

Response to Reviewer 1:

We would like to thank the reviewer for performing a thorough review and for the many helpful suggestions to improve the paper.

Below, we respond to each of the review comments. For the sake of clarity, the review comments are given in blue italics and our response is printed in normal font. Changes to the manuscript are printed in green.

I wish to thank the authors for a well-written paper. It is structured and organized, making it easy to follow. In most cases the descriptions are precise, so that the meaning is not left to interpretation by the reader. The conclusions are reasonable; the authors do not over-interpret their results.

Following are several issues I would like addressed.

Section 1 The authors state their objective is to evaluate TropOMI radiometric accuracy using the best available surface albedo data sets. This implies they wish to address its absolute accuracy and not merely the TropOMI calibration relative to that of OMI and SCIAMACHY. Given this broader objective there should be an assessment of OMI/SCIA radiometric accuracy, and a discussion of how representative the respective LER data sets are of the underlying calibrations (an accurate calibration does not necessarily mean the LER data sets are equally accurate). Section 5 discusses one deficiency of these LER databases, but it would be better if this paper addresses data set accuracy in Section 3 rather than as only a data screening issue in Section 5.

Indeed, our goal is to assess – as much as possible – the absolute accuracy of the radiometric calibration, and not just to present a comparison relative to that of OMI and SCIAMACHY. This implies the use of a radiative transfer code to minimise the dependence on external information. Nevertheless, this approach still relies on a good knowledge of the input parameters. The most important input parameter in the wavelength range that was studied is surface albedo. The surface LER databases that were used have been validated extensively in the past, with known, citeable accuracies. These are now mentioned in Section 3.4 of the manuscript:

“ The OMI and SCIAMACHY surface LER databases have been compared to each other and to other surface LER databases [Kleipool et al., 2008; Tilstra et al., 2017]. For the OMI surface LER database an overall accuracy of 0.01–0.02 was reported, with slightly increasing values towards the shorter wavelengths [Kleipool et al., 2008]. The SCIAMACHY surface LER database was shown to have an accuracy of about 0.01 for the UV-VIS-NIR spectral range [Tilstra et al., 2017]. These accuracies reflect the uncertainties caused by various error sources, such as errors in the radiometric calibration and the occurrence of cloud contamination in the databases. Errors brought about by the Lambertian-equivalent nature of the databases are not part of these uncertainties. It should also be noted that the OMI and SCIAMACHY surface LER databases are mostly representative for the time periods from which they were derived (OMI: 2005–2009; SCIAMACHY 2002–2012). Systematic changes in surface reflectivity occurring after these time periods, for instance due to changes in land use, are not covered by the databases and will result in errors. ”

Section 3.2 No mention is made of Raman scattering. Should the readers assume that it is not modeled? Since it is an important effect for evaluating the spectral radiometric response of the instrument (as much as 1-2% at TropOMI wavelengths shorter than 400 nm), perhaps it should be stated explicitly.

Raman scattering is not modelled. The DAK RTM that we use does not have the possibility to do so. But, the wavelength bands that are defined in Table 2 have a typical bandwidth of one nm in the UV wavelength range. Raman effects in the TROPOMI reflectance spectra are therefore averaged out partly by the spectral averaging that takes place in the conversion from reflectance spectra to reflectance bands.

This should indeed have been mentioned in of the paper. In the revised manuscript the following text is added to the manuscript, in Section 3.2:

“ Raman scattering is not modelled by DAK. However, Raman effects in the TROPOMI reflectance spectra are averaged out partly by the spectral averaging that takes place in the conversion from reflectance spectra to reflectance bands. ”

Section 5.1 The information provided in this section is not entirely clear. It is not immediately obvious that the FRESCO false cloud identification over land is caused by an overestimation of surface reflectivity in the underlying databases. Also, the underestimation of land reflectivity, while an established effect, has not been established for these particular databases as their sole or even primary error. In many locations the OMI LER dataset values are higher than other standards, presumably as a result of sub-pixel cloud contamination. I will presume that SCIAMACHY LER suffers similarly.

Yes, this is true. On the one hand, the surface albedo retrieved for the surface LER databases is systematically underestimated for certain geometries, because the algorithms are searching for the minimum scene LER values that are observed but not taking into account the directional dependence of the surface reflection. On the other hand, the surface albedo can be overestimated as a result of cloud contamination in the databases. The OMI and SCIAMACHY surface LER databases both suffer from such issues.

The text in Section 5.1 was rewritten to make it more clear that the FRESCO false cloud identification is caused by inaccuracies in the input surface albedo:

“ The deviations in the cloud fractions over land are primarily caused by a systematic underestimation of the surface reflectivity for certain geometries in the (Lambertian) surface albedo database that is used in the FRESCO algorithm. This explanation is in line with conclusions from a recent study by Lorente et al. [2018], who showed that traditional Lambertian surface albedo databases, like the ones described in ... ”

Section 5.2 The alternative methods presented in this section appear to be designed to eliminate scenes for which the LER databases are affected by non-Lambertian surface characteristics. Since these errors are as much a function of viewing conditions as they are surface type, isn't the TropOMI LER subject to similar errors? Is it legitimate to screen the reference data and not the measured data as well?

It would not be legitimate, but it is not done that way. The filtering is performed on a box-to-box basis. That is, if the input surface albedo for a one-by-one degree box is rejected, then so is the associated TROPOMI observation box. The measured and reference data are therefore subjected to the same screening mask.

Section 6.2 It is not clear where the referenced 772 nm slope and intercept are coming from. Is it Figure 6 or Figure 8?

The analysis that is used to determine the average slopes and intercepts is visualised in Figure 8, but not for the 772-nm wavelength band.

This is indeed confusing, so we have changed the manuscript. We now also mention the result for 758 nm in the text, because this wavelength band is shown in Figure 8. We still mention the 772-nm results. Mentioning the 772-nm wavelength band is important, because it shows the largest deviation. In the manuscript we have added the comment that the 772-nm results are not shown in Figure 8. The changes to the manuscript are:

“ ... The results were obtained for all 21 wavelength bands. In Fig. 8 the results are presented for 5 of the wavelength bands. ... ”

“ ... At 758 nm the average slope is larger than one ($\overline{m} = 1.015 \pm 0.008$) and the intercept is small ($\overline{n} = 0.004 \pm 0.003$). At 772 nm (not shown in Fig. 8) the average slope has increased further ($\overline{m} = 1.033 \pm 0.008$) and the intercept is now very close to zero ($\overline{n} = 0.003 \pm 0.002$) ... ”

Changes to the manuscript:

During the review phase of this manuscript the TROPOMI instrument kept generating new data and we took the opportunity to extend the time range that was studied originally (May 2018–February 2020) by three months. The studied time period is now May 2018–May 2020, covering two years. Some of the numbers reported in the paper have changed slightly, but not significantly. Figure 8 and Table 3 have also been updated. The changes w.r.t. the previous version are in all cases insignificant, well within the reported accuracies. The extension of the time range that was studied has led to an increased accuracy of the results. Figure 3 was also updated.

References:

- Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, *J. Geophys. Res.*, 113, D18308, doi:10.1029/2008JD010290, 2008.
- Lorente, A., Boersma, K. F., Stammes, P., Tilstra, L. G., Richter, A., Yu, H., Kharbouche, S., and Muller, J.-P.: The importance of surface reflectance anisotropy for cloud and NO₂ retrievals from GOME-2 and OMI, *Atmos. Meas. Tech.*, 11, 4509–4529, doi:10.5194/amt-11-4509-2018, 2018.
- Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, *J. Geophys. Res.-Atmos.*, 122, 4084–4111, doi:10.1002/2016JD025940, 2017.

Response to Reviewer 2:

We would like to thank Dr. Lang for performing a thorough review and for the many helpful suggestions to improve the paper.

Below, we respond to each of the review comments. For the sake of clarity, the review comments are given in blue italics and our response is printed in normal font. Changes to the manuscript are printed in green.

The paper by Tilstra et al. on the In-orbit Earth reflectance validation of TROPOMI on board the Sentinel-5 Precursor satellite is an important contribution to the meanwhile significant history of knowledge in characterising this class of instruments, since the launch of GOME-1 on ERS-2. The paper is well written and organized and it shows robust and convincing results on the assessment of the TropOMI instruments radiometric in-flight calibration accuracy using independent radiative transfer (RT) forward modelling of the expected Top-Of-Atmosphere (TOA) radiances in the region between 328 and 2314 nm.

The fact that the results are convincing and robust with respect to the provided statistical analysis of the reflectance correlations (between simulated and measured results) is not a given, since previous attempts to use RT model results in evaluating the in-flight calibration accuracy and performance of this class of sensors were limited predominantly by uncertainties in model inputs.

In this respect the statistical significance of the presented results, and the fact that they may be used as a robust evaluation of the sensor performance for most of the large spectral regions investigated here, makes it a unique contribution, generally in the field of post-launch high-spectral resolution sensor calibration in the UV to the near infra-red (NIR).

In this respect, and in view of future missions in development (Sentinel-5), or planned mission like the proposed High-Priority Copernicus Candidate CO₂ monitoring mission (HPCC CO₂M), it is unfortunate that the short-wave infra-red (SWIR) band evaluation had to be excluded from the results, probably because of systematic error contributions in the RT model surface reflectance input. In view of the significantly stronger focus of S5 and CO₂M on the SWIR, improving the analysis and associated methods in this spectral region would be very welcome (and as suggested by the authors they consider this possible provided a better surface albedo input becomes available).

It is indeed very unfortunate that no strong conclusions could be drawn from the SWIR results. It is still not understood what exactly causes the unexpected and suspicious behaviour for the 2314-nm wavelength band. Because we cannot explain the behaviour for this wavelength band we are forced to dismiss the end result for this wavelength band. We are at the same time, however, not excluding the possibility that the results that were found for the 2314-nm wavelength band are “real”. Further studies will have to investigate this.

The authors address the two key error-sources for the overall approach in quite some detail. One is the selection of real clear-sky Rayleigh scattering scenes, with no or negligible residual cloud or aerosol scattering contribution. The other is the accuracy of the used surface reflection data, and in particular their angular and spectral variation (spectral BRDF), which is a large contributor - in addition to the accurate knowledge of the surface albedo itself - to the total budget of the forward modelling RT error and has significantly hampered previous attempts for carrying out such studies.

Cloud screening is predominantly done using co-located VIIRS cloud detection, and aerosols are detected by deriving an Aerosol Absorbing Index (AAI) from the measurements themselves. Here the authors state that they avoid a vicious circular problem by “correcting” the reflectances. However it seems not quite clear if the way these values are corrected is then actually implying an overall iterative approach, i.e. going through the full procedure, coming up with a radiometric offset (or residual between measurement and RT result), then use these results to correct the radiances, and finally repeat the procedure until some convergence criterion is fulfilled. This aspect should be, from my point of view, better described or clarified in the paper.

The corrections on the reflectances that are used to calculate the AAI are not completely based on the validation study, but partly derived independently.

To explain, we independently analysed the reflectances at 340/380 nm which are stored in the TROPOMI AAI product. From this, a (time dependent) correction for degradation was derived for the reflectances following the method explained in Tilstra et al. [2012]. This correction for degradation was applied to the TROPOMI 340/380-nm reflectances that go into our calculation of the AAI. This part of the correction does not depend on the outcome of the study, because it is derived independently.

A day-zero correction was also applied to the 340/380-nm reflectances, by focusing on one day only (“day zero”, the first day studied after the end of the commissioning phase, 1 May 2018, indicated by the blue dots in Figure 8). For this particular day, the reflectance correction factors were found from the reflectance validation study and these were then used as day-zero corrections for the reflectances that go into the AAI calculation. This part of the correction *does* depend on the validation study itself.

However, the correction factors resulting from the validation study are not changing significantly if a second round is performed to calculate the day-zero correction. Note that the AAI threshold value was set to a relatively high value of 2.0 index points. The impact of the filtering on AAI is therefore rather low to begin with.

In conclusion, there is no real need for an iterative scheme.

Note that the primary reason for calculating the AAI ourselves is the fact that it is not allowed to average AAI values, but it is allowed to average the reflectances from which the AAI is derived, as explained in Section 4.3.

We have added the following sentences to Section 4.3 of the paper, to better explain the approach that was followed to correct the reflectances that go in to the calculation of the AAI:

“ The corrections consist of a time-dependent multiplicative correction to account for the linear drift observed in the 340/380-nm reflectances that are stored in the S5P AAI product, and of a day-zero multiplicative reflectance correction. The linear drift was determined using the method presented in Tilstra et al. [2012]. The day-zero correction factors were found from applying the reflectance validation study on 1 May 2018 (“day zero”) and determining the correction factors found for 340 and 380 nm. In principle this step could be repeated by applying the correction factors, calculating newer correction factors from these, until convergence is reached. An iterative scheme for this was not pursued, however, because a secondary iteration turned out to already provide almost the same correction factors as the first one. ”

The accuracy of the used surface reflectance is for sure the most critical aspect for the performance of the RT forward model results in the described validation approach. The author’s present 4 methods to validate

the Lambertian Equivalent Reflectance (LER) database derived from OMI and SCIAMACHY, the latter being probably used in order to cover the SWIR. The missing angular variation is considered as one of the largest errors in the existing OMI and SCIA LER databases, especially over vegetation, and this is why the chosen filtering on low quality LER values makes use of BRDF and albedo information as provided by Modis the Terra or Aqua satellite platforms. This however raises two questions:

1) Since the problem in the missing angular information in the LER has already been described for the existing LER databases derived from GOME-2, and therefore an angular dependent version of the GOME-2 LER database has been developed in the context of the Atmospheric Composition Satellite Application Facility (ACSAF) activities for Metop, the question arises why this database has not been used here. Especially since, in the end, the SCIA LER database in the SWIR shows deficiencies, and the analysis of this region had to be excluded anyway.

The *directional* GOME-2 surface albedo database can unfortunately not be used to provide the missing angular information, because the directional information stored in this database is representative for polar orbits with a 09:30 LT equator crossing time. The GOME-2 instruments are all kept in 09:30 LT orbits, while TROPOMI is in a 13:45 LT orbit. It is only possible to use the GOME-2 directional information for data observed from polar orbits with equator crossing times close to 09:30 LT. This includes data from GOME-1, SCIAMACHY, GOME-2 itself, the future Sentinel-5/UVNS, but not data from OMI or TROPOMI.

It is possible to use the *non-directional* GOME-2 surface LER information which is also stored in this product. In fact, this is what we did during the initial phase of the study. Unfortunately, the GOME-2 surface LER has some issues for wavelengths below 400 nm. These issues have been reported in the past and one could easily take this into account via corrections. However, this would not improve the readability of the paper and it would also not improve the results of this study because the SCIAMACHY surface LER database does not have a lower quality than the GOME-2 surface LER database. SCIAMACHY does have the advantage that its equator crossing time (10:00 LT) is a bit closer to the 13:45 LT of TROPOMI.

SCIAMACHY also provides surface albedo for the SWIR channel. Although we indeed had to exclude the results for the 2314-nm wavelength band from the manuscript, we were at least able to report that linearity problems and offset issues do not seem to be present for the SWIR channel. We would not have been able to draw these conclusions using the GOME-2 surface LER database.

In the revised manuscript we now mention in Section 3.4 that the GOME-2 surface LER database could also have been used, but that we decided to use the SCIAMACHY surface LER database primarily because of the additional information in the SWIR wavelength range:

“ As an alternative to the SCIAMACHY surface LER database we could also use the very similar GOME-2 surface LER database [Tilstra et al., 2017]. However, the GOME-2 surface LER database does not provide information for the SWIR wavelength range, which is the primary reason for adopting the SCIAMACHY surface LER database instead. Alternatively, we can make use of MODIS surface BRDF . . . ”

2) In the context of TOA test-data simulations of this kind of spectrometers, the MODIS surface albedo and BRDF is frequently used in conjunction with MODIS surface type characterisation and the ADAMs (A surface reflectance Database for ESA's earth observation Missions - <https://nebula.esa.int/sites/default/files/>

neb_study/1089/C4000102979ExS.pdf) spectral database, which provides the possibility to calculate the angular dependent BRDF at any wavelength from the UV to the SWIR (using principle components of spectral vectors for various surface types). This proved to provide realistic BRDF values in the wavelength region covered here in simulation studies for that type of sensors. Why has this option not been considered? Since this approach might turn out to be useful to solve the issue for the SWIR band radiometric performance validation for TropOMI. I am sure the authors will have some convincing answers on the few issues I have raised here, in which case I can highly recommend the paper for publication in AMT.

The possibility of using the ADAM database was considered and discussed, but in the end we did not go for this option. The principal component method used in the ADAM database & toolkit seems to be a responsible way of handling the interpolation between the MODIS wavelength bands. Certainly, the wavelength interpolation provided this way is addressing the problem that we mentioned in the manuscript (in Section 3.4) of the – for our specific purpose – inconvenient location of the MODIS wavelength bands.

However, we do have our doubts that the interpolation can be done with a high enough accuracy, because the seven MODIS bands are quite far apart. For example, the shortest MODIS wavelength band is centred at 469 nm, which would involve extrapolation (not interpolation) to get towards 328 nm. A lower quality is to be expected here, as reported in, for instance, Bacour et al. [2020]. For the longest MODIS wavelength band centred at 2130 nm we would need extrapolation to get to 2314 nm. For the wavelength region at about 700 nm, interpolation has to be done between 645 and 858 nm with no support from space-based observations. Especially here, around the red edge, interpolation of any kind seems to be challenging. The OMI and SCIAMACHY surface LER databases provide surface LER for all wavelength bands studied in our validation. The (monochromatic) wavelength bands of these databases were defined in exactly the same way as the wavelength bands defined in the manuscript.

In the recent paper by Bacour et al. [2020] the ADAM database and its toolkit are introduced extensively. The paper mentions that the ADAM database *overestimates* the BRDF near the “hot spot” by a factor of about 1.5 for barren surfaces. For vegetated and other surfaces, the ADAM database *underestimates* the BRDF near the hot spot by a factor of 1.5–2.0. This is quite a large underestimation/overestimation, and it is not clear to us whether these errors are caused by the ADAM interpolation scheme or by the underlying MODIS BRDF database. It would be an alarming fact if these errors would be present in the MODIS surface BRDF information. In any case, the reported underestimation/overestimation by a factor of 1.5–2.0 is a thing to worry about, because with the close to noon (13:45 LT) equator crossing time of TROPOMI and the relatively wide orbit swath the backscatter condition is found in a rather large part of the TROPOMI orbit.

The water BRDF in the ADAM database is not derived from MODIS, but modelled. Although modelling surface reflection for water surfaces is perfectly possible, it could lead to a difference in quality between land and water pixels. Such differences are not found in our analyses, because we treat water and land pixels in the same way. If there are deficiencies in our approach because of issues in the OMI or SCIAMACHY surface LER used, then at least the error is made in a consistent way for both land and water surfaces.

Another thing to consider is that using ADAM surface albedo input would link the validation results to the radiometric calibration of MODIS. Right now, they are linked to OMI and SCIAMACHY. One could argue that the validation results are also linked to MODIS, because of the filtering which is partly based on MODIS.

But for the most part, the link is with OMI and SCIAMACHY. This seemed to us to be the proper thing to do, especially considering the heritage of the TROPOMI instrument and of its retrieval algorithms.

In the end, for the reasons mentioned above, we decided to rely on the OMI and SCIAMACHY databases.

In the revised manuscript we now mention the ADAM database, that this approach would address the problem of the position of the MODIS wavelength bands, and explain why we decided not to use it. The text now reads:

“ Another possible source of surface albedo is the ADAM database [Bacour et al., 2020]. This database is based on MODIS BRDF, but provides interpolation to any wavelength in the wavelength range 240–4000 nm. This would address the issue related to the position of the seven MODIS wavelength bands. However, interpolation does lead to a lower quality and reliability, as explained in Bacour et al. [2020]. Next to this, the ADAM database shows both underestimation and overestimation by a factors of 1.5–2 in backscatter directions [Bacour et al., 2020, p. 18]. For water surfaces, the surface reflection is modelled, and bears no relationship with the MODIS BRDF data. This means that consistency in calibration between land and ocean is broken. For the above reasons we decided to use the OMI/SCIAMACHY approach. ”

Minor and editorial issues:

p.2, l. 13: The reference to GOME-2 on Metops could be associated to the relevant paper by Munro et al.

Yes, it should indeed. We have added the reference to Munro et al. [2016] to the revision of the paper.

We have also added the reference to TROPOMI, which was also missing.

General, the exclusion of band 8 should probably be motivated more towards the beginning of the paper.

We agree and now do this by referring to Section 4.1 already in the Introduction of the manuscript:

“ The analyses cannot be performed for TROPOMI bands 1, 2, and 8 for reasons explained in Sect. 4.1. ”

Section 3.4. The temporal aspects on using a database derived from SCIAMACHY and its application to a recent missions, should probably be mentioned and/or addressed, especially for vegetation and crop surfaces.

We agree and have added the following lines to Section 3.4:

“ It should also be noted that the OMI and SCIAMACHY surface LER databases are mostly representative for the time periods from which they were derived (OMI: 2005–2009; SCIAMACHY 2002–2012). Systematic changes in surface reflectivity occurring after these time periods, for instance due to changes in land use, are not covered by the databases and will result in errors. ”

p. 8, l.3. “From Fig. 1 it can be inferred. . .”. I find it actually quite difficult to infer it from the Figure if the differences are small. From the Figure 1 one can only for sure infer that they are in the right overall magnitude and spectral relation.

We agree and have changed the sentence:

“ However, the impact is relatively small and of little concern. ”

p. 10, l.17ff: How is the inhomogeneity of the target area actually determined? Is it just derived from coastline maps and surface type database or from the TropOMI radiance variance itself - which would be the best option I guess? In this respect, can the surface albedo as derived from the OMI and SCIA databases can be considered a true average over the 1 by 1 box as used here as a target?

The term “inhomogeneity” is meant as inhomogeneity due to the occurrence of both land and water inside the box. It is based on the surface type information that is available in the TROPOMI L2 products.

Inhomogeneity in terms of the standard deviation of the TROPOMI reflectances was calculated, but used to track down outliers which were not detected by the TROPOMI flagging. This is sometimes an issue for the SWIR channel. It was not used for filtering out scenes/boxes that were considered too inhomogeneous.

The idea behind the one-by-one degree boxes is to have the OMI and SCIAMACHY surface albedo information line up as best as possible with the TROPOMI observations.

In the revised manuscript, we now clarify how the inhomogeneity of a box is determined:

“ ... sun glint or solar eclipse events. Inhomogeneity and snow/ice presence are determined based on the surface type and surface condition indicators inside the TROPOMI L2 products. A box is only considered inhomogeneous if it contains both land and water. The number of boxes ... ”

Figure 3 shows the location and number of measurements over the defined period and number of days for clear sky scenes. However, it is not clear if this is the final statistics for all 56 days applying method 4 for cloud screening. The distribution of the locations of the latter, which actually go into the results, would here be of highest interest.

It is the final statistics for all 56 days, so Figure 3 illustrates the distribution and frequency of the clear-sky one-by-one degree boxes that were collected over all 56 days. These are all boxes which were not filtered away for reasons explained in the text, such as presence of snow or absolute latitudes above 60 degrees. Filtering such as performed in Method 4 was not applied for this image, because the other three methods are also discussed in the paper, and because Method 4 has not been defined at this point in the manuscript.

To clarify the meaning of Figure 3 we have changed the caption of Figure 3 in the following way:

“ Figure 3. Location and frequency of the $1^\circ \times 1^\circ$ cloud-free latitude/longitude boxes that were selected by the algorithm every two weeks for the period from May 2018 until February 2020. A total of 56 days contributed to the image. The selection and filtering approach is described in Sect. 4.4. ”

Changes to the manuscript:

During the review phase of this manuscript the TROPOMI instrument kept generating new data and we took the opportunity to extend the time range that was studied originally (May 2018–February 2020) by three months. The studied time period is now May 2018–May 2020, covering two years. Some of the numbers reported in the paper have changed slightly, but not significantly. Figure 8 and Table 3 have also been updated. The changes w.r.t. the previous version are in all cases insignificant, well within the reported accuracies. The extension of the time range that was studied has led to an increased accuracy of the results. Figure 3 was also updated.

References:

- Bacour, C., Bréon, F.-M., Gonzalez, L., Price, I., Muller, J. P., Prunet, P., Straume, A. G.: Simulating multi-directional narrowband reflectance of the Earth's surface using ADAM (a surface reflectance database for ESA's Earth observation missions), *Remote Sens.*, 12, 1679, doi:10.3390/rs12101679, 2020.
- Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., Kokhanovsky, A., Livschitz, J., and Eisinger, M.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing – an overview, *Atmos. Meas. Tech.*, 9, 1279–1301, doi:10.5194/amt-9-1279-2016, 2016.
- Tilstra, L. G., de Graaf, M., Aben, I., and Stammes, P.: In-flight degradation correction of SCIAMACHY UV reflectances and Absorbing Aerosol Index, *J. Geophys. Res.*, 117, D06209, doi:10.1029/2011JD016957, 2012.
- Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, *J. Geophys. Res.-Atmos.*, 122, 4084–4111, doi:10.1002/2016JD025940, 2017.

In-orbit Earth reflectance validation of TROPOMI on board the Sentinel-5 Precursor satellite

Lieuwe G. Tilstra¹, Martin de Graaf¹, Ping Wang¹, and Piet Stammes¹

¹Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

Correspondence: L. G. Tilstra (tilstra@knmi.nl)

Abstract. The goal of the study described in this paper is to determine the accuracy of the radiometric calibration of the TROPOMI instrument in-flight, using its Earth radiance and solar irradiance measurements, from which the Earth reflectance is determined. The Earth reflectances are compared to radiative transfer calculations. We restrict ourselves to clear-sky observations as these are less difficult to model than observations containing clouds and/or aerosols. The limiting factor in the radiative transfer calculations is then the knowledge of the surface reflectance. We use OMI and SCIAMACHY surface Lambertian-equivalent reflectivity (LER) information to model the reflectivity of the Earth's surface. This Lambertian, non-directional description of the surface reflection contribution results in a relatively large source of uncertainty in the calculations. These errors can be reduced significantly by filtering out geometries for which we know that surface LER is a poor approximation of the real surface reflectivity. This filtering is done by comparing the OMI/SCIAMACHY surface LER information to MODIS surface BRDF information.

We report calibration accuracies and errors for 21 selected wavelength bands between 328 and 2314 nm, located in TROPOMI spectral bands 3–7. All wavelength bands show good linear response to the intensity of the radiation and negligible offset problems. Reflectances in spectral bands 5 and 6 (wavelength bands 670 to 772 nm) have a good absolute agreement with the simulations, showing calibration errors on the order of 0.01 or 0–3%. Trends over the mission lifetime, due to instrument degradation, are studied and found to be negligible at these wavelengths. Reflectances in bands 3 and 4 (wavelength bands 328 to 494 nm), on the other hand, are found to be affected by serious calibration errors, on the order of 0.004–0.02 and ranging between 6% and 10%, depending on the wavelength. The TROPOMI requirements (of 2% maximal deviation) are not met in this case. Trends due to instrument degradation are also found, being strongest for the 328-nm wavelength band, and almost absent for the 494-nm wavelength band.

The validation results obtained for TROPOMI spectral band 7 show behaviour that we cannot fully explain. As a result, these results call for more research and different methods to study the calibration of the reflectance. It seems plausible, though, that the reflectance for this particular band is underestimated by about 6%. A table is provided containing the final results for all 21 selected wavelength bands.

1 Introduction

Satellite monitoring of atmospheric composition has evolved a lot over the last decades. Continuous monitoring started out in the year 1978 with the launch of the first Total Ozone Mapping Spectrometer (TOMS) instrument. The TOMS instrument was designed to measure the Earth reflectance at six wavelengths between 310 and 380 nm (Heath et al., 1975). With this limited spectral information, it was possible to retrieve ozone, SO₂, and various other properties globally on a daily basis. The spectrometer Global Ozone Monitoring Experiment (GOME) (Burrows et al., 1999), launched in 1995, was the first satellite instrument recording high-resolution spectra. The spectra were measured with a spectral resolution of 0.2–0.3 nm over the spectral range between 240 and 790 nm. Retrieval of atmospheric species from the spectra was achieved using the differential optical absorption spectroscopy (DOAS) method (Platt, 1994).

Since then, many more spectrometers have been launched, including Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) (Bovensmann et al., 1999), launched in 2002, Ozone Monitoring Instrument (OMI) (Levelt et al., 2006), launched in 2004, the GOME-2A/B/C instruments on the MetOp series of satellites (Munro et al., 2016), launched in 2006, 2012, and 2018, respectively, and Tropospheric Monitoring Instrument (TROPOMI) (Veeffkind et al., 2012), launched in October 2017 and the focus of this paper. Each of these satellite instruments represents an improvement with respect to its predecessors in terms of spatial resolution, spectral coverage, spectral resolution, and/or performance. Using the spectral coverage and resolution offered by the spectrometers, many atmospheric species can be retrieved successfully, including ozone, SO₂, NO₂, BrO, CH₂O, H₂O, CO, CH₄, as well as surface, cloud, and aerosol properties.

The quality of the retrieved properties is depending in one way or another on the quality of the radiometric calibration of the satellite instrument. For DOAS retrievals the absolute calibration of the reflectances is not so important, because any constant radiometric error cancels out in the DOAS method (Platt and Stutz, 2008). However, for many products the retrieval is not based on the DOAS method. Examples are the retrieval of ozone profiles (Hasekamp et al., 2002; Liu et al., 2010; van Peet et al., 2014; Shah et al., 2018), absorbing aerosol index (Herman et al., 1997; Torres et al., 1998; de Graaf et al., 2005), surface reflectivity (Herman and Celarier, 1997; Koelemeijer et al., 2003; Kleipool et al., 2008; Tilstra et al., 2017), and cloud and aerosol properties (Wang et al., 2008, 2012; Joiner et al., 2012; Torres et al., 2013; Lelli et al., 2014). For all these products the retrieval codes do rely on a correct absolute calibration of the reflectances.

The radiometric calibration of a satellite instrument starts on-ground, in a laboratory, with extensive characterisation of the instrument's response to the exposure of radiation. However, the conditions in space are different from the conditions in the laboratory environment. Also, before, during, and after launch the instrument can undergo instrumental, electronic or optical degradation. Once the satellite is in orbit and the instrument is switched on in-flight monitoring starts. This in-flight monitoring is primarily based on measuring the solar irradiance signal on a daily basis, to use it as a reference under the assumption that the Sun is a stable source. These daily measurements of the solar signal can point to shortcomings in the pre-flight radiometric calibration and they can reveal changes in the radiometric response of the instrument due to instrument degradation. They can be used to derive corrections on the pre-flight calibration of the solar irradiance.

However, this only corrects the solar irradiance, not the Earth radiance. The assumption that both signals degrade in the same manner is not justified (see e.g. van der A et al., 2002; Tilstra et al., 2012). As a result of this, the reflectance, which is a ratio of the Earth radiance and the solar irradiance, may not be corrected properly. To address this issue, in-flight calibration of the reflectance is needed, as has been done in the past in various ways, such as by intercomparison with other satellite instruments (Acarreta and Stammes, 2005; Kokhanovsky et al., 2007; Tilstra and Stammes, 2006, 2007; Jourdan et al., 2007), or by comparison with radiative transfer calculations (van Soest et al., 2005; Jaross and Warner, 2008; Tilstra et al., 2005, 2014; Cai et al., 2012).

In this paper we study and validate the radiometric calibration of the TROPOMI instrument by comparing the reflectance measured in-flight with radiative transfer calculations. This is done for clear-sky land and ocean scenes for wavelengths ranging from 328 to 2314 nm. The difference compared to earlier studies is that these were restricted to the UV wavelength range (below 400 nm). The reason for this was that for longer wavelengths the influence of the surface albedo on the reflectance is higher while at the same time the uncertainty in the input surface albedo for the simulations increases. In this study we perform the analyses for TROPOMI bands 3–7, i.e. for the much larger UV-VIS-NIR-SWIR wavelength range, using the best possible relevant surface albedo input available at this time. [The analyses cannot be performed for TROPOMI bands 1, 2, and 8 for reasons explained in Sect. 4.1.](#)

The outline of this paper is as follows. Section 2 provides a brief description of the TROPOMI instrument. Section 3 describes the radiative transfer modelling and the most important input parameters, such as the surface reflectivity. Section 4 describes the approach that was followed, including the selection of the wavelength bands, the cloud and aerosol screening, and the selection of scenes to analyse. Section 5 discusses ways to reduce the impact of surface albedo errors. Section 6 presents the results, and draws conclusions about the radiometric calibration of TROPOMI. Trend analysis, to study instrument degradation, is also part of this section. The paper ends with the most important conclusions in Sect. 7.

2 Description of TROPOMI

2.1 Instrument description

The TROPOMI instrument (Veefkind et al., 2012) is the only instrument on board the Sentinel-5 Precursor (S5P) satellite. The S5P satellite was launched on 13 October 2017 and was put into a near-polar, Sun-synchronous orbit with a mean altitude of 824 km above the Earth's surface, an equator crossing time of 13:30 LT, and a repeat cycle of 17 days. TROPOMI is the successor of OMI (Levelt et al., 2006), which is kept in a very similar orbit to allow accurate intercomparisons and cross-validation. OMI has proven itself to be important for many advances in the field of satellite remote sensing (Levelt et al., 2018). TROPOMI is meant to expand on these successes, and to fill the gap that exists between the SCIAMACHY mission, which ended in April 2012 with the loss of the Envisat platform, and the future Sentinel-5 mission, which is planned for launch in 2022.

TROPOMI, like its predecessor OMI, is a nadir-looking spectrometer that observes the spectral domain in the across-track dimension in one go using a two-dimensional detector. However, while OMI only observes the ultraviolet-visible wavelength

range (270–500 nm), TROPOMI is able to observe the ultraviolet-visible (UV-VIS, 267–499 nm), near-infrared (NIR, 661–786 nm), and shortwave infrared (SWIR, 2300–2389 nm) wavelength ranges. TROPOMI is equipped with four detectors, which each host two spectral bands. The characteristics of these eight spectral bands are found in Table 1. The extension of the wavelength range (compared to OMI) allows the retrieval of additional trace gases such as CO and CH₄ and better retrieval of cloud and aerosol information using the O₂-A and O₂-B absorption bands.

Another improvement with respect to OMI is the high spatial resolution of TROPOMI. The TROPOMI footprint size was 7.2×3.6 km², but since 6 August 2019 this has been reduced to 5.6×3.6 km². TROPOMI observes the sunlit side of the Earth when S5P is in the ascending part of its orbit. The orbit swath is 2600 km wide, allowing global coverage in one day. The TROPOMI commissioning phase ended on 30 April 2018. This is officially the date from which the data can be used. The current level-1 version is 1.0. Further information about the instrument and the derived products can be found in Veefkind et al. (2012) and Kleipool et al. (2018).

2.2 On-board calibration

For the sake of on-board calibration TROPOMI is equipped with a number of internal light sources. LED strings are mounted next to each of the four detectors (DLED), to monitor the performance of the detectors themselves. For the visible wavelength range a LED (called common LED, CLED) is placed in the calibration unit, to monitor the throughput of the internal light path from the calibration unit to the detectors (via telescope, slit, and spectrometers). Additionally, a white light source (WLS) is installed in this calibration unit. All these light sources are meant to provide stable signals for the purpose of monitoring the instrument’s throughput (Ludewig et al., 2020). Note that the CLED and WLS measurements can only provide information for part of the internal light paths that are used for the radiance and solar irradiance measurements.

Table 1. Characteristics of the TROPOMI detectors and spectral bands. The eight spectral bands are located on the four detectors. The design spectral coverage of the detectors and the full spectral coverage of the spectral bands are both listed. The spatial sampling in the along track direction was changed from the indicated 7.2 km to 5.6 km on 6 August 2019, at the start of orbit 9388.

Detector	UV		UV-VIS		NIR		SWIR	
	270–320		320–495		675–775		2305–2385	
Spectral coverage [nm]	1	2	3	4	5	6	7	8
Spectral band	1	2	3	4	5	6	7	8
Full spectral coverage [nm]	267–300	300–332	305–400	400–499	661–725	725–786	2300–2343	2343–2389
Spectral resolution [nm]	0.47	0.46	0.51	0.50	0.34	0.34	0.23	0.23
Spectral sampling ratio	7.2	7.1	2.6	2.5	2.7	2.8	2.4	2.4
Pixel size across track [km]	29.6	3.6	3.6	3.6	3.6	3.6	7.1	7.1
Pixel size along track [km]	7.2	7.2	7.2	7.2	7.2	7.2	7.2	7.2
(after 6 August 2019:)	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6

Daily solar measurements provide another reference, and from these the stability of the solar irradiance over time can be assessed. The absolute radiometric calibration of the solar irradiance can be verified easily because the Sun is a relatively stable light source. For the radiance measurements, other techniques have to be applied.

3 Radiative transfer modelling

5 3.1 Theoretical background

In this paper, the Earth reflectance is defined as

$$R = \frac{\pi I}{\mu_0 E} . \quad (1)$$

The quantity I is the Earth radiance at the top-of-atmosphere (TOA), in units $\text{Wm}^{-2}\text{sr}^{-1}\text{nm}^{-1}$. The quantity E is the solar irradiance at the TOA, perpendicular to the incoming solar beam, in units $\text{Wm}^{-2}\text{nm}^{-1}$. The parameter μ_0 is defined as $\mu_0 = \cos \theta_0$, where θ_0 is the solar zenith angle. For the viewing direction we use a similar definition, namely $\mu = \cos \theta$, with θ the viewing zenith angle. The viewing and solar azimuth angles are symbolised by ϕ and ϕ_0 , respectively.

For the radiative transfer calculations we restrict ourselves to clear-sky scenes and impose Lambertian surface reflection. According to Chandrasekhar (1960) the following relationship is then valid:

$$R(\mu, \mu_0, \phi, \phi_0, A_s) = R^0(\mu, \mu_0, \phi - \phi_0) + \frac{A_s T(\mu, \mu_0)}{1 - A_s s^*} \quad (2)$$

15 In Eq. (2), the quantity R^0 is the so-called path reflectance, which represents the purely atmospheric contribution to the reflectance. In other words, it corresponds to the reflectance of the Rayleigh atmosphere when it is bounded below by a black surface that does not reflect any of the incoming radiation. The second term represents the surface contribution to the reflectance. This term is determined by the surface albedo A_s , by the total transmission of the atmosphere T , and by s^* , the spherical albedo of the atmosphere for light coming from below. The quantities R^0 , T and s^* can be determined from radiative
20 transfer calculations.

3.2 Radiative transfer calculations

For the radiative transfer calculations we make use of the ‘‘Doubling-Adding KNMI’’ (DAK) radiative transfer code (de Haan, 1987; Stammes, 2001). This code can calculate the monochromatic Earth reflectance taking polarisation into account. It can handle molecular scattering, scattering and absorption by clouds and aerosols, and absorption by various trace gases. DAK
25 employs Lambertian surface reflection and simulates a pseudo-spherical atmosphere. [Raman scattering is not modelled by DAK. However, Raman effects in the TROPOMI reflectance spectra are averaged out partly by the spectral averaging that takes place in the conversion from reflectance spectra to reflectance bands.](#) We used version 3.2.0 [of the DAK code.](#)

The calculations were performed for the 21 wavelength bands listed in Table 2. We used a standard Mid-Latitude Summer (MLS) atmospheric profile (Anderson et al., 1986) and included absorption by ozone, NO_2 , and $\text{O}_2\text{-O}_2$. For some of the

Table 2. Definition of the wavelength bands and of the way the radiative transfer calculations were performed.

Wavelength band	328	335	340	354	367	380	388	402	416	425	440
Instrument channel	3	3	3	3	3	3	3	4	4	4	4
Central wavelength (nm)	328.0	335.0	340.0	354.0	367.0	380.0	388.0	402.0	416.0	425.0	440.0
Bandwidth (nm)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Spectral/monochromatic	S	M	M	M	M	M	M	M	M	M	M
Ozone absorption	+	+	+	+	+	+	+	+	+	+	+
NO ₂ absorption	+	+	+	+	+	+	+	+	+	+	+
O ₂ -O ₂ absorption	+	+	+	+	+	+	+	+	+	+	+
Oxygen absorption	-	-	-	-	-	-	-	-	-	-	-
Water vapour absorption	-	-	-	-	-	-	-	-	-	-	-
Wavelength band	463	494	670	685	697	712	747	758	772	2314	
Instrument channel	4	4	5	5	5	5	6	6	6	7	
Central wavelength (nm)	463.0	494.0	670.0	685.0	696.97	712.7	747.0	758.0	772.0	2314.0	
Bandwidth (nm)	1.0	1.0	1.0	1.0	0.3	0.3	1.0	1.0	1.0	0.5	
Spectral/monochromatic	M	M	M	M	S	S	M	S	S	M	
Ozone absorption	+	+	+	+	+	+	+	+	+	+	
NO ₂ absorption	+	+	+	+	+	+	+	+	+	+	
O ₂ -O ₂ absorption	+	+	+	+	+	+	+	+	+	+	
Oxygen absorption	-	-	-	-	+	+	-	+	+	-	
Water vapour absorption	-	-	-	-	+	+	-	-	-	-	

The reflectance calculations are performed using spectral band integration or monochromatically. For all wavelength bands absorption by ozone, NO₂, and O₂-O₂ is included. Absorption by oxygen and/or water vapour is included for only some of the wavelength bands.

wavelength bands we additionally included absorption by oxygen (697, 712, 758, and 772 nm) and/or water vapour (697 and 712 nm). For these wavelength bands monochromatic calculations do not suffice and spectral calculations are in order. Table 2 indicates when this is the case. Clouds and aerosols were not included in the simulations, because in this study we only focus on clear-sky scenes.

5 3.3 Look-up tables

For each of the 21 wavelength bands look-up tables (LUTs) were created. These were calculated for 7 ozone column values (50, 200, 300, 350, 400, 500, and 650 DU), for 10 surface heights (0–9 km), for two water vapour columns (0 and 4 g/cm²), and for 42 non-equidistant values for each of the zenith angles μ and μ_0 . For the path reflectance R^0 we make use of a Fourier expansion in terms of the relative azimuth angle $\phi - \phi_0$, which ends after only three terms because the simulations are

performed for clear-sky Rayleigh atmospheres over a completely black surface (Hovenier et al., 2004). That is to say,

$$R^0 = a_0(\mu, \mu_0) + \sum_{i=1}^2 2a_i(\mu, \mu_0) \cos i(\phi - \phi_0). \quad (3)$$

The reflectances themselves are not stored in the LUTs but the Fourier coefficient a_0 , a_1 , and a_2 . This has the advantage that the dependence on the relative azimuth angle $\phi - \phi_0$ can be handled analytically by Eq. (3). The quantities T and s^* are also stored in the LUTs. The LUTs therefore only contain the parameters a_0 , a_1 , a_2 , T , and s^* . This is sufficient to calculate the reflectance R using Eqs. (3) and (2).

3.4 Surface albedo input

Scenes with clouds and/or large amounts of aerosol are excluded in this study so the largest source of errors in the radiative transfer calculations is the uncertainty in the surface albedo (Tilstra et al., 2005; Cai et al., 2012). We use the OMI and SCIAMACHY surface LER databases (Kleipool et al., 2008; Tilstra et al., 2017) to provide surface albedo input to the simulations. These databases contain monthly climatological database grids with a spatial resolution of $0.5^\circ \times 0.5^\circ$. The OMI surface LER database, covering the wavelength range 328–499 nm, is used for the wavelengths up to 500 nm. The SCIAMACHY surface LER database, covering the wavelength range 328–2314 nm, is used for the longer wavelengths. The surface albedo contained in these databases is the Lambertian-equivalent reflectivity (LER) of the surface. Adopting Lambertian surface reflection can cause errors especially for the longer wavelengths. For example, neglect of the directional dependence can lead to errors of a factor of two or more in the surface reflectance of vegetated surfaces (Lorente et al., 2018).

The OMI and SCIAMACHY surface LER databases have been compared to each other and to other surface LER databases (Kleipool et al., 2008; Tilstra et al., 2017). For the OMI surface LER database an overall accuracy of 0.01–0.02 was reported, with slightly increasing values towards the shorter wavelengths (Kleipool et al., 2008). The SCIAMACHY surface LER database was shown to have an accuracy of about 0.01 for the UV-VIS-NIR spectral range (Tilstra et al., 2017). These accuracies reflect the uncertainties caused by various error sources, such as errors in the radiometric calibration and the occurrence of cloud contamination in the databases. Errors brought about by the Lambertian-equivalent nature of the databases are not part of these uncertainties. It should also be noted that the OMI and SCIAMACHY surface LER databases are mostly representative for the time periods from which they were derived (OMI: 2005–2009; SCIAMACHY 2002–2012). Systematic changes in surface reflectivity occurring after these time periods, for instance due to changes in land use, are not covered by the databases and will result in errors.

As an alternative to the SCIAMACHY surface LER database we could also use the very similar GOME-2 surface LER database (Tilstra et al., 2017). However, the GOME-2 surface LER database does not provide information for the SWIR wavelength range, which is the primary reason for adopting the SCIAMACHY surface LER database instead. Alternatively, we can make use of MODIS surface BRDF (Gao et al., 2005) as input. Strictly speaking, this is not correct, because the DAK radiative transfer model expects Lambertian surface albedo input. Also, six of the seven MODIS bands for which BRDF parameters are available are outside the TROPOMI spectral coverage. However, MODIS BRDF can be used to filter out observations for which the Lambertian OMI and SCIAMACHY surface LER databases are not able to provide adequate surface

albedo input, see Sect. 5.2. The MODIS product that we use for this is the daily global MODIS MCD43C1 product (Schaaf and Wang, 2015). The spatial resolution of this product is $0.05^\circ \times 0.05^\circ$.

Another possible source of surface albedo is the ADAM database (Bacour et al., 2020). This database is based on MODIS BRDF, but provides interpolation to any wavelength in the wavelength range 240–4000 nm. This would address the issue related to the position of the seven MODIS wavelength bands. However, interpolation does lead to a lower quality and reliability, as explained in Bacour et al. (2020). Next to this, the ADAM database shows both underestimation and overestimation by a factors of 1.5–2 in backscatter directions (Bacour et al., 2020, p. 18). For water surfaces, the surface reflection is modelled, and bears no relationship with the MODIS BRDF data. This means that consistency in calibration between land and ocean is broken. For the above reasons we decided to use the OMI/SCIAMACHY approach.

10 3.5 Other input parameters

Other input parameters needed for the calculations are surface height, total ozone column, and water vapour column. The surface height and total ozone column are already available in the TROPOMI UVAI product (Stein Zweers, 2018), and are simply copied. The surface height therefore originates from the GMTED2010 (Global Multi-resolution Terrain Elevation Data) surface elevation database and the total ozone column from ECMWF (European Centre for Medium-Range Weather Forecasts) model data. The water vapour column is taken from daily CAMS (Copernicus Atmosphere Monitoring Service) fields.

4 Approach

4.1 Wavelength bands

The reflectance validation described in this paper was performed for a selection of 21 wavelength bands (see Table 2 for a list of the bands and their specifications). Figure 1 presents three reflectance spectra measured by TROPOMI over the Sahara desert, the Amazonian rainforest, and the Atlantic ocean. All three spectra were recorded on 20 February 2018 and in each case there were no clouds present at the time of measurement. The black dotted lines indicate the positions of the 21 wavelength bands that were defined. These wavelength bands cover the TROPOMI spectral range as much as possible while positioned such that there is minimal interference from the surrounding absorption bands.

Nevertheless, for the wavelength bands at 697 and 712 nm some impact of water vapour absorption is present. ~~From Fig. 1 it can be inferred that~~ However, the impact is relatively small and of little concern. In spite of this, the dependence on water vapour column is taken into account in the LUTs, as explained in Sect. 3.2 and 3.3. Below 325 nm the dependence of the reflectance on the (shape of the) ozone profile sets in (Tilstra et al., 2005; Cai et al., 2012). This prevents using a LUT approach and requires knowledge of the ozone profile. This can in principle be handled (e.g., van Soest et al., 2005) but it would complicate matters and it would increase the computational effort (and time needed) by the radiative transfer code considerably. As a result, the shortest wavelength band listed in Table 2 is the one positioned at 328 nm, and none of the wavelength bands are located in spectral bands 1 or 2.

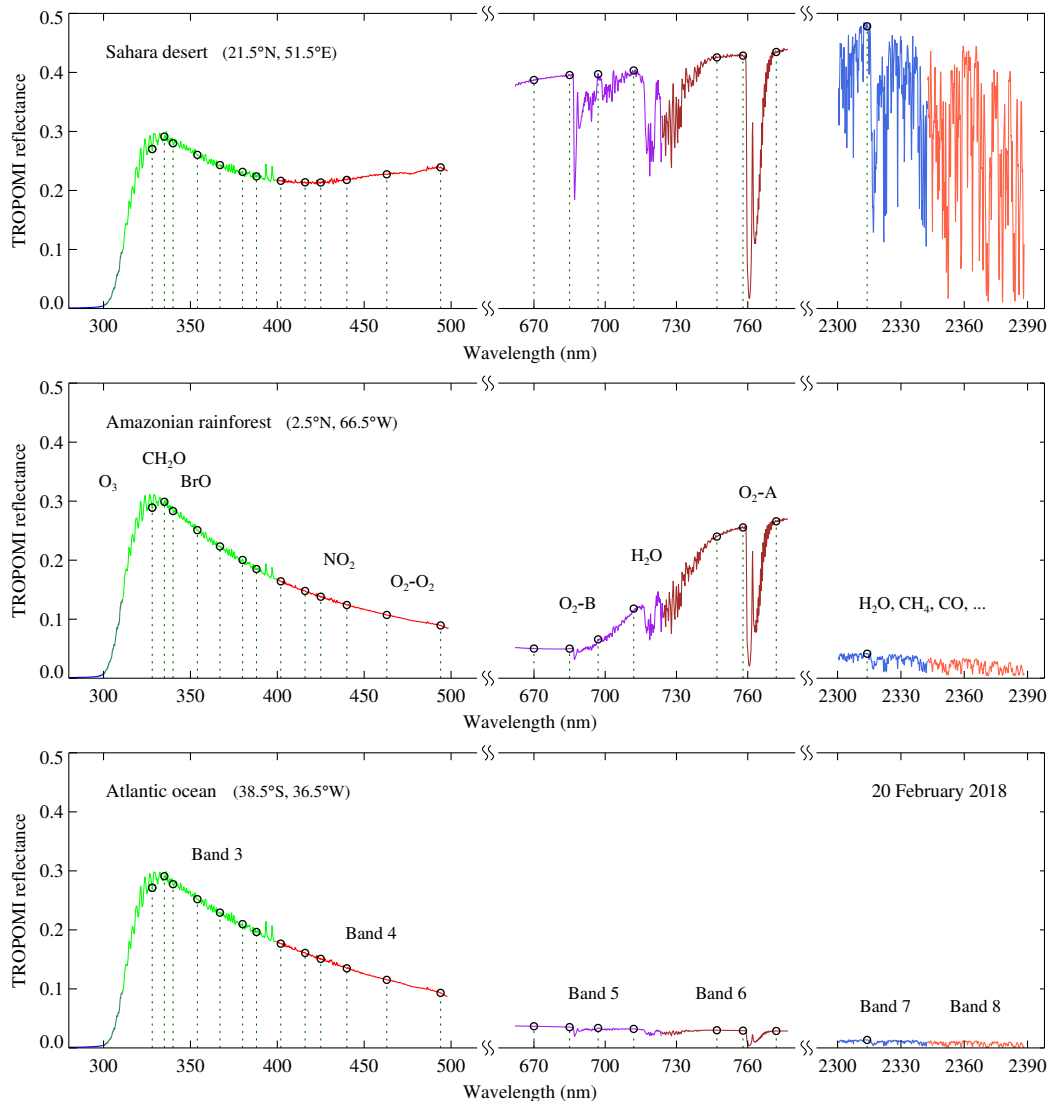


Figure 1. Reflectance spectra measured by TROPOMI on 20 February 2018 while observing the Sahara desert, the Amazonian rainforest, and the Atlantic ocean. In all three cases there was almost no cloud presence in the observed scene. Each TROPOMI spectral band was given an own colour. The black dotted lines indicate the position of the 21 wavelength bands that were selected for this study. In the middle panel, absorption bands are indicated by labels of the corresponding trace gases.

For TROPOMI spectral band 8 no suitable wavelength band can be defined because of strong absorption in the entire wavelength range, predominantly by water vapour. The reflectance validation study is, in conclusion, limited to spectral bands 3–7. Note that the wavelength band at 670 nm is situated outside the design spectral coverage of the NIR detector. This is not a problem because the detectors still perform well a fair bit outside the wavelength range for which they were designed.

4.2 Cloud screening

Cloud filtering is performed on the basis of the S5P NPP-VIIRS cloud information product. This product is based on measurements performed by the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument on board the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite. The Suomi NPP satellite is kept in an orbit very similar to that of S5P but with a headstart of about 3 minutes, resulting in a small time difference between the observations made from the two platforms. The S5P NPP-VIIRS product provides cloud information, measured by VIIRS, for each of the TROPOMI footprints. It reports, amongst other things, the number of VIIRS observations which were confidently clear ($N_{c,clr}$), probably clear ($N_{p,clr}$), probably cloudy ($N_{p,cld}$), and confidently cloudy ($N_{c,cld}$) (Siddans, 2016).

We define a geometrical cloud fraction c_f as the number of (confidently plus probably) cloudy S5P-VIIRS observations divided by the total number of S5P-VIIRS observations:

$$c_f = \frac{N_{c,cld} + N_{p,cld}}{N_{c,clr} + N_{p,clr} + N_{c,cld} + N_{p,cld}} \quad (4)$$

The thresholds that are used to filter out cloudy scenes are defined in Sect. 4.4.

4.3 Aerosol screening

Aerosol screening is performed using the Absorbing Aerosol Index (AAI) (Torres et al., 1998; de Graaf et al., 2005). We use a relatively high threshold value of 2 index points. This only removes scenes containing the highest concentrations of absorbing aerosol. We do not use the S5P AAI product (Stein Zweers, 2018), but instead calculate the AAI ourselves using the method described in Tilstra et al. (2012). Although the two methods are very comparable, in Sect. 4.4 observations are grouped in boxes and combined, and it is better to calculate the AAI from the average reflectances rather than to average the AAI values. Additionally, we ~~inserted~~ ~~insert~~ corrections to the reflectances which ~~were~~ ~~are~~ fed to the module that ~~calculated~~ ~~calculates~~ the AAI values. ~~This is to~~ ~~These corrections~~ prevent offsets and trends that would otherwise be present in the AAI.

The corrections consist of a time-dependent multiplicative correction to account for the linear drift observed in the 340/380-nm reflectances that are stored in the S5P AAI product, and of a day-zero multiplicative reflectance correction. The linear drift was determined using the method presented in Tilstra et al. (2012). The day-zero correction factors were found from applying the reflectance validation study on 1 May 2018 (“day zero”) and determining the correction factors found for 340 and 380 nm. In principle this step could be repeated by applying the correction factors, calculating newer correction factors from these, until convergence is reached. An iterative scheme for this was not pursued, however, because a secondary iteration turned out to already provide almost the same correction factors as the first one.

4.4 Scene selection

The selection of suitable clear-sky scenes is conducted in a way best explained by Fig. 2. The Earth’s surface is represented by a $1^\circ \times 1^\circ$ latitude/longitude grid and in Fig. 2 a small part of this grid is shown. The decision to scale down to $1^\circ \times 1^\circ$ latitude/longitude boxes was made with the available surface albedo databases in mind. The course-resolution OMI and

SCIAMACHY surface albedo databases (see Sect. 3.4) are not able to provide realistic surface reflectance input beyond the spatial resolution defined by their grids. The small dots in Fig. 2 represent TROPOMI measurement footprints. We look for $1^\circ \times 1^\circ$ latitude/longitude boxes, such as the one shown in the middle, which are sufficiently cloud-free. This is done by first calculating the geometrical cloud fraction determined from all S5P NPP-VIIRS footprints inside the box, using Eq. (4).
5 The box as a whole is considered cloud-free if this cloud fraction is below 0.03. After that, the geometrical cloud fraction is determined for the surrounding $3^\circ \times 3^\circ$ latitude/longitude box. If this cloud fraction is below 0.05, then the middle box is not only considered to be cloud-free, but also to be unaffected by cloud shadows from clouds in neighbouring boxes.

The next step is the calculation of the radiance spectrum for the middle box. This is done for each spectral band separately. Co-registration problems that exist between spectral bands of different detectors (Kleipool et al., 2018) are removed automati-
10 cally because for each spectral band only the footprints that fall inside the box are combined. After that, the Earth reflectance is determined using Eq. (1). Examples of spectra are shown in Fig. 1. Next, the band reflectances belonging to the 21 wavelength bands are determined. These correspond to the black circles in Fig. 1.

From the band reflectances at 340 and 380 nm the AAI is calculated. If the AAI is lower than 2 then the box is considered aerosol-free, otherwise the box is skipped. Other reasons to skip a box are inhomogeneity (i.e. the box contains both land
15 and water), snow/ice in the box, if the box is too close to the polar regions (absolute latitudes above 60°), if the solar zenith angle is above 75° , if the viewing zenith angle is above 40° , or if the observations are affected by sun glint or solar eclipse events. [Inhomogeneity and snow/ice presence are determined based on the surface type and surface condition indicators inside the TROPOMI level-2 products. A box is only considered inhomogeneous if it contains both land and water.](#) The number of boxes found this way amounts to about a thousand per day. This corresponds to about 2–3% of all latitude/longitude boxes
20 between 60°S and 60°N . This percentage is in agreement with the percentage that is to be expected for boxes with an area of $\sim 10,000 \text{ km}^2$ (Krijger et al., 2007). Figure 3 presents an example of the location and frequency of clear-sky boxes that were collected over the period from May 2018 until ~~February~~[May](#) 2020.

The choice of only accepting boxes for which the viewing zenith angle is below 40° is also related to the surfaced albedo databases. We expect the OMI database to be mostly representative for the near-nadir viewing geometries, because the al-
25 gorithm that was used to retrieve the OMI (and SCIAMACHY) surface LER is in most cases focused on the lowest scene reflectances (Kleipool et al., 2008; Tilstra et al., 2017). The SCIAMACHY surface LER database is also mostly representative for the near-nadir viewing geometries, for the simple fact that in the SCIAMACHY orbit swath the viewing zenith angles go up to 40° at most.

After the scene selection and the calculation of the 21 measured band reflectances these band reflectances are simulated in
30 the manner described in Sect. 3. The last step, reduction of the impact of surface albedo errors, is discussed in Sect. 5.

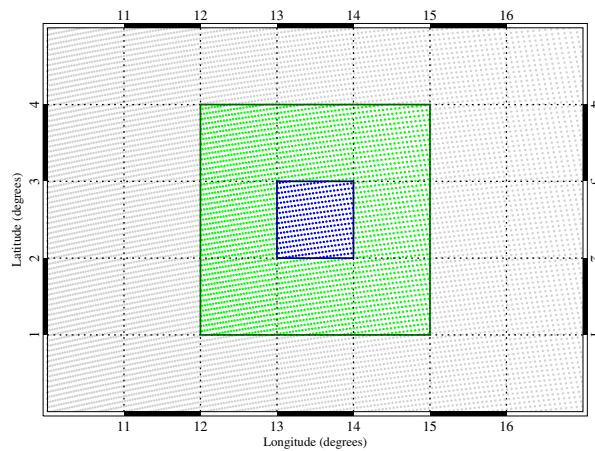


Figure 2. Illustration of the approach followed in this paper. The TROPOMI measurement footprints are represented by small dots. We focus, however, on the $1^\circ \times 1^\circ$ latitude/longitude box in the middle. If the geometrical cloud fraction derived from the S5P NPP-VIIRS cloud information of the blue footprints is below 0.03 then the NPP-VIIRS geometrical cloud fraction is determined for the surrounding $3^\circ \times 3^\circ$ latitude/longitude box. If this cloud fraction is below 0.05 then the middle box is considered cloud-free. If it is also considered aerosol free then the TROPOMI reflectance spectrum is calculated for the middle box. This is done for each of the spectral bands. More detailed information is provided in the main text.

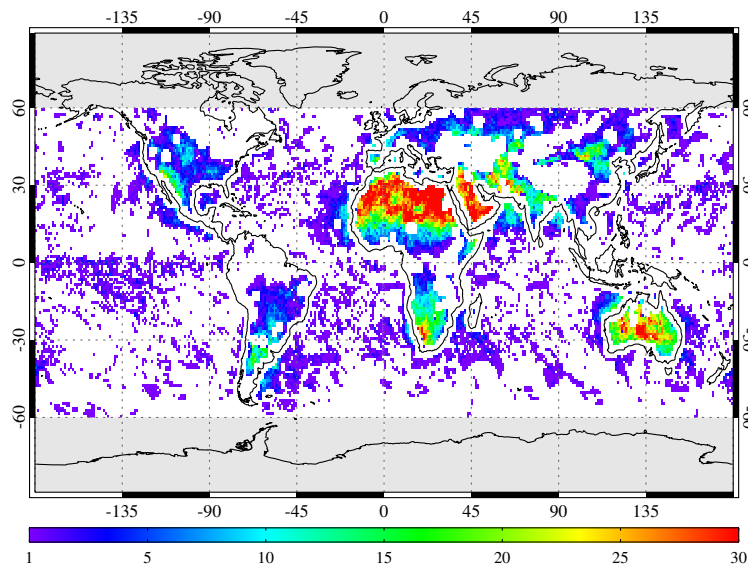


Figure 3. Location and frequency of the $1^\circ \times 1^\circ$ cloud-free latitude/longitude boxes that were selected by the algorithm every two weeks for the period from May 2018 until February-May 2020. [A total of 56 days contributed to the image. The selection and filtering approach is described in Sect. 4.4.](#)

5 Methods for reducing surface albedo errors

5.1 Surface albedo errors

As explained in Sect. 3.4, the largest error source in the radiative transfer simulations is the surface albedo. Surface albedo input is taken from the OMI/SCIAMACHY surface LER database. To study the inevitable errors brought about by adopting non-directional Lambertian surface albedo, we study the FRESKO (Wang et al., 2008, 2012) cloud information also stored while processing the TROPOMI data. In Fig. 4 we plot the (effective) cloud fraction retrieved by the TROPOMI FRESKO algorithm for all clear-sky latitude/longitude boxes on 1 May 2018. The TROPOMI FRESKO cloud fraction from boxes located over land is plotted in green, for boxes located over the ocean blue is used. The NPP-VIIRS geometrical cloud fraction is also plotted, in red. The NPP-VIIRS cloud fraction is always close to zero, because it was used for the cloud masking. The FRESKO cloud fraction over the ocean is also close to zero, as expected. However, the FRESKO cloud fraction over land deviates quite a lot from zero.

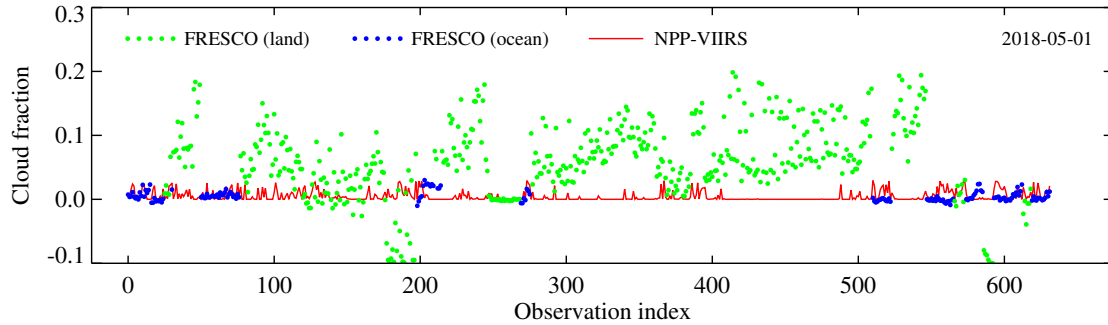


Figure 4. Cloud fractions calculated from TROPOMI FRESKO and NPP-VIIRS observations for all cloud-free $1^\circ \times 1^\circ$ latitude/longitude boxes on 1 May 2018. The FRESKO data are plotted in green over land and in blue over the ocean. The NPP-VIIRS cloud fractions, plotted in red, are all close to zero, as expected, because they were used for the cloud screening. The large deviations from zero found in the FRESKO cloud fractions over land are primarily the result of inadequate (non-directional) surface albedo information provided to the FRESKO retrieval algorithm, resulting in large errors in the FRESKO cloud fractions for certain scattering geometries.

The deviations found in the cloud fractions over land are in primarily caused by a systematic underestimation of the surface reflectivity for certain geometries in the (Lambertian) surface albedo database that is used in the FRESKO algorithm. This explanation is in line with conclusions from a recent study by Lorente et al. (2018), who showed that traditional Lambertian surface albedo databases, like the ones described in Sect. 3.4, seriously underestimate the surface reflectivity for certain geometries due to the fact that surface reflectance anisotropy is not accounted for in these databases. The error made this way can be as large as a factor of two in the surface albedo over forested scenes at 772 nm. For cloud retrieval algorithms that operate in the near-infrared wavelength range, like the FRESKO algorithm which makes use of the O_2 -A band near 760 nm, the errors in the cloud fractions can be as large as 0.2 and even higher (Lorente et al., 2018, Fig. 7b).

5.2 Selecting the best surface albedo method

Four methods are employed to model the surface reflectivity and/or to apply additional filtering:

1. Method 1 is the standard case with OMI/SCIAMACHY surface LER as surface albedo input.
2. Method 2 additionally performs filtering on the FRESCO cloud fraction. Observations are excluded if the absolute cloud fraction exceeds 0.06. In Sect. 5.1 it was made plausible that the deviations from zero in Fig. 4 are primarily the result of the usage of a non-directional surface LER database. Cloud fractions that deviate too much from zero are therefore assumed to be indicative for incorrect surface albedo input.
3. Method 3 uses MODIS BRDF as surface albedo input. As explained in Sect. 3.4, this is not possible for most of the 21 wavelength bands. We use MODIS band 1 (centred around 645 nm) to estimate the surface BRDF at 670 nm for the scattering geometry defined by θ , θ_0 , and $\phi - \phi_0$ and use it in the normal baseline which expects Lambertian surface albedo input. This is not entirely correct, but at 670 nm, where the amount of Rayleigh scattering in the atmosphere is low, surface-only reflection is dominating for clear-sky scenes and the above procedure is actually quite a fair approximation.
4. Method 4 filters out cases for which the Lambertian OMI/SCIAMACHY surface LER does not agree with the directional MODIS surface BRDF. Figure 5 presents the SCIAMACHY surface LER, determined for 645 nm, versus the surface BRDF from MODIS band 1, for all cloud-free land scenes observed on 1 May 2018. The differences can be rather large and we attribute these to the non-directional nature of the SCIAMACHY surface LER database. Method 4 calculates the parameter δ_{645} , which represents the difference between MODIS surface BRDF and SCIAMACHY surface LER for a given scene at 645 nm:

$$\delta_{645} = A_{\text{BRDF}} - A_{\text{LER}} \quad (5)$$

As we are only interested in land scenes for which the SCIAMACHY surface LER is in reasonable agreement with the MODIS BRDF we only continue with land scenes for which we have $|\delta_{645}| \leq 0.02$. These scenes correspond to the red circles in Fig. 5.

For each of these four methods the baseline method described in Sect. 4 was put to work on a full day of TROPOMI data from 1 May 2018 and complemented with radiative transfer simulations as described in Sect. 3. The results for the 670-nm wavelength band are shown in Fig. 6. The green data points represent observations taken over land, the blue data points denote scenes containing water surfaces.

For Method 1, there seems to be a fair agreement between measurement and simulation. Pearson's correlation coefficient r is 0.989, indicating high correlation. The linear fit to the data points (represented by the grey line) is quite close to the one-to-one relationship (indicated by the dotted line). However, the spread in the data points is rather large. The standard deviation of the data points w.r.t. the aforementioned linear fit amounts to $\sigma = 0.022$.

For Method 2, the scatter in the scatter plot has clearly been reduced, from $\sigma = 0.022$ to $\sigma = 0.016$. Pearson's r increased slightly from 0.989 to 0.995. The effect on the coefficients of the linear fit is small but significant.

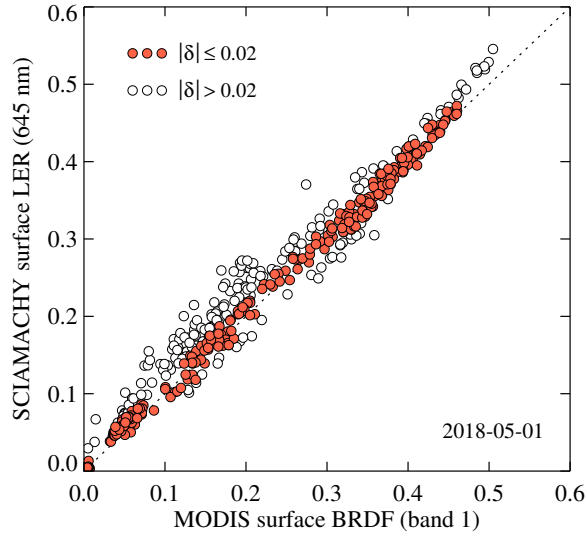


Figure 5. SCIAMACHY surface LER at 645 nm versus MODIS surface BRDF from MODIS band 1 (centred around 645 nm), for cloud-free conditions and land surfaces observed by TROPOMI on 1 May 2018. The red circles indicate observations for which there is good agreement ($|\delta_{645}| \leq 0.02$). These are used in Method 4.

For Method 3, the correlation between TROPOMI measurement and DAK simulation is very good. Pearson’s correlation coefficient r now amounts to 0.998 and the standard deviation σ dropped down to 0.008. Note that there are no water scenes present, because MODIS BRDF is available primarily over land. If water scenes would have been removed from the Method 1 and Method 2 analyses, then their σ values would have increased. The improvement of Method 3 in terms of σ is therefore substantial. However, the slope $m = 1.031$ of the linear fit is meaningless because the BRDF from MODIS band 1 is representative for 645 nm, not for 670 nm. Using MODIS BRDF directly is therefore a good illustration of the most optimal situation that can be achieved in terms of r and σ , but it is not a practical way of improving the comparison.

Finally, for Method 4, the reliability of the comparison has improved compared to Methods 1 and 2. The correlation coefficient r is 0.997 and, more importantly, the standard deviation σ went down to 0.012. These numbers are close to the numbers found for Method 3.

5.3 Final method

Method 4 essentially filtered out scenes with surface scattering geometries which were not well represented by the (Lambertian) SCIAMACHY surface LER database. This is now done for all the longer wavelength bands (670 to 2314 nm). For the shorter wavelength bands (328 to 494 nm) the OMI surface LER database is used, but now we use MODIS band 3 (centred around 469 nm) to filter out scenes for which $|\delta_{469}| > 0.025$. We use a higher threshold value, because the difference between BRDF and LER is higher due to the fact that multiple scattering has become more important.

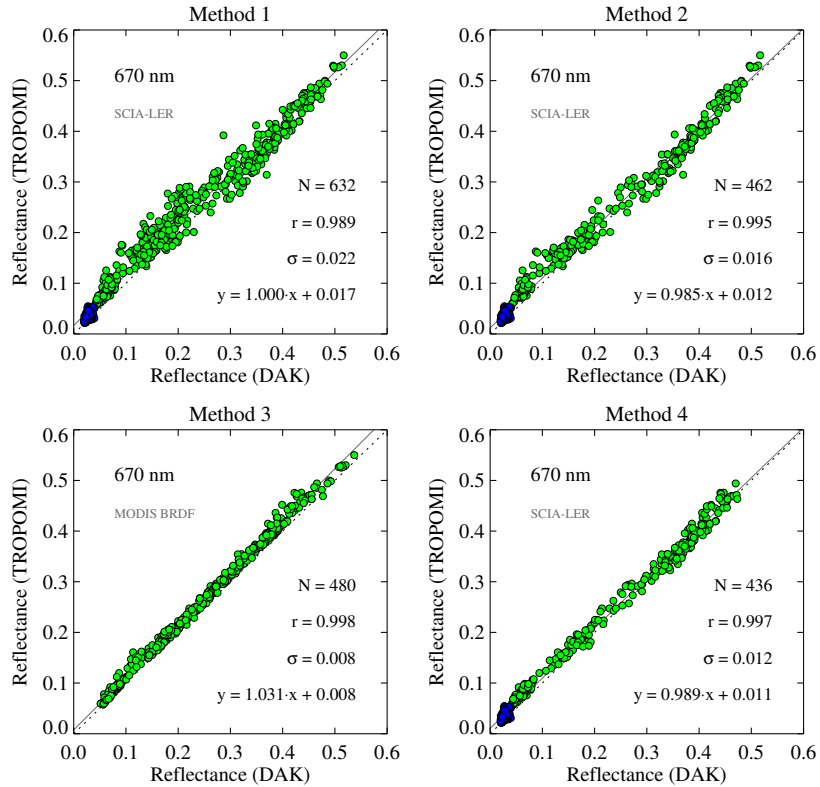


Figure 6. TROPOMI reflectance versus DAK simulated reflectance at 670 nm for cloud-free conditions on 1 May 2018. Four methods are employed to model the surface reflectivity and to apply filtering. Method 1 is the standard case with SCIAMACHY surface LER as surface albedo input. Method 2 is the same but with an additional filtering based on the FRESCO cloud fraction. Method 3 uses MODIS BRDF at 645 nm as albedo input. Method 4 uses SCIAMACHY surface LER but with a filtering requiring that the SCIAMACHY surface LER and MODIS BRDF at 645 nm are in good agreement. The green circles represent cloud-free scenes over land and the blue circles correspond to cloud-free water scenes. Further explanation is provided in the main text.

6 Results

6.1 First analysis

In Fig. 7 we present the results for the final method defined in Sect. 5.3 for 12 of the 21 wavelength bands. Each window contains a scatter plot with linear fit, fit coefficients, standard deviation σ , correlation coefficient r , and the surface LER database used. The shortest wavelength band (328 nm) shows a very low σ value of 0.003. Here the sensitivity of the reflectance to errors in the surface reflectance is the lowest (Tilstra et al., 2005, Fig. 4). The linear fit has near zero offset ($n = 0.000 \pm 0.016$) but the slope deviates significantly from one ($m = 1.119 \pm 0.056$). As the wavelength increases, the simulations become more sensitive to errors in the surface albedo and this is reflected in the increase in the standard deviation σ , which goes up to 0.008

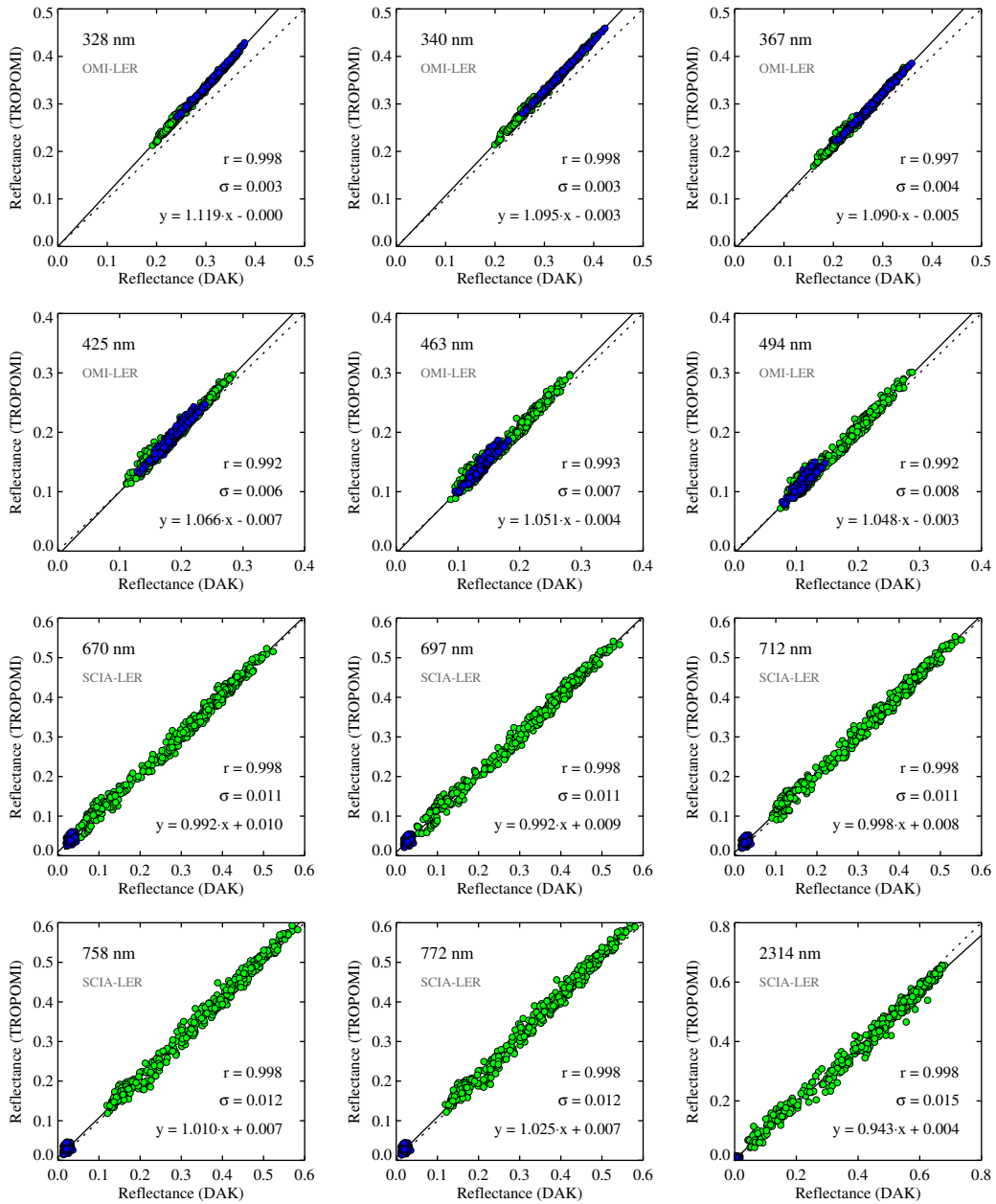


Figure 7. Reflectances measured by TROPOMI on 30 October 2019 versus simulated reflectances calculated using the DAK radiative transfer model, for 12 of the 21 wavelength bands studied in this paper. Only clear-sky observations were used in the scatter plots. As before, green circles represent clear-sky scenes over land and blue circles correspond to clear-sky water scenes. For the wavelength bands below 500 nm the surface albedo is taken from the OMI surface LER database; above 500 nm surface albedo input from the SCIAMACHY surface LER database is used. The analysis was performed using Method 4. The black lines represent linear fits $y = m \cdot x + n$ to the data. The fit results for m and n and the standard deviation σ of the data points w.r.t. the linear fit are shown in each window.

at 494 nm. At this wavelength the linear fit deviates less severely from the one-to-one relationship: $m = 1.048 \pm 0.041$ and $n = -0.003 \pm 0.006$.

For the longer wavelength bands (670 to 2314 nm) the results are notably different. There is more spread in the data points, but almost near perfect correlation ($r = 0.998$ in all cases). More importantly, the one-to-one relationship is respected for all wavelength bands except 2314 nm. The special case of 2314 nm will be discussed in Sect. 6.3. For the wavelength bands 670 to 772 nm from TROPOMI spectral bands 5/6 the slope m of the linear fit is largely equal to one within the accuracy of the comparison. For example, at 670 nm, $m = 0.992 \pm 0.014$. At 758 nm, $m = 1.010 \pm 0.013$. Conclusions about the calibration of TROPOMI will be drawn in Sect. 6.4.

Studying the results we found no dependence on viewing zenith angle, or on any of the other angles, for any of the 21 wavelength bands that were studied. The existence of such a dependence could have indicated problems with the radiative transfer calculations at the shorter wavelengths, where the atmospheric contribution is large, or it could have pointed to errors introduced mainly for the longer wavelengths by the handling of surface reflection in the simulations. This suggests that the comparisons that were performed are sound.

6.2 Time series and trends

To study whether the results are representative we repeat the analyses for the period 2018–2020 by processing one day per fortnight. This results in 56 days to be analysed. This time we determine the fit coefficients of the linear fit through the data points in the scatter plots, as before, but also the average difference between TROPOMI and the DAK simulations. The results are shown in [Fig. 8](#) [the results are presented](#) for 5 of the wavelength bands. The first column shows the slope m of the linear fit, the second column the intercept n of the linear fit, and the third column the average difference d_m between the TROPOMI and DAK reflectances. Data from the commissioning phase are shown in red and do not contribute in any way to the analyses.

At 328 nm a small upward trend is visible in both the slope m and intercept n of the linear fit and even more clearly in the parameter d_m . The value of d_m reaches 0.035, which is quite a large difference considering the fact that the typical value of the reflectance at this wavelength is about 0.3 (see Fig. 1). The upward trend in d_m goes down as the wavelength increases, until 494 nm, where it is nearly completely absent. At 494 nm there is a small downward trend in the slope m , but the small upward trend in the intercept n is comparable to the 328-nm case. The slope m is, apparently, changing with wavelength, but the intercept n is not.

For the longer wavelength bands no clear trends are visible. At 670 nm the average slope \bar{m} is very close to one (~~$\bar{m} = 0.998 \pm 0.009$~~ $\bar{m} = 0.998 \pm 0.009$) but the average intercept \bar{n} suggests a slight offset ($\bar{n} = 0.010 \pm 0.003$). The offset is present in both the land and sea portions of the data, and in Fig. 6 it was also present when MODIS surface BRDF was used as surface albedo input for the 670-nm wavelength band. As a result, the offset does not seem to be caused by surface type and/or errors in the surface albedo. At ~~772~~758 nm the average slope is larger than one (~~$\bar{m} = 1.033 \pm 0.009$~~ $\bar{m} = 1.015 \pm 0.008$) and the intercept is small ($\bar{n} = 0.004 \pm 0.003$). At 772 nm (not shown in Fig. 8) the average slope has increased further ($\bar{m} = 1.033 \pm 0.008$) and the

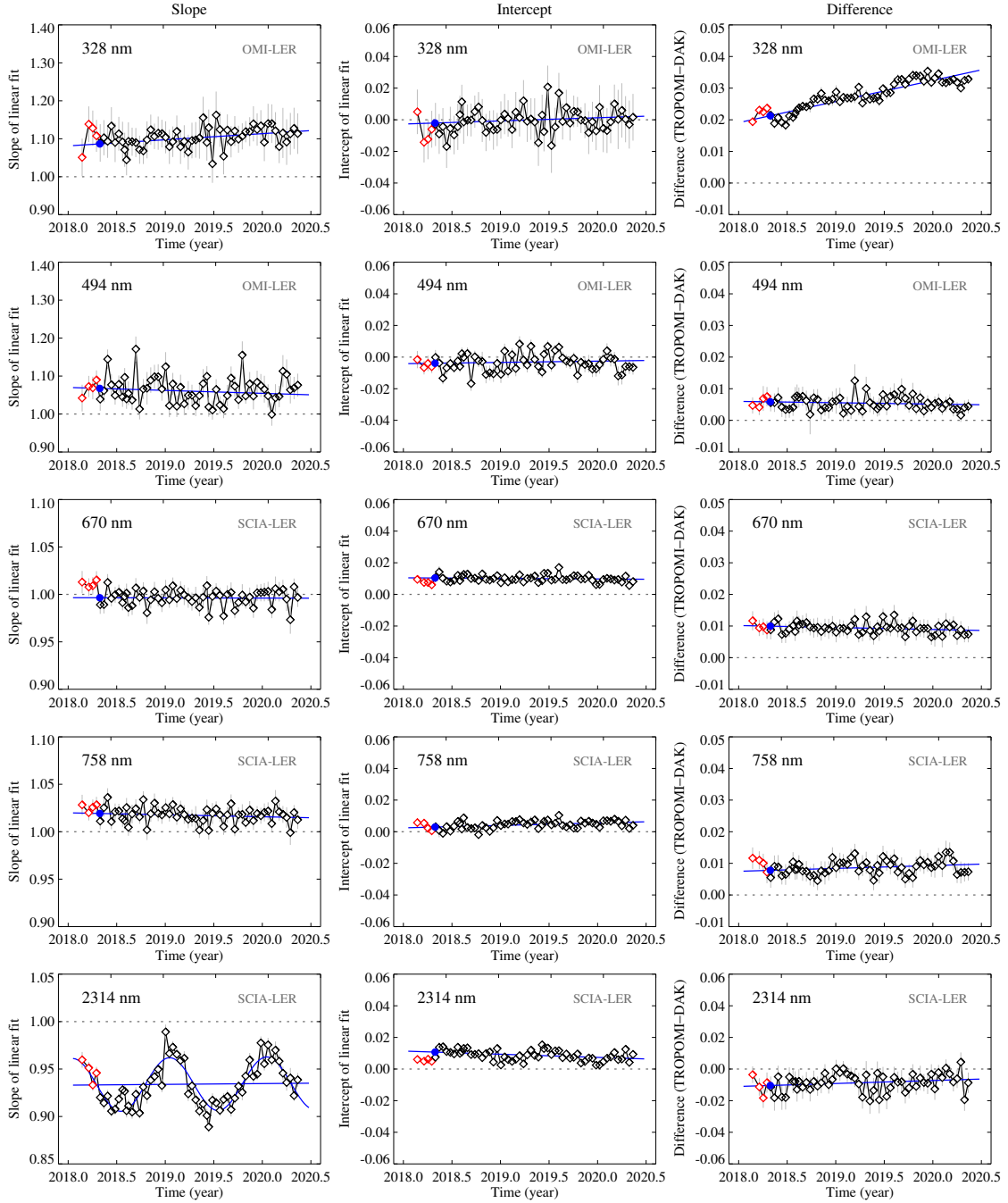


Figure 8. Time series of the slope m (left column) and intercept n (middle column) of the linear fit to the data in scatter plots such as shown in Fig. 7, for 5 of the 21 wavelength bands defined in this paper. The last column presents the mean differences d_m between the TROPOMI and DAK reflectances. The red data points are based on data from the commissioning phase. These are plotted but not taken into account in the linear fits. In all cases the horizontal dotted line represents the situation of perfect agreement between measurement and simulation. The blue lines/curves are fits to the data and are discussed in Sect. 6.2. The blue circle illustrates the fit result for 30 April 2018.

intercept is now very close to zero ($\bar{n} = 0.003 \pm 0.002$). The lower value of the average intercept \bar{n} as compared to the 670-nm wavelength band shows that the offset found at 670 nm cannot be related to residual clouds.

For the wavelength band at 2314 nm the results are very different. The time series of the intercept n shows no clear trend and there is only a modest average offset ($\bar{n} = 0.009 \pm 0.004$). The slope m , however, shows a seasonal cycle with a period T of one year. We therefore fit a function of the following form to the data points:

$$m(t) = a_0 + a_1 t + a_2 \sin(2\pi t/T + a_3) , \quad (6)$$

where t is the time in years. The fit results indicate only a negligible trend: ~~$a_1 = -0.0004 \pm 0.0049$~~ $a_1 = 0.0008 \pm 0.0042 \text{ yr}^{-1}$. The fitted function is plotted in blue in Fig. 8 for the 2314 nm case. The behaviour of this wavelength band will be discussed further in Sect. 6.3.

10 6.3 SWIR channel (2314 nm)

In Fig. 9 we present scatter plots of TROPOMI reflectance versus DAK reflectance for two days: 16 January 2019 (left column) and 13 June 2019 (right column). Focusing for the moment only on the scatter plots in the top row of Fig. 9, we see that for the first day the data are reasonably close to the one-to-one line. For the second day the data are significantly deviating from the one-to-one relationship. The two days illustrate the situation near the top and near the bottom of the sinusoidal dependence shown in Fig. 8 for the 2314-nm wavelength band. The question to address is whether the sinusoidal behaviour is originating from the TROPOMI data or from the radiative transfer modelling.

In-flight monitoring results for the TROPOMI SWIR channel have appeared in a recent paper by van Kempen et al. (2019). This paper reports a radiometric stability of the SWIR channel within the 0.1% level, a result which was based on analyses of TROPOMI's DLED and WLS observations. These calibration monitoring measurements are in fact representative for the Earth radiance light path. We found the solar irradiance to also be stable within the 0.1% level (not shown). The conclusion is that the sinusoidal behaviour found in Fig. 8 cannot be not caused by TROPOMI. It must therefore be caused by the radiative transfer calculations. The most logical explanation would then be that there is an issue with the 2314-nm wavelength band of the SCIAMACHY surface LER database. This would immediately explain the period T of a year that was reported in Sect. 6.2.

The 2314-nm wavelength band of the SCIAMACHY surface LER database was retrieved from SCIAMACHY band 8. This band was troubled by the build up of an ice layer on the detector, and by the many attempts to remove this ice layer by heating up the detector (Lichtenberg et al., 2006). This had an impact on the throughput of the detector. However, in deriving the SCIAMACHY surface LER database, necessary corrections were applied to the 2314-nm reflectances. These corrections cannot possibly have introduced the sinusoidal behaviour, and the ice-layer related problems themselves are of a completely different nature than the clear sinusoidal relationship that is found in Fig. 9. However, there could be other causes for errors in the 2314-nm SCIAMACHY surface LER. Proof that the SCIAMACHY surface LER database is responsible for the sinusoidal behaviour was not found.

Another potential explanation for the sinusoidal behaviour could be that the approach of filtering out geometries according to Method 4 (see Sect. 5.2) breaks down at 2314 nm. To test this hypothesis, we also performed the radiative transfer calculations

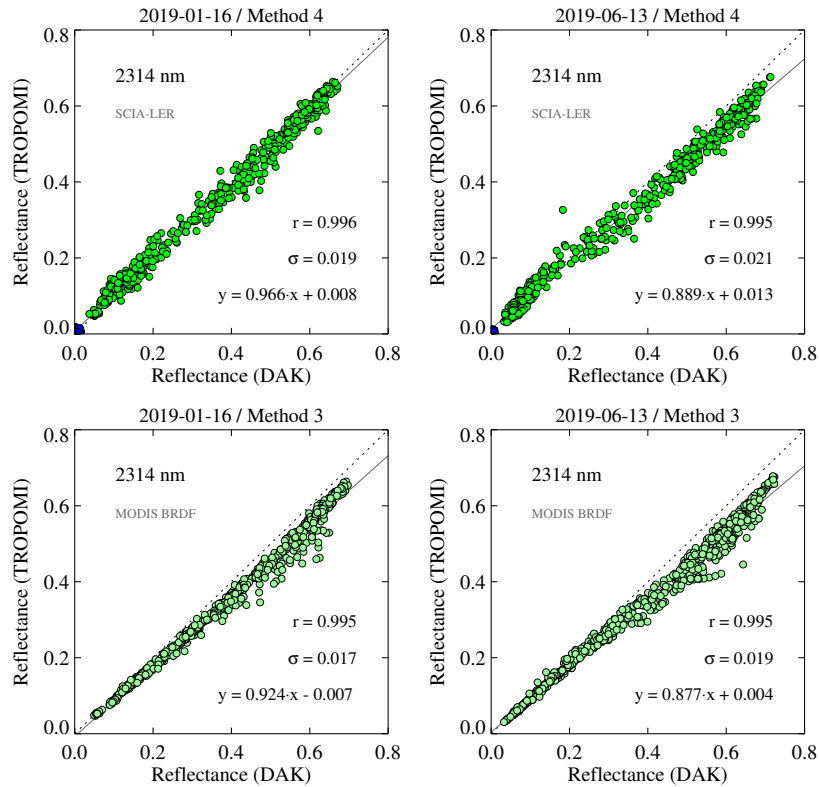


Figure 9. Scatter plots for the 2314-nm wavelength band for 16 January 2019 (left column) and 13 June 2019 (right column). In the top row the surface albedo used is that of the SCIAMACHY surface LER database, and filtering is applied according to Method 4. For 13 June 2019 the simulated DAK reflectance overestimates the TROPOMI reflectance much more than it does for 16 January 2019. In the bottom row the surface albedo is that of band 7 of the MODIS surface BRDF database (central wavelength 2.1 μm). It seems that using MODIS surface BRDF also leads to a larger overestimation of the simulated DAK reflectance for 13 June 2019 than for 16 January 2019.

using MODIS surface BRDF from band 7 (centred around 2.1 μm) as surface albedo input for the radiative transfer calculations. The results are shown in the bottom row of Fig. 9. Obviously, the surface reflectivity at 2.1 μm is somewhat different from that at 2314 nm, so a clear one-to-one relationship cannot be expected. But comparing the scatter plots of the upper and lower row it can be seen that in both cases the DAK simulations overestimate the reflectance compared to TROPOMI more for 13 June 2019 than for 16 January 2019. This suggests that the sinusoidal behaviour is (1) not caused by errors in the SCIAMACHY surface LER database and (2) also not caused by shortcomings in the additional filtering we apply in Method 4.

Looking further for potential discrepancies in the radiative transfer calculations is also not providing us with an explanation. In the wavelength region surrounding 2314 nm absorption by water vapour is an issue. However, the selected wavelength is precisely located in a continuum part of the spectrum. Tests involving alterations of the slit function and wavelength assignment

showed that the 2314-nm wavelength is well chosen and that its reflectance is stable. Unfortunately, all of this leaves the sinusoidal behaviour unexplained.

6.4 Summary and interpretation of the results

We conclude that all 21 selected wavelength bands show the correct linear response to the intensity of the detected radiation.

- 5 Table 3 summarises the end results for the validation of the absolute radiometric calibration. The reported slopes, intercepts, deviations, and their errors are all representative for 30 April 2018. For TROPOMI, this day is important because it is the first mission day after the end of its commissioning phase. The results from the time series presented in Fig. 8 were extrapolated via the linear fits to this particular day, which is indicated by the blue circles in Fig. 8.

The last row in Table 3 presents the percentage $D_{1.0}$, where $D_{1.0}$ for an analysed day is defined as

$$10 \quad D_{1.0} = 100\% \cdot (m + n - 1) . \quad (7)$$

Table 3. Final results from the reflectance validation study described in this paper. For each of the 21 wavelength bands the differences w.r.t. simulations are given. The slopes, intercepts, differences, and their uncertainties are representative for the clear-sky scenes that were studied and best describe the situation for 30 April 2018, the first day after the end of the TROPOMI commissioning phase. The percentage $D_{1.0}$ deviates from this in that it was determined for the entire time period that was studied (for which it is representative).

Wavelength band (nm)	328	335	340	354	367	380	388	402	416	425	440
Spectral band	3	3	3	3	3	3	3	4	4	4	4
Surface albedo input	OMI	OMI	OMI	OMI	OMI	OMI	OMI	OMI	OMI	OMI	OMI
Slope	1.086	1.072	1.057	1.062	1.059	1.075	1.059	1.062	1.068	1.066	1.073
Slope uncertainty	0.024	0.023	0.023	0.026	0.029	0.033	0.036	0.040	0.044	0.044	0.043
Intercept	-0.002	-0.001	-0.001	-0.004	-0.003	-0.004	-0.005	-0.007	-0.007	-0.007	-0.008
Intercept uncertainty	0.007	0.008	0.007	0.008	0.008	0.008	0.008	0.009	0.009	0.008	0.008
Difference	0.021	0.020	0.015	0.012	0.011	0.012	0.007	0.005	0.005	0.004	0.004
Difference uncertainty	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Percentage $D_{1.0}$	10.1%	8.8%	7.3%	7.4%	7.1%	8.5%	6.7%	6.5%	6.9%	6.5%	6.7%
Wavelength band (nm)	463	494		670	685	697	712	747	758	772	2314
Spectral band	4	4		5	5	5	5	6	6	6	7
Surface albedo input	OMI	OMI		SCIA	SCIA	SCIA	SCIA	SCIA	SCIA	SCIA	SCIA
Slope	1.077	1.067		0.996	0.999	0.992	1.001	1.014	1.019	1.034	0.918
Slope uncertainty	0.040	0.036		0.010	0.009	0.010	0.009	0.009	0.008	0.008	0.022
Intercept	-0.007	-0.004		0.011	0.011	0.011	0.007	0.006	0.003	0.003	0.011
Intercept uncertainty	0.007	0.006		0.003	0.003	0.003	0.002	0.003	0.002	0.002	0.004
Difference	0.005	0.006		0.010	0.011	0.009	0.008	0.010	0.008	0.012	-0.010
Difference uncertainty	0.002	0.002		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.006
Percentage $D_{1.0}$	7.0%	6.1%		0.7%	1.0%	0.5%	1.0%	2.2%	2.2%	3.7%	-5.7%

In Eq. (7) the m and n are the slope and intercept as defined in Sect. 6.1. The percentage $D_{1.0}$ therefore represents the estimated calibration error in the TROPOMI reflectance for a case with a reflectance of 1, expressed in a percentage. Contrary to the other results presented in Table 3, we did not make $D_{1.0}$ representative for 30 April 2018, but instead calculated $D_{1.0}$ for the entire time period that was studied. The $D_{1.0}$ therefore also include the effects of instrument degradation since 30 April 2018.

5 We can now interpret the results presented in Table 3 and draw conclusion about the radiometric calibration of the TROPOMI Earth reflectance. We start with the selected wavelength bands in TROPOMI spectral bands 5 and 6 (wavelength bands 670 to 772 nm). Here the accuracy of the comparison is estimated to be about 2–3%. We see that there is full agreement within the reported errors for all wavelength bands except at 772 nm. Here $D_{1.0}$ exceeds the accuracy of the comparison. We therefore conclude that the TROPOMI radiometric calibration is not meeting the TROPOMI calibration requirement of 2% at 772 nm.
10 For the other wavelength bands the radiometric calibration seems to be correct within 2%.

For the selected wavelength bands in TROPOMI spectral bands 3 and 4 (wavelength bands 328 to 494 nm) we find, depending on the wavelength, differences $D_{1.0}$ between 6% and 10%. The differences are much larger than the estimated accuracy of the method (1–3%), and therefore significant. The differences thus are to be interpreted as calibration errors in the TROPOMI reflectance. It should be noted that the magnitude of the errors is in agreement with radiometric calibration errors found recently in the TROPOMI solar irradiance product (Ludewig et al., 2020, Fig. 19). We conclude that for spectral bands 3 and 4
15 the radiometric calibration does not meet the TROPOMI calibration requirement of 2%.

The situation for the SWIR channel is described in Sect. 6.3. Based on the discussion in this section, we have to dismiss the results for the 2314-nm wavelength bands. However, we do believe that there are indications that the reflectance of TROPOMI spectral band 7 is too low, and that further checks are needed.

20 Detailed end results are provided in Table 3.

7 Conclusions

In this paper we studied the quality of the radiometric calibration of the TROPOMI instrument by comparing the reflectances observed by TROPOMI with reflectances simulated by the radiative transfer code DAK. This was done for clear-sky scenes, to avoid the (complicated) modelling of scenes containing clouds. Comparisons between satellite observations and radiative
25 transfer calculations have a limited accuracy, for a variety of reasons. For a large part of the wavelength range that was studied in this paper (328–2314 nm), the limiting factor in the comparison is the knowledge of the surface albedo.

There are a number of surface albedo databases available which provide the spectral surface reflectivity information needed as input for the radiative transfer calculations. However, these databases do not provide surface BRDF, but Lambertian, non-directional surface reflectance information. The established MODIS surface BRDF product provides an excellent kernel-based
30 expansion of the surface BRDF, but only for a few wavelength bands. The MODIS wavelength bands cannot be used for the intercomparison in this paper.

In this paper we have tried to increase the quality of the radiative transfer calculations by using a mixed approach in which we use OMI and SCIAMACHY surface LER as surface albedo input and apply filtering by comparing the surface LER values

with MODIS surface BRDF at a specific reference wavelength. This way, the radiative transfer simulation can make use of the spectral surface reflectivity information that is available and at the same time take advantage of the directional information contained in the MODIS surface BRDF product. The result is a higher accuracy for the intercomparison.

From the intercomparison between TROPOMI and the radiative transfer code DAK we can conclude that all 21 selected wavelength bands show the correct linear response to the intensity of the detected radiation. Especially the reflectances in spectral bands 5 and 6 show a good absolute agreement with the simulations. The radiometric calibration errors are on the order of 0–2%, within the accuracy estimate of the intercomparison, except for the 772-nm wavelength band, which shows larger deviations from the simulations. For this wavelength band there is a mild, but statistically significant, calibration issue.

The results for TROPOMI spectral bands 3 and 4 point to more severe problems in the radiometric calibration. For the selected wavelength bands (328 to 494 nm) we found differences with the simulations between 6% and 10%. The differences are larger than the estimated accuracy of the method (1–3%), and therefore clearly point to calibration errors leaking into the TROPOMI reflectance. In any case, for spectral bands 3 and 4 the radiometric calibration does not meet the TROPOMI requirement of 2%.

The analysis of the 2314-nm wavelength band located in the SWIR channel showed unexpected behaviour, which we attributed to the radiative transfer calculations. This means that we have to dismiss the results for TROPOMI spectral band 7. More research is needed to investigate the radiometric calibration of this spectral band, and of TROPOMI spectral bands 1 and 2.

Author contributions. LGT wrote the manuscript and performed most of the work. MdG and PW contributed to the cloud and aerosol screening. PS and PW helped with the radiative transfer modelling. All authors discussed the results and commented on the manuscript.

20 *Competing interests.* The authors declare that they have no conflict of interest.

Disclaimer. The results presented in this paper are based on version-1.0 Sentinel-5 Precursor data. All conclusions that were drawn in principle only apply to version 1.0 of these data.

25 *Acknowledgements.* The work presented in this paper was performed in the context of the Sentinel-5 Precursor Validation Team (S5PVT). We acknowledge the free use of TROPOMI data provided by ESA. The MODIS MCD43C1 data product was retrieved from the online Data Pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, <https://lpdaac.usgs.gov/tools/data-pool/>.

References

- Acarreta, J. R. and Stammes, P.: Calibration comparison between SCIAMACHY and MERIS onboard ENVISAT, *IEEE Geosci. Remote Sens. Lett.*, 2, 31–35, <https://doi.org/10.1109/LGRS.2004.838348>, 2005.
- Anderson, G. P., Clough, S. A., Kneizys, F. X., Chetwynd, J. H., and Shettle, E. P.: AFGL Atmospheric Constituent Profiles (0–120 km), Environ. Res. Pap. 954, Rep. AFGL-TR-86-0110, Air Force Geophys. Lab., Hanscom AFB, Mass., 1986.
- [Bacour, C., Bréon, F.-M., Gonzalez, L., Price, I., Muller, J. P., Prunet, P., Straume, A. G.: Simulating multi-directional narrowband reflectance of the Earth's surface using ADAM \(a surface reflectance database for ESA's Earth observation missions\), *Remote Sens.*, 12, 1679, <https://doi.org/10.3390/rs12101679>, 2020.](https://doi.org/10.3390/rs12101679)
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission objectives and measurement modes, *J. Atmos. Sci.*, 56, 127–150, 1999.
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weissenmayer, A., Richter, A., de Beek, R., Hoogen, R., Bramstedt, K., Eichman, K.-U., Eisinger, M., and Perner, D.: The Global Ozone Monitoring Experiment (GOME): Mission concept and first scientific results, *J. Atmos. Sci.*, 56, 151–175, 1999.
- Cai, Z., Liu, Y., Liu, X., Chance, K., Nowlan, C. R., Lang, R., Munro, R., and Suleiman, R.: Characterization and correction of Global Ozone Monitoring Experiment 2 ultraviolet measurements and application to ozone profile retrievals, *J. Geophys. Res.*, 117, D07305, <https://doi.org/10.1029/2011JD017096>, 2012.
- Chandrasekhar, S.: Radiative Transfer, Dover Publications, Mineola, New York, 1960.
- de Graaf, M., Stammes, P., Torres, O., and Koelemeijer, R. B. A.: Absorbing Aerosol Index: Sensitivity analysis, application to GOME and comparison with TOMS, *J. Geophys. Res.*, 110, D01201, <https://doi.org/10.1029/2004JD005178>, 2005.
- de Haan, J. F., Bosma, P. B., and Hovenier, J. W.: The adding method for multiple scattering calculations of polarized light, *Astron. Astrophys.*, 183, 371–391, 1987.
- Gao, F., Schaaf, C. B., Strahler, A. H., Roesch, A., Lucht, W., and Dickinson, R.: MODIS bidirectional reflectance distribution function and albedo Climate Modeling Grid products and the variability of albedo for major global vegetation types, *J. Geophys. Res.*, 110, D01104, <https://doi.org/10.1029/2004JD005190>, 2005.
- Hasekamp, O. P., Landgraf, J., and van Oss, R.: The need of polarization modeling for ozone profile retrieval from backscattered sunlight, *J. Geophys. Res.*, 107, 4692, <https://doi.org/10.1029/2002JD002387>, 2002.
- Heath, D. F., Krueger, A. J., Roeder, H. A., and Henderson, B. D.: The Solar Backscatter Ultraviolet and Total Ozone Mapping Spectrometer (SBUV/TOMS) for NIMBUS G, *Opt. Eng.*, 14, 323–331, <https://doi.org/10.1117/12.7971839>, 1975.
- Herman, J. R., Bhartia, P. K., Torres, O., Hsu, C., Seftor, C., and Celarier, E.: Global distributions of UV-absorbing aerosols from Nimbus 7/TOMS data, *J. Geophys. Res.*, 102, 16911–16922, <https://doi.org/10.1029/96JD03680>, 1997.
- Herman, J. R. and Celarier, E. A.: Earth surface reflectivity climatology at 340–380 nm from TOMS data, *J. Geophys. Res.*, 102(D23), 28003–28011, <https://doi.org/10.1029/97JD02074>, 1997.
- Hovenier, J. W., van der Mee, C., and Domke, H.: Transfer of Polarized Light in Planetary Atmospheres, Basic Concepts and Practical Methods, Kluwer Academic Publishers, Dordrecht, The Netherlands, 258 pp., 2004.
- Jaross, G. and Warner, J.: Use of Antarctica for validating reflected solar radiation measured by satellite sensors, *J. Geophys. Res.*, 113, D16S34, <https://doi.org/10.1029/2007JD008835>, 2008.

- Joiner, J., Vasilkov, A. P., Gupta, P., Bhartia, P. K., Veefkind, P., Sneep, M., de Haan, J., Polonsky, I., and Spurr, R.: Fast simulators for satellite cloud optical centroid pressure retrievals; evaluation of OMI cloud retrievals, *Atmos. Meas. Tech.*, 5, 529–545, <https://doi.org/10.5194/amt-5-529-2012>, 2012.
- Jourdan, O., Kokhanovsky, A. A., and Burrows, J. P.: Calibration of SCIAMACHY using AATSR top-of-atmosphere reflectance over a hurricane, *IEEE Geosci. Remote Sens. Lett.*, 4, 8–12, <https://doi.org/10.1109/LGRS.2006.881726>, 2007.
- Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, *J. Geophys. Res.*, 113, D18308, <https://doi.org/10.1029/2008JD010290>, 2008.
- Kleipool, Q., Ludewig, A., Babić, L., Bartstra, R., Braak, R., Dierssen, W., Dewitte, P.-J., Kenter, P., Landzaat, R., Leloux, J., Loots, E., Meijering, P., van der Plas, E., Rozemeijer, N., Schepers, D., Schiavini, D., Smeets, J., Vacanti, G., Vonk, F., and Veefkind, P.: Pre-launch calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite, *Atmos. Meas. Tech.*, 11, 6439–6479, <https://doi.org/10.5194/amt-11-6439-2018>, 2018.
- Koelemeijer, R. B. A., de Haan, J. F., and Stammes, P.: A database of spectral surface reflectivity in the range 335–772 nm derived from 5.5 years of GOME observations, *J. Geophys. Res.*, 108, 4070, <https://doi.org/10.1029/2002JD002429>, 2003.
- Kokhanovsky, A. A., Bramstedt, K., von Hoyningen-Huene, W., Burrows, J. P.: The intercomparison of top-of-atmosphere reflectivity measured by MERIS and SCIAMACHY in the spectral range of 443–865 nm, *IEEE Geosci. Remote Sens. Lett.*, 4, 293–296, <https://doi.org/10.1109/LGRS.2007.894159>, 2007.
- Krijger, J. M., van Weele, M., Aben, I., and Frey, R.: Technical Note: The effect of sensor resolution on the number of cloud-free observations from space, *Atmos. Chem. Phys.*, 7, 2881–2891, <https://doi.org/10.5194/acp-7-2881-2007>, 2007.
- Lelli, L., Kokhanovsky, A. A., Rozanov, V. V., Vountas, M., and Burrows, J. P.: Linear trends in cloud top height from passive observations in the oxygen A-band, *Atmos. Chem. Phys.*, 14, 5679–5692, <https://doi.org/10.5194/acp-14-5679-2014>, 2014.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The Ozone Monitoring Instrument, *IEEE Trans. Geosci. Remote Sens.*, 44, 1093–1101, <https://doi.org/10.1109/TGRS.2006.872333>, 2006.
- Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B. N., Streets, D. G., Eskes, H., van der A, R., McLinden, C., Fioletov, V., Carn, S., de Laat, J., DeLand, M., Marchenko, S., McPeters, R., Ziemke, J., Fu, D., Liu, X., Pickering, K., Apituley, A., González Abad, G., Arola, A., Boersma, F., Chan Miller, C., Chance, K., de Graaf, M., Hakkarainen, J., Hassinen, S., Ialongo, I., Kleipool, Q., Krotkov, N., Li, C., Lamsal, L., Newman, P., Nowlan, C., Suleiman, R., Tilstra, L. G., Torres, O., Wang, H., and Wargan, K.: The Ozone Monitoring Instrument: overview of 14 years in space, *Atmos. Chem. Phys.*, 18, 5699–5745, <https://doi.org/10.5194/acp-18-5699-2018>, 2018.
- Lichtenberg, G., Kleipool, Q., Krijger, J. M., van Soest, G., van Hees, R., Tilstra, L. G., Acarreta, J. R., Aben, I., Ahlers, B., Bovensmann, H., Chance, K., Gludemans, A. M. S., Hoogeveen, R. W. M., Jongma, R. T. N., Noël, S., PETERS, A., Schrijver, H., Schrijvers, C., Sioris, C. E., Skupin, J., Slijkhuis, S., Stammes, P., and Wuttke, M.: SCIAMACHY Level 1 data: calibration concept and in-flight calibration, *Atmos. Chem. Phys.*, 6, 5347–5367, <https://doi.org/10.5194/acp-6-5347-2006>, 2006.
- Liu, X., Bhartia, P. K., Chance, K., Spurr, R. J. D., and Kurosu, T. P.: Ozone profile retrievals from the Ozone Monitoring Instrument, *Atmos. Chem. Phys.*, 10, 2521–2537, <https://doi.org/10.5194/acp-10-2521-2010>, 2010.
- Lorente, A., Boersma, K. F., Stammes, P., Tilstra, L. G., Richter, A., Yu, H., Kharbouche, S., and Muller, J.-P.: The importance of surface reflectance anisotropy for cloud and NO₂ retrievals from GOME-2 and OMI, *Atmos. Meas. Tech.*, 11, 4509–4529, <https://doi.org/10.5194/amt-11-4509-2018>, 2018.

- Ludewig, A., Kleipool, Q., Bartstra, R., Landzaat, R., Leloux, J., Loots, E., Meijering, P., van der Plas, E., Rozemeijer, N., Vonk, F., and Veefkind, P.: In-flight calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2019-488>, in review, 2020.
- 5 [Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., Kokhanovsky, A., Livschitz, J., and Eisinger, M.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing – an overview, *Atmos. Meas. Tech.*, **9**, 1279–1301, <https://doi.org/10.5194/amt-9-1279-2016>, 2016.](https://doi.org/10.5194/amt-9-1279-2016)
- Platt, U.: Differential Optical Absorption Spectroscopy (DOAS), in: *Air Monitoring by Spectroscopic Techniques*, edited by: M. W. Sigrist, Chemical Analysis Series, vol. 127, John Wiley & Sons, New York, 27–84, 1994.
- Platt, U. and Stutz, Z.: *Differential Optical Absorption Spectroscopy, Principles and Applications*, Springer, Heidelberg, Germany, 598 pp., 10 2008.
- Schaaf, C., and Wang, Z.: MCD43C1 MODIS/Terra+Aqua BRDF/AlbedoModel Parameters Daily L3 Global 0.05Deg CMG V006, distributed by NASA EOSDIS Land Processes DAAC, <https://doi.org/10.5067/MODIS/MCD43C1.006>, 2015.
- Shah, S., Tuinder, O. N. E., van Peet, J. C. A., de Laat, A. T. J., and Stammes, P.: Evaluation of SCIAMACHY Level-1 data versions using nadir ozone profile retrievals in the period 2003–2011, *Atmos. Meas. Tech.*, **11**, 2345–2360, <https://doi.org/10.5194/amt-11-2345-2018>, 15 2018.
- Siddans, R.: S5P-NPP Cloud Processor ATBD, Doc. No. S5P-NPPC-RAL-ATBD-0001, Issue 1.0.0, 12 February, RAL Space, Harwell, United Kingdom, available at: http://www.tropomi.eu/sites/default/files/files/S5P-NPPC-RAL-ATBD-0001_NPP-Clouds_v1p0p0_20160212.pdf, 2016.
- Stammes, P.: Spectral radiance modelling in the UV-visible range, in: *IRS 2000: Current Problems in Atmospheric Radiation*, edited by: 20 Smith, W. L. and Timofeyev, Y. M., A. Deepak Publishing, Hampton, Virginia, 385–388, 2001.
- Stein Zweers, D. C.: TROPOMI ATBD of the UV aerosol index, Doc. No. S5P-KNMI-L2-0008-RP, Issue 1.1, 15 June, Koninklijk Ned. Meteorol. Inst., De Bilt, the Netherlands, available at: http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0008-RP-TROPOMI_ATBD_UVAI-1.1.0-20180615_signed.pdf, 2018.
- Tilstra, L. G., van Soest, G., and Stammes, P.: Method for in-flight satellite calibration in the ultraviolet using radiative transfer calculations, with application to Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY), *J. Geophys. Res.*, **110**, D18311, <https://doi.org/10.1029/2005JD005853>, 2005.
- 25 Tilstra, L. G. and Stammes, P.: Intercomparison of reflectances observed by GOME and SCIAMACHY in the visible wavelength range, *Appl. Opt.*, **45**, 4129–4135, <https://doi.org/10.1364/AO.45.004129>, 2006.
- Tilstra, L. G., and Stammes, P.: Earth reflectance and polarization intercomparison between SCIAMACHY onboard Envisat and POLDER 30 onboard ADEOS-2, *J. Geophys. Res.*, **112**, D11304, <https://doi.org/10.1029/2006JD007713>, 2007.
- Tilstra, L. G., de Graaf, M., Aben, I., and Stammes, P.: In-flight degradation correction of SCIAMACHY UV reflectances and Absorbing Aerosol Index, *J. Geophys. Res.*, **117**, D06209, <https://doi.org/10.1029/2011JD016957>, 2012.
- Tilstra, L. G., Lang, R., Munro, R., Aben, I., and Stammes, P.: Contiguous polarisation spectra of the Earth from 300 to 850 nm measured by GOME-2 onboard MetOp-A, *Atmos. Meas. Tech.*, **7**, 2047–2059, <https://doi.org/10.5194/amt-7-2047-2014>, 2014.
- 35 Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, *J. Geophys. Res.-Atmos.*, **122**, 4084–4111, <https://doi.org/10.1002/2016JD025940>, 2017.
- Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J.: Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis, *J. Geophys. Res.*, **103**(D14), 17099–17110, <https://doi.org/10.1029/98JD00900>, 1998.

- Torres, O., Ahn, C., and Chen, Z.: Improvements to the OMI near-UV aerosol algorithm using A-train CALIOP and AIRS observations, *Atmos. Meas. Tech.*, 6, 3257–3270, <https://doi.org/10.5194/amt-6-3257-2013>, 2013.
- van der A, R. J., van Oss, R. F., Pitters, A. J. M., Fortuin, J. P. F., Meijer, Y. J., and Kelder, H. M.: Ozone profile retrieval from recalibrated Global Ozone Monitoring Experiment data, *J. Geophys. Res.*, 107, 4239, <https://doi.org/10.1029/2001JD000696>, 2002.
- 5 van Kempen, T. A., van Hees, R. M., Tol, P. J. J., Aben, I., and Hoogeveen, R. W. M.: In-flight calibration and monitoring of the Tropospheric Monitoring Instrument (TROPOMI) short-wave infrared (SWIR) module, *Atmos. Meas. Tech.*, 12, 6827–6844, <https://doi.org/10.5194/amt-12-6827-2019>, 2019
- van Peet, J. C. A., van der A, R. J., Tuinder, O. N. E., Wolfram, E., Salvador, J., Levelt, P. F., and Kelder, H. M.: Ozone Profile Retrieval Algorithm (OPERA) for nadir-looking satellite instruments in the UV–VIS, *Atmos. Meas. Tech.*, 7, 859–876, [https://doi.org/10.5194/amt-](https://doi.org/10.5194/amt-7-859-2014)
10 7-859-2014, 2014.
- van Soest, G., Tilstra, L. G., and Stammes, P.: Large-scale validation of SCIAMACHY reflectance in the ultraviolet, *Atmos. Chem. Phys.*, 5, 2171–2180, <https://doi.org/10.5194/acp-5-2171-2005>, 2005.
- Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate,
15 air quality and ozone layer applications, *Remote Sens. Environ.*, 120, 70–83, <https://doi.org/10.1016/j.rse.2011.09.027>, 2012.
- Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendaal, M.: FRESCO+: an improved O₂ A-band cloud retrieval algorithm for tropospheric trace gas retrievals, *Atmos. Chem. Phys.*, 8, 6565–6576, <https://doi.org/10.5194/acp-8-6565-2008>, 2008.
- Wang, P., Tuinder, O. N. E., Tilstra, L. G., de Graaf, M., and Stammes, P.: Interpretation of FRESCO cloud retrievals in case of absorbing
20 aerosol events, *Atmos. Chem. Phys.*, 12, 9057–9077, <https://doi.org/10.5194/acp-12-9057-2012>, 2012.