Response to Reviewer 1:

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

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 wish to thank the authors for a well-written paper. It is structured and organized, making it easy to follow. In most cases the descriptions are precise, so that the meaning is not left to interpretation by the reader. The conclusions are reasonable; the authors do not over-interpret their results.

Following are several issues I would like addressed.

Section 1 The authors state their objective is to evaluate TropOMI radiometric accuracy using the best available surface albedo data sets. This implies they wish to address its absolute accuracy and not merely the TropOMI calibration relative to that of OMI and SCIAMACHY. Given this broader objective there should be an assessment of OMI/SCIA radiometric accuracy, and a discussion of how representative the respective LER data sets are of the underlying calibrations (an accurate calibration does not necessarily mean the LER data sets are equally accurate). Section 5 discusses one deficiency of these LER databases, but it would be better if this paper addresses data set accuracy in Section 3 rather than as only a data screening issue in Section 5.

Indeed, our goal is to assess – as much as possible – the absolute accuracy of the radiometric calibration, and not just to present a comparison relative to that of OMI and SCIAMACHY. This implies the use of a radiative transfer code to minimise the dependence on external information. Nevertheless, this approach still relies on a good knowledge of the input parameters. The most important input parameter in the wavelength range that was studied is surface albedo. The surface LER databases that were used have been validated extensively in the past, with known, citeable accuracies. These are now mentioned in Section 3.4 of the manuscript:

“ The OMI and SCIAMACHY surface LER databases have been compared to each other and to other surface LER databases [Kleipool et al., 2008; Tilstra et al., 2017]. For the OMI surface LER database an overall accuracy of 0.01–0.02 was reported, with slightly increasing values towards the shorter wavelengths [Kleipool et al., 2008]. The SCIAMACHY surface LER database was shown to have an accuracy of about 0.01 for the UV-VIS-NIR spectral range [Tilstra et al., 2017]. These accuracies reflect the uncertainties caused by various error sources, such as errors in the radiometric calibration and the occurrence of cloud contamination in the databases. Errors brought about by the Lambertian-equivalent nature of the databases are not part of these uncertainties. It should also be noted that the OMI and SCIAMACHY surface LER databases are mostly representative for the time periods from which they were derived (OMI: 2005–2009; SCIAMACHY 2002–2012). Systematic changes in surface reflectivity occurring after these time periods, for instance due to changes in land use, are not covered by the databases and will result in errors. ”

Section 3.2 No mention is made of Raman scattering. Should the readers assume that it is not modeled? Since it is an important effect for evaluating the spectral radiometric response of the instrument (as much as 1-2% at TropOMI wavelengths shorter then 400 nm), perhaps it should be stated explicitly.

Raman scattering is not modelled. The DAK RTM that we use does not have the possibility to do so. But, the wavelength bands that are defined in Table 2 have a typical bandwidth of one nm in the UV wavelength range. Raman effects in the TROPOMI reflectance spectra are therefore averaged out partly by the spectral averaging that takes place in the conversion from reflectance spectra to reflectance bands.

This should indeed have been mentioned in of the paper. In the revised manuscript the following text is added to the manuscript, in Section 3.2:

" Raman scattering is not modelled by DAK. However, Raman effects in the TROPOMI reflectance spectra are averaged out partly by the spectral averaging that takes place in the conversion from reflectance spectra to reflectance bands. "

Section 5.1 The information provided in this section is not entirely clear. It is not immediately obvious that the FRESCO false cloud identification over land is caused by an overestimation of surface reflectivity in the underlying databases. Also, the underestimation of land reflectivity, while an established effect, has not been established for these particular databases as their sole or even primary error. In many locations the OMI LER dataset values are higher than other standards, presumably as a result of sub-pixel cloud contamination. I will presume that SCIAMACHY LER suffers similarly.

Yes, this is true. On the one hand, the surface albedo retrieved for the surface LER databases is systematically underestimated for certain geometries, because the algorithms are searching for the minimum scene LER values that are observed but not taking into account the directional dependence of the surface reflection. On the other hand, the surface albedo can be overestimated as a result of cloud contamination in the databases. The OMI and SCIAMACHY surface LER databases both suffer from such issues.

The text in Section 5.1 was rewritten to make it more clear that the FRESCO false cloud identification is caused by inaccuracies in the input surface albedo:

" The deviations in the cloud fractions over land are primarily caused by a systematic underestimation of the surface reflectivity for certain geometries in the (Lambertian) surface albedo database that is used in the FRESCO algorithm. This explanation is in line with conclusions from a recent study by Lorente et al. [2018], who showed that traditional Lambertian surface albedo databases, like the ones described in . . . "

Section 5.2 The alternative methods presented in this section appear to be designed to eliminate scenes for which the LER databases are affected by non-Lambertian surface characteristics. Since these errors are as much a function of viewing conditions as they are surface type, isn't the TropOMI LER subject to similar errors? Is it legitimate to screen the reference data and not the measured data as well?

It would not be legitimate, but it is not done that way. The filtering is performed on a box-to-box basis. That is, if the input surface albedo for a one-by-one degree box is rejected, then so is the associated TROPOMI observation box. The measured and reference data are therefore subjected to the same screening mask.

Section 6.2 It is not clear where the referenced 772 nm slope and intercept are coming from. Is it Figure 6 or Figure 8?

The analysis that is used to determine the average slopes and intercepts is visualised in Figure 8, but not for the 772-nm wavelength band.

This is indeed confusing, so we have changed the manuscript. We now also mention the result for 758 nm in the text, because this wavelength band is shown in Figure 8. We still mention the 772-nm results. Mentioning the 772-nm wavelength band is important, because it shows the largest deviation. In the manuscript we have added the comment that the 772-nm results are not shown in Figure 8. The changes to the manuscript are:

" … The results were obtained for all 21 wavelength bands. In Fig. 8 the results are presented for 5 of the wavelength bands. … "

" … At 758 nm the average slope is larger than one ($\overline{m} = 1.015 \pm 0.008$) and the intercept is small ($\overline{n} = 0.004 \pm 0.003$). At 772 nm (not shown in Fig. 8) the average slope has increased further ($\overline{m} = 1.033 \pm 0.008$) and the intercept is now very close to zero ($\overline{n} = 0.003 \pm 0.002$) … "

Changes to the manuscript:

During the review phase of this manuscript the TROPOMI instrument kept generating new data and we took the opportunity to extend the time range that was studied originally (May 2018–February 2020) by three months. The studied time period is now May 2018–May 2020, covering two years. Some of the numbers reported in the paper have changed slightly, but not significantly. Figure 8 and Table 3 have also been updated. The changes w.r.t. the previous version are in all cases insignificant, well within the reported accuracies. The extension of the time range that was studied has led to an increased accuracy of the results. Figure 3 was also updated.

References:

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res., 113, D18308, doi:10.1029/2008JD010290, 2008.

Lorente, A., Boersma, K. F., Stammes, P., Tilstra, L. G., Richter, A., Yu, H., Kharbouche, S., and Muller, J.-P.: The importance of surface reflectance anisotropy for cloud and NO2 retrievals from GOME-2 and OMI, Atmos. Meas. Tech., 11, 4509–4529, doi:10.5194/amt-11-4509-2018, 2018.

Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, J. Geophys. Res.-Atmos., 122, 4084–4111, doi:10.1002/2016JD025940, 2017.