Response to anonymous Referee #2

Referee comment	Author's response	Proposed adaptation
General Comment 1: This paper is too long. There is a reason that scholarly journals restrict paper lengths to 15 pages, 20 pages at the most. That is because doing so forces the authors to avoid excessive detail and to summarize their findings in a way that helps the reader understand what was performed and what was concluded. The specific details of the TROPOMI analysis are of little benefit to readers outside the TROPOMI instrument team. No one will attempt to repeat the steps outlined here, so it seems these are included here as a substitute for an internal team report. It is important to describe problems and the general techniques used to address those problems, but by including too much detail the authors fail to provide a useful summary to the readers.	We thank the referee for its thorough review and comments. We understand the paper seems unusual long, however, there is a reason for this. The original documentation can never be made publically available due to their proprietary nature. This means that this is the only occasion for the calibration team to report on the results obtained, and how and why some choices were made. We went to great length to reduce the contents of the calibration analysis documentation from 1200 pages to a single paper. We feel that for such an important mission the length of this paper is justified. The level of detail has been reduced so far as possible to remain useful for user of TROPOMI data, and calibration experts of future missions.	We will try to shorten certain sections a bit further, also to make room for some additions requested by the other reviewer.
General Comment 2: The sections dealing with electronics and with spectral characteristics are well organized and written. The same cannot be said for the sections about radiometric response. These sections would benefit from some hierarchy in the discussions. As it is, the reader is presented with too much detail and not enough overview. What is the calibration philosophy/approach? Why were the measurements performed in the manner they were? Why were the characterized parameters chosen the way they were? These sections could also use more critical evaluation of the results. Do the results make sense? Are the validations sufficient to give us confidence in the error estimates?	We also agree that the radiometric section would benefit from balancing the different topics and parts within a topic.	We will take your questions and suggestions along when restructuring the radiometric part.

General technical 1: Many of the plots lack axis labels, and some do not even have a description of the axes in the caption. Reference to detector "columns" and "rows" is ubiquitous, and should be replaced more generally with "spectral" or "spatial" dimension.	agreed	We will recheck all figure and captions.

Referee comment	**Author's response**	**Proposed adaptation**
Technical Comment 1: Page 1, Line 20: I don't understand the sentence starting "In case : : :" The way this is written implies that there will not be a product problem if random errors are larger than systematic errors. I don't think the authors mean to say this, so I advise a different choice of words. Or simply delete this sentence, because I don't see its relevance in the abstract. The abstract should highlight key points of the paper, and this sentence does not seem to fit that objective	agreed	We will remove the sentence.
Technical Comment 2: Page 1, Line 39: I don't understand the term "In-compliance." Do the authors mean non-compliant	We assume you mean line 19?	Will change to 'not compliant'
Technical Comment 3: Page 11, Line 3: I don't agree with this description of full-well. Typically, an immediate flattening of the linearity curve indicates register full-well rather than pixel full-well. When the latter occurs it appears as a sharp curve, but over a finite range of integration times. To me, the term "immediate" implies a slope discontinuity in the linearity curve	We assume you mean page 17, line 1. For TROPOMI, the register full well capacity is about three times larger than the pixel full well capacity as shown in table 6. See also our response to referee #1 Specific comment 5	We will rephrase the word 'immediate flattening'
Technical Comment 4: Section 6.1: The abbreviation ISRF is not defined until later in the paper	agreed	Will be added.
Technical Comment 5: Section 6.2: This discussion is confusing, and could be clarified by better defining	agreed	We will explicitly define the different straylight terms as used

terminology. The authors use the terms in-field, in-band, out-of-field, and far-field but don't clearly explain what stray light falls into each category. This is important because the choice of terms contradicts common definitions used elsewhere. Words like "band" and "range" have subjective interpretations if left undefined. It might be simpler to use the terms spectral and spatial stray light. A schematic or detector image might help to clarify the definitions. From the section title I assume this section pertains to spatial stray light, yet other characterizations are described such as out-of-spectral range.		for TROPOMI.
Technical Comment 6: Where are the detailed descriptions of measurements? This section deserves the same level of detail as Section 6.3 has. Spatial stray light can be rather difficult to characterize, especially when the instrument is looking out of a chamber through a window. How do you know what portion of the measured SL is contributed by setup and OGSE?	We have calibrated the out-of-field straylight, and described it on page 26, line 7.	We can add some additional description of the relevant measurements in section 6.2. And also include the commissioning of the setup to address setup straylight.
Technical Comment 7: Telescope SL is also the simplest of stray light components because it is driven almost entirely by the roughness of the telescope mirrors. Therefore, it is straightforward to model this SL. Have the authors done this as a way to validate their in-band measurements?	Modelling of straylight was partially done for certain components by industrial parties. For the L01b processing this is not sufficient because it requires the total straylight response of the integrated instrument as build and not as designed.	
Technical Comment 8: The parameters v and w are poorly defined. It sounds like one is spectral and the other spatial, but I cannot tell which is which. This is important for Fig. 14 because the spatial dimension will show the slit image (the telescope stray light) as a stripe illuminating all rows at the source's wavelength. A similar stripe in the spectral dimension can be an indication of a grating defect	agreed	We will add the definition.
Technical Comment 9: The abbreviation PRF is not defined until later in the paper	agreed	We will include the abbreviation.

Technical Comment 10: The hole-in-cloud measurement and validation seem to ignore spectral stray light. How is spectral stray light characterized and how is it validated? Past experience with imaging spectrometers has shown that spectral stray light is much more important to science products than is spatial stray light.	We agree that the hole-in-cloud does not provide information on spectral straylight. The calibration showed that the straylight is dominated by near-field, which has both a spectral and a spatial component. This was measured with a laser source and is the basis of the current straylight correction in the L01b processor. Spectral ghost were sufficient small to not be corrected.	We will add our response to the text.
Technical Comment 11: Section 6.3: This type of spectral stray light is more commonly referred to as out-of-range because it is beyond the measurement range of the instrument. Rather than describing a distinct characteristic of the instrument, as is done with other sections, this one describes a separate measurement campaign. This is confusing, but if the authors feel this needs to be done they should do a better job reconciling this discussion with that of Section 6.2. For example, the authors describe in-band measurements as part of this campaign. Such in-band measurements were also part of the discussion in Section 6.2. Were these the same measurements or different ones. If different, how do they compare? Why was one technique chosen versus the other? Also, the depth of discussion in this section is in direct contradiction to that of Section 6.2. Section 6.2 has too little description of the measurements and analysis, but Section 6.3 has maybe too much.	This section indeed describes a different measurement campaign and therefore deserves a different treatment. However, we agree that more effort can be put in the consolidation with section 6.2	We will adapt both section in line with your comments.
Technical Comment 12: Page 18, Line 65: The terms in this equation are not defined	Which line number do you mean exactly?	We will recheck all equations nonetheless.
Technical Comment 13: Figure 16 requires more explanation	agreed	We will expand the caption.
Technical Comment 14: Section 6.4: This section contains multiple subsections, each describing a step in	agreed	We will restructure and shorten the section.

the data reduction. Lacking is a description that ties all these steps together. Why are each of these corrections necessary? Why is it important to separate the radiometric response into low and high frequency components		
Technical Comment 15: The Figure 20 caption is incomplete. What source are we looking at?	Agreed	We will update the caption
Technical Comment 16: Section 6.5: The distinction between ABSRAD and RELRAD is confusing. The authors provide a clear description in Page 24, Lines 4-10. However, Fig. 26 appears to be a combination of ABSRAD and RELRAD, even though the caption talks only of ABSRAD. Furthermore, the BSDF discussion in Section 6.7 is clear about using only ABSRAD, yet Fig. 31 contains row dependence. Does ABSRAD contain RELRAD or not	Figure 26 is not a combination; we understand the confusion however, and will explain better in the text. In the BSDF discussion ABSRAD is normalized with ABSIRR, which has a row dependence. RELRAD does not enter this equation. We will clarify.	We will clarify the text.
Technical Comment 17: Page 27, Lines 3,4: Doesn't this caveat invalidate the distance offset approach the authors are describing? No stray light estimates are provided to prevent the reader from drawing this conclusion	The line numbers the referee uses do not seem to match the single column manuscript online, but appear to come from a double column version. Is this correct? It makes tracking comments rather difficult. Please clarify which sentence you are commenting on, so we can respond.	
Technical Comment 18: Section 6.6: This section contains only a brief mention of diffuser feature smoothing. Other than that, there is no discussion of fitting data or separation of high and low frequency components, so the reader must assume this was not undertaken. How is this reconciled with the exhaustive analysis described in Sections 6.4, 6.5 for radiance? Aren't many of the radiance artifacts also present in the irradiance data	The derivation of ABSIRR is indeed less complicated than RELRAD. The latter needs stitching of multiple measurements, and onwards a separation into RELRAD and PRNU without smoothing. For ABSIRR no separation is needed, only a smoothing to remove diffuser features due to speckle.	
Technical Comment 19: Section 6.7: Given its importance to Level 2 products (as the authors note in lines 39, 40), this should be the primary radiometric	In section 6.7, second paragraph we explain that we would rather have measured the BSDF as a primary calibration parameter, and then to	We will clarify the section and add a figure.

description of the paper, yet it appears to be presented only as a validation. Why was so much time and effort placed on the radiance calibration, such as described in Section 6.4 and 6.5, but no effort to ensure that the BSDF calibration is smooth and represents the expected characteristics of the diffusers? The approach taken seems backward, since a smooth, physical BSDF is more important than artifact-free radiances alone. For instance, can the authors explain why the spectral dependence of BSDF has the unusual shapes exhibited in UV and UVIS? And why does it have the structure shown in SWIR? How does the derived BSDF compare to the QVD BRDF	use in onwards with ABSRAD to yield ABSIRR. The direct BSDF measurement was not possible due to stimulus failure. We were onwards forced to recover by taking the backwards approach of using the FEL lamps. We did check whether the resulting BSDF was artifact-free. On request from reviewer #1 we will already add a figure addressing the smoothness of the BSDF.	
Technical Comment 20: Page 28, Lines 57, 58: What do these numbers mean and where do they come from? They contradict Figures 30, 31	We cannot find this due to the line numbering problem. Please clarify so we can respond.	
Technical Comment 21: Page 38, Lines 71-79 Can the authors speculate why the Earth port and sun port wavelength registration yields significantly different results? This is an unexpected result, is it not	The accuracy of the measurements is about 2/3rd of the observed difference between the earth and sun port. Theoretically they should be the same. We cannot determine whether this is significant or not, and will address this after commissioning and report in a future paper.	