 330-3420 m agl	PSAP-1wave TSI neph-3wave B1999,AO1998 No f(RH) adj Dp<2-4 μm	Wavelength=native Wave_adj = none RS SSA < IS SSA	Profile matched within 2 hours of AERONET measurement. Profiles within 10 km of AERONET measurement. ~~Used~~ Used V2 AERONET Level 2.0 data from http://aeronet.gsfc.nasa.gov/ They defend the lack of f(RH) correction because (a) ambient RH values < 56% and (b) previous measurements of f(RH) of BB aerosol suggest minimal hygroscopicity Paper mostly focused on size dist comparison; SSA comparison seems like afterthought. Extrapolated from lowest altitude range to ground to account for aerosol below plane
DABEX 3 profiles Osborne et al., 2008	Africa Dust/BB AOD comparison RS AOD < IS AOD (by up to 40%) AOD_{550}~0.3-0.6 100-5000 m	PSAP-1wave TSI neph-3wave B1999,AO1998 No f(RH) adj Dp<2-4 μm	Wavelength=550 nm Wave_adj=log interpolation RS SSA < IS SSA	No details on how profiles matched with retrievals in terms of time. Profiles within 100 km of AERONET measurement Used V2 AERONET Level 2.0 They defend the lack of f(RH) correction because (a) ambient RH values are mostly low (<60%) and (b) previous measurements of f(RH) of BB aerosol suggest minimal hygroscopicity Jan 21, 23 and 30 profiles IS overpredicts AOD so IS SSA is greater

				than RS SSA Suggest it could be due to large particle correction to IS measurements using PCASP. McConnell et al., (2008) suggests problems with nephelometer sensitivity Did not specify agl or asl Altitude range is min and max over 4 flights – no flights covered that entire range
DABEX 1 profile Johnson et al., 2009	Africa DUST/BB AOD comparison: RS AOD < IS AOD (by ~10%) AOD_{550} > 0.7 150-3000 m	PSAP-1wave TSI neph-3wave B1999,AO1998 No f(RH) adj Dp<2-4 μm	Wavelength=550 nm Wave_adj=log interpolation RS SSA < IS SSA	Profile matched within 1 hour of AERONET measurement. Profile within 100 km of AERONET measurement Used V2 AERONET Level 2.0 They defend the lack of f(RH) correction because ambient RH values are mostly low (<40% with a max of 70%) Jan 19 profile Incorrectly used Mie to adjust σ_{ap} to 550 after B1999 applied Did not specify agl or asl

IS=In-situ measurements, RS=Remote sensing (AERONET) measurements. In-situ instrument corrections: B1999=Bond et al., 1999; V2010=Virkula et al., 2010;O2010=Ogren, 2010; AO1998=Anderson and Ogren, 1998; H2000=Hartley et al., 2000; A2005=Arnott et al., 2005; Ångström interpolation – indicates wavelength adjustment using Ångström exponent interpolation; f(RH) adjust – indicates the in-situ measurements were adjusted to ambient humidity conditions for the AOD and SSA comparison.
*Information about Discover-AQ flights from Crumreyrolle et al. (2014)

Figures


Cumulative fraction (AeroCom median)
AOD at 440nm
0.0
0.2
0.4
0.6
0.8
1.0
AOD
Direct RF, all comp.
Direct RF, BCFF
Global
Land
Sea

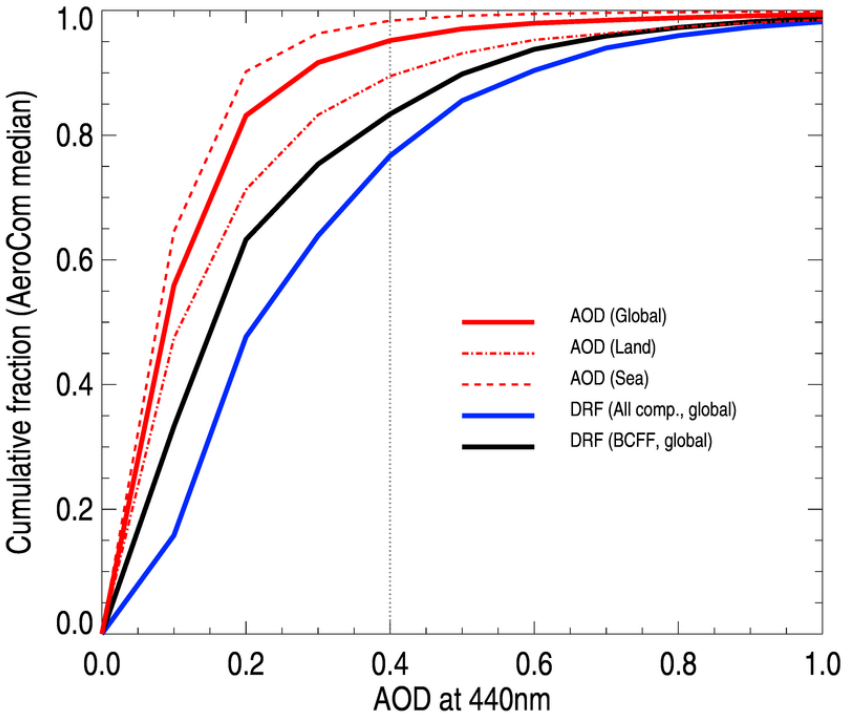

Cumulative fraction (AeroCom median)
AOD at 440nm
0.0
0.2
0.4
0.6
0.8
1.0
AOD (Global)
AOD (Land)
AOD (Sea)
DRF (All comp., global)
DRF (BCFF, global)

Figure 1. Cumulative AOD_{440} frequency distribution (red lines) based on output from five AeroCom models. Blue and black lines show contribution of total aerosol and fossil fuel black carbon, respectively, to the global radiation budget as a function of AOD_{440}. See text for details. Models used to generate the AOD lines include: GMI-MERRA-v3, GOCART-v4, LMDZ-INCA, OsloCTM2, and SPRINTARS-v385. Models used to generate the radiative forcing lines include all but the GMI-MERRA-v3 model. Model information and references can be found in Myhre et al., (2013).

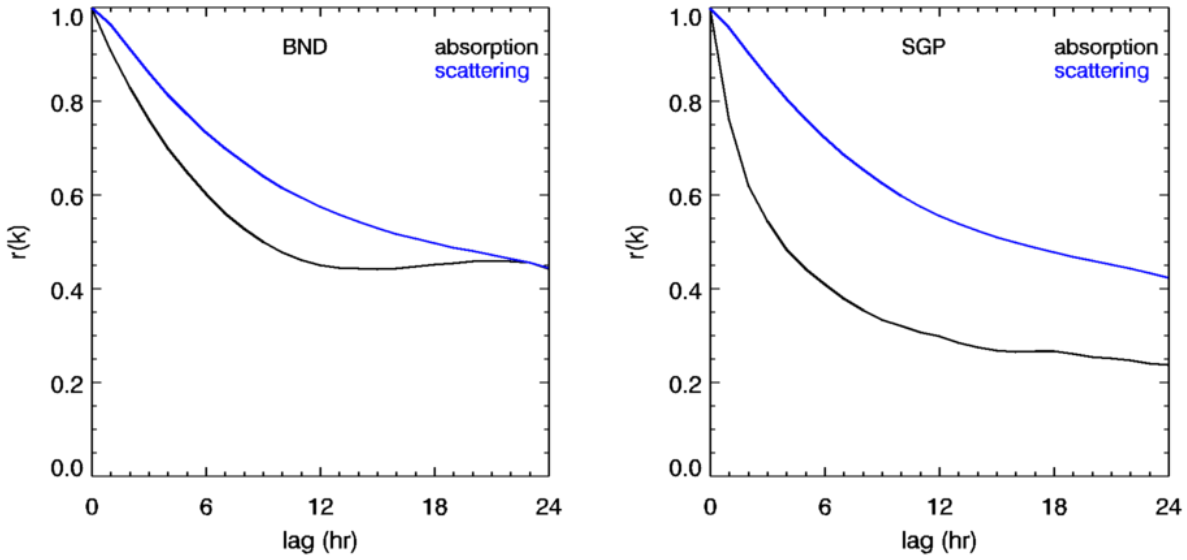
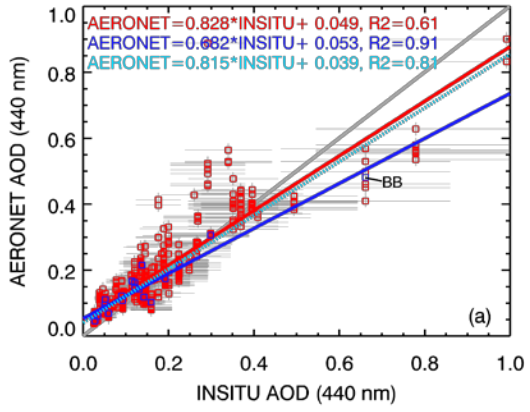
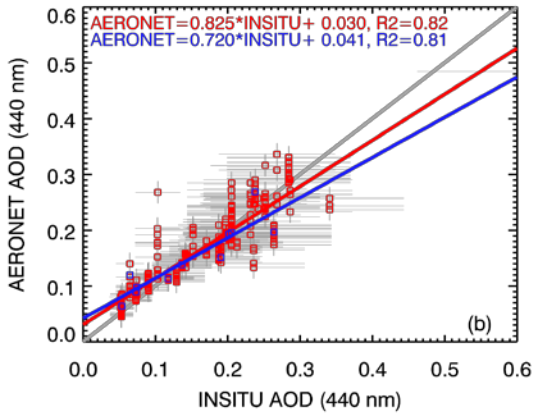
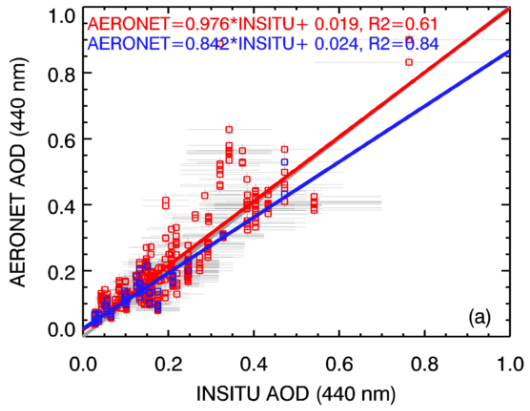
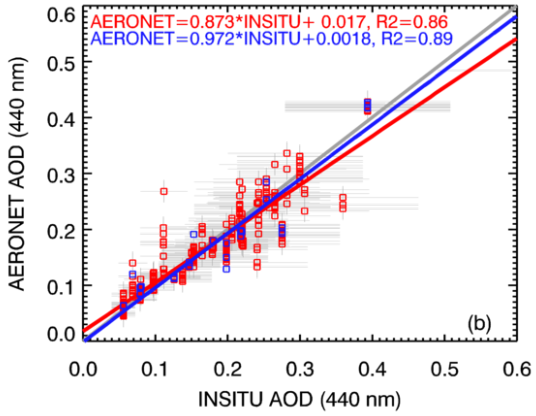


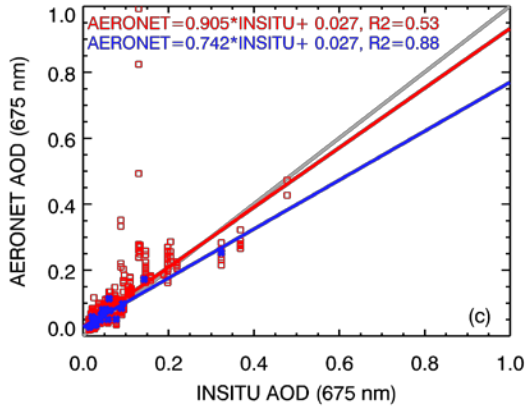
Figure 2. Correlograms for BND and SGP; wavelength = 550 nm, D_p<10 μm, based on hourly averaged surface in-situ data between 1995-2013 (BND) and 1996-2013 (SGP). The value r(k) on the y-axis represents the autocorrelation at lag time 'k'.

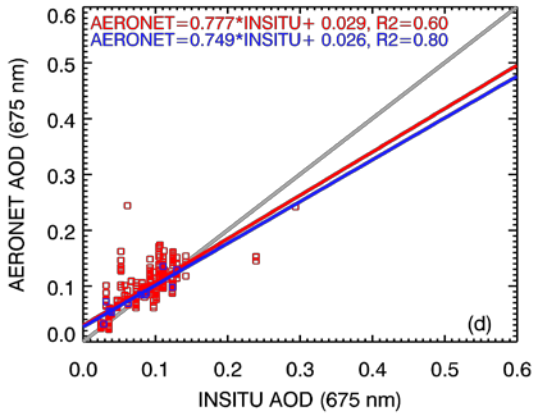

AERONET=0.828*INSITU+ 0.049, R2=0.61
AERONET=0.682*INSITU+ 0.053, R2=0.91
AERONET=0.815*INSITU+ 0.039, R2=0.81
BB
(a)
AERONET AOD (440 nm)
INSITU AOD (440 nm)

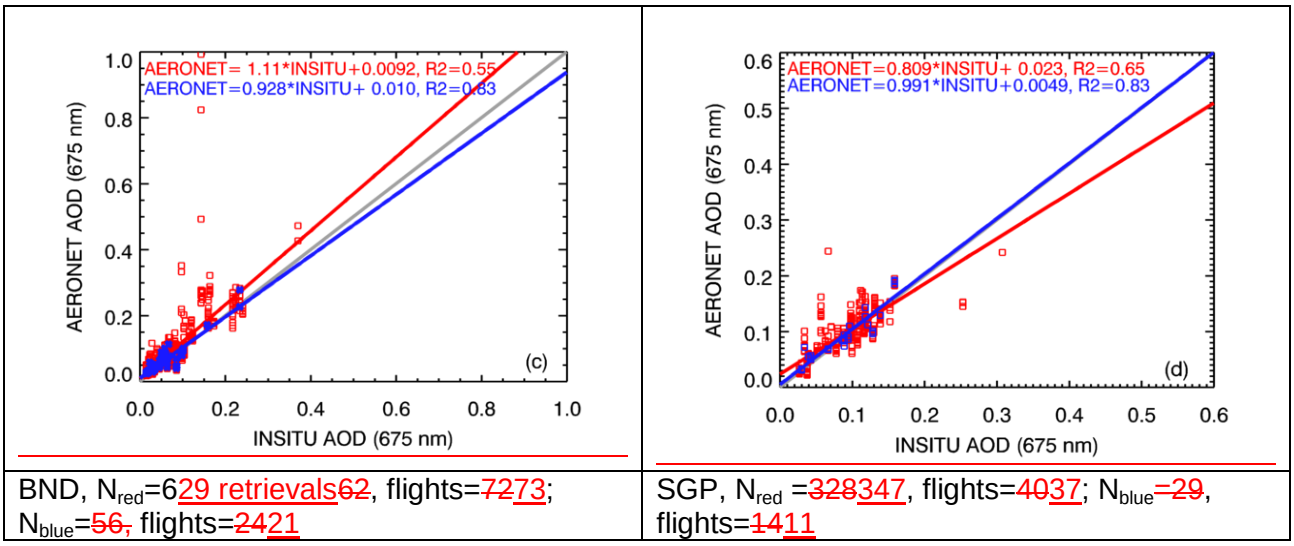

AERONET=0.825*INSITU+ 0.030, R2=0.82
AERONET=0.720*INSITU+ 0.041, R2=0.81
(b)
AERONET AOD (440 nm)
INSITU AOD (440 nm)


AERONET=0.976*INSITU+ 0.019, R2=0.61
AERONET=0.842*INSITU+ 0.024, R2=0.84
(a)
AERONET AOD (440 nm)
INSITU AOD (440 nm)


AERONET=0.873*INSITU+ 0.017, R2=0.86
AERONET=0.972*INSITU+0.0018, R2=0.89
(b)
AERONET AOD (440 nm)
INSITU AOD (440 nm)

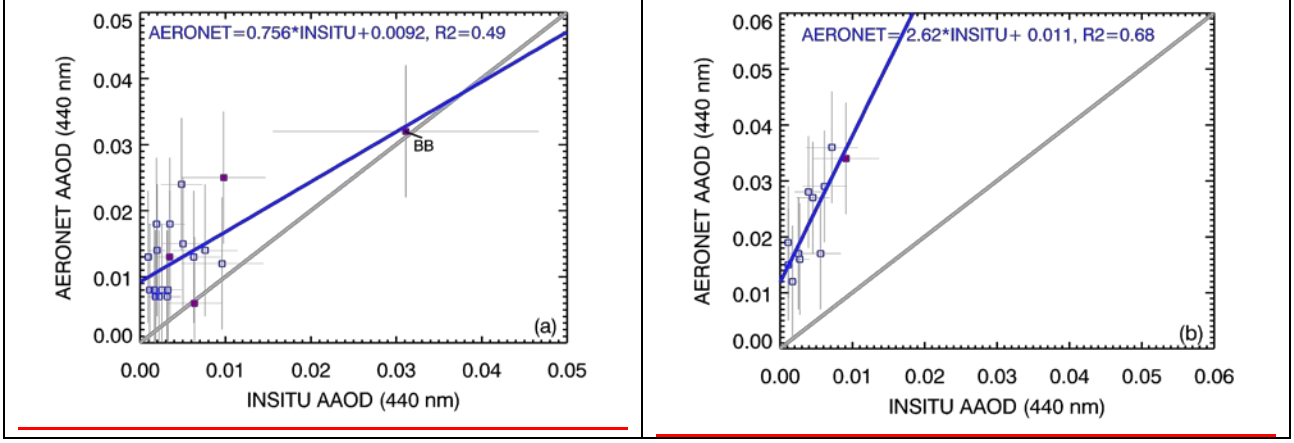

AERONET=0.905*INSITU+ 0.027, R2=0.53
AERONET=0.742*INSITU+ 0.027, R2=0.88
(c)
AERONET AOD (675 nm)
INSITU AOD (675 nm)

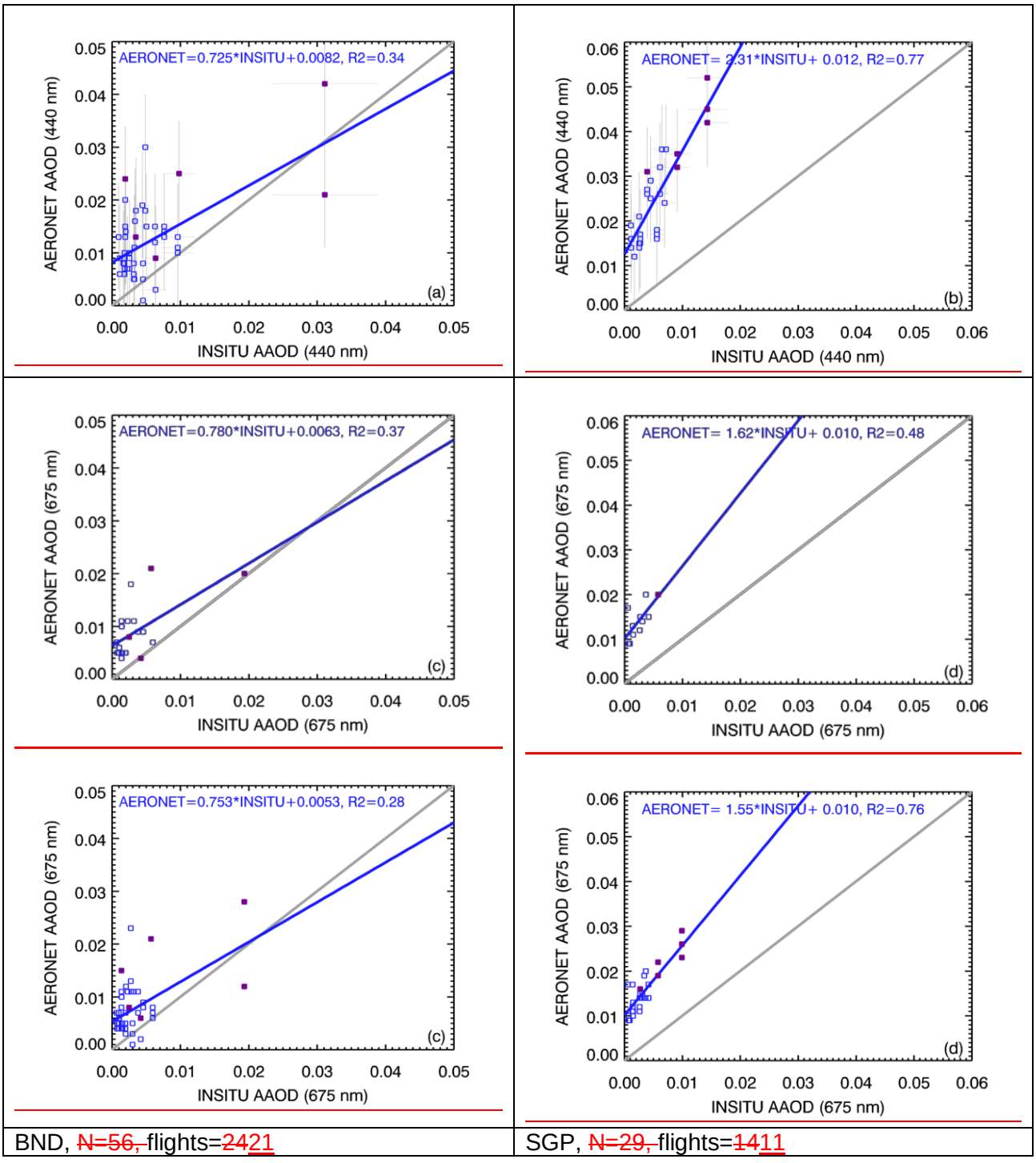

AERONET=0.777*INSITU+ 0.029, R2=0.60
AERONET=0.749*INSITU+ 0.026, R2=0.80
(d)
AERONET AOD (675 nm)
INSITU AOD (675 nm)



BND, N_{red}=629 retrievals~~62~~, flights=~~72~~73; N_{blue}=~~56,~~ flights=~~24~~21	SGP, N_{red} =~~328~~347, flights=~~40~~37; N_{blue}~~=29~~, flights=~~14~~11

Figure 3. AOD comparison (a) BND at 440 nm; (b) SGP at 440 nm; (c) BND at 675 nm; and (d) SGP at 675 nm; thick gray line is 1 to 1 line. Thin gray lines associated with each data point represent measurement uncertainties. Red points and fit line represent all AERONET direct sun Level-2 AOD measurements within +/-3 hours of end of profile. Blue points and fit line represent the average of AERONET Level-2 AOD measurements with successful almucantar retrievals within +/-3 hours of end of profile. The light blue dashed line is the fit if the BB point is excluded. Note: two BND direct sun AOD440 points corresponding to the two highest AOD675 points in the figure below are off the scale of the plot and not shown. The third high AOD440 point is partly obscured by the legend.

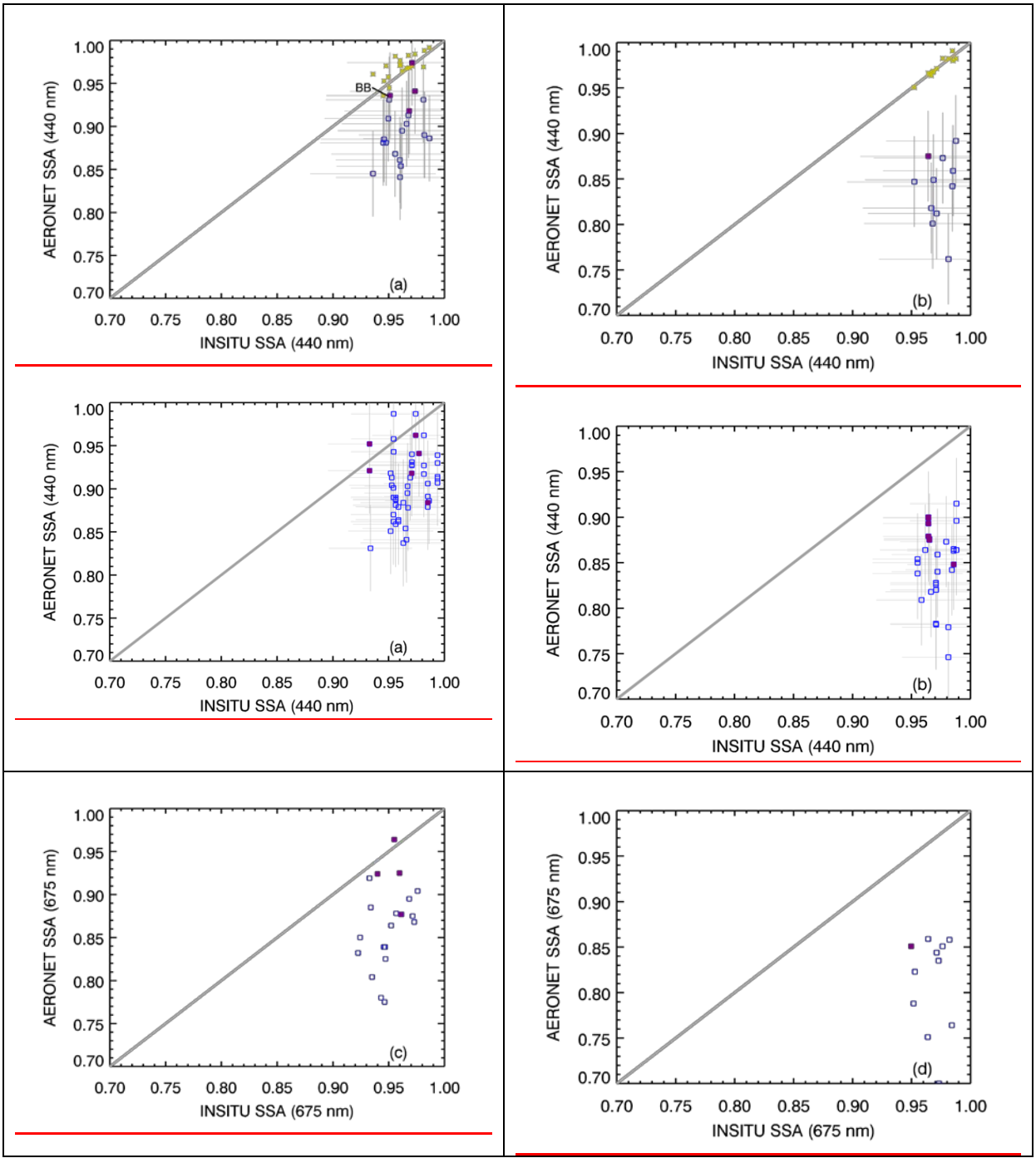


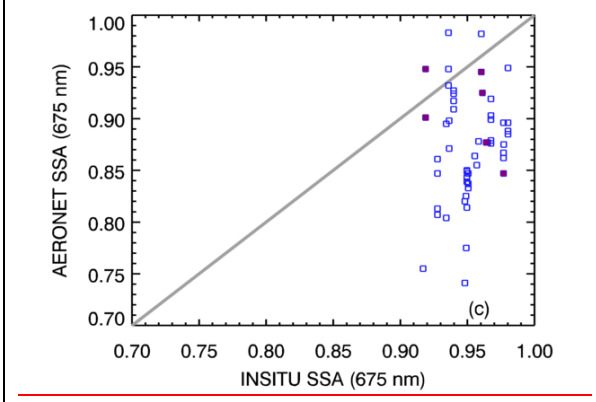


BND, ~~N=56,~~ flights=~~24~~21 | SGP, ~~N=29,~~ flights=~~14~~11

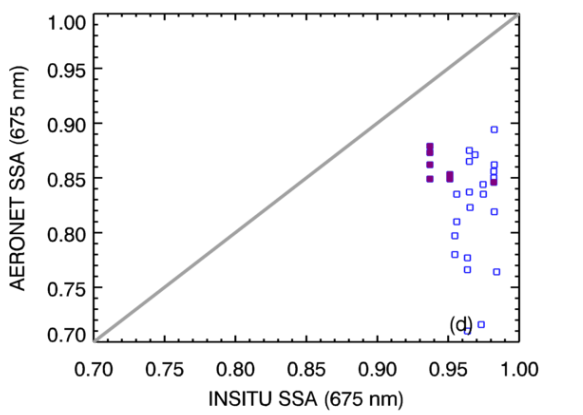
Figure 4. AAOD comparison, (a) BND at 440 nm; (b) SGP at 440 nm ; (c) BND at 675 nm; and (d) SGP at 675 nm. Blue line is linear fit for all points shown; gray line is 1 to 1 line. Thin gray lines associated with each data point represent measurement uncertainties. Points show the average of AERONET Level-1.5 AAOD retrievals for which there was a successful AERONET Level-2 almucantar retrieval within +/-3 hours of end of profile. Purple points indicate the few

comparisons points for which there are AERONET Level-2 almucantar retrievals and where the average AERONET AOD_{440} for those retrievals was great than >0.2.





BND, ~~N=56,~~ flights=~~24~~21



SGP, ~~N=29,~~ flights=~~14~~11

Figure 5. SSA comparison, (a) BND at 440 nm (b) SGP at 440 nm; (c) BND at 675 nm; and (d) SGP at 675 nm. Blue line is linear fit for all points shown; gray line is 1 to 1 line. Thin gray lines associated with each data point represent measurement uncertainties. Blue ~~P~~points show the average of all AERONET Level-1.5 AAOD retrievals for which there was a successful AERONET Level-2 almucantar retrieval within +/-3 hours of end of profile. Purple points indicate the few points for which there are AERONET Level-2 almucantar retrievals and where where the average AERONET AOD_{440} ~~></~~for those retrievals was great than 0.2. The yellow points represent the 'hybrid SSA' which utilizes the AERONET AOD and the in-situ AAOD to derive SSA as described in the text.

.

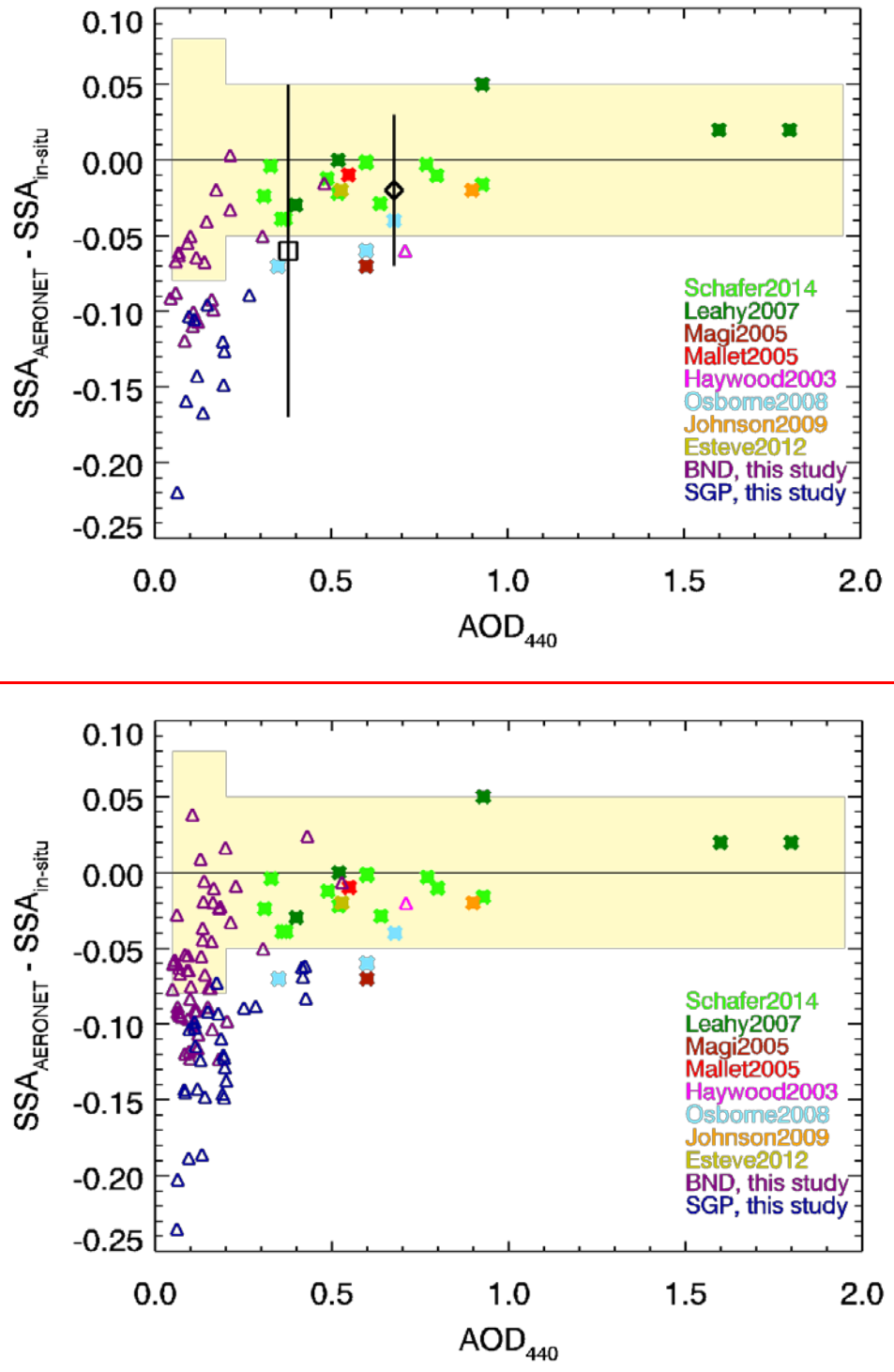


Figure 6. AOD_{440} vs [$SSA_{AERONET}$-SSA_{INSITU}] for direct comparisons studies listed in Table 4. Open symbols are for SSA_{440} difference; filled symbols are for SSA_{550} difference. AOD_{440} values for Leahy2007, Osborne2008, Johnson2009 use the Level-2 values reported on the AERONET webpage for the locations and dates of the specific profile. Shading indicates combined

uncertainty of AERONET SSA values as function of AOD as reported in Table 4 of Dubovik et al. (2000) and uncertainty in the in-situ SSA calculated using equation 2. The black square and black diamond with vertical black lines represent, respectively, the mean and 2*standard deviation for all direct comparisons (including BND and SGP) and for literature direct comparisons only.

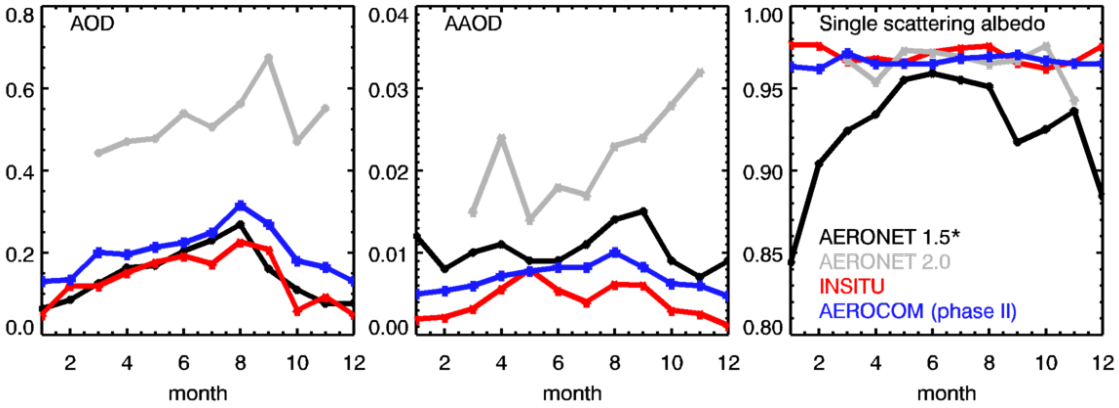


Figure 7a. Monthly medians of BND aerosol optical properties at 440 nm. AERONET medians are for 1996-2011. AERONET AOD medians are for observations with Level-2 almucantar retrievals, with corresponding AAOD and SSA retrievals at Level-1.5 (black) or Level-2 (gray). In-situ data are for June 2006-September 2009. AERONET Level-2.0 almucantar AOD and AAOD values are biased high by definition, because of the AOD_{440}>0.4 constraint. AERONET 2.0 direct sun retrievals (not shown) are similar to the AERONET 1.5 AOD values. In-situ values are derived from 365 flights over BND. AeroCom Phase II median model results cover various time periods (depending on the model) and are reported at 550 nm.

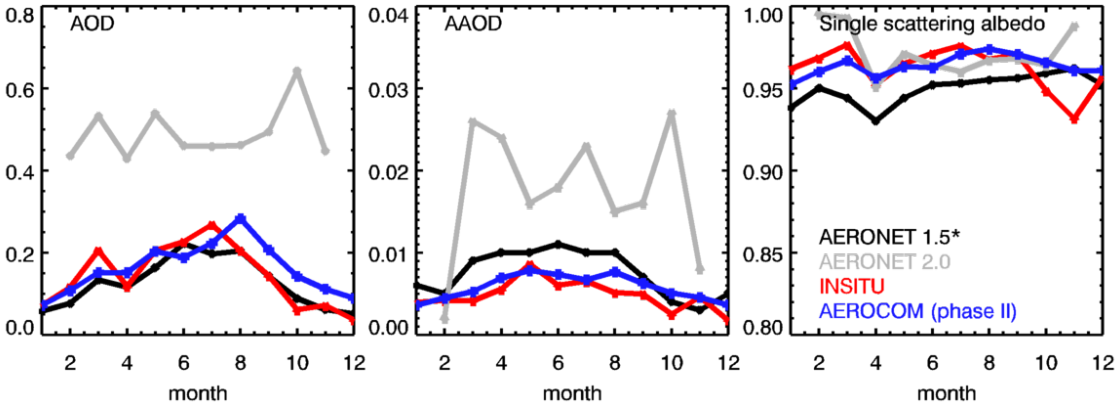


Figure 7b. Monthly medians of SGP aerosol optical properties at 440 nm. AERONET medians are for 1996-2011. In-situ data are for September 2005-December 2007. AERONET AOD medians are for observations with Level-2 almucantar retrievals, with corresponding AAOD and SSA retrievals at Level-1.5 (black) or Level-2 (gray). AERONET Level-2.0 almucantar AOD and AAOD values are biased high by definition, because of the AOD_{440}>0.4 constraint. AERONET 2.0 direct sun retrievals (not shown) are similar to the AERONET 1.5 AOD values. In-situ values are derived from 322 flights over SGP. AeroCom Phase II median model results cover various time periods (depending on the model) and are reported at 550 nm.

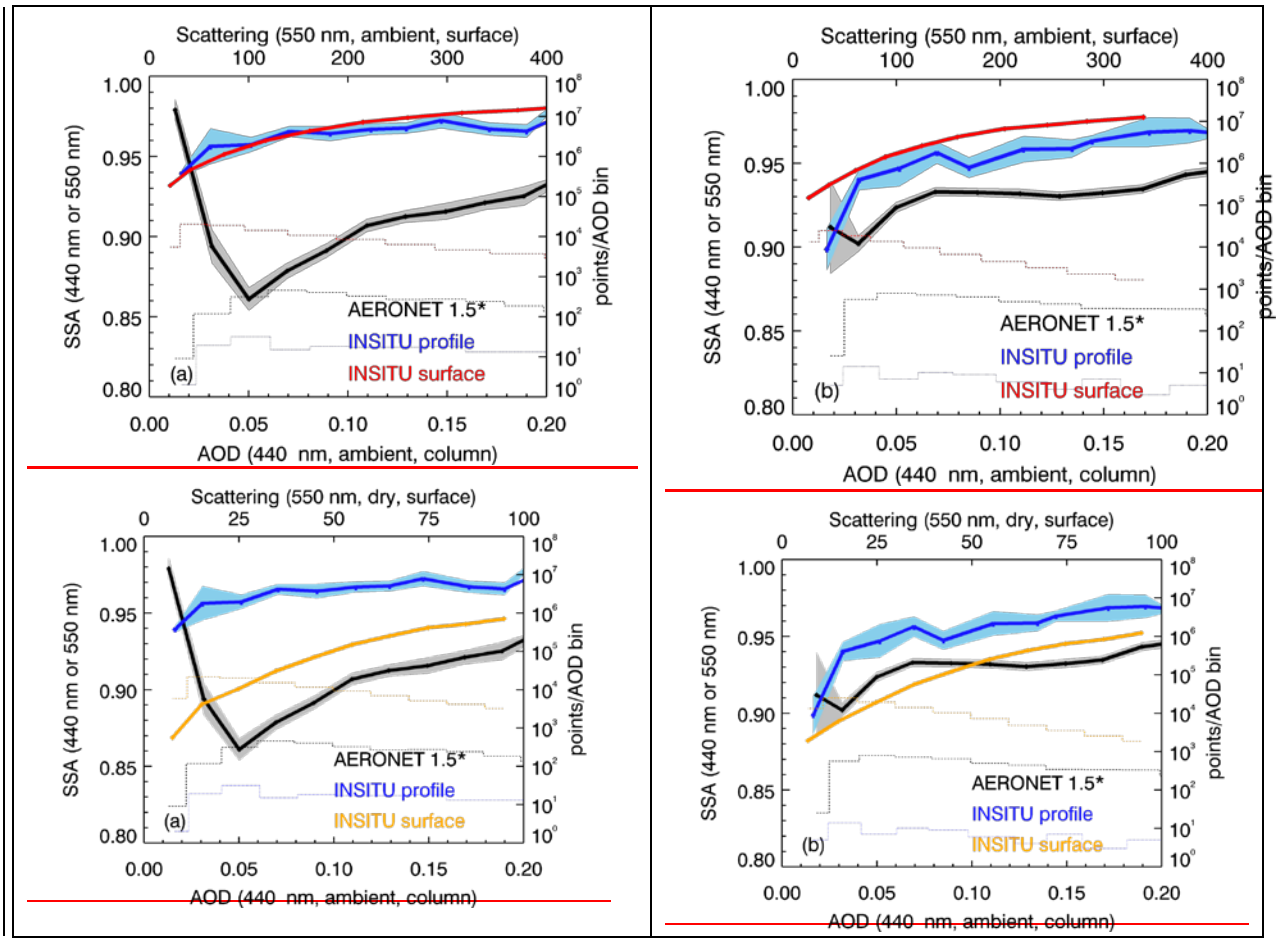


Figure 8. Systematic variability of SSA as a function of loading for (a) BND and (b) SGP for AERONET 1.5* AOD and SSA (black lines), AOD and SSA from in-situ profiles (blue lines) and in-situ scattering and SSA from surface measurements (~~orange~~ red lines). Solid lines indicate mean values of SSA and AOD for each 0.05 AOD bin (10 Mm^{-1} scattering bin). Shaded areas represent mean standard error (mean standard error for surface data is within thickness of red~~orange~~ line). Histograms indicate the number of points in each AOD (or scattering) bin. Plot based on BND and SGP AERONET data (date range: 1996-2012) and BND INSITU profile data (date range: 2006-2012); SGP INSITU profile data (date range: 2006-2007). Surface data (orange lines) are for 550 nm, low RH, hourly in-situ data from the surface sites at BND (date range: 1997~~6~~-2013) and SGP (date range: 1998~~7~~-2013). AERONET 1.5* is from Level-1.5 retrievals with a corresponding Level-2 almucantar retrieval.