Our responses to anonymous reviewer's #1 comments are detailed below. Reviewer's comments are in *italics* and our responses in standard font.

General comment:

However I think that a better discussion of the uncertainties on the AERONET data and the analysis itself is required. Many of the uncertainties here are not the authors' fault: there are uncertainties associated with the AERONET inversions, and the small data counts mean that regression relationships presented are limited in what they can say. For example, about 73% (going by Figure 1) of the inversions sampled smoke 0 or 1 days old, and only about 6% sample smoke 4 or more days old. In this sense the prominence of 'ageing effects' in the paper title is perhaps getting our hopes up too much. Indeed, perhaps these limitations are the reasons why a study on this scale has not been performed before. Perhaps it would be better to look at the events contributing to these 6% in more detail, to separate out near-source variability and ageing effects better.

Reply

Thank you for a very valid and constructive review. We have attempted to address all of the issues raised in this, and in following specific comments.
A discussion of uncertainties associated with individual AERONET retrievals has been added to the data section:

"AERONET uncertainty for individual SSA retrievals in these cases is approximately 0.03 (Dubovik et al., 2000). Particle fine mode volume median radius has an uncertainty of 0.01μ m , the spread of the fine mode particle distribution – 0.06. Estimated uncertainty in asymmetry parameter ranges from 0.015 at AOT(440) to 0.04 at AOT(1020) (Dubovik et al., 2000; Sayer et al., 2014)."

To account for uneven sampling along the time axis we have changed the way the results are presented. Instead of using linear regression coefficients, the ageing effects are discussed comparing differences in distributions of plumes grouped into broader age categories. The small number of very old estimates indeed limits what we can say about differences beyond the first 4 days of ageing. Therefore plumes older than 72 hours (more than 3 days) were pooled into one category and compared to smoke aged for less than 1, 2 and 3 days. We believe that such analysis better suits the data and represents ageing effects more realistically avoiding the pitfalls of the linear regression model.

Specific comments:

1.

P6451 L15-16: Strictly, the statement that level 2 inversions contain only retrievals for AOT(440) of 0.4 or higher is wrong. Cases of lower AOT are still included in the data product as provided on the AERONET website, it is just that information related to e.g. SSA is removed because this is thought to be unreliable. The size distributions are still there. One possible way to increase the data volume of aged plumes in the analysis would be to look at the inversions below this AOT=0.4 limit. It would not help the SSA analysis, but could help the analysis about particle size, Angstrom exponent, precipitable water, and AOT changes.

This was a misleading statement indeed. The authors were aware that the level 2 inversions for retrievals at AOT(440) lower than 0.4 contain information on many parameters except SSA and complex refractive index. The AOT(440)=>0.4 data selection threshold was used throughout because analysis of ageing effects on SSA was one of the main interests of the study when conceived. We do agree that extending the analysis with AOT(440)<0.4 cases could be beneficial in reducing the uncertainty in our estimates. However, exploration of these observations indicated that source attribution was much more difficult for the optically thin plumes. We were gaining a small number of identifications at disproportionally increasing processing costs. Importantly, looking at the additional points it was obvious that the scatter was still there and we felt that further processing would not have a significant effect on the main conclusions of the study.

The statement has now been removed. Instead it is stated that "AERONET level 2 inversions contain SSA retrievals only at AOT(440) levels of 0.4 or greater." Section 2.1.1 (Data selection) was updated stating that only retrievals at AOT(440))=>0.4 were used in the analysis.

2.

P6458, L6: Here and elsewhere the authors state that the AERONET SSA uncertainty is 0.02. I am puzzled where this number comes from (no references are provided by the authors here). The 'canonical' number (Table 4 of Dubovik et al., JGR 2000) is 0.03 for biomass burning aerosols, for cases where AOT(440) is 0.5 or higher. It's also not apparent how much of that is systematic, and how much is random, error. I realise that this will probably not affect the interpretation of results since a main conclusion appears to be that the results for most surface types are not significantly different from SSA=0.95, but nonetheless I think this should either be corrected, or a source for the SSA uncertainty being 0.02 rather than 0.03 provided here.

3.

We apologize for this human mistake. The number has been corrected where applicable, providing the reference at the first occurrence.

4.

Figure 6: A lot of the discussion on ageing effects in the text relates to this figure. Confidence intervals (CI) are drawn based on linear regression fits to the data. However it is clear that the ageing effects are small compared to the scatter in the data. So these relationships are explaining only a fraction of the variability. In this sense I think discussions of the CI could mislead a casual reader – these are the CI on the gradient, but could be mistaken to imply something about the spread of the underlying data as well. For this reason I think it would be useful to also provide the coefficient of determination for these plots (and in the text).

Reply

As stated above, we have since abandoned the linear regression analysis, hence this does not apply any more.

5.

Related to the CI, how exactly where these calculated? This is related to the assumptions about the uncertainty on the y-axis data going in. Were these taken to be the AERONET uncertainties, or weighted equally, or what? Were data points assumed to be uncorrelated? Were uncertainties on regression coefficients scaled by the reduced chi-square value (which is the equivalent to assuming that the regression model is 'correct') or not? This information should be provided in the paper.

Reply

Confidence intervals for any quantity discussed (and previously for regression coefficients) were derived using bias-corrected non-parametric bootstrap method (Efron, 1993). During the resampling y-axis error was modeled as normally distributed AERONET uncertainty given as 1 standard deviation. We have now clarified this in the methods section.

6.

I would also point out that statistical significance does not necessarily correspond to scientific significance (for example, if an effect can be discerned, but it is so small as to be negligible for practical purposes). Equally a number may be large but have a huge error bar, in which case all you can say is that you don't know what's going on.

Reply

We have reviewed the manuscript and removed the relationships (AAE, precipitable water), which are not only weak, but difficult to explain as well. The remaining ones (fine mode median radius, AE, asymmetry parameter, SSA) are interesting 1) either way if differences comparing young and well aged plumes are significant or not, and 2) agree with existing estimates. We do not present changes in ageing plumes for all land cover types, as indeed the uncertainties were large because of small number of plumes attributed to most of the classes. We have integrated AERONET uncertainties in our confidence interval calculations, and provide them for all of the estimates.

7.

I'm also curious about the two-sided p threshold of 0.01 being used to denote significance when CI are presented as 95% bounds (which I would think corresponds to p values of 0.025, or am I wrong)? Why the inconsistency here?

Reply

The mismatch has been corrected. 95% confidence intervals are given throughout.

8.

As hinted earlier, one way to cut through some of this scatter would be to look at some case studies in more detail. Figure 1 says there are 6 cases with an estimated smoke age of 6 days. I am not completely clear if this means that there are 6 cases where the same plume was observed more than once during these 6 days, or just that there are 6 cases where a plume that was 6 days old was observed. If the former (or if some of these events were sampled more than once), analogous plots to Figure 6 could be created for each of these case studies (rather than just looking at all points at once). Then we could see whether the noise in these plots is still present or not. That will say something about whether ageing effects are for example stronger or weaker than the AERONET inversion uncertainties (which is not really discussed for e.g. size distribution)

Reply

We perhaps didn't make it clear enough, but we initially did not include discussion of lagrangian pairs -- observations of the same plume at two or more AERONET stations. However, there were a few cases of identified paired observations of the same smoke, the discussion of which has been now added:

"The trajectory analysis indicates that in several cases the same plume was transported over more than one AERONET station allowing to infer changes in plume properties between the two observations. Unfortunately, only 13 of such events were identified preventing a more robust analysis (Fig. 6a). In 10 out of the 13 cases older particles are larger, while hree of the pairs suggest a decrease in R fv between the observations. The median R fv change rate is 0.0075 (-0.001–0.03) μ m per day. The estimate agrees well with the growth rate suggested by differences in particle distributions between young and aged plumes. However, it has large confidence intervals due to low number of paired observations and uncertainty in individual AERONET retrievals."

We did not present these cases initially, and do not want to give it too much weight now, because of their small number and several problems associated with such identifications. For very large plumes it is somewhat easier to determine source and approximate age than it is to identify paired AERONET retrievals of the same air parcel. The trajectory uncertainty analysis indicates that the trajectory window can grow to hundreds or thousands of kilometres in a matter of hours or days, and thus it is not possible to say that the two observations along the trajectory are of the "same" smoke with any degree of certainty. Allowing bigger trajectory windows yields more identifications, increasing false positives as well. We avoided having to inspect individual cases and selectively pick the good looking ones. As a result there were only a handful of such identifications. Importantly, AERONET uncertainty on individual retrievals is more important comparing a pair of observations (or a few pairs), and results in huge confidence intervals on the estimate. For the 13 paired observations that we had, the median slope in volume median radius is reasonable and agrees well with inferred particle growth. However, when AERONET uncertainty is factored in, the estimate is not significant. Comparing larger groups of observations helps to deal with this, and we are able to say with reasonable confidence that on average, older plumes have larger particles. We believe that such interpretation of the data is more appropriate, and have revised the manuscript accordingly.

9.

I also worry from Figure 6 that the data are not suitable for linear regression analyses (due to non-uniformity of sampling along the age axis, and non-Gaussian behaviour of departures from the best-fit line); reducing to a subset of data corresponding to these few case studies would help with this issue.

Reply

This is obsolete as the linear regression analysis has been abandoned.

10.

Figure 7: For panels (e-g), it would be good to show the 'all classes combined' data in a heavier font/line, as at the moment it is hard to see among all the other lines. I also think that the data from this plot should be placed in a table, in addition to the Figure.

Reply

The figure 7 has been replaced by another as we no longer present change rates derived from linear regression analysis. The new figure should be clearer as the number of parameters and classes we present has been reduced.

11.

General throughout: ‘angstrom’ is not typeset correctly. In LaTeX, I think this should be \AA{}ngstr\'{o}m.

Reply

Thank you for picking this up. The word has been changed throughout.

12.

Conclusion: Here (and in the introduction) the authors mention the need for better aerosol optical models for satellite retrieval algorithms. This is undoubtedly true. However as a practical matter, it is not clear how this new information could be used in (for example) a satellite AOT retrieval algorithm, because the instantaneous satellite snapshot will not ‘know’ how old the smoke is or where it came from, and it’s not practical (might not even be possible) to use ancillary data to successfully identify smoke age and/or origin on a case-by-case basis for a global algorithm. Even if it were possible, the authors do not present evidence that including these ageing effects would make a significant difference to these retrievals (for example there are no radiative transfer simulations provided in this study), especially because (as mentioned) the ageing effects appear to be smaller than the scatter in the data (Figure 6). Sayer et al (ACP 2014) did some analysis in this regard looking at intra/inter-site variability in smoke aerosol models, and AOT errors resulting from the assumption of the wrong aerosol model, although this was mainly driven by SSA considerations (from virtually nonabsorbing through to strongly absorbing) so those simulations are not directly transferable to the case here (where there is little SSA variability and the main ageing effect, or difference between sources, is a change in fine mode particle size). On the modelling side of things, similarly, do these small ageing changes or differences in properties between different smoke sources result in meaningfully different calculations of e.g. shortwave flux or other relevant quantities? I realise that doing this in detail would probably be out of the scope of the manuscript, but found the authors’ comments here to be a bit careless given the lack of substantiation, and the large error bars on the results of the analysis. The authors say that these new results can help but don’t give a specific or quantitative look at how. I would prefer to see a deeper discussion here, or else remove the statements from the manuscript.

Reply

The statements have been removed.