Anonymous Referee #1

General comments and recommendation

This study looks at trends in aerosol optical thickness (AOT) from MODIS and MISR, as well as trends in aerosol shortwave direct radiative effect (DRE) from CERES data. CALIOP data are also used. This is in part an update of earlier work by some of the authors, updated using newer versions of the MODIS data, and in part a new analysis. The study is within scope of ACP and the methodology is fairly standard and reasonable. The topic is of relevance and interest.

I did however find it a bit hard to read. Some sections are quite verbose and hard to pick out the key take-away messages. This is however in part the authors being thorough in comparing this analysis to their previous MODIS analysis, as well as in noting some limitations of one of the CERES data products. So it's hard to give advice on how to remedy this while keeping the analysis thorough (which is an aspect I definitely like). As a result I recommend publication after minor revisions, listed below, mostly to address writing style. There is however also one important statistical error in terms of discontinuous trends in Figure 11 which needs to be addressed to make the manuscript technically correct.

Response: We thank the reviewer for his/her constructive suggestions and comments. We tried to re-organize the paper as suggested and details are shown below. Also, we have also modified Figure 11 as suggested, to include piecewise linear regressions as suggested.

Specific comments:

Title: MISR should be added here. Maybe CALIOP too? Or the authors could remove the specific sensor names and say "various satellite products" or something.

Title: "Longer term variation" is a bit clunky and, to me at least, implies longer than single-sensor records (which isn't what is discussed in this study). I guess the authors chose this wording to make a contrast with their previous studies, which were decadal? Perhaps "21st century variations" would be better, since the data start in 2000 or later?

Response: These are nice suggestions. We have revised the title to:
A Study of 15-Year Aerosol Optical Thickness and Direct Shortwave Aerosol Radiative Effect Trends Using MODIS, MISR, CALIOP and CERES

Lines 103-104: a reference for MISR should be added here. I'm not sure what the best one is. Perhaps Kahn et al (JGR, 2010), which I think is the main validation study for this version of the data?

Response: Kahn et al., 2010 has been added to the reference list.

Line 109: As a minor point, the MODIS product doesn't do "spectral AOT retrievals". It retrieves AOT at 550 nm and the weighting between fine and coarse aerosol modes, for various mode combinations. Spectral AOT is derived from these parameters. I suggest something like "provides spectral AOT at seven wave lengths" or even just removing the bit about wavelengths, since only 550 nm (the main data product) is used in this study anyway.

Response: Done. We have changed the sentence to "provides spectral AOT at seven wave lengths" as suggested.

Line 110: "increased resolution" isn't quite right here, since the data are coarser at the edge of the swath. I think the authors either mean "increased pixel size" or "decreased resolution".

Response: Thanks for the suggestion. We have changed to "increase pixel size" as suggested.

Line 160: This line says only data with CP > 95% are used, while line 182 says CP > 99% are used. Is this inconsistent or am I misunderstanding something here? If these are for two different parts of the analysis, why the different thresholds?

Response: The first threshold (CP > 95%) is used for the initial collocation step. This would allow us to perform a sensitivity study to evaluate the impact of cloud fraction on the analysis as shown in Table 5. Only collocated pairs with CP > 99% are used in the final analysis. We have revised the sentence as follows to avoid confusion.
"Note that only CERES pixels that have a MODIS reported cloud fraction of 1% or less are used in the final process. A more relaxed CP threshold of 95% is adopted here, partially for studying the impact of cloud contamination on CERES derived SWAREs as shown in Table 5"

Lines 167-169: I'm not sure why the first part of this sentence is needed. I think it's fine just to say the arithmetic mean MODIS AOT is used.

Response: We have made the change as suggested. Thanks.

Line 186: There have been a large number of studies into cirrus contamination of MODIS AOT data, not just Toth et al (2013), and many were well before that paper. I suggest rewording this to make it clearer that was not the first study, and maybe cite some of those other ones too.

Response: Thanks for your suggestion. We have revised the sentence as "Several studies have suggested that MODIS AOT retrievals may be contaminated with optically thin cirrus clouds (OTC, e.g. Kaufman et al., 2005, Huang et al., 2011, Feng et al., 2011, Toth et al., 2013)."

We have added the papers to the reference list.
Kaufman, Y. J., Remer, L.A., Tanre, D., Li, R.-R., Kleidman, R., Mattoo, S., Levy, R., Eck, T., Holben, B.N., Ichoku, C., Martins, V., and Koren, I.: A critical examination of the

residual cloud contamination and diurnal sampling effects on MODIS estimates of aerosol over ocean, IEEE Trans. Geosci. Remote Sens., 43, 2886–2897, 2005.

Huang, J., Hsu, N.C., Tsay, S.C., Jeong, M.-Y., Holben, B.N., Berkoff, T.A., and Ellsworth, J.W.: Susceptibility of aerosol optical thickness retrievals to thin cirrus contamination during the BASE-ASIA campaign, J. Geophys. Res. 116, D08214, doi:10.1029/2010JD014910, 2011.

Feng, Q., Hsu, N.C., Yang, P., and Tsay, S.-C.: Effect of thin cirrus cloud on dust optical depth retrievals from MODIS observations, IEEE Tran. Geosci, Remote Sens., 49, No.8, 2011.

Lines 200-207: This paragraph doesn't really fit in this Section, which is otherwise describing the data sets used. I think it should be broken out into a new section summarising how trends are calculated and assessed (i.e. construction of time series of monthly deseasonalized AOT anomalies). It would be useful to add a bit of brief information about these two significance methods here as well. For example the Weatherhead approach attempts to account for autocorrelation, which is important in some areas for monthly AOT time series.

Response: As suggested, we have moved this paragraph to a later section and added additional discussions.

Section 3.1: I think I understand what was done here but from the discussion and tables it isn't always clear what results apply to what bit. My understanding is the authors (1) compare C5 trends to C6 trends (for 2000-2009) and (2) compare C5 trends to the Zhang and Reid (2010) trends, which used a 'data assimilation (DA) grade' version of the MODIS products. So in this way they assess whether differences are more because of the C5/C6 change or the fact that Zhang and Reid (2010) used the DA-grade product and there isn't a C6 equivalent (that I know of) DA-grade product. To help with this I suggest restructuring this section as follows:

1. Move the bit about how trends are calculated to a new section earlier in the paper (see prior comment about lines 200-207). This will help streamline the text by putting the methodology in a methodology section.

Response: Done. Thanks for the suggestion.

2. Remove the text defining regions from the main body, since regions are already defined in Table 2, where they're easier to read.

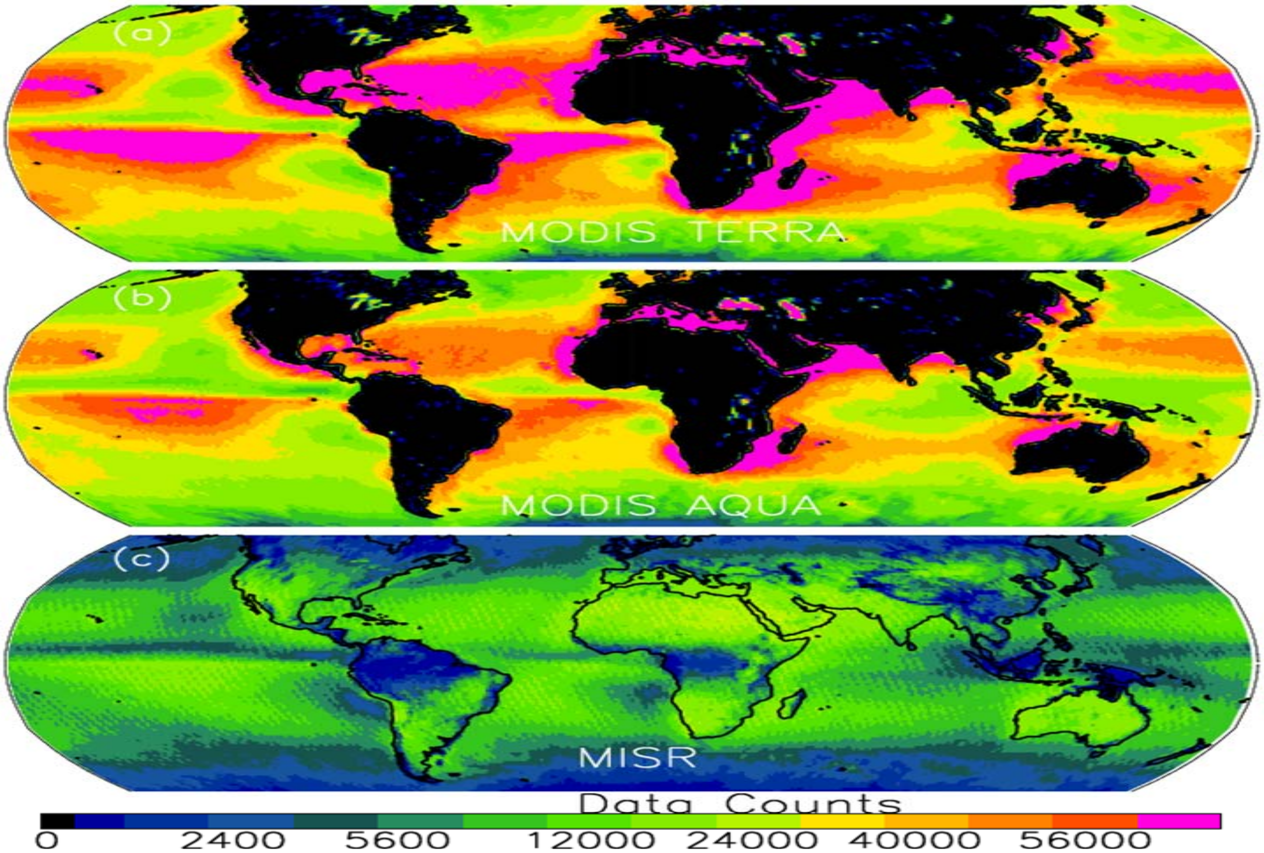
Response: Done.

3. Split out the analysis into two separate subsections, one to compare C5 vs. C6 trends for the 2000-2009 period, the other to compare C5 trends with and without the DA process. (Alternatively, since the conclusion seems to be that the differences are mostly minor, you could *put in a few sentences that you looked at it but didn't find that things had changed much, and then just cut out the rest of the section.)*

Response: Note that DA data are not used in this study and the comparison of C5 MODIS and DA datasets has already been reported by Zhang and Reid, 2010. To avoid confusion, we have revised the following paragraph:
"Here regional and global mean C5 AOTs are derived using similar steps as were used in constructing the C6 AOT data, which are differ from the data-assimilation quality C5 MODIS DT data as used in Zhang and Reid (2010). Still, as suggested from Zhang and Reid (2010), although QA steps could lower the mean global over ocean AOTs from ~0.15 to ~0.11, in part due to the removal of cloud contaminated retrievals, minor impacts on the AOT trend analysis are reported."

Line 273: I guess the authors use 1000 data counts because the MODIS level 3 aerosol products don't provide a count of number of days per month, despite various requests over the years. It would be good to indicate briefly the main areas where this removes data, and what the typical variations of data volume are in other grid cells (e.g. are the results sensitive to the threshold choice, or do most grid cells have many times more than 1000 retrievals?). From Figure 2 it appears that for MODIS it doesn't remove (m)any ocean grid cells in the studied latitude range. For MISR the gaps are roughly where I'd expect from e.g. cloud patterns in the tropics.

Response: The data count is rather an arbitrary number used to remove some over land water retrievals over scenes such as lakes. This is also partially used to ensure sufficient data are included in the trend analysis. Attached below is the data count plot associated with Figure 2. This plot is not included in the paper, as the paper is already very long. We have added the following sentence to clarify the concern:
"(this is an arbitrary threshold selected for removing some over land water retrievals over scenes such as lakes. It is also partially used for ensuring sufficient data are included in the trend analysis)"



Line 275: Remer et al (2006) was before the MODIS Collection 5 release was complete, and you are using Collection 6 data. I don't know of a similar study to Remer et al (2006) using Collection 6 data, so it's probably still fine to cite that study here, but may be worth noting that was for an older data product version.

Response: Done. We have added the following discussion as suggested: "Remer et al. (2006) using 3 years of C5 MODIS data"

Lines 282-285: are these area weighted or simple mean? This should be stated. 1 degree grid cells at high latitudes are a lot smaller in real terms than those at the Equator. It may not affect the offset and trends shown in the figure too much, but may affect the baseline global-average AOT, since AOT tends to be higher in the Equatorial belt due to continental outflow.

Response: Area weighting is not applied and arithmetic averages are applied to compute means. We have added "simple arithmetic mean" into the text as suggested.

Lines 300-316: This is an interesting and I think pretty reasonable way of addressing/correcting for potential calibration drift, so that's good that the authors have done so. The basic idea is that *if there's a trend in a region that's expected to be stable, one can subtract that trend from apparent trends elsewhere. However a caveat here is that assumes that the calibration degradation propagates linearly into AOT. That is probably fine for areas with AOT close to that of the remote region used as a baseline. But for example a 3% change in reflectance may cause a certain change in AOT when the true AOT=0.1 as compared to at e.g. AOT=0.5, since the radiative transfer isn't linear in AOT. The correction might therefore be an under/over-correction in those higher-AOT areas. Again, there's probably no simple better way of approaching this so the method is reasonable to use here. But since many readers of the article might not be familiar with the underlying radiative transfer and retrieval algorithms, I think this caveat should be mentioned.*

Response: Thanks for the suggestion. We have added the following paragraph as suggested: "A caveat here is that we assume that the calibration degradation propagates linearly into AOT. The correction might therefore be an under/over-correction in those higher-AOT areas."

Lines 339-340: the authors state that "the rates of increase of aerosol loading have slowed down over the last five years" because trend estimates over the period 2000-2015 are less positive than those for 2000-2009. That is certainly one possibility, but the statement is unsupported by the evidence. The trends for both periods may be statistically distinct from zero, but are they statistically different from each other? That is the relevant factor here. Only if so can one say that that the trend has slowed. The reader can't tell if this is the case, since uncertainty estimates for the trends are not shown. I suggest the authors look into this and either add text supporting it (if the trends are statistically distinguishable from each other) or remove this text (if they're not).

Response: We are unclear about the comments. For trends estimated for the periods of 2000-2009 and 2000-2015, data that are used for estimating trends are the same for the first ten years (2000-2009). Thus the trend changes by adding 5 more years of data are likely linked to new data added to the analysis. The flattening of the last 5 years of trends for the Bay of Bengal and Arabian Sea can also be seen from Figure 4. We revised the sentence as below to avoid confusion:
"However, the rates of increase of aerosol loading have plausibly slowed down over the last five years for both regions, indicated by ~20-30% reductions in AOT trends when estimated using the near full Terra data records. Flattening of AOT trends with respect to time can also be observed in Fig. 4 for both regions for 2010-2015.

Sections 3.2, 4: As a general comment related to the above, it would be good if the estimates of trend precision could be given in the text when specific numbers are mentioned. For example on line 428 the SWARE trends in a region are given as 39.8 and 43.7 W/m2/AOT for Aqua and Terra. Without uncertainty estimates on those numbers, we don't know if the 4 W/m2/AOT difference between the two sensors is significant or within the uncertainty of the data sets used. This is just one example, the comment extends throughout the paper. It doesn't necessarily need to be given for every statistic in the paper but when it is a key result or comparison between two quantities, it makes sense to consider the uncertainty estimates. I realise that often both WH and MK methods are used to estimate significance in this study; it probably doesn't matter too much *which method is used when you're quoting these uncertainties for the above points (as I'm guessing they will be similar).*

Response: Trend uncertainty analysis has been included in section 4.2. Trend significances are also discussed with the use of the WH and MK methods. The numbers referred from this comment are aerosol SW direct forcing efficiencies, not trends. To estimate the uncertainties in aerosol SW direct forcing efficiencies, in situ observations may be needed, and the evaluation process can be a paper of its own. Thus, we leave this topic for a future paper.

Section 5: This section says it compares results to other trend studies, but really it only compares results to other trend studies published by the same authors. There are a number of other regional/global trend analyses using satellite aerosol data which could be considered. For example various Mischenko group papers for AVHRR over ocean, Thomas (ACP 2010) for ATSR over ocean, Hsu (ACP 2012) for SeaWiFS land and ocean, Yoon (ACP 2011) for SeaWiFS regionally, Yoon (AMT 2012) for AERONET, Dey and Girolamo (JGR 2011) for MISR in India, Babu (JGR 2013) for Indian surface observations. It would be good to include some of these more independent studies in the discussion here. The point is there's a lot of work which has been done and is relevant to the discussion here but isn't acknowledged. Maybe there isn't space to include anything but the authors only self-citing here is a bit of a let down.

Response: In fact, we have tried to compare region trends from other studies but only realized that each study has its own domain defined differently, making the inter-comparison less intuitive, as sampling differences also need to be considered. Thus, we only selected studies that report trends with similar geographic domains.

Figures 6, 7: I couldn't find a mention of how the black lines in panels a, c here were calculated. This should be added. Also, it seems like results like this are the basis for quoting an aerosol forcing efficiency in units of W/m2/AOT. From the shape of these curves it looks a bit more like a logarithmic fit with a kink around AOT=0.15. I know people like to think in units of W/m2/AOT but perhaps this paper is a good place to point out that the relationships aren't really that linear. This is something which could be highlighted again in the Conclusions (either in list items 5, 6, or a new item).

Response: The black lines are global means. We have added a line in the figure caption to clarify this:
"Color lines are for selected regions and the black thick line is for global oceans."

Also, based on the figure, aerosol forcing efficiency is a non-linear function of AOT. We have added discussions as suggested in the text:

"Figure 6b shows the Aqua MODIS AOT and Aqua SW_{ssf} relationship (non-linear) for 5 selected regions"
"It also worth noting that a non-linear relationship is found between SWARE and AOT."

Figure 11 and associated text: This bit needs further work. It is fine to show trends split by periods, but the discontinuity at the breakpoint is not physical; it implies a sudden jump in the system. Having a breakpoint discontinuity is a sign that the derived values are not robust. There are methods to identify breakpoints in trends, and fit a piecewise continuous trend, rather than an unphysical broken trend. (I think the Weatherhead paper mentioned may discuss this? If not then some of her other work.) The authors should repeat this part of the analysis using a continuous piecewise fit. It's quite possible that this may affect the conclusions. Even if you get a similar answer, it will be on firmer theoretical ground, so it is necessary to do otherwise the manuscript contains methodological errors.

Response: We have implemented a method as mentioned in Tomé and Miranda (2004) to detect the breakpoints and details of the approaches have been included in Zhang et al. (2017).

We have added the following discussions in the text as well:
"Here a piecewise linear fit method from Tomé and Miranda (2004) is applied to detect turning points in trends, similar to what is suggested by Zhang et al. (2017). Also, similar to Zhang et al. (2017), we assume a minimum of 36 months between any two detected turning points."