**This paper evaluates the AOD retrieval from geostationary platform GOES ABI and proposed an empirical bias correction scheme to improve the AOD accuracy. The GOES AOD product is potentially very useful in radiative forcing and air quality studies, in that it offers the diurnal variability of AOD on large scale. However, the existence of bias in the diurnal cycle is a significant drawback that limits its use. Therefore, the bias correction scheme offered in this paper is both important and useful. However, I hope the authors can give more analysis proving and explaining that surface reflectance is responsible for the bias, and that the bias correction is effective under all AOD loading and surface conditions. These I think are major issues, although they should not be too difficult to address. My detailed comments are listed below.**

**Major comments:**

**I agree with the authors that surface reflectance parameterization is the most likely cause of the AOD bias. However, in the paper the authors seem very definitive on this point. For example, in the abstract, it says "ABI AOD has diurnally varying biases due to errors in the land surface reflectance relationship between the bands used in the ABI AOD retrieval algorithm". Therefore, I wonder if they can offer more detailed analysis proving this point and explain how the relationship between surface reflectance of different channels vary with geometry?**

In the revised paper, we give a detailed case study at GSFC site for different geometries. Specifically, in the case study, the surface reflectance relationships used are closer to the real relationships in the afternoon than at noon, and therefore, the afternoon AODs retrieval are closer to the AERONET AODs.

**The difference between the test position and the current operational position does not seem large enough to account for such high AOD bias.**

We did not say it is the main reason. But it is one reason, although the effect may be small.

**One possibility is that the NDVI also varies with solar zenith angle. Do the authors use MODIS NDVI? They are calculated from polar orbiting satellites and the NDVI only represent one solar zenith angle. Although NDVI should be a normalized quantity that is not affected by the angle, the large different solar position between polar orbit and geostationary orbits may cause MODIS NDVI not representative of all angles.**

No, ABI AOD retrieval algorithm doesn't use MODIS NDVI. The algorithm uses ABI top of atmosphere reflectance of 0.64 μm and 0.86 μm bands to calculate it, independent from MODIS. NDVI is defined by red and NIR bands at TOA as

$$\mathrm{NDVI} = \frac{\rho_{0.86}^{\mathrm{TOA}} - \rho_{0.64}^{\mathrm{TOA}}}{\rho_{0.86}^{\mathrm{TOA}} + \rho_{0.64}^{\mathrm{TOA}}}$$

The geometry dependence of NDVI is an issue, but not so large for the main cause of the AOD bias, as shown in the case study at GSFC in the revised paper. We are aware that this NDVI is not an "aerosol-resistant" NDVI. The choice of wavelength was dictated by the availability of ABI channels. MODIS AOD algorithm uses 1.24 μm and 2.12 μm band pair, i.e.
$\mathrm{NDVI} = \frac{\rho_{2.12}^{\mathrm{TOA}} - \rho_{1.24}^{\mathrm{TOA}}}{\rho_{2.12}^{\mathrm{TOA}} + \rho_{1.24}^{\mathrm{TOA}}}$. However, ABI does not contain the 1.24 μm band. Based on the available ABI bands, we analyzed the dependence of 0.47 μm and 2.25 μm surface reflectance relationship to the NDVI from (0.64,0.86) μm pair and from (0.86,2.2) μm pair. It turned out that NDVI from (0.64,0.86) μm pair better separates the soil-based and vegetation-based. Based on this NDVI, the surface is classified into 4

different NDVI ranges and the surface reflectance parameterization are derived using ABI reflectances (Table 3-12 in ATBD), independent from MODIS.

**2. The bias correction assumes that the difference between 30-day minimum AOD and background AOD is the systematic error, and subtract this error from every AOD retrieval. I wonder if the bias also depends on AOD itself, i.e., aerosol loading, so that the systematic bias derived as above does not represent all AOD conditions?**

In the revised paper, an evaluation is performed for the bias correction algorithm for different AOD loading. Figure 7 in the revised paper shows the ABI AOD error and standard deviation in different AERONET AOD bins, with equal number of matchup data in each bin. For high quality AOD, bias correction reduces bias in the highest two AOD bins, with center around 0.3 and 0.57. In the range [0.1, 0.3], bias correction over corrects and introduces negative mean bias with slightly larger magnitude than the original mean bias, around 0.01 in magnitude differences. In the range [0,0.1], AOD mean biases are close to zero both before and after correction, but the bias correction AOD error has smaller standard deviation. For the top 2 qualities ABI AOD, bias correction reduces the bias in the whole AOD range with slight over corrections of magnitude of about 0.02 when AOD is greater than 0.1.

**The validation set seems somewhat small (only 6 days of data) and all days have low AOD (<0.1). I thus wonder how the bias and correction algorithm may perform for high AOD cases?**

Those are for case studies. The scatter plots in Figure 5 (in the revised paper) includes all the 5 months matchup data of AERONET sites over CONUS. Figure 7 (in the revised paper) and the corresponding discussion is also added to answer your question.

**Another issue is that the effect of correction is not obvious for top quality data, mostly because the bias data are already removed from top quality (see Figure 1). Is this because these retrievals have high residual error so that they are removed from top quality set? Investigating the reason may offer some clue for the causes of the bias or algorithm improvements.**

A lot of them are due to the relatively large standard deviation of 3x3 box in the 0.47 µm band, which is used in the ABI AOD retrieval algorithm to remove residual cloud contamination with a standard deviation threshold 0.006 for high quality AOD retrieval. This method was adopted from VIIRS retrieval. However, VIIRS retrieval uses a different band 0.41 µm, in which surface reflectance is much lower than 0.47 µm band. As a result, the standard deviation test likely erroneously removes clear pixels with high standard deviation caused by surface. Because the standard deviation information is only available in the intermediate product, which were not archived for long term use, we examined several granules of ABI AOD retrieval from off-line algorithm run and found that 65-80% in medium quality land pixels have standard deviation of 0.47 µm band above the threshold of 0.006.

**Minor comments:**

1. **Section 2.1: What cloud screening scheme is used?**

ABI has a cloud mask product (ABI Cloud Mask ATBD, 2012, [https://www.star.nesdis.noaa.gov/goesr/documents/ATBDs/Baseline/ATBD_GOES-R_Cloud_Mask_v3.0_July%202012.pdf](https://www.star.nesdis.noaa.gov/goesr/documents/ATBDs/Baseline/ATBD_GOES-R_Cloud_Mask_v3.0_July%202012.pdf) , last accessed 5/3/2020), which is used in the ABI AOD retrieval algorithm. In addition, several internal tests are performed to further remove contamination from cloud(ABI AOD ATBD, 2018) : (1) internal cloud test; (2) internal cirrus test; (3) internal inhomogeneity test.

**And which NDVI data is used, MODIS?**

No, not MODIS NDVI. The algorithm uses ABI top of atmosphere reflectance of 0.67 μm and 0.86 μm bands to calculate it: NDVI is defined by red and NIR bands at TOA as

$$\mathrm{NDVI} = \frac{\rho_{0.86}^{\mathrm{TOA}} - \rho_{0.64}^{TOA}}{\rho_{0.86}^{\mathrm{TOA}} + \rho_{0.64}^{TOA}}$$

**2. Section 2.2: Is there any quality control performed on AERONET Level 1.5 data? What is estimated AOD error?**

Level 1.5 AERONET AOD data is cloud screened and quality controlled, with a + 0.02 bias and one sigma uncertainty of 0.02 (Giles et al., 2019).

**3. Line 304, the following reference also points out the poor VIIRS aerosol model selection over China:**

Thanks. We included this reference in the paper

**4. Comparison with PM2.5 seems not very relevant, and removing it does not impair the integrity of the study. There are a lot of factors affecting the AOD-PM2.5 relationship and I think this comparison may complicate the analysis.**

One of the main applications of NOAA AOD product is to operationally derive surface PM2.5 for air quality monitoring and forecasting applications. Arguably, there are many factors that affect the AOD to PM2.5 relationship (aerosol composition, aerosol layer height, relative humidity, time of observation, accuracy of AOD, etc.). It is intuitive that an accurate AOD gives a better estimate of surface PM2.5 given that other factors influencing this relationship the way they are. Therefore, demonstrating that the relationship improves with improved AOD is quite important for our studies and work we do with user community.

**5. Figure 6: could the authors also compare with MODIS to demonstrate the effect of bias correction? The peak of the bias happens at 17UTC, which is 1PM US east time and is close to Aqua overpass.**

The MODIS AOD from Aqua dark target and deep blue algorithm are added (Figure 9 in the revised paper). The bias corrected ABI AOD compares very well with deep blue MODIS AOD in both magnitude and data coverage.

Reference

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmos. Meas. Tech., 12, 169–209, https://doi.org/10.5194/amt-12-169-2019, 2019.