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.

General Comments

The majority of the results section focuses on the analysis for the Pittsburgh region. The goal of the paper is to assess the utility of low-cost sensors in deriving satellite AOD conversion factors, however, the results for Pittsburgh seem to suggest that ground monitor data overall performs poorly as a data source for the conversions over the region, at least in terms of correlations. As the authors note, this is likely due to the low concentrations being within the range of signal-to-noise in the sensors. This makes the results less meaningful, because it is difficult to determine whether the results are reflecting the ability of the low-cost sensors to be data sources for the satellite AOD conversion, or whether the results are just overwhelmed by the uncertainties in the measurements, and undermines the authors' conclusions that low-cost sensors perform just as well if not slightly better than the regulatory grade monitors in this region.

One of the major motivations for including the results from Pittsburgh is to present a baseline case for a densely monitored (with both regulatory and low-cost monitors) region in order to contrast with results from more sparsely monitored locations in Rwanda and elsewhere. In particular, although we agree that overall performance of the satellite AOD to ground $PM_{2.5}$ conversion is rather poor in the conditions of Pittsburgh, it is at least consistent for both ground data sources (regulatory reference instruments and low-cost monitors). Note that the typical $PM_{2.5}$ concentrations in Pittsburgh (an inter-quartile range of 6 to 12 μg/m^3) are still above the hourly-average measurement uncertainty (3 to 4 μg/m^3) of the low-cost sensors. Considering the reasonable agreement between low-cost and regulatory-grade monitors identified in previous work, together with the observation from this work that performance is not noticeably disadvantaged by the substitution of regulatory-grade for low-cost monitors, we believe it is reasonable to assume that most of the poor performance of the satellite AOD to ground $PM_{2.5}$ conversion is due to the inherent difficulties of this problem and the low-concentration regime of Pittsburgh, rather than the data quality of the ground source. We have restated the conclusion based on our comparative analysis of low-cost and regulatory-grade instruments in Pittsburgh to better emphasize this (lines 389-396):

> "In all cases, performances using low-cost sensor data are comparable to that of the same conversion approaches utilizing the regulatory-grade instruments. Note that the low-cost monitors used here have been carefully corrected by collocation with regulatory-grade monitors (Malings et al., 2019b) which accounts for known artefacts with low-cost sensors. Thus, there is no evidence from this analysis of any inherent disadvantage to the use of carefully corrected low-cost sensors to provide ground data as compared to more traditional instruments. Rather, based on these results, any additional uncertainty due to data quality differences between low-cost sensors and regulatory-grade instruments are seen to be negligible compared to the difficulties

associated with relating satellite AOD to surface-level $PM_{2.5}$, and therefore have had no systematic impact on the performance of the assessed linear conversion method, at least for this study area."

The analysis over Africa appears to be more promising, but much less time is spent discussing those results. The authors may be better suited by more prominently presenting the analysis over Africa. Low-cost sensor data would provide more benefit over regions such as Africa where the regulatory grade monitors are sparse; there already exist dense regulatory grade monitors over North America, so focusing more on the analysis over Africa would be of greater interest. Describing in detail the comparison of low-cost sensors and regulatory grade monitors in Pittsburgh would make sense if the results were meaningful, as they would provide a meaningful evaluation of the ability of the low-cost sensors to be used to convert satellite AOD in general, but in this case the results seem to suggest the method just doesn't work over Pittsburgh, and does little to provide confidence in the low-cost sensor only analysis over Africa.

We thank the reviewer for recognizing the potential benefit of low-cost sensors for Africa. This is a point we seek to make and support quantitatively through the results presented in this paper. We have expanded our discussion of results in Africa to increase the relative emphasis placed on these results. We have also reorganized the paper somewhat and restructured the discussion of the results, including a new figure related to this discussion (Figure 6) to better emphasize the relative significance and importance of the results for Africa. However, we feel it is also important to present the "weaker" results for Pittsburgh as a basis of comparison for the more promising results for sub-Saharan Africa. Furthermore, the analysis of the potential benefits of high spatial density low-cost sensor networks (the "how many sensors are needed" question) can only be performed using the Pittsburgh data, where such a network has been operational since 2016.

Specific Comments

- Several of the figures are difficult to decipher. Figure 2 is difficult to read because the labels on the y-axis are clustered so close together. Figure 7 is extremely difficult to interpret, because it is hard to see the shades of red. Supplemental figures S6-S9 are very hard to follow and do not help to clarify the methods.

The vertical spacing of Figure 2 has been increased. The color scale of Figure 7 has been changed to improve interpretability. Numerical values corresponding to these colors have also been provided in the supplemental information (Table S8). Supplemental Figures S6 to S9 were augmented with a more detailed narrative description of the methods (including new figures, with all figures now numbered S11 to S18), which we believe makes these points more clearly.

- In addition to Figure S5, the authors should have map plots for each region with the monitor locations over-laid, with a better indicator for the distance between monitors than just latitude and longitude. It is very difficult from Fig S5 to discern where the monitors are positioned throughout the cities, which would provide insight into the results. It is very difficult to tell which monitors are low-cost and which are regulatory without looking extremely closely.

Map plots depicting the locations of the monitors have been included in the supplemental information as Figures S4 through S9. The markers are much larger and are overlaid on geographical maps which should help better illustrate the monitor locations.

- It is unclear how the satellite AOD and ground monitor data are being sampled; are the authors using pixels co-located to the ground monitor sites, or are they comparing a broader area of AOD to the ground monitor points? Also at which time-scales are the data points being sampled? At satellite-overpass time? This information would have important implications for the results.

A more complete description of the sampling method 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."

- In several instances more "methods" type descriptions are mixed in with the results. Having all methods descriptions in the methods section would make the presentation of the results clearer.

We thank the reviewer for this suggestion. These descriptions have been moved into their own subsection (2.6) within the "Methods" section.

Minor comments:

- Line 70: what is a "good" correlation? No range of values from the studies is given.

A representative value from the reference has been provided (lines 72-75):

> "Nevertheless, early examinations of AOD data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument, launched aboard the Terra and Aqua satellites in 1999 and 2002, showed good correlation (e.g. correlation coefficient r about 0.7 for Jefferson County,

> Alabama in 2002) with surface PM2.5 concentrations in the United States, although these relationships varied from region to region (Wang and Christopher, 2003; Engel-Cox et al., 2004)."

- Throughout the manuscript the authors refer to "satellite AOD measurements", when technically they are retrievals and not direct measurements.

References to satellite AOD "measurements" have been modified throughout the paper. We now refer to these as AOD data or AOD retrievals.

- In the introduction the second paragraph on page 3 is confusing. It is structured as though they are discussing studies that use models combing satellite AOD with CTMs to estimate PM2.5, but then all of a sudden they are discussing satellite AOD and ground monitor PM2.5 agreement over Africa.

This paragraph has been split into two to better present these different topics.

- When discussing the yearly/monthly conversion factors on page 11, it is unclear whether the monthly conversion factors are applied on a monthly basis, or if they are calculated on a monthly basis then applied on an annual basis: "the 'monthly' case, data from the previous month are used to develop conversion factors used in the current month; the median performance across months is presented".

These factors are applied on a monthly basis. This has been clarified in the text (lines 295-299):

> "For a "yearly" conversion, data from the entire calendar year are used to develop the conversion factors, while in the "monthly" case, data from the previous month are used to develop conversion factors that are then assessed in the current month (e.g. January data are used to develop conversion factors that are applied in February, then the February data are used to develop conversion factors that are applied in March, etc.). For the "monthly" case, the median performance across months is presented."