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.

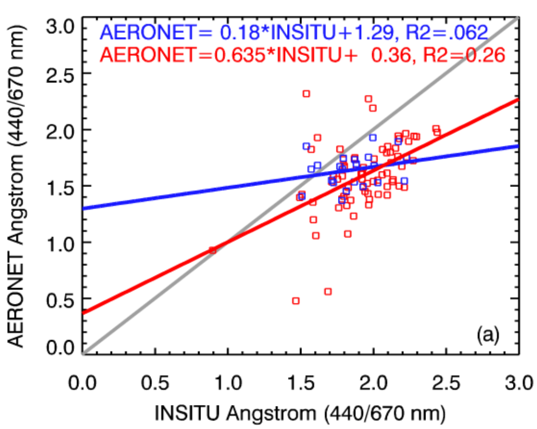
We should note that the AERONET and in-situ AOD are in fair agreement, whereas the AAOD comparisons look much, much different. Suppose the AOD discrepancy were entirely due to particles above 4.2 km agl – what SSA would those particles need in order to eliminate the AAOD discrepancy? We suspect that 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)

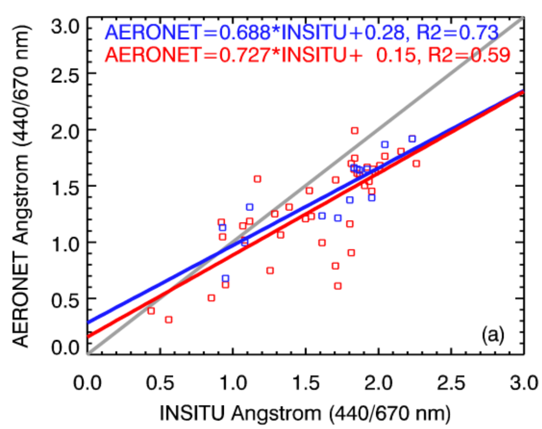
At the Davos, Switzerland site the comparison of the time co-located and matched 500 nm AOD differences between AERONET and GAW-PFR from 2007 through 2010 resulted in a mean AOD difference of -0.0024 and a root-mean square error of 0.0071. These issues should be included in the discussion on AERONET 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 $\lambda = 500$ 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 (CO_2) 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."

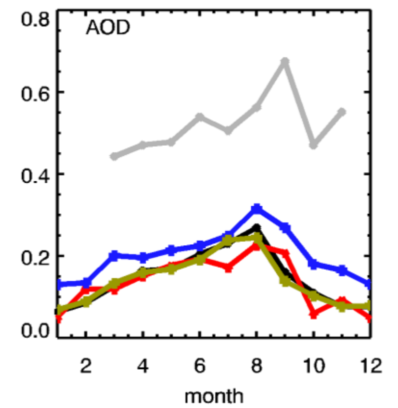
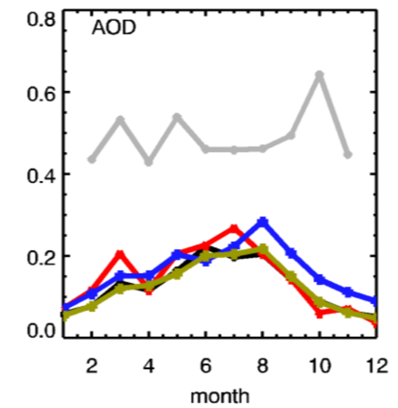
Section 3.3 (line 754-768): Please note that AOD sampled by AERONET in the Level 2 dataset (not just for the subset that have L2 retrievals) includes many more days of data than the in situ flights, and is therefore a much more statistically robust data sample. Please note that the Level 2 AOD climatology 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 $AOD_{440}\sim0.1$."

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.