Response to anonymous Referee #1

Referee comment	Author's response	Proposed adaptation
General comment: The paper is well written and of good quality, with a considerable number of new interesting topics and techniques, and shall certainly be published. However, I am of the opinion that the quality of the paper can be much improved to be more useful with a comparatively small additional effort, in line with the comments and suggestions provided below. After these comments and suggestions have been adequately addressed, the paper shall certainly be published.	We thank the referee for its thorough review, and hope that our proposed changes will address the comments.	
Comment 1: On page 1, lines 11-14, a new and promising methodology is introduced in the abstract to quantify residual uncertainties/errors at L1b after 0-1b correction. This point also comes back to some extent in the conclusion section 9. This methodology is to me one of the new important and interesting aspects described in this paper. This methodology can be applied to individual correction factors, as also mentioned in the paper. However, the methodology is not always used consistently throughout the paper, and results of applying this methodology for individual parameters and corrections are not always clearly shown. I feel that the quality of the paper can be improved by improving these aspects and perhaps showing/discussing more results of applying this methodology.	We agree that this new methodology is highly interesting, and we have demonstrated its benefits in a few examples in the paper. We would have liked to show all results, but this would make the paper excessively long. Especially because approximately half of the analysis work on onground calibration went into validation and verification using this method. Thus reporting on these as well would make the paper too long.	We propose to extent the validation and verification analysis on a few extra topics, namely: electronic non-linearity and PRNU. Furthermore we can add a few lines at each section identifying additional validation performed.
Comment 2: The paper discusses the TROPOMI calibration. However, I am of the opinion that the paper would benefit from (briefly) describing a number of critical performance parameters such as signal-to-noise ratio as function of wavelength (for low albedo scenes), spectral/spatial features (from diffusers, coatings, polarisation scrambler, etc.) and polarisation behaviour, even when these parameters are not direct calibration	agreed	We will update table 3 and the instrument overview with additional parameters.

parameters used directly in 0-1b data processing.		
Comment 3: The title of the paper suggests that the full TROPOMI calibration is described. However, for many parameters the paper focuses on the UV-VIS-NIR spectral range, not on the SWIR wavelength range (there are some exceptions). I propose that the title of the paper is changed to refer to UVVIS-NIR (preferred), or that a clear reference is given to the remaining parts for the SWIR calibration parameters. See also the examples provided below.	This paper covers the calibration of the entire TROPOMI instrument, with the exception of the SWIR detector characterization [Hoogeveen 2013], the SWIR straylight correction [Tol 2017] and SWIR ISRF [van Hees 2017]. All other SWIR calibrations are part of the work presented in this paper (PRNU, RELRAD, ABSRAD, ABSIRR, RELIRR, BSDF, LOS, PRF…). We therefore feel that the title is justified, and propose to leave it as is.	We will update all tables to include the numbers for the SWIR channel as derived in the mentioned references.
Comment 4: Some more comparisons with respect to realistic earth atmosphere low-albedo scenes and signals within absorption peaks shall be presented and included for quantifying stray light at L0 and L1b.	Unfortunately, we cannot do this with the data available; measuring realistic earth scenes (e.g. zenith sky measurements) was not feasible during onground calibration. Therefore we were forced to restrict the analysis to establishing compliancy with the requirements. These requirements were formulated as the hole-in-the-cloud scene, the closest similarity we can achieve is the scene constructed from EWLS measurements.	We can add some extra detail on why and how the EWLS hole-in-the-cloud validation scene was created and used.
Comment 5: The radiometric error budgets presented in table 9 seem somewhat unbalanced / unjustified and in some cases too optimistic. The error budgets in table shall be justified or modified in line with the comments provided below.	We can see that this is unclear. The numbers in the table refer to the error in the calibration key data *only*. This error is used in the L01b processor to propagate the total error in the L1b products Radiance and Irradiance. Because the end-user is mostly interested in Reflectance, we have excluded errors (identified with an asterisk) from the CKD as they will cancel out when calculating the Reflectance.	We will adjust the text in the relevant sections to clarify this. It is clear that some extra explanation is needed how the final error in the L1b products is calculated and handled; we will add a paragraph on this.
Comment 6: The intra-band and inter-band co-registration errors don't seem to make sense in view of the spatial sampling distances. This shall be explained in	Due to the instrument design not all detector pixels observe the same ground scene at the same time. This co-registration mismatch can	We will add some clarification

more detail.	be large while the spatial sampling distance is small for each individual pixel.	

Referee comment	Author's response	Proposed adaptation in manuscript
Specific comment 1: Page 2, line 4: This is not correct, see also http://www.copernicus.eu/main/overview I propose to replace this by a quote on that website: "The Programme is coordinated and managed by the European Commission. It is implemented in partnership with the Member States, the European Space Agency (ESA), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), the European Centre for Medium-Range Weather Forecasts (ECMWF), EU Agencies and Mercator Océan."	agreed	We will double check with ESA and change the text.
Specific comment 2: Page 2, line 14: The Sentinel-4 FM1 launch is now planned for 2022. Please correct.	agreed	We will change the text.
Specific comment 3: Page 4 line 18 / page 5 line 1: Please quantify more accurately: "The difference in flight time between the two positions is about 2 seconds"	agreed	We will provide the exact time difference at nadir.
Specific comment 4: Page 16, lines 24+25: Is a non-linearity knowledge of 0.6% compliant with the requirements at L1b? It seems to be rather large. Why is that? Please show some more results from the residuals between measured and fitted curves to quantify the 0.6% (additions to figure 7), also to stress the importance of the new methodology introduced in the abstract (page 1 lines 11-14).	This is indeed an error; the error after validation is a few hundred electrons, far smaller than the 0.6% mentioned.	We will correct the text.
Specific comment 5: Page 17: Pixel full well capacity. I guess detector pixel full well capacity in the detector pixels is reached before ADC saturation? Please mention this explicitly. Is this true for all wavelength ranges? Why are the SWIR results not included? If possible, include also SWIR in this section / table.	PFW capacity varies per CCD, but is more or less equal for all detector pixel on a CCD. The electronic gain in each band is chosen such that Register Full Well occurs before ADC saturation. The only exception is band 1, in which the fixed gain is so high that PFW can	We will add a comment on the ADC saturation in section 2.7.3. We will also add/quote the results for SWIR.

	never be reached, but ADC saturation can. The SWIR PFW was calibrated on unit level by SRON.	
Specific comment 6: Section 4.6, detector pixel quality calibration: Why is SWIR not included? If possible, include also SWIR in this section / table	The SWIR DPQF was calibrated on unit level by SRON.	We will add the SWIR results in table 5.
Specific comment 7: Page 20, lines 11+12: Same question as earlier for non-linearity, now for PRNU. Is a PRNU knowledge of 0.6% compliant with the requirements at L1b? It seems to be rather large. Why is that? "Several validation tests" are mentioned, but no results shown. Please show some more results from the residuals to quantify the 0.6% (additions to figure 8), also to stress the importance of the new methodology introduced in the abstract (page 1 lines 11-14). Please explain in the text if the PRNU is a purely detector pixel linked effect, or a wavelength linked effect, and why.	This is also an error; the error after validation is a smaller than the 0.6% mentioned. PRNU is a difficult subject to quantify. Fortunately PRNU cancels out in the calculation of the Reflectance.	We will add more validation results and discussion on the accuracy.
Specific comment 8: Page 22, line 7: Please quantify the temporal drifts in offset, and the residual errors in L1b for not correcting this effect	agreed	We will update the text
Specific comment 9: Figure 11: Please explain what the source is for the blue curves, and why the blue curves seem to have more noise than the red curves for all wavelengths.	The source of the blue curves is the integrating sphere. These do not have higher noise than the red curves. The cyan curves do; these stem from QTH2 measurements that had severe problems due to the stimulus shape and output.	We will update the text
Specific comment 10: Section 6.2, in-band stray light calibration. Usually signal-to-noise requirements are formulated for low-intensity scenes, i.e. for low albedo scenes in absorption lines. It is fine to report the stray light fractions in the way this is now done in the paper, but these stray light fractions at L0 and L1b shall also be reported with respect to these minimal signals for low albedo and inside the spectral absorption lines, in order to appreciate (quantify) the relative errors in the signals	See also comment 4. We agree that the straylight correction performance with realistic earth spectra and various albedos is interesting. However, this is out of scope for this paper due to the lack of measured realistic earth scenes, and because all applicable requirements were formulated as a linear fraction at L1b level using the hole-in-cloud scene. This validation scene has no spectral	We will explain in more detail the character of the observed straylight and that spectral features only play a minor role.

used for fitting L2 data products. Please report stray light fractions at L0 and L1b also (in addition to what is reported now in the paper) with respect to the signals for low albedo, also at wavelengths in the atmospheric absorption lines. Describe clearly (and distinguish between) the various different signal levels used for quantifying stray light fractions at L0 and L1b. It is acknowledged that the above request is fulfilled to some extent by the hole-in-cloud assessments on pages 28+29, but for these assessments it is not clear what the cloud and hole-in-cloud radiances are and if the radiances in the absorption lines are also accounted for. For example, in the NIR channel significantly higher stray light fractions at L0 and L1b were expected in the O2 absorption bands, but this does not seem to be the case (on the contrary, the stray light fraction at 765 nm is lower). Please explain and quantify and assess what the impact of a hole in the cloud scenario would be on L0 and L1b stray light with a real earth absorption spectrum (low albedo). in addition, page 29, line 1: Please explain what the spectral / spatial stray light requirements are at L0 and L1b and how they compare with scenes of low albedo and wavelength-dependent signals, also including signals within atmospheric absorption lines.	structure, only spatial. Some L0 performance is presented though. During the inflight commissioning phase the straylight performance will be assessed as suggested, and we plan to report on this in a future paper.	
Specific comment 11: Section 6.2, in-band stray light calibration. Please include an overview with quantitative assessments for: in-field and in-spectral-band (correctable) stray light at L0 and L1b. in-field and out-of-spectral band (correctable) stray light at L0 and L1b. out-of-field (uncorrectable) stray light at L0.	agreed	We will add a table with these numbers.
Specific comment 12: Section 6.2, in-band stray light calibration, table 8, page 28 line 15. The results in table 8 are applicable for what appears to be a TBD EWLS spectrum. It would be interesting to know what the corresponding numbers would be for a real low-albedo	Also see comment 4 and 10; this is out of scope for this paper due to the lack of measured realistic earth scenes. During the inflight commissioning phase the straylight performance will be assessed as suggested,	

earth spectrum, what stray light correction factors would be obtained. This would also quantify statements as “a very strong out-of-spectral range straylight contribution” and “This contribution is expected to be smaller in-flight than it is in the on-ground calibration measurements”. Please add some relevant assessments for quantifying L0 and L1b stray light for a real low-albedo earth spectrum	and we plan to report on this in a future paper.	
Specific comment 13: Section 6.3, out-of-spectral-band straylight. It would be interesting (essential) to add a number of comparisons between the NIR stray light measurements in TV conditions and ambient conditions: signal-to-noise, dynamic range between measured stray light signal-to-noise and source illumination, stray light as measured between the two.	Under TV conditions we only measured with a Xenon lamp with high-pass filter. The source out-of-band spectrum and its power is not known, and therefore only a qualitative assessment is possible.	We can add some extra information regarding dynamic range and noise for the ambient campaign.
Specific comment 14: Section 6.3, out-of-spectral-band straylight, also figure 16. Please add a plot of the relative stray light (percentage as function of signal at the source wavelength) as function of wavelength in the range 600-1100 nm. It seems virtually all out-of-band stray light in NIR is originating from 620-650 nm and 807-828 nm. Please explain briefly what is causing this, if possible. Quantify the stray light at L0 and L1b for a hole in the clouds scenario for a low albedo scene from a real earth spectrum, also in earth absorption lines in the NIR wavelength range, for the stray light as shown in figures 15 and 16 (referring to the importance of the new methodology introduced in the abstract (page 1 lines 11-14)). Quantify the error at L1b in stray light correction accuracy in the NIR wavelength range due to errors in radiance knowledge (since this is out of band) between 620-650 nm and 807-828 nm	It is correct that all straylight originates from these wavelengths, see figure 16. The instrument prime has not given a conclusive reason where the straylight originates in the optics. During the inflight commissioning phase the straylight performance will be assessed as suggested, and we plan to report on this in a future paper.	We will add explicitly where the source wavelengths are.
Specific comment 15: Page 41, figure 23. The noise shown in these plots is about 1%, suggesting a signal-to-noise ratio of 100. Clarify in the text why this signal-to-	This is not noise but diffuser features.	We will clarify this in the text.

noise ratio is so low		
Specific comment 16: Page 44, figure 25. Clarify in the text if the gradient observed at e.g. column 512 is also observed in the radiance measurements, which should be the case if it originates from detector quantum efficiency.	The observed gradient is the combined result of detector quantum efficiency and optical throughput of the spectrometer. The caption is not explaining this clearly.	We will update the caption.
Specific comment 17: Page 44, lines 3+4. This statement is not agreed / understood, because the distance is referenced with respect to the crosshair installed in the lamp socket that is used in the same way during calibration at NIST and use during TROPOMI calibration. Please clarify	We agree, we mean that the coil of the FEL lamp extends a few millimeter in the vertical direction. Therefore it is not the ideal point source as we treat it. Therefore the 1/r^2 law will not yield a unique distance for the optical pathlength to and within the internal diffuser.	We will explicitly mention that we cannot locate the exact point inside the volume diffuser due to this problem.
Specific comment 18: Page 45, lines 22-26. The advantage of the sun simulator would have not been only signal-to-noise, but also a much more flight-representative illumination geometry than a FEL lamp, that emits light to everywhere, because the sun simulator, as the name suggests, would illuminate diffusers more as the sun does. Please clarify.	Agreed.	We will add the field geometry to the sentence.
Specific comment 19: Page 47, lines 13-15. The quoted accuracies seem questionable in view of the limitations as described in this paper. It would be interesting (essential to support the statements on accuracy) to show also comparisons between the FEL, integrating sphere and sun simulator measurements for wavelength ranges where this is most useful (also in terms of signal-to-noise). Since for integrating sphere and sun simulator the absolute radiometric scales are not calibrated this exercise would have to include also the BSDF calibration, obviously	We do not have a reliable measurement of the instrument BSDF due to instabilities with the Sun Simulator and SNR issues with the integrating sphere. Therefore the BSDF is calculated as the fraction between ABSRAD / ABSIRR. None of these three methods give the same result within the error bars. We are forced to use the FEL measurements, also because they have good SNR. The errors presented are realistic from our point of view, but, these do not include the geometric errors, which we cannot validate due to lack of suitable measurements. We plan to validate this with inflight measurements and report it in a future paper.	We can add a figure with this comparison.

Specific comment 20: Page 48, lines 3+4, and lines 18-20. It is written that for bands 1 and 3 the snr (integrating sphere) was too low, but it would still be useful (essential to support statements on accuracy) to show the comparisons for the other bands. It is not clear how the uncertainties quoted in lines 18-20 are derived / justified. The range in UV is rather large. Clarify how these uncertainties are derived in view of the various FEL, integrating sphere and sun simulator measurements	See comment 19.	We can add a figure with this comparison.
Specific comment 21: Page 49, figures 31+32. The instrument BSDF should be a property of the differences between earth and sun paths only, i.e. diffusers plus maybe some mirrors. All other contributors drop out in the BSDF. Therefore the BSDF is a smooth function of wavelength. To show this, please plot the FEL-BSDFs in figure 31 as function of wavelength rather than column number, and quantify the differences in the wavelength-band overlap areas. In addition, compare the FEL BSDF results with those of the integrating sphere for wavelength ranges where this can be done (all bands, except bands 1 and 3?). These assessments/comparisons should also flow into the uncertainty budgets	The captions and the figures have gotten mixed up.	We will plot the BSDF as a function of wavelength as suggested.
Specific comment 22: Table 9. There are some questions with respect to table 9. - Errors are probably 1-sigma. Please indicate this. Clarify if non-linearity errors (0.6%, page 16) should be included. Clarify if PRNU errors (0.6%, page 20) should be included. Clarify if stray light errors (0.811% UV, 0.527% UVIS, 3.314% NIR, page 28) should be included. The uncertainties quoted for the diffuser calibration are in my view unrealistically low. I would have expected 1-sigma numbers of about 0.5% in UV, 0.4% in UVIS and NIR. Please provide a justification for these low numbers or modify them if	See general comment 4. We understand that this is unclear. The numbers in the table refer to the error in the calibration key data *only*. This error is used in the L01b processor to propagate the total error in the L1b products Radiance and Irradiance. Because the end-user is mostly interested in Reflectance, we have excluded errors (identified with an asterisk) from the CKD as they will cancel out when calculating the Reflectance. We will double check the reported accuracies	We will adjust the text in the relevant sections to clarify this. It is clear that some extra explanation is needed how the final error in the L1b products is calculated and handled; we will add a paragraph on this.

necessary. - It is not clear why the unexplained measurement discrepancy is given as a rather large range, e.g. 0.0-1.5% in UV, where the high number exceeds by quite a bit the low number given in the total uncertainty ABSRAD and FEL-BSDF. This is not very credible. Please provide a justification for this approach or modify the numbers if necessary (for example by providing a single number of e.g. 1.0% for UV, 0.3% for UVIS and 0.7% for SWIR, similarly to the NIR case). Furthermore, this table applies to the on-ground calibration (as the paper title suggests, of course), but it is not clear how the numbers given in table 9 would translate into the case for a realistic low-albedo earth spectrum. Please clarify	for the diffuser calibration. The unexplained measurement discrepancy range is the range over the detector; we will clarify this.	
Specific comment 23: Section 6.8, relative irradiance. The conclusion of this section is that the on-ground calibration measurements were not good enough and that the calibration will have to be (re)done in orbit (page 55, lines 5+6). Is there really an added value for this section? I propose to remove it, or at least shorten it drastically to a few sentences	agreed	We will shorten this substantially.
Specific comment 24: Page 58, figure 40. Figure 40 shows that the coregistration error increases to 4.0 km in UV, 2.0 km for UVIS, 5.0 km in NIR and 3.5 km in SWIR towards the swath edges. Table 2 gives the across-track and along-track spatial sampling distances for UV, UVIS and NIR of 0.50 degrees (7.2 km) and 0.059 degrees (0.8 km) and 0.16 degrees (2.3 km), respectively, for SWIR. In view of the numbers given in table 2 the coregistration errors as shown in figure 40 seem to be huge. Please clarify / describe in the text, also highlighting compliance (or not) with the applicable requirements	Due to the instrument design not all detector pixels observe the same ground scene at the same time. This co-registration mismatch can be large while the spatial sampling distance is small for each individual pixel.	We will clarify the definitions.
Specific comment 25: Pages 59+60, figure 41. See also the previous comment. Interband coregistration	See comment 24.	We will clarify the definitions.

errors going in some cases to 10, 20 or 30 km are shown in figure 41. How do these numbers compare with the numbers given in table 2 for across-track and along-track spatial sampling distances and with the applicable requirements (and compliance to those)? Please clarify this in the text		
Specific comment 26: Section 9, conclusions. The conclusion section is too short, given the large amount of information presented in this paper. Expand the conclusions with descriptions of what worked well and which accuracies were obtained (or generic) and which problems were encountered and why. The abstract discussed a new methodology (page 1, lines 11-14), but this concept is not optimally exploited (at least not described) in this paper, not in the conclusions. Consider to expand this. The statement on "In addition, the out-of-spectral-band straylight correction for the NIR detector has to be validated using in-flight measurements." comes out of the blue, and could have been quantified using the methodology of using the 0-1b processor with real earth atmospheric low-albedo input data. It is not clear how this validation will be done. This sentence is more for section 6.3, where it should be worked out in more detail (see also comment #14), not for the conclusions	Agreed	We will rewrite the conclusion, and include future work during the commissioning phase.

Referee comment	**Author's response**	**Proposed adaptation**
Technical correction 1: Page 49, figures 31 and 32. The legends and the figures don't seem to match, because QVD1 seems to be in the left 4 figures, QVD2 in the right 4, unlike the legend states (top vs bottom). Please correct if necessary	This is correct, manuscript versus article style in latex.	We will adapt this.
Technical correction 2: Page 61 shows some equations that are a bit distorted. Please consider	This is correct, manuscript versus article style in latex.	We will adapt this.

correcting this		