Comment on amt-2020-502
Ruediger Lang (Referee)

The paper by Tilstra et al., *Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth’s surface measured by the GOME-2 satellite instruments*, presents the next evolution of the Lambertian-equivalent reflectivity (LER) surface databases, as derived from high spectral resolution grating spectrometers covering the UV, visible and towards the near and short-wave infrared spectral region. LER surface databases derived from such instruments have the advantage to provide their data at significantly more atmospheric window or well controlled absorption wavelengths, i.e. at higher spectral resolution, than the familiar surface databases derived from band imagers like AVHRR, MODIS, Meteosat, or Sentinel-3. However, up to now, the original LER approach assumed homogenous, non-directional reflection of the surface, which is known to lead to significant biases in particular in the backscatter direction.

The directional evolution of the LER surface retrieval approach (DLER), applied to the meanwhile considerable GOME-2 data record of more than 10 years from two Metop platforms at 9:30 LT and over an observation angle range of -55 to 55 degrees, is therefore a very significant improvement to the currently existing (and frequently used) LER databases (Tilstra et al., 2017). The results show that the anisotropy is considerable depending on surface types (in particular for vegetation), as is expected, and that at least for such surface types and at large observation angles the previously used LER databases introduces significant biases.

The paper presents a comparison to the principally more accurate bi-directional reflectance distribution function (BRDF) approach, e.g. as applied to MODIS observations, which however is limited in its available spectral resolution. The comparison of synthetic data from radiative transfer calculations shows a good correspondence between the two approaches above 500 nm improving towards longer wavelength. A validation comparing BRDF reflectivity values with DLER values for the MODIS 640 nm band confirms the significant improvements of DLER with respect to LER in the backscatter regime (West-viewing for GOME-2 daylight descending orbits).
The scientific results presented here are significant and will be of high interest to users of grating spectrometer data in the UV to near infrared. The paper is well written and I can therefore recommend it for publication in AMT, noting a couple of aspects for the authors to consider.

One of the main advantages using DLER (and LER) with respect to imager derived BRDF databases is its higher spectral resolution. While the comparison of BRDF and DLER values derived from synthetic Top-Of-Atmosphere (TOA) data identifies the spectral regime in which both perform similar and where not, the validation results of section 7 provides results only at 640 nm. A tabulated statistics of slope, intercepts and correlation wavelength at other MODIS wavelength (in particular towards the blue) would be very helpful for users to decide where to use or not to use DLER for their applications. In particular, since close to 50% of the provided DLER wavelength are in the <500 nm regime. In this respect, a comparison of DLER performance with respect to the frequently used combination of MODIS BRDF and spectral principle components provided by the ESA ADAM surface reflectance database would be of value for follow on studies.

Also the paper is only discussing in passing DLER results over persistently snow covered (high mountains and polar regions) regions, and is not discussing ocean surfaces (or/and water bodies in general) at all. Both surface types are either missing or filtered out in BRDF land databases like the ones derived from MODIS, because of considerable uncertainties in the BRDF coefficients for snow surfaces so far, or are generally neglected (like ocean colour variation and potentially associated directional effects apart from glint). Appendix B seems to indicate that DLER (like the previous LER database) also provides values over oceans, although this is never explicitly mentioned or even discussed in the body text of the paper, as it seems. It would surely be very interesting to understand how well DLER performs for these two surface types, which are (or seem) both included in the discussed database.

Finally, it is not very clear to me why in the “Case studies” part of the validation section (Section 7.1) the authors emphasize the need for focussing on largely homogenous surfaces. A proper averaging of MODIS BRDF sub-pixels to the DLER grid pixel should in principle provide an accurate comparison independent of sub-pixel surface variations. And it would be also interesting to provide the corresponding averaged MODIS BRDF results in Figure 8 for comparison with the DLER grid pixel results along with the individual ones (and ideally show similar comparisons for non-homogeneous cases too).