Review comment amt-2021-430-RC3

Reviewer: Ruediger Lang

Dear Ruediger,
Thank you for your detailed review of our article. Our responses to your remarks, questions and considerations can be found below. The responses also include the planned actions for the revised manuscript.

General comment

"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."

Response to general comment

We thank the reviewer for his thoughtful commentary on our approach to long-term radiometric calibration of OMI, and rather agree with his conclusion that the uncertainty in our assumptions has not been fully addressed. The reviewer is correct that all scene stabilization techniques in the UV are limited due to ozone absorption. This problem is universal, though it is more pronounced in high latitude ice surfaces due to the increased atmospheric path length. The high degree of predictability of such ice surfaces counterbalances the increased uncertainty due to ozone. Jaross et al. 2008 establishes a 2% absolute uncertainty in the technique for wavelengths as short as ~330 nm. Because we include ozone absorption in our atmospheric model this uncertainty is mostly a result of uncertainty in the surface BRDF and not caused by ozone variability. All ozone instruments for which we have used this technique derive ozone concentrations using wavelengths shortward of 330 nm with much larger ozone sensitivity. The ozone-related error at 330 nm is therefore a second-order effect.

Furthermore, we are not attempting to establish or verify an absolute radiometric calibration, rather a time-dependent calibration. The viewing conditions over Antarctica for a stable orbit are very repeatable each solstice. The ozone overburden is mostly repeatable from year to year with variability becoming a source of noise over a 15+ year record. There is of course an error related to long-term secular change in ozone. The drift error at 330 nm is related only to the mean change in total column ozone over the OMI record multiplied by the ozone sensitivity ratio between the shorter ozone-absorbed wavelength (used in retrievals) and 330 nm, approximately 10x. And 330 nm is merely the shortest wavelength we consider. As indicated in Section 5.5, our evaluation of the wavelength-dependent degradation is based on 340 nm and longer.

We do not claim the OMI radiometric changes are spectrally flat. This is rather unlikely. Instead, we are saying that the wavelength-dependent response change determined through the ice radiance technique is statistically consistent with zero (see Fig. 10). We will add a quantitative assessment of Fig. 10 to demonstrate this. As the reviewer correctly observes, the ALU1 change shown in Fig. 4 provides strong circumstantial evidence for spectrally flat sensor degradation. The small change seen in that figure is primarily a result of folding mirror degradation, an element not present in the Earth radiance measurements. An unfortunate consequence of the OMI design is the lack of a means to directly measure sensor change that includes the primary telescope mirror. What we can do is argue that this mirror's optical degradation as a result of photopolymerization is likely much less than that of the folding mirror because the latter sees a significant number of photons shortward of 300 nm and the former does not. If we use the ALU1 change as an upper bound to optical degradation at ozone-absorbed wavelengths, the resulting change in the measurement vector used in total column ozone retrievals is less than 0.1% over the mission (1% change between 250 nm and 500 nm; 317/331 nm pair used for ozone retrievals). We will include these additional points in the Section 5.5 discussion.

Response to specific comments

Item	Referee comment	Author’s response
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)?	We will add a statement on the stability of the Sun relative to our analysis method. The analysis was based (smoothed) over the entire (continuum) wavelength range and not based on variable Fraunhofer lines.
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?	The original 32 possible categories was an approach which did not work very well and was not user friendly. We indeed reduced the criteria that then result in one flag.
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.	On the fully integrated instrument, the diffuser BSDF cannot be measured separately from the rest of the optics and electronics. This is why we chose this terminology. We will consider alternate names for this sensitivity
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.	We will add which orbit we have used as start of the mission. It is an orbit at nominal operational temperature after the thermal testing at the beginning of the mission. We will address the question of the Sun’s stability relative to our analysis method. We have not performed a sensitivity analysis as such. However, per year references were used and together with the binning in wavelength dimension the random fault of a specific measurement should not translate into a bias which is significant compared to the uncertainty associated with the absolute radiometric calibration from on-ground calibration.

Item	Referee comment	Author’s response
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?	Yes, this is what we observe. The higher degradation might be related to the row anomaly and caused by higher exposure.
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?	As far as we are aware no usable temperature dependence measurements from on-ground are available. Collection 3 already used in-flight measurements for this.
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.	Agreed, we can change the order of the text here and provide some typical numbers.
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?	This is correct, the transient flagging is performed for all measurements. The SAA is the most likely region for transients and is therefore flagged separately (even if no transient is detected for a specific pixel).

Response to editorial comments

Item	Referee comment	Author’s response
All	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).	We will point to the relevant sections in the ATBD.
All	Generally, on figure captions: Captions often refer to top/down panels where there are only left/right panels	This will be corrected.
Section 2.1	OB, or OBP or OPB?	It should be OPB. We will check that this abbreviation is used consistently.

Item	Referee comment	Author's response
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?	We will add the date (beginning of October 2017).
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)	We will add a comment on the effort to streamline product formats.
Section: 4.1	So what is the correct value then? Since only the "erroneous" conversion is reported.	We will add an explanation.
Section 4.2, l. 282	Check sentence	We will rewrite this and the following sentence.
Figure 19	The terminology difference between "terrain height" and "surface altitude" is nowhere explained.	Will add an explanation. In both cases it's the height with respect to the reference geoide.