Response to Interactive Comment from Michael Mishchenko

We thank Dr. Mishchenko for his review of our paper. Our responses are embedded below.

The paper attempts to address the important question of the influence of sampling on aerosol retrievals from space. As explained below, this study appears to have major drawbacks which significantly reduce its scientific value.

General comments

It is known from statistics that a random sampling estimate of a dataset's mean is al- ways unbiased (i.e., the expected value of the estimate equals the mean of the dataset) and consistent (converges to the dataset's mean as sample size increases). The variance of such an estimate depends on the sample size, and is proportional to the inverse of the number of measurements (pixels) in a typical sample. AOT values at each pixel of a track crossing a certain region can be considered a random sample and used to estimate the regional mean. In order for this estimate to be sufficiently accurate the "region" should include a number of pixels large enough to reduce the estimate's variance (i.e., the "region" should not be too small relative to the sampling spacing). Quantitative evaluation of the relationship between the region size and the sampling spacing can be done easily using existing statistical techniques, but this is not among the authors' goals.

One may be concerned whether the AOT sampling by, e.g., a curtain track is suffi- ciently random for the above-mentioned statistical methods to be applicable. Indeed, too sparse a set of orbits can systematically miss significant localized features (such as, e.g., dust or pollution plumes). However, the curtain track spacing used in the paper appears to be sufficient to study AOT, which is known to have large variability scales (hundreds of km). This can be seen in Figs. 9 and 10 as each high-AOT region is crossed by multiple curtain tracks (so nothing significant "falls through the net"). Also, the AOT values in "observed" pixels are not so different from those in "not observed" nearby pixels (except for the glint-affected C3 case). We should also note that if the regional biases due to non-random sampling indeed existed they would have led to a global bias, since all AOT features have higher than average value (so they cannot cancel each other). However, no global bias is seen in Fig. 14c.

We agree that the impact of spatial sampling can be reduced by the inverse-square-root of the number of samples, e.g., for sufficiently long time averaging and sufficiently large spatial averaging. However, this only applies to a population distributed randomly about a mean value. As you acknowledge, there are regionally varying sources and sinks, prevailing wind (transport) patterns, and other spatially varying features (e.g., clouds, surface conditions, etc.) that play into how aerosols are observed for different spatial sampling strategies. We show in our paper the limitation of reduced spatial sampling on meaningful regional and temporal scales. Figures 9 and 10 in the original manuscript (see Figures 7 and 8 in the revision), showing the "observed" and "not observed" pictures for seasonal-regional sampling, are made quantitative in Figure 11 in the original manuscript (see Figures 9, 10, and S1 in the revision). There we show that indeed sampling errors are present. That the global annual mean time series shown in the original Figure 14c shows no bias with respect to the individual samplings does not seem so surprising. Some samplings will overestimate the AOT relative to the full-swath sampling in some regions and underestimate in other regions. We make this explicit in the original Figures 11 and 14d,e, although we do not show explicitly these over- and underestimates, but show rather the spread about the full-swath value. (Figures 9, 10, and S1 in the new manuscript do present the range about the full swath value more explicitly.)

It follows from the above discussion that any reasonable sub-sampling strategy is ex- pected to provide an unbiased AOT climatology, spatial and temporal resolution of which depends on the sampling density (less frequent sampling yields a coarser- resolution AOT map and/or seasonal time series). This means that the sampling den- sity cannot be "good" or "bad", but is a quantitative parameter that should be chosen depending on the given climatological resolution requirements. Thus, the authors estimate that a perfect along-track instrument that "could retrieve aerosol properties with no cloud exclusions... would still be sampling only about 10% of the globe" is arbitrary as it depends crucially on the required climatological resolution.

For their study of "observability" the authors chose a 0.5o×0.625o grid essentially studying the observability of each grid cell. These cells are too small for curtain-track sampling, so one would expect the lack of statistical significance of the results. How- ever the findings are presented as applicable to regions which are orders of magnitude larger than the cell size.

Any aerosol instrument we might consider will be making retrievals on a footprint smaller than the 0.5 x 0.625 degree grid we used in this study. The fundamental unit used in this study is the MODIS AOT retrieval footprint, with a nominal resolution 10 x 10 km2 at nadir, which is far smaller than the grid aggregation resolution chosen. Regional aggregations must be built from the individual retrieval footprint, and statistics assembled accordingly.

Finally, the paper's main conclusion that sub-sampling results in "significant regional and seasonal biases" is not supported by the data at all. There is not a single plot comparing regional seasonal means from a sub-sample to that from the full swath. Instead, the authors show plots of their own non-standard metric delta-AOT, which does not have a precise mathematical meaning.

The delta-AOT spatial sampling artifact is the difference between the maximum and minimum regional AOT across all of our sampling strategies. It is therefore the range in uncertainty we find for a given seasonal-regional mean AOT among the sampling strategies chosen. It is a lower bound, because including additional samples might make that range larger, but the addition of other samples to the mix would not make the artifact smaller. Figure 10 in the revised paper does present the range in AOT of all the samples considered about the full swath value.

The "test" MODIS dataset itself has significant gaps and uncertainties

The authors use MODIS level 2 aerosol product to evaluate the influence of sampling on the accuracy of narrow band and curtain-like instruments. The choice of this dataset is less than optimal, since the retrieval errors in MODIS aerosol product are an order of magnitude larger than the reported sampling uncertainties (and hence contribute much more significantly to the total uncertainties in the mean AOT values). The sampling of the aerosol field by MODIS itself is limited. Only around a quarter of all available MODIS Aqua pixels over ocean between 60oS and 60oN are actually used to retrieve aerosol properties, and this fraction has a pronounced seasonal dependence. Over land the fraction of pixels suitable for AOT retrievals is even less: 10%. Clouds, sun- glint, bright or variable surface may lead to gaps in MODIS retrievals. It has been shown using a representative set of model aerosol fields that full Dark target MODIS dataset may produce negative biases in global monthly mean AOT as large as 0.07- 0.12 (30% – 45% of the long-term mean) during boreal summer over land and 0.015- 0.03 (10% - 20%) over ocean (Model-based estimation of sampling-caused uncertainty in aerosol remote sensing for climate research applications, I. Geogdzhayev, B. Cairns, M. I. Mishchenko, K. Tsigaridis,T. van Noije, QJRMS, in press). In large part these bi- ases are caused by the gaps in coverage of the MODIS dataset. Regionally these biases may be significantly greater. Compared to these deficiencies the effect of re- duced sampling of pixel-wide along track instrument is small for global mean AOT.

Although the MODIS AOT product has its own issues, it provides the best dataset available by way of a long time series of global, wide-swath, satellite-derived AOT over both land and ocean. To bolster confidence in our conclusions, we have in the revision added results of an aerosol transport model, which is not subject to the same sorts of uncertainties as the MODIS AOT product. In all cases, the actual AOT field must be sampled sufficiently to derive statistically meaningful trends on decadal time-scales. We find that the conclusions from our analysis of the MODIS data are supported by our analysis of the model fields.

<u>Real narrow-swath instruments can retrieve AOT where MODIS fails</u>

Given the limitations of the original dataset any conclusions drawn from sub-sampling of it may at best be applied to narrow-band or pixel-wide along track

sampling instruments which have the same measurement accuracy and limitations as MODIS itself. This fact is disregarded by the authors who extend the conclusions from sub-sampling MODIS data far beyond their applicability to include any narrowband or pixel-wide along track sampling instrument. Yet the whole purpose of designing and flying such instru- ments is to include advanced capabilities, which would allow more accurate retrievals in places where MODIS fails. For example the CALIOP lidar has a much higher spatial resolution making many measurements for each MODIS footprint. This makes it possible to observe aerosols in scenes with broken cloudiness where no MODIS retrievals would be possible. Unlike MODIS, a lidar can observe aerosols above clouds. Consider Sahara or Asian dust outflow regions where MODIS has difficulties because of glint, underlying clouds and particle non-sphericity. These problems, however, can be addressed by an APS-like instrument with high accuracy polarization channels and multi-angle viewing capabilities.

The point of the paper was not at all a critique of other retrieval approaches, which we hope yield a better AOT product and additional aerosol information than the current MODIS retrievals. The paper addresses the issue of spatial sampling, which unfortunately cannot be addressed by any other existing satellite dataset (or, at any rate, certainly not by the CALIOP dataset). Again, our confidence is bolstered by the results of the aerosol transport model used in the revision, which is not subject to the MODIS retrieval errors. And it has yet to be determined how high the accuracy of an APS-like instrument would actually be at viewing the real, heterogeneous Earth, given the instrument's 6 km pixel footprint at nadir that grows with view angle.

MODIS cross-track bias is not adequately addressed

In his recent presentation at the 2013 AGU Fall Meeting, the lead author admitted that along-track sampling results are affected by the MODIS cross-track bias and crossed over Fig. 5 as no longer valid. However, in the paper this figure and along-track sam- pled data are still present. This especially concerns the C3 sample, which is also affected by sun glint. Fig. 9cd is particularly misleading since neither Sahara dust nor equatorial biomass burning regions are "observed" in C3 track. The authors also claim that MODIS "scan angle biases in the AOT field ... will not affect the statistical signif- icance of the derived trends". This is wrong: such biases will increase the apparent variability of the aerosol load above its actual natural level reducing the statistical sig- nificance. Since these biases are likely to depend on aerosol composition and height, cloud cover frequency, and season, their effect may be more pronounced for narrow scan samples as they view a given location less frequently.

We worked through the view angle artifact to demonstrate our process. We collapsed the two figures in the original document into a single figure (Figure 4) so that the “sample-then-average” mean AOT (i.e., with the view angle artifact) can be directly compared with the results of the “average-then-mask” sampled AOT. For the comparison of the two sets of global, annual mean AOT values we have:

“It is clear that in the global, annual mean, the AOT for the sub-samples generally differs from the full swath value by much less than 0.01. However, *the “average-then-mask” approach for the curtain-like cases provides spatial sampling that is much greater than would ever be acquired by an actual curtain instrument*, because the aggregated samples come from different parts of the broad MODIS swath. So although this approach minimizes the view-angle bias, it includes far more of the broad-swath data than would be available from a curtain instrument.”

The figure referenced just illustrates an issue in the MODIS retrieval with C3 and N3 sampling, namely glint contamination. Those samplings are omitted in deriving the AOT sampling artifact shown in Figures 9, 10, and S1 in the revised manuscript. So these figures show that even excluding those sub-samples that clearly perform poorly due to glint, the spatial sampling artifact remains.

Finally, we computed the trends by two methods: the “sample-then-average” method used in the original document (Figures 11 and 12 in the new) and the much more generous “average-then-mask” method (Figures S2 and S3). We write:

“For comparison, the trends and statistical significance were also computed from monthly means constructed using our “average-then-mask” method. Recall that this provides far greater sampling than the actual along track sampling would provide. The patterns, signs, and magnitudes of the trends (Figure S2) are quite similar to those shown in Figure 11. The effect of the more favorable “average-then-mask” approach is most notable for the patterns of statistical significance (Figure S3), which are broadened considerably for the N1 and C1 samples in this approach. Still, considerable spatial sampling gaps remain, especially over India, the Arabian Sea, Sudan and Ethiopia, Brazil, and the Central U.S., where statistical confidence in the derived trends is not assured in the narrower samplings.”

Use of uncommon statistical methodology

The delta-AOT metric used by the authors is a non-standard statistic which is mis- leading since it conceals the difference between bias and noise contributions to the sub-sampling estimates. The authors should instead use the standard statistics ac- cepted in sampling techniques, such as PDF of the sampling estimate of the regional mean AOT (with the expected value and variance of this estimate).

For the along-track sampling approach, what is shown in the original Figure 11 (Figures 9, 10, and S1 in the revision) are the results of our “average-then-mask” approach. All samples are composed of the seasonal-regional mean full swath averages; we are only excluding **from the seasonal average** those grid boxes that were **never** observed by the given sampling strategy. So the grid boxes commonly observed have exactly the same bias and noise statistics. The resulting delta-AOT metric is then the spread in the seasonal-regional mean AOT value among the remaining grid boxes for any given sampling strategy. For the across-track sampling (now in Figures 9, 10, and S1) we show both the "average-then-mask" approach, which has the same consideration as above, and the "sample-then-average" approach, which possibly has some imprint of the reviewer's concern. But the main conclusion is solid, and is supported by the analysis of the model, which is not subject to those same bias/noise considerations. And assessing the spread in an ensemble of options is a standard statistical technique, applied to advantage in situations where the distribution of error is not random, so expected value and variance are not meaningful. This is frequently the case for model inter-comparisons, and it is also true of the AOT measurements considered here.

Misinterpretation of decadal trends

The authors' choice of the 0.5ox0.625o grid-cell size to assess statistical significance of the AOT trends is highly questionable. Calculating trends independently in each grid-cell puts reduced sampling strategies at a disadvantage, while there is no com- pelling physical reason for the decadal AOT trends in two points 600 km apart to be wildly different. In fact Fig. 12 clearly shows that the trend patterns exhibit features of a regional scale, which are spatially much larger than the 0.5ox0.625o grid. This suggests that considering a few regions instead would improve the confidence for all types of sampling. This conclusion is confirmed by Figs. 17 and 18, where the use of coarser 10ox10o resolution significantly improves the agreement between various subsamples and the confidence levels.

Yet the authors chose to refer to the higher resolution Fig. 13 to state that "our sub- samples could not assign significance at the 95% confidence level to any decadal- scale trends over Amazonia or the central United States, and had reduced confidence in western Africa and India" inconsistently switching from local to sub-continental scale.

The observing system footprint for any conceivable instrument will be smaller than either of the aggregation grids chosen. The point we are making is that aggregation at the higher spatial resolution does not permit assigning statistical significance to the "curtain"-like sampling in many locations *at that resolution*. The point of showing the coarser aggregation grid was to show that if you are satisfied to resolve trends and significance at that scale then perhaps the single pixel wide sampling is sufficient. In our conclusions we write:

"Spatial sampling affects the derived magnitude and assignment of statistical significance in aerosol trends. Along-track curtain sampling results in reduced trend magnitude and essentially eliminates statistical confidence in the derived, decadal-scale trends when the data are aggregated at high spatial resolutions.

Trend magnitudes and statistical significance were more similar to full swath values for the narrow-swath sampling.

"Aerosol trends and statistical significance were found to be similar across sampling strategies when the trends were composed from coarsely gridded aggregates of the sub-sampled MODIS AOT data, suggesting that single pixel width sampling may be sufficient to detect and attribute trends at spatial scales of order 1000 km."

Further, there are certainly compelling physical reasons that decadal AOT trends in two points 600 km apart can be wildly different; at any time, some locations are aerosol sources, others are downwind of aerosol sources, and yet others are neither. Such locations are not distributed randomly in space or time.

Mixing land and ocean MODIS pixels

It is unfortunate that many of the selected aerosol regions mix over land and over ocean pixels in different proportions as the MODIS retrieval approaches and coverage are completely different over land and ocean. Combining them obscures the biases and gaps of the land and ocean retrievals. Also, the authors' statement that "DARF depends strongly on the reflective properties of the surface over which the particles reside, most of which would be unobserved by the curtain instrument" is irrelevant to the study of climatological effects of aerosols, which require statistical properties of aerosol composition and load – tasks for which advanced along-track aerosol sensor is well suited. In addition, ocean surface reflectivity can be modeled and extensive observational data exist on the statistical properties of land surfaces.

The MODIS AOT product is indeed different over both land and ocean. In terms of area, the only one of our chosen regions that is significantly a mix of land and ocean cells is Southern Africa, where in any case the sampling artifact is quite small. The results are consistent for the aerosol transport model, not subject to land/ocean differences regarding the way the aerosol fields are computed. The issue of uncertainty of inferred aerosol forcing we present is well established, and is made more explicit in the revision due to the inclusion of aerosol model results. Spatial sampling implies sampling of aerosol loading, aerosol type, surface type, and clouds. As addressed in this study, we find uncertainty on the TOA SW all-sky aerosol forcing can be as much as 2 – 3 W m2 for certain seasons and regions.