Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-438-RC1, 2018

Atmospheric Measurement Techniques Discussions

AMTD

Interactive comment

Interactive comment on "New method to determine the instrument spectral response function, applied to TROPOMI-SWIR" *by* Richard M. van Hees et al.

Anonymous Referee #1

Received and published: 25 January 2018

#############

General comments

#############

The paper addresses the determination of Instrument Spectral Response Functions (ISRF) of the recently launched Tropomi/Sentinel-5P mission. ISRF uncertainty is a notorious limitation of past and future space-borne atmospheric chemistry missions (e.g. GOME-2, OMI, Sentinel-4, Sentinel-5), as well as missions targeting greenhouses gases (OCO-2, MicroCarb, Sentinel mission). The SWIR band of the Tropomi/Sentinel-5P is used for retrieval of CO and CH4, where ISRF knowledge is most critical (together

Printer-friendly version

Discussion paper

with the NIR band) because of strong, narrow absorption features. In this sense, the paper addresses a critical aspect of trace and greenhouse gas retrieval and therefore the topic is of high scientific relevance.

The introduction of a new method for ISRF characterisation by on-ground calibration measurements, as promised by the title, would be of high interest for planned future missions. However, the manuscript fails to deliver key elements for introducing and making a case for a new method. Neither do the authors describe other techniques, nor do they motivate the introduction of a new one, explain the difference, or compare it with previous calibration/validation measurements. If the objective of the paper is the introduction of novel method for ISRF determination, as suggested by the title, the differences to previous instrument calibrations (e.g. SCIAMACHY, OMI, OCO-2) have to be pointed out. Ideally, a comparison shall be offered pointing to the advantages of the new approach.

In fact the reader is left wondering which are the new elements of the proposed approach: On the instrumental aspect: Is it the first time an OPO was used ? If so, why is it expected to yield better performance (than e.g. monochromator) ? What are the instrumental limitations of the approach ? The impact of key parameters of the measurement data (e.g. signal-to-noise ratio) is not discussed. No details on the instrumental setup are provided, and the quoted reference does not contain them either. On the modelling aspect: Is the novelty of the approach the mathematical model for the ISRF in terms of a peak and a tail function ? Then, why were these particular functions chosen and not any other ? Are there physical reasons why the ISRF tails should follow a Pearson type VII distribution ? All these questions are not addressed in the manuscript. In large parts, the text resembles an technical document (or ATBD), which describes a mathematical algorithm without explaining, why certain steps are taken.

The fitting procedure to determine the ISRF seems to suffer from an under-determined equation system, although this is never mentioned. This is mitigated by “fixing” some parameters selectively in the inversion during four iterative steps. Again, the authors do not justify the presented sequence of partial fits, other than it results in “good fits” (defined by low residuals). This is particularly worrying as the approach is verified with two synthetic ISRF shapes, which were presumably computed using the same mathematical model as in the inversion (albeit this is not clear from the text). In theory, if the information content of the measurements would be sufficient to estimate all model parameters, this should result in perfect agreement between fit and forward model. However, the fit residuals clearly exhibit systematic features, indicating a weakness in the fit procedure. Nevertheless, the authors conclude “compliance” as the fit residuals’ amplitude is below the accuracy threshold (1% requirement).

The iterative fit procedure yields ISRF shapes for all detector pixels, which are not uniform and clearly show high-frequency variation. At this point the authors argue that the ISRF is only determined by the spectrometer optics (PSF), which varies smoothly with wavelength and swath angle. This assumption is not in line with the definition of ISRF used in other publications, and even with the introduction of the present manuscript. Defined as the spectral response of a single pixel as a function of wavelength, the ISRF includes the detector PSF (convolved with the slit boxcar and the optical PSF). The detector PSF is mainly driven by cross-talk, which may indeed introduce a pixel-to-pixel or column-to-column variation. While such systematic features are clearly visible in the estimated parameters across the detector (e.g. Fig. 5), the authors indirectly dismiss them as artefacts (by the above assumption) and eliminate them with an elaborate “smoothing procedure”.

The smoothing equations (bi-variate polynomial fit) are reported, without justifying the choice of parameters (merely stating to yield “smoother and better” results). It is impossible to judge the impact (and therefore significance) of the smoothing, since no statistics are provided as of how much the individual parameter have been smoothed. The difference w.r.t the original fit results is not presented. Pixel-dependent effects (e.f. cross-talk) are “smoothed out” by this procedure, but presence of such effects does not mean that the individual ISRFs determined in the previous step were less accurate.

The use of smoothing procedures gives the impression that the authors do not have much confidence in their technique. They regard the smoothed ISRF parameters as more accurate, and report the observation that the individual rms deviation from the measurements have increased as “counter-intuitive” (although seems to be expected).

It is understood that for practical purposes (operational Level-2 processing), calibration key data cannot take into account ISRF variations at the pixel-to-pixel level. However, the authors should clearly state where such compromises are made (e.g. between level of detail and computational resources) and justify them by quantifying or estimating the impact on Level-2 accuracy. Currently, the ISRF is defined such as to match the fit and smoothing procedures, effectively establishing a definition which is only valid for the presented approach. This compromises the ability to compare with other instrument calibrations and the applicability for future instruments.

Any local variation is interpreted as resulting from not further specified “laser artefacts”, which are smoothed out by a bi-variant polynomial fitting. This needs to be further justified, answering the following questions: - Are the laser artefacts repeatable ? Have measurements been repeated for some of these “bad data” ? - Do they occur at given angles and wavelengths (patterns) or are they randomly distributed ? - What is the likely instrumental root cause (e.g. speckle or wavelength instability ?) ? - Why are such instrumental effects absent in the “good data” ? Could these also be affected by “laser artefacts” The paper lacks a discussion of error sources of the new method including instrument effects (only fit residuals are considered). A true error analysis of the technique would involve a rigorous analysis of instrumental error sources, such as - SNR of the laser measurements - laser stability - speckle amplitude (of integration spheres, diffusers) - straylight correction efficiency - non-linearity (resp. knowledge thereof) - pixel-to-pixel cross talk variation

The methodology for evaluating the suitability of the approach is questionable. The only figure-of-merit considered is “good fit quality”. The amplitude of fit residuals is interpreted as the accuracy of the method, compared with the requirement of 1% of the

ISRF peak. However, fit residuals only provide information about the consistency between the mathematical model and the measurement. If the measurement is affected by systematic instrument error (say, from straylight), the fit “absorbs” this error into the estimated parameters. This does, however, not mean that the true ISRF was determined more accurately. In fact, the authors identify “different treatment of straylight” as the likely cause of discrepancies between the ISRF for radiance and irradiance ports (although the correction technique and its accuracy is not presented), which in theory should be identical. But instead of interpreting this as a limitation, the two different straylight backgrounds are fitted by the model and both ISRFs are regarded as true ones. It is clear that instrument effects are fundamentally unavoidable (and obviously present here). They should be identified as such, and not “absorbed” into parameters of the model and declared part of the “true” ISRF.

Due to a poor (not further explained) laser performance, the irradiance measurements were used to compute key data , while the radiance fits classified as “good” were used for validation. Some validation results are reported, but only median values for selected columns, from which the conclusion is drawn, that radiance and irradiance ISRF are identical. This raises the question, if ISRF characterisation (which is typically a cost driver for imaging spectrometers), can be restricted to irradiance measurements only. This would greatly reduce effort (one wavelength scan versus $\sim$ 100 fpr each spatial sample) and cost. A discussion and quantitative analysis would greatly enhance the impact of the paper.

The paper includes a short section on in-flight calibration of the ISRF with Tropomi’s on-board calibration system, comprising five tuneable laser diodes. This part of the paper has the potential to significantly raise its impact, as this aspect is relevant for several upcoming missions (see list above). However, this opportunity is missed as no comparison between on-board and external diffuser/laser ISRFs is provided. The authors merely the state, that “The ISRF measured with the diode lasers is in close agreement with the ISRF calibration data,. . . ”, without presenting any plot, table, or statistics. Since data using the on-board calibration system were acquired and are available, it is strongly suggested to extend this part of the paper by providing quantitative comparison.

Finally, the authors do not give adequate credit to previous work. The reference list is rather short and limited to Tropomi-related publications. Tropomi/S5P not the first grating spectrometer for which extensive ISRF calibration has been performed, and also not the first covering the SWIR spectral range (e.g. SCIAMACHY, OCO-2). A more extensive discussion on previous work (actually needed for the introduction of a new method) should include a literature review covering the following missions (non-exhaustive list): - SCIAMACHY - GOME-2 - OMI (Dobber et al. 2004) - OCO-2 (Crisp et al., lasers used for ISRF calibration in SWIR and NIR) They should also mention upcoming missions for which a new technique may become relevant, like e.g. - Sentinel-4 - Sentinel-5 - FLEX - MicroCarb - future Copernicus mission for anthropogenic CO2 Also, the definition of ISRF and ISSF functions in the introduction, although accurate, needs to make reference to previous discussions in the literature (see detailed comments).

Overall, the draft paper in its current status cannot be considered for publication in a peer-reviewed scientific journal. However, as the topic under study is highly relevant for interpreting results from Tropomi/S5P, as well as for a wealth of future space-borne push-broom imaging spectrometers, they are encouraged to thoroughly rework the manuscript along the lines indicated in the above comments, and the more detailed ones below.

#############

Detailed comments

############# Abstract: From the abstract it is not clear what is the difference between “deriving” and ISRF, as proposed in this paper and applied to the SWIR band of the TROPOMI instrument on one hand, and “measuring an ISRF”, which is reported to be done for all bands. Please rephrase or add a sentence, clarifying the topic of the work presented here.

It is not clear from the abstract (and in large paerts of the paper), if a new generic technique for determining an ISRF is proposed, or if the it is simply reported how it has been done for the Tropomi instrument. If the main topic is the introduction of a new technique for ISRF derivation, which can be used for future instruments, then the abstract shall contain quantitative statements about the advantages of the new technique.

Will the proposed new method or be used to derive the in-flight ISRF ?

#############

1.) Introduction

L. 13: incomplete sentence '"The latter. . ." L. 14: "entrance slit"; also add that light from ground scene is collected by a common telescope (not mentioned)

L. 19: Why does CO "have to be measured" with this acc. ? If this is a scientific requirement, please clarify and reference it (e.g. from a study)

- add a sentence clarifying the term ISRF (in US literature instrument line shape) and how it is defined

- Better clarify the link between the requirement for XCH4 accuracy (1%) and the ISRF requirement. Is this error the only contributor to the budet ?

The discussion of the difference between ISRF and ISSF on p. 2 (L. 2-17) is important. It is basically paraphrasing a similar discussion in: J. Caron, et al.: THE CARBONSAT CANDIDATE MISSION: RADIOMETRIC AND SPECTRAL PERFORMANCES OVER SPATIALLY HETEROGENEOUS SCENES ICSO 2014 Tenerife, Canary Islands, Spain International Conference on Space Optics 7 - 10 October 2014 Please cite this paper in the introduction

- It shall be pointed out what is the role of the ISSF, which is composed of of ISRFs of neighbouring detector pixels. Is only the ISRF useful for

L. 27-28: Previously it has not been mentioned that the instrument also measures solar irradiance. Please clarify that in the introduction, so that the reader understands the difference in “irradiance ISRF” and “radiance ISRF”.

“In the spectral dimension, about 4–5 points have significant signal. This is the spread function of the instrument for this wavelength.’ If “This” refers to 4-5 points of signal, it must be replaced by “these” ? It shall be added, that an ISSF cannot be measured continuously, but only sampled, while the ISRF can be measured continuously.

Recommendation: Cite a publication about the SCIAMACHY instrument, which pioneered space based measurements in the SWIR spectral range. Also cite the NASA mission OCO-2, which also deployed tuneable lasers in ISRF calibration.

#############

GENERAL REMARK The introduction (and abstract) fail to motivate a new in the technique for ISRF determination, as is promised by the title. Why is a new approach needed ? Is the 1% accuracy requirement not reachable by previous techniques ? What is the basic idea of the new approach ? Please add a section motivating the new approach and its basic idea (and why it is only applied to the SWIR). #############

#############

2.) Calibration measurements

- Please provide a figure depicting the calibration setup (incl. OPO, integrating sphere, collimating optics, etc.). This would make it easier to follow the given description

- Please explain in the text why the “irradiance ISRF” is expected to be different from the “radiance ISRF” of the same spectrometer. Theoretically, the ISRF is a spectrometer property determined by the optics after the entrance slit. The only difference is the pre-slit illumination via the Sun diffuser (which btw. is not mentioned in the instrument description given in the introduction). The cited paper (Tol et al. 2017) does not provide a sketch neither.

- The sequence of measurements is not clear from the sentence in L. 11-12. With 165 sec. of measurements at 10 Hz we have 1650 acquisitions, so 20 for each wavelength ? What is the spectral sampling of the calibration measurement (2 nm / 80 = 0.025 nm, equally spaced) ? Is the wavelength adjusted in steps or continuously (by temperature variation ?). Please clarify in the text.

- The wording "100 wavelength scans per manual wavelength setting...to cover the swath..." is somewhat confusing. Is the manual setting a coarse adjustment which is then scanned in finer steps ? Or is it that 100 swath positions are manually aligned and then scanned over wavelength ?

- 1.1 deg. of the FoV corresponds indicates that more then one detector pixel is illuminated (216 pixels / (108 deg / 1.1 deg) = 2.2 pixels). In the introduction it is stated that the ISRF is determined for individual pixels. How does image distortion (smile, frown) influence the ISRF measurements over >1 pixel ?

- P. 3, L. 14: "Calibration measurements via the radiance port and irradiance port have been performed to verify that they are identical." Were the irradiance and radiance ISRF measurements identical ? To what extent ? If this is reported later, indicate it in the text.

- P. 3; L. "...was performed up and down" -> better "...was performed with increasing and decreasing wavelength scan direction."

#############

3.1 ISRF shape

p. 3; L. 25: "block distribution" is not a commonly used expression, recommended to use the term "boxcar function" instead.

p. 3; l. 28: “In the end, the Pearson type VII distribution resulted in the best fit.” What criteria determines what is the best fit ? In general, the criteria for the quality of the determined ISRF are not clearly defined at this stage.

Define or reference “Pearson type VII distribution”

- Please clarify why the particular representation of the ISRF was chosen. The idea of the Peak function is sketched (convolution of a perfect slit image with a function representing image blur by the optics. However, a Gaussian is not a perfect representation of the spectrometer PSF and the detector cross talk, and other functions may be used. Please comment on how the shape components were chosen (e.g. why a Pearson distribution). How robust are the results w.r.t. other representations ?

It is also not shown (nor stated) that eq. (2) and (3) represent a convolution of a skew-normal distribution with a boxcar function (which is suggested on p. ; L. 1-2). Please clarify.

#############

3.2 Data preparation

p. 4, l. 21: “The measurement data are corrected for. . .and stray light (irradiance only)” Why are the irradiance data straylight-corrected, and radiance data (apparently) not ? Generally, the straylight correction raises the question of definition of straylight, which is linked to the definition of the ISRF. In some ESA and NASA missions, straylight is defined by the spectral extent of the ISRF. E.g. for OCO-2, spectral straylight is defined as the light detected at wavelengths beyond 6 times of the FWHM of the ISRF. Inside this range it is part of the ISRF. The authors should highlight the relation between straylight and ISRF definition, and explain the definition used for Tropomi.

Question: Have the data been corrected for detector non-linearity as well ? This could be important for measurement of the far wings of the ISRF, where signal levels are low.

p. 4 l. 22: “Readouts from bad pixels are discarded in the analysis.” The manuscript often uses terms from requirement definitions, which may not be familiar to all readers. Please briefly clarify the term “bad pixels” (e.g. reduced spectral detection efficiency).

p. 4; l. 23-24: “Frames where the light source was off or very weak are discarded.” SwitchedâĂŤoff light source is not expected in s nominal calibration procedure. Specify how many data were lost or, if insignificant, remove sentence.

p. 4; l. 24-25: “In each remaining frame, the column with the maximum average signal is determined and the columns up 25 to 7 pixels from this peak column are selected, to include the faint signal of the tails.”

Are 7 pixels sufficient to capture the “faint signal of the tails” ? The ISRF could be deliberately saturated to increase the sensitivity further out to the wings. The criteria should be the definition of the ISRF boundary (1% of the peak). Please comment (and add in text) that 7 pixels cover that range.

This short (5 lines) section should be merged with the next one, or even the next two, which can be combined in a Section “ISSF and ISRF fitting”.

#############

3.3 ISSF fit

p. 5; l. 4: “The ISSF is assumed to be the mirrored version of an ISRF,. . . “ With growing wavelength distance (and changing optical PFS), the measured ISSF should differ from a mirrored ISRF centered at c0. Deviations from the symmetry assumptions could potentially affect the far wings of the ISRF. Please comment on the assumption and justify be demonstrating negligible error. If possible, present data (plots) supporting the validity of the assumption.

p. 5; l. 4: “which can be modelled with the function AR(c;d,$-$s,w,η,γ,m,c0) using Eq. (1); only its skew parameter s has the opposite sign” The introduction of the function AR in this sentence is confusing. Just define it by R with reversed skew parameter to reflect the mirror shape w.r.t. the ISRF. Maybe even add the defining equation AR(s) =

R(-s).

p. 5; l. 5: “In each frame, the ISSF of an illuminated row is fitted to the ISRF shape to normalize the signals and to find the wavelength peak position expressed in pixel units (Fig. 1b).” The process of fitting the ISSF to the ISRF shape involves the estimation of the 8 parameters indicated in Eq. 1-4. In a previous sentence, the number of useful pixels of a measured ISSF frame is only about 5: p. 2; l. 9: “In the spectral dimension, about 4–5 points have significant signal.” Therefore the inversion of an ISRF profile seems underdetermined. Please explain how the ISRF parameters can be still be estimated (assumptions on constant parameters ?).

p. 5; l. 18: “The square root of the fit variance is the rms value.” This can be assumed to be known by the reader (recommended to be removed).

In general, the description of the fit procedure is rather sparse. It reports processing steps, which are not further justified, leaving the reader wondering why a certain step is taken in a particular way: For example: 1) “As the laser-wavelength scan is not regular, the ISRF data points are not on a regular grid. Therefore, the points in the scan range are collected in bins of 1/32 of a spectral pixel and a median is applied to the data points in each bin. Empty bins are discarded.” Does the sampling (1/32) automatically follow from the non-regularity of the grid ? How irregular are the wavelength steps of the laser ? Why can’t the fit be performed on a non-regular grid, when the functional shape and the relative wavelength are prescribed anyway ?

2) “The quality of the fit is determined by calculating the fit variance, the sum of the squared fit residuals where the fit function is larger than 6% of the maximum, divided by the number of degrees of freedom (number of points minus the free fit parameters).” Please explain (or reference) the fit quality parameter. In absence of explanation it appears to be an arbitrary choice (e.g.why 6% threshold).

Since a new method shall be introduced here (according to the title), please extend the discussion of the fit procedure, explaining and justifying all steps.

#############

3.5 ISRF parameter smoothing

p. 5, l. 21: “However, the ISRF fits are valid locally (at location (r, c)) and not available for all pixels” A previous sentence seems to indicate that the fit is done for all pixels: p. 2, l. 24: “A description of the method and algorithm used to derive the ISRF for all illuminated pixels is presented in Sect. 3” Please clarify.

p. 5; l. 21-22: “It is expected that the fit parameters that define the local ISRF vary only smoothly over the surface of the detector as this is determined by the spectrometer optics. ” -> This is only true if detector effects can be neglected. This assumes that all pixel-to-pixel effects (like PRNU, DRNU) are perfectly calibrated. Cross-talk effects, which determine the detector PSF (one component of the ISRF and therefore not eliminated by calibration), may vary from pixel to pixel. Please justify the assumption that the ISRF is determined by optics only. If possible, provide empiric evidence for constant cross-talk.

p. 5; l 25, Eq. 5: Similar to the previously described fit procedure, this particular smoothing function seems to “fall from the sky”. Please provide justification for this particular function and how its parameetrs are determined (e.g. are the constants 255 and 999 the detector pixels in spatial and spectral dimension ?)

p.6; l. 1-3: “To obtain good results for the ISRF parameter fitting, obvious outliers in the individual ISRF-fit results should be rejected before the bivariate polynomial fit is performed. Given the distribution of outliers (in columns at the same wavelength), it is judged that most of them are caused by laser artefacts”. - Typo: Replace “artifacts” by “artefacts”

- Comment: Each ISRF has been determined by a multitude of ISSF measurements at many wavelengths, which have been used to fit a composite shape function with relatively few parameters. Random laser effects (which ones?) should already be

Printer-friendly version

Discussion paper

smoothed out by this procedure. It is somewhat surprising, that these obtained parameters need to be further smoothing. Please elaborate (in the text) on: - the size and distribution of outliers - why it is judged that the outliers are caused by laser artefacts - the likely cause of "laser artefacts" (e.g. can better lasers improve the method ?)

p.6; l. 8: "unrealistic curve-fit solutions are rejected. . ." Please report the fraction of rejected fits.

p.6; l. 9-10: Please report the fraction of rejected fits. And again, please justify the numbers of the rejection filter (why rms <= 0.0065 and not any other number ?)

Question: What was the impact of "bad" and "dead" pixels in the procedure ? This may provide important guidelines for detector cosmetics requirements

p.6; l. 12: "all automated scans are performed twice: scanning up and down in wavelength. " Please report if a systematic difference was observed between the two scan directions (hysteresis effect).

p. 6; l. 16-17: "The irradiance data has a better coverage in both spectral and spatial directions,. . ." -> Why is this the case ? The difference between Sun and Earth ports is the diffuser before the slit. This may extend the illumination in the spatial direction (full slit illuminated), but not in the spectral. ". . .so a higher order M = 7 could be applied on the parameters d and s, which show much more structure than the other fit parameters." -> It would be useful to compare (in plots and by statistics) the variability of the different ISRF parameters/

p. 6; l. 21-25: "The quality of the parameter fitting is determined by comparing the measured ISRF data points with the ISRF that results from the parameter model." -> This implies that the measurements are independent from the chosen mathematical model of the ISRF. However, each "measurement point" is already the result of fitting the ISRF shape model to the measured ISSF. " In general, the parameter smoothing will result in better and smoother ISRF calibration key data due to averaging and interpolation" -> This seems to imply that "smoother" automatically means "better". However, exaggerated smoothing could systematically affect the accuracy of the derived ISRF shapes. Please comment on the chosen balance between smoothing and accuracy. Have measurements been repeated at outlier positions to verify they are due to random instrumental artefacts ? "Possibly counter intuitive, the rms value will be slightly larger as the ISRF data points are now compared with a smoothed ISRF instead of an optimized local ISRF that might be influenced by measurement imperfections." -> What does "slightly" mean ? Please provide numbers. This "counter-intuitive" observation actually may indicate exaggerated smoothing. Since the indivudual fits are closer to the truth, I would certainly expect the rms to increase.

Please report on the overall statistics of the smoothing procedure. What is the scatter of the original ISRF parameters around the smoothed value used in the key data ? Please provide a table for all ISRF parameters.

#############

3.6 ISRF parameter iteration

p. 6; l. 31: "block width" -> "boxcar width"

1) "Once the ISRF has been fitted, the skew and tails are known approximately, and can be included as fixed properties. . ." 2) "Therefore, the refitted block width as a function of row and column is smoothed and used as a fixed property in the final ISRF fit"

This section provides some explanation for the question raised above: How can 8 parameters be estimated from ISSF of 4-5 significant pixels? However, the approach of the "passes or stages" cannot solve an under-determined measurement problem. Parameters are estimated in stages, limiting the number of unknowns in each step to avoid under-determination. However, the results of each stage impact the ones of the next one (by fixing parameters estimated from an incomplete model). Each stage necessarily yields errors (incomplete description of then shape profile by fixing parameters), and these are propagated into the next stage. In addition, smoothing seems to be applied between the 4 steps (see 2)), which also introduces systematic errors, propagated into the next step.

Please include a discussion, justifying the choice of the sequence, in which parameters are estimated.

p. 7; l. 3-8: How have the “realistic ISRFs” been simulated > Using the same parametric model ? This paragraph seems to describe a convergence test of the fit procedure for ideal data (no noise, no parametrisation errors, same ISRF). What are the start values chosen before stage 1 ? How far from the known true values can they be ? Please perform a test to demonstrate the convergence range of the technique.

p. 7; l. 9-14 The discrepancy of the derived ISRF from the known true one may indeed indicate the numerical problem highlighted above and result from the approach (underdetermined problem “solved” by fixing parameters and repeated iterations).

p. 7; l. 14: “However, the differences between the true ISRF and the derived ISRF are less than 0.25% and are considered acceptable” Acceptability could be stated if this were the maximum possible error. However, it seems to be assumed that the true ISRF is computed by Eq. 1-4, so it is consistent with the mathematical model and no modelling error is included. Please test the approach with ISRF profiles deviating from the chosen mathematical model to demonstrate robustness of the approach. Please also quantify the impact/sensitivity to measurement noise. This latter would give useful information on the required quality of the calibration system.

#############

4 Discussion of results

p. 7; l. 22: “A median has been taken over all rows illuminated. ” What does this mean (a medium of what parameter) ?

p. 7; l. 22: “From visual inspection of the displayed ISRFs, one can conclude that:

(i) the ISRF is sharper and higher at higher column number (longer wavelength)" A large ($\sim$20% change) in ISRF width (2.3-2.7 pixels) should have been predicted by the optical design analysis. Is the magnification changing in spectral direction ? See also comment on Figure 3 and 4.

"(ii) the ISRF fit resembles the ISRF data very well, e.g. the residuals are very small, except where small artifacts can be identified in the ISRF data" Correct typo "artifacts". The log plots show significant discrepancy in the wings and the residual plots show periodic structures, whose peak-to-peak amplitude correspond to almost half of the requirement (1%). I would change "very well" to "satisfactory" to "compliant". Since a new method is proposed (according to the title), the question arises if it provides superior performance over previous calibration campaigns.

"(iii) the fit residuals of the irradiance ISRF are nearly a factor 2 smaller compared with the radiance ISRF." Please provide explanation.

p. 7; l. 27: "The difference is likely due to differences in stray light in these measurements." -> It was stated before that all measurements (both for radiance and irradiance) are based on straylight corrected data. Now straylight is identified as the cause for a discrepancy between radiance and irradiance ISRFs. This needs to be commented to avoid confuciotn. How accurate is the straylight correction ? Is the observed discrepancy (apparently averaged across the entire detector) explainable by the limitation of straylight correction. All this would be part of an accuracy analysis of the "new method", which is currently missing.

p. 7; l. 28: "In all subsequent fitting, shape parameter m is fixed to 1.25 to enhance convergence of the curve-fitting routine. . ." -> Why is this value fixed whereas all others are (partially, sequentially) fitted ? If it represents the "straylight level", as suggested in the text, why should it be constant ? How much do the fitted values of m vary and deviate from the median value ? In absence of this discussion this appears an arbitrary reduction of parameters for better convergence. Please comment and justify.

“It has to be noted that the contribution of the tail to the ISRF is small (< 10 %) and only significant 1-2 pixels away from the peak” -> This seems in contrast to studies analysing the impact of ISRF shapes on CH_4 retrieval, for which ISRF far wings are very important.

“It has been verified that fixing the m parameter has negligible effect on the resulting ISRF and the fit residuals expressed in the rms value.” Please provide evidence. Level-2 processing (not visual inspection) defines when an effect on the resulting ISRF is negligible. Has this been verified by retrieval simulations ?

p. 8; l. 10: “Block width w of the ISRF is determined by the projection of the slit onto the detector and therefore decreases as a function of wavelength.” Again, please explain why a 20% reduction of slit width image is expected. In fact, the spectral sampling requirement (>2.5 pixel) seems to be violated across a wide spectral range.

p. 9; l. 19: “However, width parameter d has been designed such that no errors are introduced by the ISRF parameter fit.” What does it mean to “design” a parameter ? Is it the choice of xi in Eq. 2 ?

p. 8; l. 28: “The quality of the ISRF fits as determined with the parameters from the bivariate parameter-fitting models shown in Fig. 6b.” Typo, “is” is missing.

p. 8; l. 30: “There are a few small regions which coincide with the fine-scale structures visible in the skew-normal width, see for example around row 50 at columns 525 and 610.” It is very difficult (if not impossible) to see the described fine-scale structures in Fig. 5 and 6 (a single row is not visible).

p. 9; l. 3-4: “In general, the laser performed worse during the radiance measurements, yielding radiance ISRF measurements of poorer quality than the irradiance measurements.” Please provide details as to why and how much the laser has performed worse

p. 9; l. 13-14: “On the left side of the detector, the block width of the radiance ISRF tends to be smaller than that of the irradiance ISRF. This subtle difference is attributed to the non-optimal scanning of the laser at these wavelengths.”

Please indicate in which way the laser-scanning was non-optimal. Provide recommendation for optimum laser scanning.

Based on the poorer quality of the radiance fits, irradiance measurements are used for key data generation, while the radiance measurements merely serve for validation. What was the calibration time partition between irradiance and radiance measurements ? Noting that - significantly more time was spent on radiance (100 scans versus 1) and - the differences are stated to be negligible one conclusion could be that ISRF calibration can be reduced to measuring irradiance only. Please comment.

#############

5 In-flight Monitoring of ISRF

It is appreciated that a section on in-flight calibration is included in this paper. However, more detailed information shall be given here. As a minimum, the text should provide - type of laser diodes (Distributed Feedback (DFB) - their distribution over the SWIR range This is particularly important since the ISRF changes significantly over the spectral range - scan range in nm (not roughly pixels) - mention of a negligible laser bandwidth Also, reference to publications shall be made here, which describe the instrument design of Tropomi.

The comparison between ISRFs from the ISRF calibration campaign with external laser sources and measurements using the on-board lasers is of high interest in the context of future missions (e.g. Sentinel-5). Therefore the authors should elaborate on the quantitative comparison between the ISRFs derived form the two sources. The curent discussion is too qualitative and does not allow an evaluation of on-board ISRF monitoring. Adding a quantitative discussion (with plots) would significantly enhance the impact of this paper.

p. 9; l. 26-27: “ The scanning range is about 6 spectral pixels so that the ISRF can be monitored for one or two wavelength pixels per laser." This is confusing, since the term "wavelength pixels" is not defined, or at least not discriminated to "spectral pixels" in the same sentence. I assume the authors mean the spectral range corresponding to two FWHM of the ISRF. Please rephrase.

p. 9; l. 25-26 "The laser wavelengths are scanned by tuning the temperature of the laser using a built-in thermo-electric cooler" Built in what: the (DFB?) laser or the calibration unit? Please describe more precisely.

p. 9; l. 28: "As the diffuser is not moved during the measurements, there will be speckle." Inadequate wording for a science paper: - Replace "there will be speckle." by more precise formulation, e.g. "…the measurements will be affected by speckle patterns due to the coherent laser light." Also, the connection to "moving diffusers" might not be clear to every user. Please add a sentence like: "Such patterns can be reduced by moving (e.g. rotating) the detector during the acquisition. However, this is not foreseen for in-flight calibration due to mechanical contraints (e.g. micro-vibrations)."

p. 9; l. 29: "Most speckle is removed by taking the median of the data of all illuminated rows." This is not understood and needs more explanation: I assume that the illumination of the on-board diffuser illuminates the Is the median taken over all rows (swath direction) to yield only one ISRF for the entire focal plane ? Why the median and not the mean (affected by outliers) ? The latter makes more sense for reasons of energy conservation. Shouldn't the on-board calibration enable the determination of the ISRF across the entire swath width at 5 spectral positions ?

p. 9; l. 29: "During the commissioning phase, in-flight measurements with the on-board lasers will be performed with a moving and a fixed diffuser" Before it was mentioned that the on-board diffuser cannot be moved in flight. This may be different during commisioning phase, but deserves a sentence of explanation. Please provide details to improve clarity.

p. 9; l. 31: “The ISRF obtained from these measurements can be compared with the ISRF measured on ground using the external laser and the on-board diode lasers to detect any possible changes” - Remove “any”, as this suggests infinite accuracy

p. 9; l. 31: “The monitoring ISRF is of sufficient quality to check for any degradation of the instrument but cannot be applied in trace-gas retrieval.” - Remove “any”, as this suggests infinite accuracy - Please explain why the on-board ISRF “cannot be applied in trace-gas retrieval”. It may actually be useful to correct for launch effects and thermo-mechanical effects (de-focus). If such correction is made, it will be indirectly used in Level-2 processing.

In general, be more quantitative in the comparison and evaluation of the on-board ISRF measurements. What is the expected and obtained accuracy, and over which spectral range ?

p. 10; l. 3: “With an oscillating diffuser. . .” - Please explain what is meant by “oscillating” . I suppose that a calibration disk is moved back and forth by a few degrees (which is not quite oscillation), but the reader has to guess. Provide more details (see comment above).

p. 10; l. 4-5: “. . .except that ISRF parameter smoothing (Sect. 3.5) is calculated from the ISRF fits of the few columns scanned per diode laser.” - Why is smoothing necessary here ? I would assume that the ISRF is determined for the five ISRFs corresponding to the center wavelengths of the diode laser scan ranges. The ISRF fit procedure probably takes the on-ground parameters as start values, so large outliers should not be expected.

p. 10; l. 4-5: “The column dependence of the shape parameters is neglected and the row dependence is smoothed by a second order polynomial.” Does a square law (second order polynomial) decribe the variation of optical effects in swath direction ?

p. 10; l. 6-7: “Then the median ISRF is calculated from all ISRF data of the central one/two fully-scanned columns, neglecting any row dependence." - replace "one/two" by "one to two" - Why is row dependence neglected (it has even been "smoothed" by a polynomial anyway) ? Why taking a median ISRF, not a mean ? Why are the 5 in-flight ISRFs not determined for all spatial samples across the swath ? The approach to in-fight ISRF characterisation appears somewhat immature.

p. 10; l. 9-10: Insert "the" between "moving" and "on-board"

p. 10; l. 10-11: "The ISRF measured with the diode lasers is in close agreement with the ISRF calibration data, thus proving the usability of the method and validating the calibration data." -> Be more quantitative here. No plot nor table is provided for this important comparison. Please plot the 5 ISRFs measured with the on-board diffuser together with the on-ground ISRF, corresponding to the same detector pixels. Perform the comparison for both, moved and stationary on-board diffuser to quantify the impact of speckle patterns.

p. 10; l. 12: "The monitoring ISRF deviates from the ISRF calibration data as could be expected." -> This is in contradiction to the sentence before.

p. 10; l. 12-13: "However, it is believed the method is sensitive enough to be used on board for long-term monitoring, being able to distinguish between changes in the real instrument ISRF and changes in the speckle pattern." -> Again, please provide a quantitative comparison, to substantiate this "believe".

Editorial: Straylight is written inconsistently in two ways: "stray-light" and "stray light". Any of them is fine (as well as one word), but be consistent.

############

6 Conclusions

p. 10; l. 15-16: "A new and accurate method using a scanning OPO has been developed and applied to characterize the TROPOMI-SWIR ISRF. " -> Remove "accurate", as this is a qualitative statement, that's need to be substantiated (How accurate? More accurate than other methods ?). In fact, its accuracy should be a results reported quantitatively here.

p. 10; l. 18-20: “An iterative scheme to derive the SWIR ISRF has been developed, where the ISRF determined in a previous iteration is used to improve the ISSF model in the current iteration. The required accuracy of the ISRF is obtained within 4 iterations.” -> It shall added here that the “iterative scheme developed” is not estimating free parameters in every iteration (as would be expected for an over-determined problemS selectively fixing part of the parameters in every iteration is characteristic to the proposed new approach, so it has ot be repeated here (and the consequences as well).

p. 10; l. 18-20: “The ISRF measured through the irradiance port using the solar diffuser has been compared with the equivalent ISRF measured via the radiance port. The differences between the ISRFs derived from both data sets are very small,. . .” ->Please be quantitative here

“. . .and largely due to differences in stray-light treatment and laser scan imperfections.” -> The statement that the discrepancy in ISRFs is “largely due to differences in straylight treatment and laser scan imperfections”, is an assertion, not a finding. It is suspected, but not demonstrated in this paper (it is even unclear, in which way straylight was treated differently).

p. 10; l. 23-25: “The derived ISRF meets the requirement on ISRF knowledge and should thus be sufficient for methane retrievals.” The claim that the derived ISRF meets the requirement on ISRF knowledge is based only on the claim that the fit residuals are smaller than 1% of the ISF peak. However, this only means that the parameters of the chosen mathematical representation can tuned to match the observed shape. It does not mean that the observed shape is accurate. An example is the straylight, which apparently affects the measurements differently in the radiance and irradiance ports. Does it mean that the true ISRF of a system depends on the quality of straylight correction ? By fitting the straylight into the line shape (parameter m), it becomes a feature of the true ISRF.

It is proposed to include a critical appraisal of the approach and results in the conclusions. This should outlined also the limitations of the approach. Accuracy (in terms of deviation from a true ISRF) shall clearly be distinguished from consistency between a fit and a measured curve.

#############

Figures and Tables

Table 1: Why is the parameters m kept at 1.25

Table 3: Reference (Beers et al.). Please explain why a a publication on “Measures of Location and Scale for Velocities in Clusters of Galaxies” is relevant, resp,. applicable to describing the ISRF variation.

Fig. A1 and A2: - Axis labels are missing (“Pixel No.”) - Figure captions should be understandable without reading the text. Please extend the figure captions, briefly explaining the difference between “ISRF fit” and “ISRF parameter fit”.

Comment: These plots indicate that there are systematic (not random) features being smoothed by the polynomial fit procedure, especially in the spectral dimension. Without evidence it is not obvious that they result from “laser artefacts”. Speckle effects (not mentioned in the text) should affect the spatial component stronger due to smoothing by spectral dispersion.

The lower panel of Fig. A2 suggests that the “block width”, representing the image width of the entrance slit, varies from 2.3 to 2.7 (pixels ?). This $\sim$20% change over the spectral range (for the entire swath) should be readily verifiable by optical analysis (diffraction and spot size PSF). Please check (and report) the plausibility of the result with the optical performance analysis.

Fig. 3 and 4: Reporting the parameter values in the caption is difficult to associate to the 15 plots and does not provide useful information. Better include in them into the plot legends.

The fit residuals (bottom row) clearly exhibit systematic (periodic) structure. This indicates a shortcoming of the ISRF shape model, which does not allow for periodic components. Are there physical reasons why periodic components in the ISRF shape are ruled out ? Please comment on this and possibly propose an improvements.

- Please include plots showing the difference between radiance and irradiance ISRF and discuss the reasons for differences.

Fig. 5: Unclear Fig. caption: "In the white area, the ISRF fit failed (vertical stripes), the light is blocked by the entrance slit of the spectrometer (top and bottom) or a shield at the detector (left and right)." It does not seem logical, that an entrance slit blocks light. Improve clarity by adding "white area at the edge" or changing color. The term "white areas" is confusing with most of the middle panel being white (not only the edge). Proposed to change color scale.

Fig. 7: The number of "good" fits is drastically lower for radiance than for irradiance. Please provide explanation why this is the case.

Fig. A3 and A4: The plots show large variability of the resulting tail fraction and width from the ISRF fits. However, the ISRF parameter fit ("model") seems to assume a single value across the detector. Has this also been fixed to the median value (as parameter m) ? What is the justification, given the large, systematic variability ? Convergence ? It should be clarified (already in Section X) which parameter have been fixed to avoid the impression that the ISRF shape model has 8 free parameters.

#############

References Buscaglione: ESA-SRDs should have no author name

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-438, 2017.