The goal of this paper is to correct for AOD retrieval biases in GOES ABI AOD product using an empirical approach. The surface reflectance in the current GOES AOD algorithm is estimated based the relationship between 0.47 and 2.2 um and 0.64 and 2.2 um since most aerosols are 'transparent' in the 2.2 um. This is based on the Kaufman et al (1997, IEEE) paper that many MODIS algorithms use to estimate surface reflectance. In this paper, the authors look for 'clear days' (based on AERONET AOD values less than 0.05) to assess the GOES AOD (for both high and medium quality retrievals) for a few selected sites. They report that the GOES AOD is biased since the GOES AOD is much larger than the AERONET AOD for these clear days. The authors note that the biases appear to be centered around 1700 UTC and it is due to surface reflectance parametrizations at various sun-satellite viewing geometries. The authors then attempt to correct this bias based on the premise that it is the surface reflectance that is the issue in the GOES algorithm. Then they use a 30-day composite of GOES AOD to estimate the minimum AOD and subtract that with the background AOD (a fixed value of .025) to correct for the bias. They use two polynomial fitted relationships to estimate biases. They then correct the AOD using these relationships and then validate the results with AERONET AOD and show improvement in these biases.

First I need to note that the paper needs to go through some editorial clean up since several sentences are awkward; key references (Kondragunta et al 2020) are missing;and some references are really old.

We removed several old references and checked the references.

I find several problems with the paper and most importantly it is the use of AOD to make these corrections rather than working with the reflectances. The algorithm retrieves AOD based on apriori assumptions of aerosol model, surface parametrizations based on NDVI , cloud clearing approaches and a host of thresholds for cloud cover, inhomogeneity, etc (ATBD, 2018). Now this paper indicates that the surface parametrizations are a problem and then to remove the biases the authors use the retrieved AOD to make bias adjustments. The original algorithm uses reflectance ratios to arrive at surface values and now this paper goes back to the older GASP approach to obtain the 30-day composite minimum (not reflectance) AOD values. Looking at this from an algorithm perspective it is not the correct solution for an operational algorithm to go through retrieval using one set of processes, retrieve AOD's and then use the retrieved AOD values to make corrections for parameters that are part of the original retrieval process (in this case surface reflectance). The authors need to think about having the correct algorithm as part of the retrieval process rather than adjusting it after the retrieval is done.

The purpose of the bias correction algorithm is to correct the bias in an already existing aerosol optical depth product. It is not intended to substitute for the original AOD algorithm. We agree, and has been fully aware of, that ideally reduction of biases should be dealt with in the AOD algorithm itself. As discussed in the paper, the deviation of the real spectral surface reflectance relationships from the parameterization used in the retrieval can cause the AOD retrieval bias. Improving the spectral surface reflectance relationships is the subject of an independent, parallel work, and thus it is not discussed in the current paper. Once such an improvement becomes available and is shown to satisfactorily reduce the AOD bias, the bias correction may be turned off. But before that happens, we plan using the bias correction algorithm to provide users an AOD product with improved accuracy and coverage. On the other hand, surface reflectance relationship parameterization is derived at AERONET sites and is assumed to be valid over all other areas. This assumption may not hold everywhere. Actually there are very few evaluations over areas other than AERONET sites. The bias correction algorithm can evaluate AOD bias over areas other than AERONET sites and reduce the bias there. The empirical bias corrections to retrieved AODs is not new. The NASA MODIS Dark Target AOD algorithm corrects

AOD using a bias correction algorithm over urban areas using post processing of AODs for areas where urban land percentage is greater than 20% (Gupta et al., Atmos. Meas. Tech., 9, 3293-3308, 2016). There are other MODIS AOD correction algorithms as well developed by users for their own applications (e.g., Lary et al. 2009). In fact, compared to these bias correction algorithms, our approach is better because it is internally consistent and does not rely on any external dataset.

The other issue is the relaxation of quality flags to allow more data. There were strong reasons for picking all the metrics for high and medium quality flags in the first place (ATBD, 2018) whether it is cloud/snow cover or inhomogeneity. Line 90 to 95 provides the various reasons for selecting the pixels for the retrievals and this paper now allows all the medium quality flags in the process but does not address cloud contamination issues.

There is always a tradeoff between better data coverage and reducing cloud contamination. From the scatter plots in Figure 4 (Figure 5 in the revised paper), the bias corrected high and medium qualities AODs have statistics close to that of the high quality alone, which suggests that once the bias correction is applied we can use data that were assigned either high or medium quality data in the original AOD without sacrificing accuracy. In other words, the expectation is that the AOD at pixels that are subjected to potential cloud/snow contamination, and thus are labeled as medium quality, are corrected (at least partly) for this contamination as a result of bias correction.

The paper needs to be more convincing that it is indeed surface issues and not cloud cover that causes these problems. The results need to be discussed in terms of scattering angles (see She et al, Remote Sensing, 2019). This will allow more quantitative analysis rather than statements like those in 160-161.

We adopted your suggestion and plotted the scattering angle dependence of the error, comparing the AOD errors before and after correction. The original ABI AOD errors have a scattering angle dependence in the plots. After applying the bias correction algorithm, the scattering angle dependence of the bias is reduced. (Figure 6 in the revised paper and corresponding discussions).

There are several reasons that the bias is not caused by cloud contamination: (1) the diurnal pattern of retrieved ABI AOD on clear days always has a peak at around noon and the peak gradually reduces away from noon; cloud contamination is not expected to produce such a pattern; (2) cloud contaminations are random errors instead of systematic errors shown in the paper. Random errors from cloud contamination won't be corrected by our algorithm. The effectiveness of our algorithm in removing systematic errors indicates that the main reason of the bias is not cloud contamination. (3) If in some cases cloudiness at a given location has its own diurnal cycle and introduces a systematic bias, the bias correction corrects it too. The bias correction algorithm does not differentiate where the bias comes from and it corrects the bias as long as the bias is systematic.

Also for Figure 1 and Figure 2 what were the histograms of actual reflectance's from the GOES channels for the various peaks. This can help explain Figure 2 better.

Instead of histogram of the surface reflectances, we plotted scatter plots of the 0.47 µm and 2.2 µm surface reflectances, surface reflectance relationship used in the ABI AOD retrieval algorithm, and the histograms of NDVI for six observations around the GSFC site (Figure 3 in the revised paper). The analysis shows that the surface reflectance relationship used in the retrieval algorithm is directly connected to the ABI AOD retrieval biases. The change of the peak ABI AOD bias amplitude is related to the different relationship used because of the differences in NDVI in the three days.

The paper uses two sets of parametrizations for adjusting the biases and then in line 255 back tracks the approach by stating that this could have large uncertainties.

We don't expect a parameterization to fit every situation. As long as it works for the majority of the locations and/or geometries, it can be used. Notice that even for the worst case in the early morning, for the University of Houston, the peak bias is reduced from 0.4 to 0.3 (Figure 1e and Figure 8e in the revised paper).

Figure 8 and Figure 9 appears as a complete afterthought since the aerosol model discussion is not complete or convincing.

We respectfully disagree with the reviewer that this is an afterthought. One of the challenges of aerosol remote sensing is the representation of aerosol optical and physical properties in the models used to generate Look-up-Tables for retrievals. Aerosol model selection over land is a known problem in MODIS/VIIRS type sensors. We don't expect to be able to solve it with the bias correction algorithm either. Here we just want to point out that the problem exists.

I have no idea why the PM2.5 discussions (Figure 9) is relevant for this paper.

One of the main reasons why NOAA generates near real time AOD retrievals is for user applications related to air quality monitoring and forecasting. Users use AOD as a proxy for surface PM2.5 and among many things that impact this relationship, accuracy of AOD itself is very important. Better AOD retrieval means better PM2.5 estimates from satellite, which is an important application of satellite AOD product.

The AERONET data used is from 2018 and the authors need to be using Level 2 not 1.5. This data should be available.

Some sites still don't have Level 2 data yet. For example, as of April 20 2020, GSFC still does not have level 2 data available for the days after September 2018. In our daily work, we routinely do our analysis with both Level 2 and Level 1.5 data and we are quite comfortable in using Level 1.5 data.

Other issues. Define accuracy and precision and be quantitative rather than merely stating that one product is better than the other.

We added the corresponding numbers into the places where we discuss the accuracy and precision.

Line 50, Deemed to have quality sufficient is rather vague.

Sentence removed.

Line 73: The word transparent to most aerosols is rather vague. Describe why this is possible briefly based on aerosol size and extinction

We removed this sentence. This is the assumption of the original MODIS algorithm. It was abandoned later on so that the retrieval is more accurate. 2.2 µm band is approximately transparent to small sized particles such as smoke, urban aerosols, but it is not as transparent to large particles such as dust. The extinction is determined by the ratio between the wavelength and the particle size. Based on Mie theory, the larger the ratio, the smaller the extinction.

Line 74. Again, poor phrasing. It is not linear reflectance BETWEEN channels if it is three channels. Be specific.

Changed to "The algorithm assumes linear relationships exist between the surface reflectance of 0.47 µm band and 2.2 µm band, and between those of 0.64 µm band and 2.2 µm band."

Line 75-80 is awkward phrasing. The algorithm does not make retrievals? Describe the algorithm clearly but briefly. Line 81-84 is not clear at all.

While I understand how the algorithm works this type of writing will not help all readers understand the algorithm and methods used in this paper.

Revised the paragraph as follows.

Over land, three ABI channels are used in the retrieval, i.e. 0.47 µm, 0.64 µm, and 2.2 µm. The algorithm assumes linear relationships exist between the surface reflectance of 0.47 µm band and 2.2 µm band, and between 0.64 µm band and 2.2 µm band. The coefficients of the relationships are functions of NDVI (between 0.86 and 0.64 µm channel) and solar zenith angle (GOES-R ABI AOD ATBD, 2018). Other atmospheric and geographic parameters needed for the retrieval are also inputted, such as surface pressure, surface height, total column ozone, etc. The algorithm only retrieves AOD over dark surface, when the TOA reflectance in the 2.2 µm band is less than 0.25. The retrieval algorithm contains two steps. In the first step, one of four aerosol models is assumed, i.e. dust, smoke, urban, and generic, and AOD for each of the aerosol model is retrieved using the 0.47 µm and the 2.2 µm bands. The algorithm uses a Look-up-Table (LUT) to perform radiative transfer calculation. The LUT stores reflectances, transmittances and other quantities for discrete states of atmosphere and Sun-satellite geometries. For each AOD in the LUT, the algorithm performs atmospheric correction in 2.2 µm band to obtain surface reflectance in that band, and uses the 0.47 µm and the 2.2 µm band relationship to obtain 0.47 µm band surface reflectance. TOA reflectance in the 0.47 µm band can then be calculated using the LUT. The AOD for the assumed aerosol model is obtained through interpolation of the two AODs that give TOA reflectances in the 0.47 µm band closest to the satellite measurement. At the end of this step, there are four AOD solutions from the 0.47 µm band and 2.2 µm band, one for each aerosol model. In the second step, one of the four solutions is then selected as the final retrieval using the 0.64 µm channel by looking for the aerosol model that gives a TOA reflectance in that channel that is the closest to the observed TOA reflectance. In this step, 0.64 µm band TOA reflectance is calculated with 2.2 µm band surface reflectance from last step, relationship between 0.64 µm band and 2.2 µm band and AOD of corresponding aerosol model. The algorithm does not make retrievals over bright land pixels, pixels covered by cloud or snow, etc. The AOD retrieval range is [-0.05,5] and any retrievals greater than 5 are marked as out of range.

Line 89: Usually very small? What does that mean? Need some numbers.

Added: "For example, the ratio between the number of the top 2 qualities and the high quality matchup with AERONET is about 2 (see the following section), while the ratio is 1.2 for VIIRS AOD (Laszlo and Liu, 2016). "

Lines 89-94 needs to be clearer with brief discussion rather than listing the problems.

They are just a list of criteria used to degrade AOD quality in the current algorithm. The starting sentence was revised as: "Following criteria are used to degrade a pixel from high quality to medium quality: …"

The problem is the standard deviation test. We did discuss it in the next sentences.

The reasons for the other criteria are out of the scope of this paper and are not discussed in the paper. Following are the reasons about the cloud/snow adjacency criteria:

Pixels close to clouds or snow can be potentially impacted by radiation scattered from them into the cloud-free and snow-free columns (e.g. Marshak and Davis, 2005; Lyapustin and Kaufman, 2001). For clouds, there is also the issue of transition from clear to cloudy, which is gradual. Cloud detection may not label these pixels as cloudy because they are not bright enough. At the same time these pixels have cloud droplets mixed with aerosol, and/or a humidity that results in aerosols, if they are hygroscopic, which are not well represented by any of the models in the LUT (e.g. Jia et al., 2019; Tang et al., 2019) .

Line 98: If the surface reflectance issues are so different between 0.41 and 0.47 micron then the authors need to show or discuss this for certain land types. Otherwise these statements are vague.

Added: Over CONUS region, from VIIRS data, the 0.41 µm surface reflectance is 0.3-0.4 times the 0.67 µm band surface reflectance and the 0.47 µm surface reflectance is 0.5-0.6 times the 0.67 µm surface reflectance (Zhang et al., 2016). Therefore, 0.41 µm surface reflectance is 20%-50% lower than 0.47 µm surface reflectance.

115-120 discussion is not "technical" enough. What does air mass movements mean? You need to then state what wind speeds at what height provide the 27.5 km radius.

We removed "the air mass movements" in the sentence. For this matchup, we did not do temporal matchup with AERONET and just plotted the time series of ABI AOD and AERONET AOD. To our knowledge, the air mass movements argument first appeared in Ichoku et al. (2002) as follows: " the average travel speed of an aerosol front is of the order of 50 Km/h. This was visually estimated from animated daily sequences of TOMS aerosol index images (http://jwocky.gsfc.nasa.gov/aerosols/aermovie.html) for July to September 1988, where aerosol fronts are seen crossing the Atlantic from the west coast of Africa to the East coas t of America (approximately 6000 Km) in about five or six days. Therefore, the 50x50 Km window would match a 1-hour sunphotometer data segment. All references to MODIS spatial statistics in the rest of this paper imply those based on the 50x50 Km (5x5 pixel) subset grid". They did not mention the height of the aerosol layer.

Line 140+: How about retrieval biases due to sun-satellite viewing geometry in radiative transfer code?

We are not aware of any report in the literature of AOD retrieval errors with magnitude $\geq$ 0.1 due to radiative transfer model within the range of ABI AOD retrieval geometry. Errors may be present at the edge of the disk due to plane parallel assumption but those retrievals are not recommended for even qualitative use, and they were excluded from the current analysis.

Line 147: We need to see these relationships between two channels for the solar geometries.

They are added in the paragraph of case studies at GSFC site.

I find the two reasons in 152-155 to be problematic. Why should the test position issue matter if these relationships are established for certain solar viewing geometries/NDVI?

The parameterization is a simplified model that assumes the relationships depend only on solar zenith angle and NDVI. However, in reality, the relationships depend on all the angles, i.e. solar zenith angle, satellite zenith angle, solar azimuthal angle and satellite azimuthal angle, and surface type (not only NDVI). In addition, NDVI is also a function of those angles. When the satellite moved, the satellite angles changed. Unless the relationships and NDVI are independent of satellite angles, the relationships should change.

Plus there are reasons why the quality flags were established for high, low, medium in the first place (cloud cover, snow cover etc). Of course one would use the best quality flags for establishing surface reflectance relationships because of contamination issues. Now if you are using medium quality flags to get more data into the analysis then of course your surface reflectance relationships are going to be different.

This is the problem of surface reflectance relationships parameterization: they cannot be generalized to other pixels without losing AOD retrieval accuracy. With bias correction, we can correct those biases caused by this problem.

Since this paper is about surface reflectance issues the authors need to show these relationships that currently exist for various angles/NDVI first to make their case stronger.

We understand what you are saying but we think the paper is not about surface reflectance issues. It is about correcting the bias that, we think, happened to be caused primarily by deficiencies in the way we parameterize the relationship between spectral surface reflectances. The detailed surface relationships are available in GOES-R ABI AOD ATBD (2018) and is out of the scope of the paper.

But for your information, following is a summary of the relationships:

The surface reflectance relationships used in the above retrieval algorithm are derived through studies of ABI pixels near AERONET sites, where AODs are accurately measured from the ground and are considered as ground truth. A set of stringent pixel selection rules are applied to build a matchup dataset between ABI pixels and AERONET AOD in order to reduce cloud contamination and uncertainties in aerosol models (GOES-R ABI AOD ATBD, 2018). If AERONET AOD is less than 0.2 of a matchup dataset, surface reflectance of the pixels at the three channels are retrieved through atmospheric correction. With surface reflectance of all such pixels, the relationships are then derived and parameterized as functions of the solar zenith angle for different ranges of the normalized difference vegetation index (NDVI, between 0.86 and 0.64 µm channel) through linear regression analysis of the spectral surface reflectance. The current surface reflectance relationships are derived from ABI full disk matchup dataset in the time period of 04/29/2017 – 01/15/2018.

The surface reflectance relationships obtained are described in the following equations:

$$\rho_{0.47}[\rho_{0.64}] = (c_1 + c_2\theta_s) + (c_3 + c_4\theta_s)\rho_{2.2} \quad (1)$$

Where $\rho_{0.47}$, $\rho_{0.64}$, $\rho_{2.2}$ are surface reflectance at the three bands, c_1, c_2, c_3, c_4 are constants depending on NDVI as shown in Table 3-12 of the ATBD (shown in the following), θ_s is the solar zenith angle. NDVI is defined by red (0.67 µm) and NIR (0.86 µm) bands at TOA as

$$\mathrm{NDVI} = \frac{\rho_{0.86}^{TOA} - \rho_{0.64}^{TOA}}{\rho_{0.86}^{TOA} + \rho_{0.64}^{TOA}} . \quad (2)$$

Table 3-12. Coefficients in the spectral surface reflectance relationship for different ranges of NDVI.

Channels (μm)	c_1	c_2	c_3	c_4
	NDVI ≥ 0.55			
0.47 vs. 2.25	1.436330E-02	2.060893E-04	1.749239E-01	-2.859502E-03
0.64 vs. 2.25	1.374160E-02	-5.128175E-05	2.761044E-01	1.034823E-03
	0.3 ≤ NDVI < 0.55			
0.47 vs. 2.25	4.163894E-02	-2.147513E-04	1.598440E-01	7.401292E-04
0.64 vs. 2.25	2.990101E-02	-1.873911E-04	4.602174E-01	9.658934E-04
	0.2 ≤ NDVI < 0.3			
0.47 vs. 2.25	5.154307E-02	5.679386E-05	2.048702E-01	-7.064656E-04
0.64 vs. 2.25	5.179930E-02	-1.043257E-04	4.937035E-01	4.310074E-04
	NDVI < 0.2			
0.47 vs. 2.25	-4.990575E-02	2.138207E-03	8.498076E-01	-1.179596E-02
0.64 vs. 2.25	-3.397737E-02	1.640336E-03	1.087497E+00	-9.538776E-03

In the revised paper, we provide a detailed analysis and the surface reflectance relationships used over the GSFC site.

The authors should also show the reflectance values on these plots so we can interpret the results better.

We added them in the surface reflectance discussion for GSFC case study.

References

GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Suspended Matter/Aerosol Optical Depth and Aerosol Size Parameter, NOAA/NESDIS/STAR, Version 4.2, February 14, 2018, https://www.star.nesdis.noaa.gov/smcd/spb/aq/AerosolWatch/docs/GOES-R_ABI_AOD_ATBD_V4.2_20180214.pdf, accessed 02/24/2020.

Gupta, P., Levy, R. C., Mattoo, S., Remer, L. A., and Munchak, L. A.: A surface reflectance scheme for retrieving aerosol optical depth over urban surfaces in MODIS Dark Target retrieval algorithm, Atmos. Meas. Tech., 9, 3293–3308, https://doi.org/10.5194/amt-9-3293-2016, 2016.

Ichoku, C., Chu, D.A., Mattoo, S., Kaufman, Y.J., Remer, L.A., Tanré, D., Slutsker, I. and Holben, B.N.: A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophys. Res. Lett., 29(12), 8006, doi:10.1029/2001GL013206, 2002.

Jia, H., Ma, X., Quaas, J., Yin, Y., and Qiu, T.: Is positive correlation between cloud droplet effective radius and aerosol optical depth over land due to retrieval artifacts or real physical processes?, Atmos. Chem. Phys., 19, 8879–8896, https://doi.org/10.5194/acp-19-8879-2019, 2019.

D. J. Lary, L. A. Remer, D. MacNeill, B. Roscoe and S. Paradise, "Machine Learning and Bias Correction of MODIS Aerosol Optical Depth," in IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 4, pp. 694-698, Oct. 2009, doi: 10.1109/LGRS.2009.2023605.

Lyapustin, A. I. and Kaufman, Y. J.: Role of adjacency effect in the remote sensing of aerosol, J. Geophys. Res., 106, 909–916, 2001.

Marshak, A., and Davis, A. (Eds), 3D Radiative Transfer in Cloudy Atmospheres, Springer Science & Business Media, 2005.

Tang, M., Chan, C. K., Li, Y. J., Su, H., Ma, Q., Wu, Z., Zhang, G., Wang, Z., Ge, M., Hu, M., He, H., and Wang, X.: A review of experimental techniques for aerosol hygroscopicity studies, Atmos. Chem. Phys., 19, 12631–12686, https://doi.org/10.5194/acp-19-12631-2019, 2019.