The authors thank the reviewers for their helpful comments. Reviewer comments are reproduced here in red, while our responses are indicated in blue. Where applicable, passages from the manuscript have been reproduced.

We would also like to note a slight change in methodology compared with the initial version of this paper. In the revision, we use AOD at either 470nm or 550nm as independent variables in the regression, whereas before both were used. Because of the high correlation of these two variables, there was little added value to including both, and in fact this led to worse performance for certain datasets with minimal initialization data. By instead using only one of the two variables, the results are generally more robust.

The goal of this paper is to assess the conversion of satellite AOD values (not measurements)…

References to satellite AOD "measurements" have been modified throughout the paper; we refer to these instead as AOD data or AOD retrievals.

Most of my comments are for PA since the paper focuses on this region.

The study region is very small 0.7 degrees by 0.7 degrees. The paper definitely needs a map of some sort showing the location of the regulation grade monitors and the location of the low cost sensors since I have no idea how close are far away these sensors are!

Additional maps for the ground calibration sites have been included in the supplemental information as Figures S4 through S9, with background maps including local landmark and scale information.

The paper never discusses as to how space-time collocation was done for the ground versus satellite data. The results vary depending upon the width of the time and space windows. The paper also does not provide the slope/intercept values for these linear correlations.

This discussion has been included in Section 2.4 (lines 216-218):

> "Satellite AOD data are considered to be collocated in space with data from a ground site when the center of the AOD pixel is within 1 km of the ground site. Data are considered concurrent if the satellite overpass occurs within the hour interval over which ground site data have been averaged to arrive at the hourly-average $PM_{2.5}$ concentration value used."

Slope and intercept values have been included in the supplemental information, Section S3.1.

The range of annual values in PA was low and the satellite data and the low cost sensors have larger uncertainties in this range and therefore the results may not be robust. Given this backdrop I am not sure how meaningful the PA results are. This is probably the main reason that the correlations are low – Page 11 (Line 325+). Not sure about the usefulness of an offline approach where only a single conversion factor is used. Why report these values when we know that this is not relevant?

In the Pittsburgh area, we are able to analyze the effect of ground monitoring network density, which is not possible with the currently sparsely-monitored African locations. Section 3.2 shows that the satellite AOD to surface PM conversion uncertainty reduces meaningfully for up to about ten low-cost sensors over the 600 square kilometer area, which is useful guidance for future low-cost sensor deployments, including those planned for African cities. Further, the results for Pittsburgh are presented to provide a baseline and contrast for the results obtained for Sub-Saharan Africa, where the low-cost sensor and satellite data combination is thus seen to be quite valuable. A single conversion factor is used as it represents the simplest and most robust calibration method, while more sophisticated calibrations might be subject to over-fitting to the calibration data sets. Results are presented for the offline approach as a baseline to compare with an online approach, to assess what benefit if any the online approach provides.

Page 11, Line 319. What is the cloud cover for each site and how does it affect annual average AOD? Given some of the issues mentioned above I am not sure that page 12 (line 24-244) conclusion is acceptable. Also given that the linear correlation has so many problems, using satellite data and ground monitors to assess the linear relationship is fraught with uncertainties.

Information on satellite data coverage is included in the supplemental information, Tables S2 and S3. However, since we do not consider long-term average values in this work, we only compare cloud-free AOD to surface PM measurements taken at approximately the same time (i.e. as the hourly average value for the period in which the satellite overpass occurs). Therefore, there will be no issues related to sampling bias for only using data from cloud-free days in these comparisons, as would be the case if we were looking at longer averaging periods.

Although we agree that the methods presented have numerous inherent uncertainties, a major goal of this paper is to assess whether, even with such uncertainties, useful results can be obtained by combining low-cost sensor and satellite data. We find that this is the case, at least in the context of Sub-Saharan Africa where signal-to-noise ratios can be higher and there is very little ground-based monitoring.

In summary, I believe that low cost sensors play an important role for $PM_{2.5}$ research but unless calibration issues and comparisons with ground monitors of regulation grade are made carefully as a function of space, time, meteorology we cannot be sure how useful the data can be for quantitative monitoring, assessment, and research (e.g. epidemiology). It is also not fair to state that (Page 16, line 482) that using the nearest monitor is better than using satellite data because none of the meteorological factors have been taken into account for estimating $PM_{2.5}$ from satellite data.

Careful corrections by collocation-based comparison of low-cost PM sensors with regulatory-grade monitors and different methodological approaches in the Pittsburgh context have been the subject of a previous paper (Malings et al. 2019b as cited in the paper, DOI 10.1080/02786826.2019.1623863). We have included, where possible, performance assessments for the low-cost monitors in other contexts, but this is a subject of ongoing work and beyond the scope of this paper. Rather, this paper represents a preliminary attempt to quantify the usefulness of simple linear relationships between AOD and ground PM from low-cost sensors, even taking into account any inherent uncertainties these instruments may have.

We did not mean to assert that the use of nearby sensors was always better than using satellite data in all contexts, but merely within the current high-spatial-density monitoring network in Pittsburgh and the confines of the linear conversion method applied. This statement has been clarified in the text (lines 501-504):

> "However, it was found that for Pittsburgh, with a relatively dense low-cost sensor network (median inter-site distance of about 1 km) and low $PM_{2.5}$ concentrations, use of the nearest ground measurement sites outperformed the use of satellite AOD data to estimate surface $PM_{2.5}$ using linear conversions."

Minor comments

Wang and Christopher, 2003 – Not Wang, 2003

We apologize for the oversight. This has been corrected.

Some of the references are outdated. E.g. Zhang et al 2009 for correlation coefficients.

This particular reference has been removed in Section 3.1, but has been retained in the Introduction for its value in providing general background information on AOD to surface PM correlations.

Page 3 : What spatial/temporal scales did Murray et al used

This paper made use of 12-km spatial scale data at daily temporal resolution. This has been noted in the text (lines 86-88):

> "Methods incorporating the outputs of chemical transport models (in this case at lower spatial resolutions of 12 km compared to the 1 km AOD resolution, and at daily temporal resolution) can further improve these results (e.g. Murray et al., 2019)."

Page 3 : Not all studies find 'anti-correlation' in India.

Thank you for pointing this out. Since our present work does not cover India, this information is not strictly relevant, and so we no longer reference it in the paper.

Page 3: Last sentence needs a reference

Low-cost sensor and reference monitor typical prices are based on manufacturer prices in our experience from the past several years. This has been stated (lines 103-107):

> "Low-cost air quality monitors have much lower purchase and operational costs in contrast to traditional or regulatory-grade monitors (Snyder et al., 2013; Mead et al., 2013). For example, a lower-cost multi-pollutant monitor (measuring gases and PM) costs a few thousand US dollars; single-pollutant PM sensors can cost just a few hundred US dollars. A comparable multi-pollutant suite of traditional air quality monitoring instruments would cost a hundred thousand US dollars or more; a regulatory-grade PM monitor can cost tens of thousands of US dollar (based on recent manufacturer quotations)."

Page 3: The cloud cover problems needs to be addressed and referenced. Christopher & Gupta (2010) Satellite Remote Sensing of Particulate Matter Air Quality: The Cloud-Cover Problem, Journal of the Air & Waste Management Association, 60:5, 596-602, DOI: 10.3155/1047-3289.60.5.596

This reference has been added to the introduction section (lines 67-70):

> "Cloud cover also makes AOD retrievals impossible; the frequency of cloudy days in an area can therefore make it difficult to establish reliable relationships between AOD and surface PM, although this is not likely to be a concern for the continental US (Christopher and Gupta, 2010; Belle et al., 2017)."

The cloud cover problem can be important for long-term averages. As noted previously, cloud cover is not an issue in our comparisons because we focus on hourly data during cloud-free periods (lines 218-223):

“As we compare data from individual satellite passes directly to temporally collocated ground site data, we do not need to consider (as would be essential for long-term averages) the potential impact of the fraction of time where satellite measures are missing (due to cloud cover or other factors). Likewise, we do not consider the biases associated with the fact that satellite passes occur at certain times of day (required when comparing with daily-averaged ground monitoring data) since here we only compare AOD to surface $PM_{2.5}$ during the same hour when the satellite pass occurs.”

Page 4: Errors cannot average out and it depends on the range of PM2.5 values and a host of other factors.

This was a conjecture as to possible future applications of satellite and low-cost sensor data. The sentence has been removed.

Section 2.1.1 to 2.1.3 belongs in a Table rather than a few sentences of text

We thank the reviewer for this suggestion. The information presented in these sections, as well as basic details of the study areas, have been presented in Tables 1 and 2.

Page 5: Line 1 : Here not hare

This has been corrected.

Page 7 says ‘as summarized in 2.1.4’ but 2.1.4 does not describe calibration in any detail. Erroneous data screening for negative values is easy but doing this manually for the entire low cost network is not possible.

A full presentation of the calibration methods is beyond the scope of this work, and is more fully covered in the cited publication (Malings et al., 2019b). While it is true that manual error detection and elimination for a large network of sensors is difficult, it can be aided through the use of certain automatic processes. While we seek to present data that has been calibrated and validated to the best of our abilities, we acknowledge that fool-proof error detection and correction is not possible. Such errors are a source of uncertainty in the present work, and one of our major goals with this paper is to demonstrate and quantify the extent to which low-cost sensor data, even with these uncertainties, can provide additional information to support the conversion of AOD to surface $PM_{2.5}$.

Additional details have been provided in the text (Section 2.1, lines 160-170):

“Collected data are down-averaged from their device-specific collection frequencies to a common hourly timescale. Erroneous data identified either automatically (e.g. negative concentration values or unrealistically high or low values) or manually (e.g. devices exhibiting abnormal performance characteristics identified during periodic inspections) are removed. To correct for particle hygroscopic growth effects (i.e. the impact of ambient humidity on the PM mass as measured by the low-cost sensors), previously developed calibration methods (Malings et al., 2019b) were implemented for the NPM and PA-II sensors. Briefly, first, a hygroscopic growth factor is computed using the local humidity and temperature as measured by the low-cost monitor itself, along with an average or typical particle composition. Then, a linear correction is applied to the data based on past collocations with regulatory-grade monitoring instruments. Utilizing these methods, the uncertainties on hourly average $PM_{2.5}$ concentration are about 4 $\mu g/m^3$ (Malings et al., 2019b). For the Alphasense OPC sensors, raw bin count numbers were integrated to produce a new concentration estimate for $PM_{2.5}$, and a similar relative humidity correction was applied (Di Antonio et al., 2018)."

Page 6: Line 180-183 says the data are scaled for workdays and non work days. This type of scaling may work for this study but how about other regions?

Indeed, different scaling factors may be necessary in other regions, and this is the subject of ongoing research on the generalizability of low-cost sensor calibration approaches across the vast continent of Africa. For the purposes of this paper, we seek to use data from low-cost sensors which represent the best available practices in each instance. Therefore, we have included scaling factors in Rwanda based on applicable local comparisons and calibration. Since we are using linear methods, the presence or absence of linear scaling factors that are equally applied to both training and testing sets of low-cost sensor data should not influence the assessment of the methodology.

Page 8: The satellite data needs some description with a proper journal reference. Briefly, how was AOD retrieved, what are the uncertainties, how much cloud cover for the analysis, what quality flags were used, etc.

A more complete description of the satellite data has been provided in Section 2.4 (lines 205-223):

> "The satellite data product used in this paper is the MODIS MCD19A2v006 dataset (Lyapustin and Wang, 2018) available through NASA's Earth Data Portal (earthdata.nasa.gov). This dataset consists of AOD information for the 470nm and 550nm wavelengths from the MODIS system, processed using the Multi-angle Implementation of Atmospheric Correction (MAIAC) algorithm, and presented at 1 km pixel resolution for every overpass of either the Aqua or Terra satellites (Lyapustin et al., 2011a, 2011b, 2012, 2018). This represents a Level 2 data product, meaning that it includes geophysical variables derived from raw satellite data, but has not yet been transformed to a new temporal or spatial resolution, as is the case for data derived from multiple satellite passes, e.g. monthly average AOD data. Data from identified cloudy pixels are masked as part of the data product; possible misidentification of cloudy pixels is one source of error in relating surface $PM_{2.5}$ and AOD. As per recommendations in the User Guide for this dataset, only data matching "best quality" quality assurance criteria are used. This dataset was chosen as it represents the highest possible spatial and temporal resolution for AOD, thus providing the most points for comparison with the high spatio-temporal resolution low-cost monitor data.
>
> Satellite AOD data are considered to be collocated in space with data from a ground site when the center of the AOD pixel is within 1 km of the ground site. Data are considered concurrent if the satellite overpass occurs within the hour interval over which ground site data have been averaged to arrive at the hourly-average $PM_{2.5}$ concentration value used. As we compare data from individual satellite passes directly to temporally collocated ground site data, we do not need to consider (as would be essential for long-term averages) the potential impact of the fraction of time where satellite measures are missing (due to cloud cover or other factors). Likewise, we do not consider the biases associated with the fact that satellite passes occur at certain times of day (required when comparing with daily-averaged ground monitoring data) since here we only compare AOD to surface $PM_{2.5}$ during the same hour when the satellite pass occurs."

Additional details on the cloud cover and uncertainty analysis are included in the supplemental information, Tables S2 and S3.