The paper by Kleipool et al. on the “OMI Collection 4: establishing a 17-year long series of detrended L1b data” details improved methodologies (with respect to earlier updates to the processing, in particular collection 3) applied to the processing of OMI level-0 to 1b measurements, including the correction of sensor degradation and calibration key-data deficiencies. In particular, low level, but important aspects like detector pixel response flagging has been improved and streamlined in collection 4, in particular to what is currently done for the Sentinel-5p mission. It is however expected that future missions like Sentinel-4, 5 and CO2M will take benefit of the presented approaches used here. Also because more harmonised approaches in these aspects will likely improve the understanding and will encourage the use of the data (and corresponding error flagging) by the users of level-1b data of these missions.

The paper is well written and clearly structured, and provides the latest version of the OMI Algorithm theoretical baseline Document (ATBD) for OMI level-0 to 1b processing as a supplement material. The latter is appreciated but also provides some challenges for non-expert readers of the paper (see below).

Next to the important aspects of detector pixel level performance monitoring and flagging, Kleipool et al. also present their methodology to address and correct the observed long-term degradation of Solar irradiance and Earthshine radiance signal levels (as expected for these type of sensors over such long time scales, in particular towards the UV). The solar irradiance degradation is larger than the Earthshine degradation due to the degradation of the involved solar diffusers and additional mirrors in the optical path of the solar port. The authors split the correction in the multi-dimensional change of the be-directional scattering distribution function (BSDF) for different illumination conditions normalised to certain reference angles. Separating it from the long-term degradation of the absolute...
irradiance levels. This approach also corrects expected deficiencies in the on-ground key-data of the BSDF because of the complexity of such measurements to be carried out on ground.

While this approach has also been used for other missions with this type of diffusers, here the BSDF changes are also evaluated as a function of time, which clearly improves the quality of the long-term time series.

For the correction of the Earthshine part the authors apply a “stable ground target” approach, also used by other missions for this purpose, where the target surface reflectance can be expected to be stable over the year and atmospheric variability is not too large. The choice of the target by the authors is snow/ice surfaces over Antarctica. While those surfaces should be quite stable (although snow BRDF function can be changing in a complex way as a function of temperature and solar illumination conditions) I am wondering if this is actually a good choice for a mission where ozone is contribution to a significant extend to the spectral variation of the measurements, in particular below 350 nm. Variability of Ozone is very large over the year in Antarctica, and arguably much more significant than at mid-latitudes. While in the latter case line absorber variability is larger (and stronger) like water vapour, these are usually covering only a small subsection of the spectrum and can therefor much better be filtered out. So I would have considered the Libyan desert being a better target, with an even more stable surface over the year (and well characterised), and less interference by ozone variability.

I particular, and as a consequence of the strong interference and variability of ozone below 335 nm, a correction of the radiances in this important region (with many level-2 products derived from this part of the spectrum) based on actual measurements, has not been carried out. Instead it has been assumed that the degradation is spectrally neutral for the Earthshine port, so can be based on the degradation coefficients derived in the region between 335 to 360 nm for band 2 and 390 to 500 nm in band 3. However, the exact regions considered usable and used for Earthshine degradation evaluation for target area measurements (and extended across the full spectrum I guess) are not explicitly stated, since other spectral regions are suffering from atmospheric absorption features, Fraunhofer lines and interference of a dichroic.

I consider the assumption on spectral neutrality a critical one and I find it has not been addressed in full by the authors. The results presented for Earthshine port correction could potentially be significantly biased because of this assumption. On the other hand, the results derived from the AU1 and AU2 diffusers, which indicate that the spectrometer and the detector assembly’s contribution to the degradation seems to be indeed spectrally quite neutral (and there can be physical arguments also made for such an observation) have not been explicitly applied to support the hypothesis, e.g. by comparing it to the observed degradation in the 335 to 360 nm region and make some interference from such comparison.

But most important I think the first mirror, which seems to be bypassed by the solar port optical path, cannot simply be ignored, in particular in the case that the region below
335nm is not addressed directly by Earthshine observations. Obviously any mirror in the light path could exhibit relative spectral neutrality in its degradation in the visible while exhibiting a strong spectral dependence in its degradation for shorter wavelength.

Acknowledging the fundamental difficulty in assessing the Earthshine port degradation in this shortwave spectral region, while at the same time also acknowledging the larger number of users of collection 4 data using particular this spectral region, I would strongly recommend to include some (at least initial) analysis applying level-2 retrievals, or applying (ozone) cross-section spectral dependency information to support the assumption on spectral neutrality.

The paper is very well suited for a publication in AMT and of high significance not only to future users of OMI level-1b data but also to users and developers of re-processed collections of current mission and future mission level-1b data processing for instruments of that type. I therefore can highly recommend it for publication providing the authors can address the issues raised before and in the specific comment sections below.

Specific comments:

Section 3.5: "In addition, a static irradiance measurement used over a 17 year mission ignores the subtle changes in the solar output, an effect that could enter the L2 products in the long term." Can we really assume that the solar variability over a timescale of 17 years is negligible in terms of signal variation observed (in particular in the UV)?

Section 4.5 on flagging: Can you confirm then that a pixel qualification using originally 31 categories have been mapped down to 3 – and finally to only 1 in the end – with RTS being separated out? Was this mapping unique or were there some ambiguities to overcome?

Section 5.3 on relative irradiance: I would consider it clearer for the reader to talk about the diffuser BSDF – after having properly defined it – and its correction (which changes over time as a function of azimuth angle, elevation and time). So I would consider to replace "relative irradiance" with "diffuser BSDF" variation/correction.

Section 5.4: The choice of the normalisation point is an extremely sensitive and delicate matter for deriving such a degradation correction. First of all because data at the day of the launch (as “start of the mission”) cannot be used. But second also since the selected normalisation point (and its inherent biases) can significantly amplify biases in the normalised time series of correction coefficients for later periods.
So what is characterized here as start of the mission? Ideally this should be the first irradiance measurement of the instrument, which can be solidly and fully calibrated (irrespective of commissioning periods or SIOV).

On the other hand, various normalisation spectra should be tested to find out the sensitivity of the choice of the reference spectrum on the degradation correction coefficients. Has such a sensitivity test been carried out?

Again I find the assumption on a completely stable sun over a 17-year period a bit tricky without further qualifications, in particular in the UV.

Section 5.5: There seems to be a systematic dependency (at first order) of the degradation over the rows with higher degradation for the middle rows and lower for the edges. Is this potentially a systematic effect?

Section 6.2.2. On the wavelength temperature correction: I would assume that the dependency of the dispersion on OPB has been measured on-ground. How do the results obtained in this study compare to the on-ground measurement temperature dependency of the spectral calibration stability?

Section 6.5: On the “transient” signal flagging. How often are pixels flagged for this "transient events"? Can some statistics be provided, and are these events unknown in their nature/origin, and therefore not categorized as any of the pixel effects before? Only in the next section it becomes clear that cosmic particle impact is one of those transient effect. So a list of potential causes would set the scene here.

Section 6.6: High latitudes are of course also very significant regions of cosmic particle impact. Here only the (important) SAA area and its evolution is shown and discussed. I would assume that transient effects also accumulate and are accounted for at a global scale (ie including high latitudes). Can you confirm?

**Editorial comments**

Generally, I think it would be very helpful to point to specific section in the supplement (ATBD) document, which is referred to throughout the paper at a time. This will help the reader (in particular the not so expert ones) to find their way through the vast amount of supplemental information provided in the ATBD (naturally not all relevant to the scope of this paper).

Generally, on figure captions: Captions often refer to top/down panels where there are only left/right panels
Section 2.1: OB, or OBP or OPB?

Section 2.1, l. 132: “it was observed that the duty cycle of the PWM of the UV detector heater occasionally dropped to zero...”. Can a concrete date be added to this?

Section 3.3: It might be worth mentioning in this context that product format porting and restructuring to netcdf is part of a wider effort to streamline the product format of instrument of that type (GOME, SCIA, GOME-2, S5p, S5 and the future S7) and with the same AC and CLIM community in terms of output format (netcdf CF-standard)

Section: 4.1, 1st paragraph: So what is the correct value then? Since only the “erroneous” conversion is reported.

Section 4.2, l. 282: Check sentence.

Caption Figure 19: The terminology difference between "terrain height" and "surface altitude" is nowhere explained.