Response to RC1 by Reviewer 1:

We would like to thank the reviewer for performing a thorough review and for the many helpful suggestions to improve our manuscript.

Below, we respond to each of the review comments. For the sake of clarity, the review comments are given in blue italics and our response is printed in normal font. Changes to the manuscript are printed in green.

I found this paper to be carefully constructed and well written. The authors provide a clear motivation for the work, and did a good job explaining details that might not be obvious to all readers. For the most part the reader need not be familiar with previous work to understand and follow the discussion in this paper.

Section 1

The authors fail to discuss the version of the TropOMI Level 1B product used in this work until Section 6.3. The version should be cited early in the paper along with the doi of the data. As the authors note, the TOA reflectance changes between product versions so it's important to state key facts early.

We agree and have added the version number of the TROPOMI L1 data to the introduction of the manuscript. The DOI of the L1B data (10.5270/S5P-kb39wni) is now also available as a link.

The text in section 1 of the manuscript now reads:

" For the generation of the DLER database we used TROPOMI level-1b data version 2.1.0 (doi:10.5270/S5P-kb39wni). "

A similar criticism can be made regarding accuracy requirements. The requirements for this work best belong in the introduction where their origin can also be described.

We now mention the accuracy requirements already in the introduction. Their origin is described in the Final Report of the S5p+ Innovation project which was set up by ESA to support the development of several TROPOMI products. We refer to this report in the revised version of the manuscript.

The text in section 1 of the manuscript now reads:

" For the validation study we used accuracy requirements on the DLER of 0.03+10% (0.03 plus 10% of the value, below 500 nm) and 0.02+10% (above 670 nm). These target requirements were taken from the final report of ESA's Sentinel-5p+ Innovation AOD/BRDF project [Litvinov et al., 2022]. "

Use of a DLER product from TropOMI is pretty much limited to satellite observations in a 1330 orbit, possibly a few others if reciprocity is assumed. While there are a significant number of instruments orbiting at this time of day it still limits the application of these results. If the authors were to assume surface BRDF models (e.g. from MODIS) it should be possible to derive a total hemispheric reflectance (THR) from these measurements. A THR product can broaden the reach of these data, allowing a larger pool of potential comparisons including instruments in morning polar orbits and geostationary instruments. The authors have effectively already performed such a comparison between afternoon orbit TropOMI data and the morning orbit MODIS data. Just a thought for a future paper.

The use of the TROPOMI DLER is indeed limited to satellite observations in a 13:30 orbit, or orbits close to that. Using a kernel-based approach such as was used for MODIS BRDF was considered by us in the past, but this is difficult to do properly with observations originating from only one overpass time. For the derivation of the MODIS BRDF product, data from morning (Terra) and afternoon (Aqua) orbits are used, such that the derivation of the kernel coefficients is supported by at least two domains of solar geometries. This is not possible for TROPOMI. Ideas to combine future Sentinel-5/UVNS data (equator crossing time of 09:30 LT) with TROPOMI data exist, but the launch of Sentinel-5 is not foreseen until 2025.

Section 4.5

It would be useful to know how sensitive LER is to the AI screening level as a way of evaluating the chosen threshold. Have the authors performed such a study? Can they provide some more justification for the screening threshold of AI = 2?

The choice for setting the threshold is partly subjective. In the paper introducing the OMI GLER paper [Qin et al., 2019], a threshold of 1.0 is used. The decision to use a more relaxed threshold of 2.0 index points was made by us because we think it is not essential to remove all levels of aerosol, especially not when this is achieved at the cost of a reduction in quality. In other words, filtering out too much will lead to gaps, large uncertainties, or biases which we want to avoid at all costs.

There are other reasons for being conservative. A trend in the AAI values could potentially harm the retrieval or bias it to certain years. The TROPOMI AAI product shows increased values at the edges of the swath. Also, the calibration of the TROPOMI AAI product has changed from version to version, so the AAI threshold to be used depends on the actual version of the TROPOMI AAI product used. It should be mentioned, though, that the latest version of the TROPOMI AAI product (v2.6.0) includes a degradation correction and is better calibrated than earlier versions.

In any case, we performed tests to see which threshold would work best, just by inspecting the fields that were returned. In the end we opted, more or less by coincidence, for the same threshold we use for the GOME-2 surface DLER database (2.0 index points). This particular threshold will remove the strongest cases of aerosol presence, but it will, admittedly, not remove small background levels of aerosol.

A plan for the future is to improve the filtering on aerosol by using AOD information. This filtering on AOD could be performed in addition to the more conservative filtering on AAI.

Sections 4.6 & 4.7

In these two sections the authors are perhaps a bit too reliant on the reader having read and remembered their previous publication on this topic. The reader is left wondering about the general approach. For example, an elaborate method of selecting a representative LER for each grid is described in Section 4.6. No mention is given to view angles, so one must assume that all angles are included in the final distributions. However, a selection of values is then made based on the lowest 10% (or the mode of the distribution in the case of snow/ice). Such a selection is necessarily biased toward view angles where the BRDF is at a minimum, meaning the Section 4.6 LER is dependent on viewing conditions. In Eqn. 8 a quantity ALER is introduced for the first time with no explanation of where it comes from. The reader can deduce from Lines 237-239 that ALER is *the reflectivity of water, which is not a Lambertian quantity except in ideal conditions. How does the LER of Section 4.6 relate to Eqn. 8? Section 4.7 stands out in its need for clearer explanation when compared with the others parts of this paper.*

Thank you for pointing this out to us. Indeed, the text in section 4.6 does not mention the separation in viewing angle regimes that is introduced later in section 4.7. In principle section 4.6 only introduces the method used for the standard non-directional LER. In section 4.7 the DLER approach is introduced, which involves calculating the LER for individual viewing angle segments.

Indeed, the non-directional LER is biassed towards viewing angles for which the BRDF is at a minimum. This is a fundamental property of the non-directional LER as retrieved in the past from instruments like TOMS, GOME-1, OMI, and SCIAMACHY. This is explained in the introduction of the paper. The DLER does not suffer from this limitation, because it takes the angular dependence into account by cutting up the viewing angle range into segments and calculating the LER for each of these viewing angle segments. After this, the directional dependence can be determined.

To clarify the situation, we have decided to change section 4.6, by already mentioning the possibility to cut up the swath in viewing angle segments and making the link with section 4.7, where the DLER is introduced. We also make clear that section 4.6 introduces the traditional, non-directional LER database.

The text in section 4.6 of the manuscript now reads:

" The traditional, non-directional surface LER database is calculated in the following way. For each calendar month, the observations from all available mission years which are considered cloud-free, clouds shadow-free, and aerosol-free by the screening steps are mapped onto a 0.125 by 0.125 degrees latitude/longitude grid. In this step, all viewing angles are accepted, although the code can also be instructed to only take a certain viewing angle range into account. The latter possibility is not used here, but it will be used for the DLER calculation introduced in Sect. 4.7. The distribution of the scene LER values of each grid cell is then analysed … "

Additionally, we reworked section 4.7 to better explain the meaning of equation (8).

The text in section 4.7 of the manuscript now reads:

" … In the retrieval code, the viewing angle range available for this is cut up into nine viewing angle containers and the normal surface LER retrieval introduced in Sect. 4.6 is performed for each of these nine containers. This results in nine surface LER values, which, as a function of viewing angle θ_{v}, are fitted by a third-order polynomial. The DLER can then be parameterised as a function of the viewing angle θ_{v}, as was done in Tilstra et al. [2021], however, with a third order term:

$$A_{\mathrm{DLER}} = A_{\mathrm{LER}} + c_0 + c_1 \cdot \theta_{\mathrm{v}} + c_2 \cdot \theta_{\mathrm{v}}^2 + c_3 \cdot \theta_{\mathrm{v}}^3. \tag{1}$$

In Eq. (8), the directional surface LER A_{DLER} is expressed as an extension on top of the non-directional surface LER A_{LER}. The values of A_{LER} and of the polynomial coefficients c_0, c_1, c_2, and c_3 are contained in the database file. For water surfaces, the polynomial coefficients are set equal to zero. … "

Section 5.4

The authors use a visual example to demonstrate the need for and the effectiveness of their cloud screening *method. In the example shown in Figure 6 it is not immediately obvious that every feature in black is a cloud shadow that should be removed. Shadows at this location should appear to the north north-east of the actual cloud, and it's rather difficult to imagine clouds that could produce some of the shadows seen to the north and north-east of Iceland in Figure 6a. It would be helpful if the authors could include a VIIRS RGB image or a TropOMI reflectivity image for the same time period to show the actual cloud field. There may be clever ways of including this as a transparency overlayed on the minimum reflectivity maps.*

Figure 6 presents retrieved surface LER, so clouds cannot be seen in this figure, because clouds have been filtered out by the two-step cloud filtering described in sections 4.3 and 4.6. As a result, it is impossible to know where the clouds responsible for the cloud shadows were located. It is, therefore, not possible to relate the position of the black features to the position of clouds. What we see are projections of the cloud shadows, not the clouds that produced them.

Moreover, the surface LER field presented in Figure 6 is determined from one month of data, so it is not possible to provide a map of the cloud field.

The cloud shadows in this location are generally located to the north of the clouds. However, because of the parallax effect, they can be found on the east and on the west side of the clouds as well. The parallax effect, caused by the east and west viewing geometries of the TROPOMI instrument, is demonstrated in Figure 9 of Trees et al. [2022]. Apart from the parallax effect, the shape of the cloud also determines the shape of the cloud shadow projected onto the surface.

Figure 6 does, therefore, not tell if the clouds shadows are located at the north, east, or west of the clouds. This cannot be deduced from the shape of the black features.

We have changed the text to clarify that the black features in Figure 6 are caused by cloud shadows. The text in section 5.4 of the manuscript now reads:

" … In the left window, individual cloud shadows can be seen as black features. These cloud shadows are taken into account by the retrieval algorithm which is focused on the lowest scene LER values … "

Section 6.1

The authors should cite the product versions used in all the comparisons described in this section.

We agree and now mention the product versions of the TROPOMI (v2.1), GOME-1 (v1.0), OMI (v3), SCIAMACHY (v2.6), GOME-2 (v4.0), and MODIS (v6.1) surface reflectivity databases in the revised manuscript.

Section 6.2.1, Lines 408-411

The authors cite Rayleigh scattering effects as a reason for not comparing DLER in the UV to the MODIS BRDF. To first order Rayleigh scattering should already be taken into account via the derivation of surface reflectivity given by Eqn. 2. The more important reason for not comparing in the UV is that MODIS makes no measurements at these wavelengths and few, if any, estimates of BRDF exist at these wavelengths. It's worth noting that a UV comparison in conjunction with a long-wave VIS comparison to the MODIS BRDF could provide a useful assessment of how BRDF might change between VIS and the UV.

We mentioned the Rayleigh optical thickness merely to point out that there is considerable multiple scattering for the shorter wavelengths. The increase in multiple scattering is the reason why DLER and BRDF are different below ∼500 nm, as explained in a previous paper by us [Tilstra et al., 2021]. For this reason, we do not attempt to compare the DLER with MODIS BRDF for its shortest (469-nm) wavelength band.

We'd like to point out, though, that we have, in the mean time, added a comparison with the OMI GLER product for this particular wavelength band. This comparison was suggested by Reviewer #2. The results of this comparison are now discussed in the new section 6.3 of the revised manuscript.

Section 6.3, Lines 481, 482

The authors state that TropOMI calibration is much improved in the v2.1 Level 1 product compared to v1.0. Can they provide a reference to substantiate this claim? There exists evidence that the TOA reflectance in the latter version is actually less accurate than in the earlier version. Figure 10d in particular is suggestive of a calibration difference of 2%.

The absolute radiometric calibration was studied by several parties, also by us. Results for the previous version of the L1B data were reported in, for instance, Tilstra et al. [2020]. One of the conclusions was that there is a ∼10% radiometric calibration error for TROPOMI bands 3/4 in version 1.0.0 of the L1B data. We have since then concluded that this calibration error went down to ∼0–2% in the latest version (v2.1.0) of the L1B data.

We were not aware that there exists evidence that the reflectance actually decreased in quality with the release of the latest version. It would be worthwhile to share this information with the TROPOMI calibration team.

References:

Litvinov, P., Chen, C., Dubovik, O., Fuertes, D., Bindreiter, L., Lanzinger, V., de Graaf, M., Tilstra, G., Stammes, P.: S5p+ Innovation AOD/BRDF Final Report, GRASP/KNMI, Issue 1.0, 10 February, available at: https://d37onar3vnbj2y.cloudfront.net/static/surface/albedo/documents/S5p%2BInnovation_AOD_BRDF_Final_Report_v1.1.pdf, 2022.

Qin, W., Fasnacht, Z., Haffner, D., Vasilkov, A., Joiner, J., Krotkov, N., Fisher, B., and Spurr, R.: A geometry-dependent surface Lambertian-equivalent reflectivity product for UV–Vis retrievals – Part 1: Evaluation over land surfaces using measurements from OMI at 466 nm, Atmos. Meas. Tech., 12, 3997–4017, doi:10.5194/amt-12-3997-2019, 2019.

Tilstra, L. G., de Graaf, M., Wang, P., and Stammes, P.: In-orbit Earth reflectance validation of TROPOMI on board the Sentinel-5 Precursor satellite, Atmos. Meas. Tech., 13, 4479–4497, doi:10.5194/amt-13-4479-2020, 2020.

Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth's surface measured by the GOME-2 satellite instruments, Atmos. Meas. Tech., 14, 4219–4238, doi:10.5194/amt-14-4219-2021, 2021.

Trees, V. J. H., Wang, P., Stammes, P., Tilstra, L. G., Donovan, D. P., and Siebesma, A. P.: DARCLOS: a cloud shadow detection algorithm for TROPOMI, Atmos. Meas. Tech., 15, 3121–3140, doi:10.5194/amt-15-3121-2022, 2022.