We would like to thank the reviewer 1 for their helpful comments that improved this manuscript. Below in *italics* please, find our replies to the reviewer's comments. Following their comments, we have revised the manuscript as outlined below: (1) We have added two new figures (and related discussion) comparing the two versions over source regions and highly polluted cities; (2) We have addressed all comments raised by the reviewer.

I have three major concerns with this manuscript:
1. While this document would make a very good ATBD for the new OMI product, it is not a very good paper in my opinion. The reason is that most of what is described in section 2 is a summary of what the authors have already published elsewhere, and while it is good to summarize everything in one place for data users, I did not see anything new here. If I missed something and the algorithms as now implemented in the OMI processor deviate from what was published in Vasilkov et al., 2017, Vasilkov et al., 2018 and Qin et al., 2019, then this should be highlighted.

We disagree with those characterizations regarding this manuscript. This is the first version of the OMI NO_2 Standard Product (OMNO2) with global implementation, extensive evaluation, and mission-long processing utilizing 1) a new concept of geometry-dependent surface reflectivity product (GLER) as introduced in Vasilkov et al. (2017); 2) new cloud product first introduced in Vasilkov et al (2018); 3) surface reflectivity product over land discussed in Qin et al (2019) and over water in Fasnacht et al (2019). The first two manuscripts presented conceptual demonstrations and case studies, and the latter two manuscripts provided detailed descriptions and validation of the GLER product at 466 nm. This work, while building on our prior works, expands them to the NO_2 spectral window (centered at 440 nm) as described in the manuscript. In this manuscript, our intension was to initially provide a brief summary of the relevant new developments in a single document as pointed out by the reviewer and then elaborate on additional new implementations that are relevant for NO_2 retrievals. The NO_2 relevant implementations are discussed on pages 10-11 for GLER, pages 13-14 for cloud parameters, and pages 14-15 for treatment over ice/snow surfaces.

2. What data users need to know is how the product changed relative to the last version. Some nice analysis is done on this as shown in Figs. 3 – 6, and I found this very interesting. However, this is only based on one day of data and does not differentiate by region, and I actually took a wrong message from this analysis, namely to expect a very significant (20 – 40 %) and consistent increase in tropospheric NO2 columns, in particular for large NO2 columns. However, as can be seen from Figure 12, this is not universally true, the differences for all Pandora stations been much smaller than what is expected from Figure 6! This is also evident from Fig. 11, where the version differences for Greenbelt have a clear seasonality. I think that the authors should pick a couple of regions representative for different NO2 scenarios (polluted places, very polluted places, biomass burning regions, soil emission regions, lightning regions) and present differences between the two data versions as a function of season as done for Greenbelt in Fig. 11. This would give the reader a much better idea of which changes to expect where and when, and such an analysis should be relatively simple to do. It would also be nice to see an example of BRDF effects on the NO2 columns outside of sun glint regions – this is a major improvement of the new data version and it would be interesting to see if it has a noticeable effect on the tropospheric NO2 columns.

This is indeed an excellent suggestion. We have included 2 additional figures (Figures 11 and 12) comparing V3.1 and V4.0 over source regions, and over highly polluted cities. We have expanded Section 2.5 discussing the impact of the changes as follows:
"Figure 11 shows some examples of how changes in the algorithm from V3.1 to V4.0 affect monthly average tropospheric NO_2 columns over areas affected by various NO_x sources. In contrast to minor changes over the pristine Pacific Ocean, month-to-month changes over source regions vary considerably. The differences in tropospheric NO_2 columns between V4.0 and V3.1 range from -11 to 15% over Beijing, China and from 0 to 29% over the Ruhr area in Germany, suggesting variations in relative differences among cities and industrial areas. The changes over a major biomass burning area of Democratic Republic of Congo, Angola, and Zambia range 13-56% during the biomass burning season of May through August, but are <5% in other months. Differences between the two versions are small over areas influenced by lightning NO_x emissions. In Figure 12, we examine monthly variation of tropospheric NO_2 columns from the two versions over five highly populated and polluted cities that vary in terrain types ranging from coastal (e. g., Shanghai, Tokyo) to mountainous (e.g., Mexico City). NO_2 columns in V4.0 are generally higher than V3.1 by 0-30%, but the difference can occasionally reach up to 50% in some months. Changes of that order of magnitude in highly polluted areas have implications for estimation of NO_x emissions and trends using these data."

3. The authors call this OMI NO2 data version "improved", and I tend to agree that the GLER surface treatment is an improvement over the use of a static reflectivity database not covering angular effects. However, the validation data shown is inconclusive, and to me it looks as if any changes in the product are within the combined uncertainties of retrievals, validation measurements and representation errors. Based on these results, there is little reason to move to the new data version! It would therefore really be nice if the authors could find an example of where the new NO2 product performs clearly better than the last version.
The validation data sets for NO_2 are scarce and are limited in space and time. In addition, validation data have their own issues, such as representativeness error and retrieval issues with Pandora observations and lack of measurements in the lowest few hundred meters in case of aircraft spiral measurements. Therefore, validation of the global product as presented in this manuscript is limited in scope by spatial and temporal coverage and retrieval conditions, and obviously are not representative of other locations and seasons. This is evident from the wide range of variation in results presented in Figures 8-11. We believe that the new results presented in Figures 11 and 12 as suggested by the reviewer have helped address the concerns.

1. Add product version number to title
*We have included version number in the title. The title now reads as "**OMI/Aura Nitrogen Dioxide Standard Product Version 4.0 with Improved Surface and Cloud Treatments**"*

2. line 21: Not sure what the authors refer to by "regional" here – as far as I can see, the improvements presented here are for the global product while the most important improvement for regional products (high resolution a priori NO2 profiles) has not been addressed. I would suggest rephrasing.
Removed "regional" and "on a global scale" from the statement. It now reads as "This version incorporates the most salient improvements for OMI NO_2 products suggested by expert users and

enhances the NO_2 data quality in several ways through improvements to the air mass factors (AMFs) used in the retrieval algorithm."

3. line 24: While the GLER was conceptually new when proposed by Vasilkov et al., 2017, it is not in this manuscript. I would suggest rephrasing.
Removed "a conceptually new," from the statement. It now reads as "The algorithm is based on geometry-dependent surface Lambertian equivalent reflectivity (GLER) operational product that is available on an OMI pixel basis."

4. line 31 / 32: I would hope that all inputs to the AMF scheme are of high quality! I also don't think that a "new NO2 AMF scheme" is presented just because the AMF module reads other inputs. I would suggest rephrasing.
Modified the statement as "The GLER combined with consistently retrieved oxygen dimer (O_2-O_2) absorption-based effective cloud fraction (ECF) and optical centroid pressure (OCP) provide improved information to the new NO_2 AMF calculations"

5. line 36 / 37: Nothing is said in the manuscript on emission and trend analysis of NOx, let alone of other trace gases. I therefore suggest removing this sentence.
Removed.

6. line 43 – 45: I think this sentence fits better to an outreach leaflet than to a scientific paper.
The statement is modified as "The Dutch/Finnish-built Ozone Monitoring Instrument (OMI) has been operating on board the NASA EOS-Aura spacecraft since July 2004 (Levelt et al., 2006, 2018)"

7. line 71: bseen => been
Corrected.

8. line 92: "day-to-day (orbital) variability in surface reflectance" – I find this formulation confusing as in my view, it is not the surface reflectance which is changing from day to day but the viewing geometry which leads to a variation in reflectance at TOA.
Bidirectional Reflection Distribution Function (BRDF) is an inherent property of any surface, but apparent surface reflectance, not just top of atmosphere (TOA) radiance, does depend on sun-view geometry. The statement is correct. However, we modified the statement as follows: "In addition, the OMLER approach neglects significant variabilities inherent of surface bidirectional reflectance resulting from day-to-day (orbital) variation in sun-satellite angles."

62: What was done for SZA > 70 where use of MCD43GF is not recommended?
Thank you for pointing out the data quality issue in the MCD43 product for SZA > 70. In the MCD43GF product, data at high SZA areas are interpolated linearly using retrievals over the same geographical area observed at lower SZAs. Therefore, these data are expected to be of inferior quality, and cautious interpretation is needed. We have clarified this in the revised manuscript.

469: is => are
Done.

474: is => are
Done.

69: delete “retrieved”
Done.

16: Differences in vertical sensitivity – isn’t that already corrected for by the AMF?
Sorry for the confusion. This refers to the difference in vertical sensitivity between satellite and ground-based observations as stated in the manuscript. The vertical sensitivity is accounted for through scattering weights and assumed profile shapes used in the AMF calculations for OMI, but Pandora uses a type of geometric correction as discussed in Herman et al. (2009). The difference in approach is still relevant for the observed difference between OMI and Pandora retrievals.

68: “to relatively OMI’s large pixels” => “to OMI’s relatively large pixels”
Done.

Figure 1: What are the Ps coming from the GLER module?
Ps represents calculated surface pressure over OMI pixel. This is clarified in Figure 1.