Changes in atmospheric aerosol loading retrieved from space based measurements during the past decade

J. Yoon\textsuperscript{1*}, J. P. Burrows\textsuperscript{1}, M. Vountas\textsuperscript{1}, W. von Hoyningen-Huene\textsuperscript{1}, D. Y. Chang\textsuperscript{2}, A. Richter\textsuperscript{1}, and A. Hilboll\textsuperscript{1}

\textsuperscript{1}[1]{Institute of Environmental Physics, University of Bremen, Bremen, Germany}

\textsuperscript{2}[2]{Atmospheric Chemistry Department, Max-Planck-Institute for Chemistry, Mainz, Germany}

\textsuperscript{*}[*]{now at: Max-Planck-Institute for Chemistry, Mainz, Germany }

Correspondence to: J. P. Burrows (burrows@iup.physik.uni-bremen.de)

Dear Anonymous Referee #2,

We thank you for the constructive comments, which replies are listed on the supplement.

The study by Yoon et al. is interesting because it documents changes in aerosol optical thickness (AOT) throughout the world on a timescale of 5 to 10 years as measured from space, and validated with independent ground-based measurements. To their credit, the authors try to take into account the temporal representativeness of various satellite datasets. However, they miss the opportunity to provide useful, quantitative statements on the effects of temporal sampling on AOT trend estimates, and for this and other reasons (below) the paper is not as useful as I would have hoped.

Furthermore, the paper does not fulfil the ambition to attribute the observed changes in AOT to their underlying causes. Meaningful statements about attribution of the AOT changes cannot be expected from the study by Yoon et al. because the relation between particulate matter emissions, aerosol formation from secondary sources,
and the ultimate aerosol loading is complex, and the relation with AOT even more complex. The paper simply does not address these relationships, but only provides some rough sketches on what may have caused the changes in AOT, mainly calling on papers in the literature. To my opinion, the problem with the study by Yoon et al. is that the authors use too strong language that suggests that actual attribution has been achieved by the authors, whereas the reality is that they only cite a suite of peer-reviewed papers to guide interpretation of their results. That clearly falls short of ‘attribution’. For instance I find the abstract suggesting that AOT trends have been successfully attributed (“increase … over East China is observed and attributed to both the increase in industrial output and the Asian desert dust”) claiming too much. Related to this, the complete discussion of Fig. 8 (section 6) is long-winded and merely anecdotal:

-> As you pointed out, this study rather fails to establish what leads to the changes in atmospheric aerosol because the main goal of this study is not to provide the evidence to explain the change, but to estimate more convincing trends of satellite-retrieved AOTs by minimizing the uncertainty effect of the unrepresentative sampling. Since there has been no study about the uncertainty effect of the unrepresentative sampling caused by limited/different sampling time and cloud disturbance, it is worth to discuss how significant it is in this study and to test a way to minimize it in trend estimate of the cloud-free AOT. We agree that we used too strong words in the manuscript to cite some attributions from the references. Actually, to identify directly the major cause leading to the trend, further study based on modelling and in-situ observation is needed. Therefore, to focus on the main goals of this study, we have modified the manuscript as you suggested. Furthermore, we have improved it by performing additional uncertainty test to show clearly how significantly the different/limited sampling influences the trend estimate.

1. In Europe the authors report a reduction in AOT and associate that with reducing emissions: “AOT from industry and traffic sources decreases significantly (Marmer et al., 2007; Karnieli et al., 2009). This is attributed to the success of environmental regulation in the EU countries (Streets et al., 2003; Yoon et al., 2011, 2012; Hilboll et
But what is the relative reduction in European AOT (%/yr) and how does this compare to reported reductions in direct aerosol emissions (%/yr) over Europe? And what is the role of secondary aerosol formation from precursors such as sulphur and nitrogen oxides? What about changes in the composition of the aerosol mixture over time? Such changes may also have led to changes in AOT because of the differences in effective optical properties, apart from loading! Neither of these important aspects is being discussed in sufficient detail to warrant the statement that “This is attributed to the success of environmental regulation”. The authors should either come up with a thorough investigation of the full emissions-to-AOT chain, or simply refrain from claims about attribution, and just report on the observed changes.

2. In China, it is unclear to what extent the increase of desert dust with time can explain the trend in overall AOT. For Japan (P26015) reductions in precursor emissions have been reported from the Bremen-group, but we see here an increase in AOT over Japan. The authors should investigate and demonstrate what has driven the increase, or refrain from vague attribution statements drawing on literature.

3. For the eastern USA, only papers are cited that handle changes in AOT, yet the claim is that the reported reductions in AOT are due to reduced emissions (P26016, lines 3-7). The authors provide no evidence that aerosol (precursor) emissions have actually decreased in the eastern USA. The authors should include Figures or tables to support the claim that the reported reductions are actually due to reduced aerosol emissions, or rule out other possibilities.

-> As you suggested, we have modified the manuscript by just reporting our observed changes. We have compared them with the results and cited the attributions from the relevant publications.

Some other important concerns

4. What is the bias in satellite AOT? The retrievals surely not just suffer from random errors. Does the bias (and random errors) remain constant over time? Different trends near coastlines are particularly suspect (i.e. as in Fig. 3(a)-(b)), especially since aerosol sources in these areas are the same for aerosol over land and sea.
The assumed albedo climatologies in the retrievals may provide information on such biases. Can the authors rule out that the jumps in trends are not due to albedo artefacts?

-> The main error sources in AOT retrieval using satellite observation are badly calibrated sensor, unscreened-out clouds, incorrect surface albedo, and inadequate a priori of aerosol optical properties assumed in the aerosol retrieval. They have been discussed in many previous studies (Kim et al., 2014; Kaufman et al., 1997; Remer et al., 2005, 2008; Levy et al., 2010; Kahn et al., 2005, 2010; von Hoyningen-Huene et al., 2003, 2006, 2011; Yoon et al., 2007, 2011) and are still challenging to be perfectly separated from pure aerosol signal. As you pointed out, it is possible that the error sources in AOT retrieval are still significant in trend estimate. However, by comparing ground-based observation and refining the algorithm, the accuracy has been improved and enough to use in trend estimate (Li et al., 2009; Yu et al., 2009; Zhang and Reid, 2010; Karnieli et al., 2009; Yoon et al., 2011; Hsu et al., 2012; de Meij et al., 2012). Even though a significant discontinuity between land and ocean surface is observed in Figures 3 (a) and 3 (b) in spite of a common aerosol source, it can be attributed to the cloud disturbance reducing retrieval number of cloud-free AOT. It is indirectly identified that the problematic regions (e.g. South Africa and Southeast Asia) are overlapped with the cloudy regions shown in Figures 3 (c) and (d). Therefore, we have attempted to minimize the impact of cloud disturbance by applying the weighted trend model and reported more convincing trend estimates.

5. Why haven’t the authors determined the local trends for the ground-based instruments for all data and for the ground-based coincident with satellite retrievals only (fewer samples)? Such a comparison would immediately put a number on the differences in the trend from differences in temporal sampling!

-> As you suggested, we have performed additional uncertainty test to show clearly how significant the uncertainty effect from limited/different sampling is in trend estimate using resampled and all-available AERONET AOTs. As shown in Figure 2, there is good chance of deriving different trends from the
different/limited samplings over such a large urban agglomeration like Beijing. For a more detailed test at the station Beijing, Figure S1 shows the linear trends derived using monthly AERONET AOTs (550 nm), which are calculated either by the all-available or resampled AERONET data. It shows clearly that the trend estimates from different/limited temporal samplings significantly differ from each other and ideally “actual trend”. Table S1 lists the relative percentage errors to the “actual trend” at the selected AERONET stations. It ranges from -156.3% to +399.2% and is an inevitable bias in trend estimate of the “ideal” polar-orbiting satellite observation to the “actual” trend.

Figure S1. Time series of monthly anomaly of AERONET AOTs (550nm) (i.e. all available data or sampled at 10:30 15 a.m±30 min, 12:20 p.m±30 min, and 01:30 p.m±30 min.) from 2003 to 2007 and corresponding trend estimates at Beijing station.
Table S1. Trend estimates of AERONET AOT (550 nm) from different sampling times (10:30 a.m. ±30 min for Terra, 12:20 p.m. ±30 min for OrbView-2, and 01:30 p.m. ±30 min for Aqua, and all available samplings) and corresponding relative errors.

<table>
<thead>
<tr>
<th>AERONET Stations</th>
<th>Linear Trends of AERONET AOT (550 nm) in Different Sampling Times [yr(^{-1})] and (*Relative Percentage Errors [%])</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Available</td>
</tr>
<tr>
<td>Avignon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00120</td>
</tr>
<tr>
<td>Banizoumbou</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00538</td>
</tr>
<tr>
<td>Beijing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00537</td>
</tr>
<tr>
<td>Dakar</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00834</td>
</tr>
<tr>
<td>GSFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00219</td>
</tr>
<tr>
<td>Ispra</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00496</td>
</tr>
<tr>
<td>Mauna_Loa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00014</td>
</tr>
<tr>
<td>MD_Science_Center</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00225</td>
</tr>
<tr>
<td>Mongu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00002</td>
</tr>
<tr>
<td>Ouagadougou</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.02895</td>
</tr>
<tr>
<td>SEDE_BOKER</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00143</td>
</tr>
<tr>
<td>Sevilleta</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00232</td>
</tr>
<tr>
<td>Shirahama</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.00107</td>
</tr>
<tr>
<td>Skukuza</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00463</td>
</tr>
<tr>
<td>Solar_Village</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.01965</td>
</tr>
</tbody>
</table>
The trend estimates from all available and limited/different temporal samplings at the selected AERONET stations are denoted by $T_{\text{Actual}}$ and $T_{\text{Resampling}}$ respectively, and the relative percentage errors ($RPE$) between them are defined as the following:

$$RPE = \frac{(T_{\text{Resampling}} - T_{\text{Actual}})}{T_{\text{Actual}}} \times 100\%.$$ 

The relative errors over $|\pm 100\%|$ are shown as **bold** type. Please note that the error less than -100% shown as **bold-underline** type indicates the opposite tendency (sign) of resampling to the actual trend.

6. Exclusive attribution of differences between ground-based and satellite retrievals to temporal sampling differences has not been sufficiently justified. The paper suggests a number of times that the AOT trend differences between MODIS Aqua and Terra are due to their differences in overpass time. But this is never actually shown. The authors should find out whether the AOT is structurally different at noon (Aqua) than in the morning (Terra).

> There have been some papers (Smirnov et al., 2002; Kocha et al., 2013; Arola et al., 2013) discussing about AOT diurnal cycle. Especially over a large urban agglomeration, the AOT diurnal cycle is significantly influenced by the emission source of anthropogenic aerosol. Figure S2 shows the diurnal cycle of AERONET AOT at Beijing station. As you see, the significant diurnal cycle can be an important attribution influencing that “ideal” MODIS-Terra and -Aqua AOTs considerably differ over Beijing.

![Figure S2. Diurnal cycles of AERONET AOT (550nm) at Beijing station.](chart.png)
The significant difference between MODIS-Terra and -Aqua AOTs used in this study is also observed in Figure S3. Since there is no difference in retrieval accuracy, cloud filtering method, and spatial resolution between MODIS-Terra and -Aqua, therefore it can be attributed to different sampling.

**Figure S3.** Plots of the time series of MODIS-Terra (MOD) and -Aqua (MYD) AOTs, and MYD AOT normalized to MOD AOT from 2003 to 2008.

7. Why is the positive trend in Fig. 3(a)-(b) for high northern latitudes not explained?

It is highly significant according to Fig. 6(a). Why not put a number on the ocean (section) trend by averaging over a larger area?

-> Yes, a significant trend is observed over high northern latitudinal oceans. However, we did not discuss it in the manuscript because it is difficult to draw a reasonable conclusion over the oceanic regions. To consider the impact of cloud disturbance in this study, we have used a new trend model (i.e. weighted least squares regression). It has been shown that the new model can provide improved results over the region where high variation of cloud fraction is located (see Figure 3 (c)). However, the method is expected to be less robust over regions, where frequent cloud occurrence persists throughout the year (e.g., most of the marine areas and tropical rain/cloud forests in the equatorial zone) and thereby small number of cloud-free retrieval is observed. Therefore, before selecting the regions for regional trend estimate, we have firstly
checked where the significant results are located by using three criteria as follows:

1. To avoid the retrieval uncertainty larger than 50%, the trends with total mean of AOT < 0.1 are removed.

2. To minimize the uncertainty effect of large and persistent cloud all year round, the trends with total means of CF (cloud fraction) > 0.8 and standard deviation ($\sigma_{\text{CF}}$) < 0.06 are discarded.

3. To get more significant result at 95% confidence level, the trends with significance ($|B_g/\sigma_{B_g}|$) < 2 are ignored.

Based on these criteria, we have carefully selected the regions as shown in Figure 1 and the trend estimate over high northern latitudinal oceans has been excluded in this study.

8. The conclusion section states (P26017, line 26) that “The positive impact of legislation in reducing AOT and improving air quality is unambiguously documented.” This is not a conclusion following from the work done in this paper, but rather expresses an opinion based on the literature. Since very little has been done by the authors to make the link between reducing emissions and aerosol loading, I don’t think such a sentence should be presented in the Conclusions section. The same holds for all of the text on page 26018, which reads as a personal opinion on the need to study climate change and develop new space-borne sensors, but really does not follow from the research described in the paper. I think the authors should revise this part thoroughly, cut the general statements, and focus what we have actually learned from their research.

-> As you suggested, we have modified the manuscript.

Specific concerns

P26008, L11-14: the non-perfect temporal correlation between AERONET and satellite measurements and the differences between the retrieved AOTs is fully attributed to the ‘limited sampling times’. I don’t think the authors have made clear
that ground-based AOT retrievals and satellite retrievals have no differences in ‘spatial resolution’. AOT retrieved from the ground depends on the relative azimuth angle, so that also differences in spatial representativeness may have contributed to the AERONET-satellite differences. The authors should explain why they rule out differences in spatial resolution in causing trend differences.

-> Of course, in the real comparison between AERONET and satellite-retrieved AOTs, the correlation can be influenced by not only spatial resolution, but also retrieval accuracy, cloud-filtering method, and so on. However, in this pattern correlation analysis, we have used only AERONET data and there is no difference in retrieval accuracy, cloud-filtering method, and spatial resolution between different samplings (at 10:30 a.m.±30 min, 12:20 p.m.±30 min, and 01:30 p.m.±30 min, and all available samplings). Therefore, we can conclude that the different or contradictory trends can be estimated from only different and limited temporal sampling.

P26009, L1-2: the authors should explain here how cloud screening leads to a reduction in the number of observations. What exact quantitative cloud filter criterion was used to screen the cloudy measurements?

-> In aerosol retrieval algorithm based on visible spectrum, the observed pixel is firstly divided into cloudy or cloud-free pixel (Kim et al., 2014; Kaufman et al., 1997; Remer et al., 2005, 2008; Levy et al., 2010; Kahn et al., 2005, 2010; von Hoyningen-Huene et al., 2003, 2006, 2011; Yoon et al., 2007, 2011, 2012). The AOT is only retrieved for cloud-free pixels, so it is called cloud-free AOT. We would provide more detailed information about the cloud masking by citing relevant papers.

P26009, L13: explain what the ‘climatology’ means here. Is it the climatological monthly mean derived from all instruments, or is it a per-instrument value?

-> It means the climatological monthly varying pattern and it is a per-instrument value.
P26011, L12-13: it is unclear how representative Figure 4 is as an example of the outlier test. For which region and month does the Figure hold? How many measurements are on average retained after the outlier test?

-> It is an example of the outlier test for monthly AERONET data at GSFC from 1993 to 2009. In this case, nine outliers are filtered out and 187 weights remain after outlier test.

P26011, L21: 5000 resampling iterations out of how many (±) samples in the total ensemble?

-> 5000 bootstrap-resample-sets were used to derive 5000 bootstrap-resample trends. We have calculated the standard deviation using the 5000 bootstrap-resample trends for each grid.

P26012, L9-19: a proper comparison of the capabilities of both the AERONET and satellite-based analyses would ensure consistent sampling (in time and space) of AERONET and satellite retrievals. It is unclear if this has been attempted. From the text, it appears as if AERONET daytime mean AOT has been used vs. satellite cloudfree AOT at particular overpass times. The authors should make clear what is exactly understood as an AERONET “actual trend”. If AERONET is sampled in a consistent manner with the satellites, differences in temporal sampling can be largely ruled out, and remaining differences can be quantified and attributed to differences caused by errors in AERONET and satellite retrievals, and to differences in spatial representativity. The differences now shown in Fig. 5 may well be due to these issues, but also due to differences in temporal representativity.

-> The AERONET AOT trend, so called “actual trend” in the manuscript, is the ideal one the satellite-derived trends should be close to. However, if we resample the AERONET AOTs at specific times and derive trend using them, it cannot be called “actual trend” any more. Furthermore, the comparison between satellite-derived and AERONET-cosampled trends doesn’t show how much satellite-derived trend is close to “actual trend” and doesn’t provide a way to get out of the uncertainty from limited/different sampling in satellite-
derived trend in Section 6. Of course, there are other uncertainties included in Figure 5. Nonetheless, it is a better way to show exactly how convincing satellite-derived trends are. Therefore we would keep Figure 5 as it.

P26012, L21-23: please substantiate why and how the smaller MISR swath might lead to a worse correlation between the MISR and the AERONET trends.

-> Although MODIS and MISR fly on the same platform and therefore have the same local equatorial crossing time, they are different in aerosol sampling, as a result of the different spatial resolutions (i.e., different revisit cycle: 1 ~ 2 days for MODIS and 7 ~ 9 days for MISR at same site) [MODIS Webpage, http://modis.gsfc.nasa.gov; MISR Webpage, http://www-misr.jpl.nasa.gov]. It means that the MISR monthly AOT product for a same region is calculated using samplings at few days per month. Therefore, the MISR trend from the insufficient samples can be different to the MODIS and AERONET trends.

P26013, L3: please specify how the orbital drift would contribute to a low correlation between the AERONET and SEAWIFS/BAER-derived trends? To my opinion, such a cause could easily be resolved by consistent co-sampling of AERONET and SEAWIFS in time.

-> Over region showing a strong diurnal variation as shown in Figure S2, the orbital drift can cause a significant bias in trend estimate. As mentioned above, it is not useful to show the comparison between SeaWiFS and AERONET-cosampled trends because the comparison doesn’t show how much satellite-derived trend is close to “actual trend” and doesn’t provide a way to get out of the uncertainty from orbital drift in SeaWiFS trend in Section 6.

P26034, Fig. 5: the Figure could be improved if the some more information would be included, e.g. by colour-coding desert-trends vs. biomass burning region trends. Such a characterization could shed more light on why trends are sometimes positive in the satellite record, and negative in AERONET.
Many thanks for your constructive suggestion. Since the comparison in Figure 5 contains the uncertainty effects from not only limited/different sampling, but also retrieval accuracy, different spatial resolution, orbital drift, and so on, it is difficult to show a direct cause for the discrepancies even including the information about dominant aerosol types. Instead, we have added a figure showing the discrepancies from only limited/different sampling with the aerosol type information as you suggested. It is more clearly show how significant the uncertainty from only limited/different sampling is and which aerosol type region is more vulnerable to it in trend estimate as shown in Figure S4.

Figure S4. Scatter plots of comparison between the linear trends of all available sampling, (a) 10:30±30 a.m., (b) 12:20±30 p.m., and (c) 01:30±30 p.m. using AERONET AOT (550 nm) data.

P26013, L18: it is not clear from Figure 8 whether the trend for different instruments hold for the same periods. The time interval should be indicated in the text, and also in the caption of Fig. 8.
As you suggested, we have indicated the research periods for each instrument in the caption of Figure 8.

P26040, Figure 11: what do the individual data points represent? (Winter) monthly means in the 2003, 2004, 2005, 2006, 2007, 2008? There are 12 data points so they cannot be seasonal (3-month) means.

We have used Winter monthly means (i.e. DJF) from 2003 to 2008. However, some AERONET fine-mode dominant data are not available in some winter months due to insufficient number of observations per month (Yoon et al., 2012).

Reference


Yoon et al., 2007


