Copernicus Atmospheric Mission Performance Cluster Service

Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #24: April 2018 – August 2024

Prepared by	Copernicus Atmospheric Mission Performance Cluster Service
Reference	S5P-MPC-IASB-ROCVR-24.00.00-20240916
CI identification	DI-MPC-ROCVR / TD-VALREP
Document update	#24
Issue	24.00.00
Date of issue	16 September 2024
Status	DRAFT
Distribution	ATM-MPC, ESA

Disclaimer

The information provided in this document has been produced in the context of the Copernicus Atmospheric Mission Performance Cluster Service (ATM-MPC). The activities leading to those results have been contracted by the European Space Agency (ESA) and coordinated by the Royal Belgian Institute for Space Aeronomy (BIRA-IASB) under supervision of the Royal Netherlands Meteorological Institute (KNMI) as the prime contractor. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. For the avoidance of all doubts, ESA, BIRA-IASB and KNMI expressly disclaim any and all warranties, express or implied, including without limitation warranties of merchantability and fitness for a particular purpose, with respect to the information. In no event shall ESA, BIRA-IASB and KNMI be liable for any direct, indirect, incidental, punitive, or consequential damages of any kind whatsoever with respect to the information. The user thereof uses the information at its sole risk and liability.

Document Information

Contractual Milestone	
This document will be attached to the payment request to ESA, regarding:	ESA contract No. 4000137010/21/I-BG
	MS5 ATM-MPC Regular Service Review 3 – **15/10/2024**

Approval Record		
Checked by:	Jean-Christopher Lambert (BIRA-IASB) Jacques Claas (KNMI) Deborah Stein-Zweers (KNMI) Antje Ludewig (KNMI) Diego Loyola (DLR) Maarten Sneep (KNMI)	MPC VAL Service Lead MPC Service Manager MPC MCC Lead MPC L0-1B Service Lead MPC L2-ALG Service Lead MPC L2-PRO Service Lead
Checked and approved by:	Angelika Dehn (ESA)	ESA Data Quality Manager
Signatures:	Jean-Christopher Lambert (BIRA-IASB)	Jean-Christopher Lambert (Signature) Digitally signed by Jean-Christopher Lambert (Signature) Date: 2024.10.01 16:24:13 +02'00'
	Angelika Dehn Digitally signed by Angelika Dehn Date: 2024.09.30 17:15:47 +02'00' Angelika Dehn (ESA)	

Document Identification	
Title	Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #24: April 2018 – August 2024
Type of document	ATM-MPC S5P Routine Operations Consolidated Validation Report (ROCVR)
Document ID	S5P-MPC-IASB-ROCVR-24.00.00-20240916
ROCVR update Issue number Date of issue Status Distribution	#24 version 24.00.00 16 September 2024 DRAFT ATM-MPC, ESA
Available on:	https://mpc-vdaf.tropomi.eu
Editors	A. Keppens and J.-C. Lambert (BIRA-IASB)
Contributors	S. Compernolle (BIRA-IASB), K.-U. Eichmann (IUP-UB), M. de Graaf (KNMI), D. Hubert (BIRA-IASB), A. Keppens (BIRA-IASB), B. Langerock (BIRA-IASB), A. Ludewig (KNMI), M.K. Sha (BIRA-IASB), T. Verhoelst (BIRA-IASB), T. Wagner (MPI-C), C. Ahn (NASA/GSFC), A. Argyrouli (DLR), D. Balis (AUTH), K.L. Chan (DLR), M. Coldewey-Egbers (DLR), I. De Smedt (BIRA-IASB), H. Eskes (KNMI), A.M. Fjæraa (NILU), K. Garane (AUTH), J.F. Gleason (NASA/GSFC), J. Granville (BIRA-IASB), P. Hedelt (DLR), K.-P. Heue (DLR), G. Jaross (NASA/GSFC), Q. Kleipool (KNMI), ML. Koukouli (AUTH), R. Lutz (DLR), M.C Martinez Velarte (SRON), K. Michailidis (AUTH), S. Nanda (KNMI), S. Niemeijer (s[&]t), A. Pazmiño (LATMOS), A. Pseftogkas (AUTH), G. Pinardi (BIRA-IASB), A. Richter (IUP-UB), N. Rozemeijer (KNMI), M. Sneep (KNMI), D. Stein Zweers (KNMI), N. Theys (BIRA-IASB), G. Tilstra (KNMI), O. Torres (NASA/GSFC), P. Valks (DLR), J. van Geffen (KNMI), C. Vigouroux (BIRA-IASB), P. Wang (KNMI), and M. Weber (IUP-UB)
Citation	Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #24: April 2018 – August 2024. Lambert, J.-C., A. Keppens, S. Compernolle, K.-U. Eichmann, M. de Graaf, D. Hubert, B. Langerock, A. Ludewig, M.K. Sha, T. Verhoelst, T. Wagner, C. Ahn, A. Argyrouli, D. Balis, K.L. Chan, M. Coldewey-Egbers, I. De Smedt, H. Eskes, A.M. Fjæraa, K. Garane, J.F. Gleason, F. Goutail, J. Granville, P. Hedelt, K.-P. Heue, G. Jaross, Q. Kleipool, ML. Koukouli, R. Lutz, M.C Martinez Velarte, K. Michailidis, A. Pseftogkas, S. Nanda, S. Niemeijer, A. Pazmiño, G. Pinardi, A. Richter, N. Rozemeijer, M. Sneep, D. Stein Zweers, N. Theys, G. Tilstra, O. Torres, P. Valks, J. van Geffen, C. Vigouroux, P. Wang, and M. Weber. S5P MPC Routine Operations Consolidated Validation Report series, Issue #24, Version 24.00.00, 212 pp., 16 September 2024

Executive Summary

This document reports consolidated results of the Routine Operations Validation Service for the Sentinel-5 Precursor (S5P) Tropospheric Monitoring Instrument (TROPOMI) [ER_TROPOMI]. S5P TROPOMI contributes to the space component of the European Earth Observation programme Copernicus [ER_CoperESA]. The S5P Routine Operations Validation Service is provided by the Atmospheric Mission Performance Cluster (ATM-MPC) for Level-1 and Level-2 data products generated by the Near Real Time (NRTI), Offline (OFFL), and reprocessing (RPRO) processors since the first public data release in July 2018. The present Routine Operations Consolidated Validation Report (ROCVR) integrates results from the MPC Validation Data Analysis Facility (VDAF) [ER_VDAF] with *ad hoc* support from S5P Validation Team (S5PVT) AO projects [ER_S5PVT]. The S5P Routine Operations Validation Service details and complements the conclusions and features described in the Product Readme Files (PRF) delivered with the S5P products, in which users can find practical recommendations on S5P data usage. The present report covers the period of S5P operation from April 2018 until August 2024. It includes validation results for versions 01.00.00, 02.00.00 and 02.01.00 of the L1B processor, for versions 02.04.00 to 02.06.00 of the NL-L2 processor suite (O_3 profile, NO_2, CO and CH_4 columns, and AAI and ALH data), and for versions 02.04.01 to 02.06.01 of the UPAS processor suite (total and tropospheric O_3 columns, HCHO and SO_2 columns, and CLOUD data). For all products, this report considers reprocessed data or later versions only. Validation results for previous processor versions remain available in version 17 and lower of the ROCVR archived on https://mpc-vdaf.tropomi.eu/.

Radiance and Irradiance

The validation of the wavelength assignment of the S5P L1B_UVN v02.00.00 products concludes to an agreement to within 0.01 nm, which is within the pre-launch calibration uncertainty. The radiance in bands 1-3 is up to 5 % smaller than OMPS-nadir radiance; above 320 nm, this is a wavelength independent bias. Below 320 nm, the wavelength dependence seems to vary with the latitude. In band 1 around 280 nm, the radiance deviates more than 10% from OMPS values. The absolute radiometric calibration for UV radiance lacks accuracy and as a result may be updated in the future. In the spectrally overlapping regions of bands 2 and 3 there is a discrepancy of about 2 % in the L1b radiance signals. The radiance in band 6 was compared to modelled spectra in the continuum around the O_2-A band. The signal of TROPOMI is 1-2 % lower than the modelled radiance. For bands 1 to 6 (UV, UVIS and NIR) degradation has been observed for the radiance. The degradation is the largest at short wavelengths. The decrease in radiance signal per 1000 orbits is between 0.31% in band 1 and 0.02 % in band 6. The correction of the radiance degradation is active in the forward stream from version 02.01.00 on (orbit 24688, 19.07.2022). The absolute and relative radiance radiometry of the SWIR bands were validated using reference stations in Railroad Valley and in the Saharan desert. Current validation results give upper limits of < 5 % for the absolute calibration and < 0.8 % for the relative calibration.

The absolute irradiance calibration of TROPOMI has been compared to other published solar reference datasets. After an update to the calibration based on OMPS-nadir data, the UV and UVIS spectrometers agree within 0-5 % with the references. For extreme swath angles, the deviations are larger in the UV. For the NIR spectrometer the irradiance spectrum is approximately 1.5-3.5 % lower than the reference spectra. The SWIR spectrum is approximately 0.6 % lower than the closest reference spectrum.

The irradiance for L1B version 02.01.00 in TROPOMI bands 1, 2, and 3 agrees well with corresponding OMPS-NP and OMPS-NM measurements from June 2018 to June 2023. In bands 2 and 3 no wavelength dependence is observed. The ratio is very stable over the entire period.

The stability of the reflectance in TROPOMI band 4 has been obtained from cloud-free nadir measurements over Pseudo-Invariant Calibration Sites in the Saharan Desert from May 2018 through August 2024. The drift is < 0.05% per 1000 orbits.

Ozone Column

The S5P L2_O3 NRTI and OFFL total ozone column data are in good overall agreement with correlative ground-based measurements from the GAW and NDACC Brewer and Dobson networks, the PGN Pandora network, and the NDACC ZSL-DOAS/SAOZ network, and with the MetOp-B GOME-2, Aura OMI, and Suomi-NPP OMPS-nadir satellite instruments. Across the networks the mean bias is about +0.5 % (NRTI) and +1.3 % (OFFL). This bias and the standard deviation of the relative difference both comply with mission requirements, that is, a bias lower than 5 % and an uncertainty due to random errors (dispersion) better than ±2.5 %. The instrumental switch to smaller (along-track) ground pixels on the 6th of August 2019 did not affect the agreement with the ground-based reference data. A first analysis of the long-term stability of the RPRO+OFFL record against the ground-based data reveals drifts below the GCOS 2022 Implementation Plan threshold of 3 %\decade at almost all stations with sufficiently well sampled time series.

The comparison of S5P TROPOMI total ozone column data between processors (NRTI versus OFFL) and with other nadir UV satellite data sets (GOME-2B and GOME-2C, OMI, OMPS) shows agreement within 1-2 % at all but the highest latitudes, where retrievals are most difficult because of both low solar zenith angles and strong surface albedo gradients. Differences with GOME-2C do contain a systematic component of -2 %, but ground-based validation suggests the latter to be biased high by such an amount.

Tropospheric Ozone Column

The S5P L2_O3_TCL RPRO+OFFL tropospheric ozone column data (CCD algorithm) are in good general agreement with correlative measurements from the SHADOZ ozonesonde network. Across the network the mean bias (around +19 % or +3.4 DU) and the mean dispersion of the differences (about 26 % or 4.6 DU) comply with mission requirements, that is, a bias lower than 25 % and an uncertainty (dispersion) less than 25 %. Comparisons of S5P tropospheric ozone columns and other satellite CCD data sets (GOME-2B and OMI) are consistent with the ground-based results. S5P bias is positive and varies from +1.2 DU / +6 % (w.r.t. OMI), +3.5 DU / +19 % (w.r.t. GOME-2B) to +3.8 DU / +21 % (w.r.t. GOME-2C). Dispersion in all satellite comparisons lies around 3.1 DU or 16 %.

The bias of S5P exhibits a seasonal cycle. A pattern of more elevated positive biases (7-10 DU or 25-60 %) during the biomass-burning season emerges at stations around the Atlantic equatorial basin. It is not clear whether there is a causal relationship. The interplay of satellite orbit and cloud coverage leads to two types of sampling error of up to 1 DU and about 5 DU, correlated in time and space (latitudinal stripes, patterns progressing along satellite orbit). Users can reduce sampling uncertainty by lowering the sampling resolution.

Ozone Vertical Profile

Comparison of the S5P L2_O3_PR ozone profile data (v02.04-6, May 2018 to August 2024) with ozonesonde and lidar measurements concludes to a median agreement better than 5 to 10 % in the troposphere and up to the upper troposphere/lower stratosphere (UTLS). The bias goes up to -15 % in the higher stratosphere (35-45 km), but with vertical oscillations. The comparisons show a dispersion of order of 30 % in the troposphere, and 10 to 20 % in the UTLS and upper stratosphere. Chi-square tests demonstrate that on average the observed differences confirm the ex-ante satellite and ground uncertainty estimates in the stratosphere, above about 20 km. Around the tropopause and below (around 15-20 km and lower), the mean chi-square value increases up to about four. Here, the retrieved satellite uncertainty is smaller than what is actually observed.

The information content of the ozone profile retrieval is characterised by about five to six vertical sub-columns of independent information (estimated from the Degrees of Freedom) and a vertical sensitivity nearly equal to unity at altitudes from about 20 km (UTLS) to 50 km and decreasing rapidly at altitudes above and below. The altitude registration of the retrieved profile information usually is close to the nominal retrieval altitude in the 20-50 km altitude range and shows positive and negative offsets of up to 10 km below and above the 20-50 km altitude range, respectively. The effective vertical resolution of the profile retrieval usually ranges within 10-15 km, with a minimum close to 7 km in the middle stratosphere. Increased sensitivities and higher effective vertical resolutions can be observed for higher solar zenith angles, as can be expected, and correlates with higher retrieved ozone concentrations. On the other hand, one can also observe some lower-DFS profiles with nearly-zero surface sensitivity in combination with a highly overcompensating sensitivity around the UTLS, ranging up to three and above. These retrievals occur for scenes that have both high SZA and high surface albedo, mostly around the Antarctic (latitudes from 60 to 90 south). A solar-zenith angle dependence is observed for the lowest ozone sub-columns, which translates into a seasonal and meridian dependence of their bias.

The more than six years of TROPOMI ozone profile data show a slight DFS degradation throughout the mission (next to a jump from the ground pixel resolution change). Comparisons with ozonesonde data reveal significant positive drifts near 2 %/year in the tropics and mid-latitudes from the surface to the UTLS, while 1-2 % per year negative drifts are observed above. This makes the current operational TROPOMI ozone profile product and its subcolumn derivatives unsuitable for vertically resolved trend studies. However, no significant drift is detected for the vertically integrated profile. This agrees with the operational TROPOMI total ozone column retrieval, although the latter is consistently about 5 % higher than the integrated ozone profile.

Nitrogen Dioxide

The three S5P L2_NO2 data products (tropospheric, stratospheric, and total column) of version 02.04.00 (RPRO) and 02.04.00-02.06.00 (NRTI, OFFL) are in good overall agreement with correlative ground-based measurements. Reference measurements from the MAX-DOAS (troposphere), the NDACC ZSL-DOAS/SAOZ (stratosphere), and the Pandonia Global Network (total), as well as correlative satellite data products (OMI), are used for validation. Generally, the negative bias between S5P and ground-based data increases with higher NO2 columns of the total and tropospheric L2 products. Similar biases and uncertainty estimates are detected for the L2_NO2 NRTI and OFFL/RPRO datasets validated with ground-based instruments.

The L2_NO2 tropospheric column data is compared to ground-based MAX-DOAS data from 31 stations. The overall negative median bias is -28.7 % (-1.5 Pmolec/cm^2), which is within the mission requirement of 50 %. The bias depends on the pollution level at the station. It is positive (13 %) over clean areas (< 2 Pmolec/cm^2) and negative (-40 %) over highly polluted areas (>15 Pmolec/cm^2). The bias estimate can be reduced absolutely by 20 % when MAX-DOAS profile data are vertically smoothed using the S5P averaging kernels. The overall median dispersion of about 3.1 Pmolec/cm^2 exceeds the mission precision requirements (0.7 Pmolec/cm^2), but is within limits for clean stations.

The L2_NO2 stratospheric column data is compared to ZSL-DOAS UV-visible ground-based measurements at 25 NDACC stations that are distributed from pole to pole. The large horizontal smoothing of stratospheric NO_2 in the zenith-scattered-light (ZSL) geometry and the NO_2 diurnal cycle is taken into account. The stratospheric NO_2 columns are generally lower by approximately 0.1 Pmolec/cm^2. The median bias of -3 % is within the S5P mission requirements (10%) and so is the dispersion of 0.3 Pmolec/cm^2, considering the combined random errors and irreducible co-location mismatches (mission requirement: 0.5 Pmolec/cm^2). The comparison of OFFL L2_NO2 stratospheric columns with ground-based FTIR data shows a positive median bias of +4.7 % and a dispersion of 0.3 Pmolec/cm^2 for 26 NDACC stations. Even larger biases are observed at high-latitude (9 to 12 %) and tropical stations (13 to 16 %).

The L2_NO2 total NO_2 column data are compared to ground-based Pandora column data at 70 PGN stations. The median bias between S5P and PGN data is -8.6 % (-0.6 Pmolec/cm^2) with a dispersion of 1.6 Pmolec/cm^2. The bias varies with the total amount of NO_2: a positive bias of 5.9 % is seen over cleaner stations and high mountain areas (< 6 Pmolec/cm^2), and a negative bias of -17.0 % over polluted areas.

Formaldehyde

The S5P L2_HCHO column data up of version 02.04.01 (RPRO) and 02.04.01 - 02.06.01 (NRTI/OFFL) is in good overall agreement with independent ground-based measurements from the NDACC FTIR, PGN and MAX-DOAS monitoring networks and to corresponding Aura OMI satellite data. Ground-based validation concludes to similar bias and uncertainty (dispersion) estimates for the L2_HCHO NRTI and L2_HCHO OFFL/RPRO dataset.

Comparisons with 29 FTIR stations show a negative bias of -30 % for high emission stations (> 8 Pmolec/cm^2) and a positive bias of 30 % for clean stations (< 2.5 Pmolec/cm^2) when vertical smoothing differences are minimized by application of the averaging kernels.

The bias of -37 % with respect to MAX-DOAS HCHO measurements is slightly higher. It reduces to -20 % using averaging kernels. Also, here biases differ for clean (+27 %) and polluted stations (-10 %). All biases are within the S5P mission requirements (40-80 %). The dispersion at clean stations versus FTIR data of about 9 Pmolec/cm^2 and versus MAX-DOAS of 10 Pmolec/cm^2 is within the S5P uncertainty mission requirements of 12 Pmolec/cm^2. These values make use of median deviations to reduce the influence of larger outliers and vertical smoothing. Comparisons with 36 PGN instruments show a negative median bias of -33 % and a dispersion of 9.6 Pmolec/cm^2, in agreement with the MAX-DOAS results.

The bias in comparison to OMI is less than -10 % for most regions with some larger negative biases in Europe, Northern America and China (< 20 %). The dispersion of differences is about 2 Pmolec/cm^2 when considering regionally averaged columns.

Sulphur Dioxide

The S5P L2_SO2 (NRTI and OFFL) sulphur dioxide column data are found in general good agreement with ground-based MAX-DOAS and PGN measurements and with other satellite observations from OMI and S-NPP OMPS. The bias and dispersion with respect to validation data are typically below 0.2 DU. From these comparisons it can be concluded that over polluted regions the S5P mission requirements are fulfilled. Over volcanic plumes, the requirement on the bias is fulfilled, while the dispersion can exceed slightly the requirement on the random component of the uncertainty, which is not considered as a substantial restriction of the data quality. In April 2023, the whole SO_2 data set was reprocessed with the new processor version 02.04.01. Reprocessed SO_2 data agree well with previous versions.

Sulphur Dioxide Layer Height

The validation of the S5P TROPOMI L2_SO_2 sulphur dioxide plume height is based on satellite-to-satellite comparisons, during large enough volcanic eruptions during the September-2023 to May 2024 period. For this purpose, the Metop-B/C IASI “SO2_vertical_level” product, i.e., the retrieved vertical level of the SO_2 plume, for the morning (AM) observations were utilized. Overall, TROPOMI shows lower SO_2 layer height retrievals compared to the combined IASI/Metop[B&C] SO_2 plume height observations by ~30 %, equivalent to a mean SO_2 layer height difference of 3 km. In more significant volcano eruptions with robust and overlapping SO_2 plumes between the involved instruments, the discrepancies are lower (~20 %) and equivalent to approximately 1.5 km. HYSPLIT airmass forward and backward trajectories are also employed as a validation tool of the S5P TROPOMI SO_2_Layer_Height observations. Results

indicate a good agreement between the height of trajectories, the height, and location of the plume with discrepancies of approximately 500 m.

Carbon Monoxide

The S5P L2_CO (NRTI or RPRO concatenated with OFFL) carbon monoxide total column data is in good overall agreement with correlative measurements from the NDACC and TCCON FTIR monitoring networks. It exhibits a positive bias of approximately +2 % for NDACC and -2.5 % for TCCON on average and all individual stations' biases fall well within the mission requirement (bias of maximum 15 %). The bias is higher (max of 7 %) for the high latitude stations. The standard deviation of the relative bias for the de-striped columns is on an average lower than 8 % against NDACC and TCCON, which is also within the mission requirement for precision (better than 10 %). The averaged correlation coefficient reaches 0.9 for NDACC and TCCON. The bias shows a limited dependence on the solar zenith angle and seasonal dependence estimated at 2 % difference between spring and autumn and can be reduced when taking the satellite averaging kernel into account. These dependences are considered not significant compared to the reported measurement uncertainties.

Methane

The S5P L2_CH4 (OFFL concatenated with RPRO) methane total column averaged data is in good overall agreement with correlative measurements from the TCCON and NDACC FTIR monitoring networks. The standard and bias-corrected S5P xCH_4 column data exhibit a negative bias against TCCON of -0.34 % and +0.22 % respectively, and against NDACC of -0.82 % and -0.06 % respectively, which falls well within the mission requirement (bias of maximum 1.5%). The standard deviation of the relative bias for TCCON is on an average 0.75 % which is also within the mission requirement for precision (<1%). The averaged correlation coefficient 0.75 is rather low, partly because of short time series at some station, presence of some mountain station, and not all outlying pixels are filtered with the `qa_value` above 0.5. The sun-glint pixels is evaluated separately and shows similar behaviour as the standard product. The validation against TCCON shows a mean bias of 0.21 % and against NDACC shows a mean bias of -0.4% for the bias-corrected product for the limited co-locations found. Comparison of the S5P XCH4 product against GOSAT proxy XCH4 data show a mean bias TROPOMI-GOSAT of -7.7 ppb ± 18.3 ppb (−0.41 ± 0.97 %) and a Pearson's correlation coefficient of 0.83. The S5P L2_CH4 data when compared to S5P WFMD-DOAS XCH4 product over ocean shows a mean bias of -3.0 ppb ± 16.4 ppb (-0.16 ± 0.9 %) and a Pearson's correlation coefficient of 0.81.

Clouds

The S5P L2_CLOUD CRB radiometric cloud fraction was compared to S5P FRESCO. S5P L2_CLOUD and S5P FRESCO capture similar meridian variations. The filter guideline of qa_value≥0.5 is too strict for some applications, but we found that the less strict qa_value≥0.25 alternative (see CLOUD PRF) does not change the overall picture on meridian variation. Comparing spatially co-located measurements, the CLOUD CRB cloud fraction is slightly below that of S5P FRESCO, but there can be higher deviations over island and/or coastal stations. Since CLOUD v2.6.1, an increase in CRB cloud fraction seems visible for island/coastal sites Graciosa Island, Soverato and Granada.

The S5P L2_CLOUD CAL cloud top height, cloud mean height, and CRB cloud height data were compared to respectively cloud top height and cloud mean height derived from ground-based measurements from the CLOUDNET and ARM networks, and with the S5P FRESCO satellite product, after filtering out the lowest cloud fractions. Note that the sensitivity of the TROPOMI NIR observations to clouds differs significantly from the sensitivity of CLOUDNET/ARM lidar/radar instruments used as a reference, and the error associated with the reference observations is also not yet included in those comparisons.

Given their different nature, we look at low clouds and high clouds separately to derive representative quality indicators (distinction set at CLOUDNET cloud top height 4 km). For low clouds, the bias of CAL cloud top height vs. CLOUDNET cloud top height is within mission requirements, but not for high clouds, where CAL is 3 km (35 %) below CLOUDNET's cloud top height. The bias of CAL cloud mean height vs CLOUDNET cloud mean height is within mission requirements for low and high clouds. The bias of CRB cloud height vs. CLOUDNET cloud mean height is negative, bordering on the mission requirement. For high clouds, the difference dispersion of CAL and CRB with CLOUDNET is almost 2 km, exceeding the S5P mission requirement. For low clouds, the difference dispersion is close to the requirement of 0.5 km. CRB cloud height is below that of FRESCO. CAL cloud mean height is below FRESCO cloud height, but with a dependence on season and location. FRESCO deviates more strongly from CRB and CAL at low cloud fractions, making the comparison depending on the exact cloud fraction filter settings.

Since CLOUD version 2.6.1, CAL cloud mean height seems to have increased compared to Cloudnet and FRESCO.

Aerosol Index

The S5P L2_AER_AI (NRTI and OFFL) UV Aerosol Absorbing Index is in good overall agreement with similar satellite data products from EOS-Aura OMI and Suomi-NPP OMPS. The whole aerosol index data set was reprocessed with processor version 02.04.00. The reprocessed UVAI is within mission requirements due to the L1b data including a correction for observed degradation of the irradiance since the beginning of the time series combined with application of an offset.

Aerosol Layer Height

The S5P L2_AER_LH (OFFL) data product shows a very good agreement with two other satellite aerosol layer height estimates, from MISR (stereoscopic imagery) and CALIOP (active lidar sensing of the aerosol vertical distribution). S5P TROPOMI AER_LH shows a systematic difference with MISR aerosol plume height of about 600 m (lower for TROPOMI). This is mostly due to the difference in the sensitivity of the instruments and the differences in the algorithms. A difference of about 500 m (lower for CALIOP) is expected from simulations, TROPOMI ALH being sensitive to the centroid aerosol layer height. For very thick plumes the difference between TROPOMI ALH and CALIOP layer height decreases to only 50 m. This is well within the requirements of 100 hPa for the bias.

This report contains validation results using the upgraded version of the ALH algorithm (v02.06.00) implemented since November 2023. The new version introduces the surface albedo in the Optimal Estimation feature vector, producing more elevated ALH for both land and ocean. The quantitative validation analysis [May 2018 – August 2023], based on the latest version 02.06.00 (Level 2, RPRO+OFFL), over the EARLINET stations illustrates that TROPOMI ALH over ocean is consistent with calculated lidar ALH, with a correlation coefficient R=0.66 and mean bias of 1.46 km. In addition, over land, an agreement of R = 0.56 and a mean bias -0.49 km is found. These differences can be attributed to the different sensitivities of the TROPOMI and Lidar ALH measurements, as well as the applied spatiotemporal collocation criteria.

Moreover, as part of the investigation into the performance of version 02.06.00, a comparative analysis with previous versions was performed, against EARLINET stations for a selected number of cases. The quantitative analysis illustrates that TROPOMI ALH over ocean is consistent with ALHbsc, with a correlation coefficient R=0.68 and mean bias 1.41 km for the latest version (Prev. version: R: 0.74 /MB: -0.62km). Over land, an improved agreement of R = 0.51 and mean bias -0.47 km is found (Prev. version: R: 0.43 /MB: -2.2km). It is clearly seen that with the implementation of the new version (v02.06) the number of processed pixels has increased significantly, both on land and at sea. It is also important to note that for the pixels retrieved, the distribution shows a significant increase in height levels. The

mean layer heights are on average higher for TROPOMI over ocean and lower over land compared against the lidar measurements.

The S5P L2_ALH dispersion is large due to cloud contamination and surface effects. With rigorous cloud screening, 50 % of the pixels are already within 1 km of the CALIOP weighted extinction height. Accounting for the expected bias, this is within the requirements of 50 hPa. However, this preliminary conclusion needs further investigation and confirmation.

A limitation of the S5P TROPOMI ALH product has become apparent following the severe bushfires in New South Wales during the 2019-2020 fire season, which produced very high-altitude smoke plumes (altitude > 20 km). These heights were not anticipated and ALH values are limited to about 13 km altitude. An update to include these very high altitudes is not foreseen for the near future.

Processing Baseline Identification

This document reports consolidated validation results for the S5P TROPOMI data products listed in **Table 1**. Validation results for previous processor versions remain available in version 17 of this document, archived at https://mpc-vdaf.tropomi.eu/

Product ID	Stream	Version	In operation from (orbit #, date)	In operation until (orbit #, date)
L1B_RA1/2/…/8	RPRO/ OFFL	02.01.00	2818, 2018-04-30	current version
L1B_IR_UVN/SIR	RPRO/ OFFL	02.01.00	2818, 2018-04-30	current version
L2_O3	NRTI	02.04.01 02.05.00 02.06.01	24697, 2022-07-20 29878, 2023-07-20 31750, 2023-11-29	29861, 2023-07-19 31750, 2023-11-29 current version
	OFFL	02.04.01 02.05.00 02.06.01	24655, 2022-07-17 29818, 2023-07-16 31705, 2023-11-26	29817, 2023-07-16 31704, 2023-11-26 current version
	RPRO	02.04.01	2818, 2018-04-30	24898, 2022-08-03
L2_O3_TCL	OFFL	02.04.01 02.05.00 02.06.01	24858, 2022-07-28 29696, 2023-07-04 31597, 2023-11-15	29682, 2023-07-03 31583, 2023-11-14 current version
	RPRO	02.04.01	2867, 2018-04-30	24844, 2022-07-27
L2_O3_PR	NRTI	02.04.00 02.05.00 02.06.00	24697, 2022-07-20 28078, 2023-03-15 31750, 2023-11-29	28074, 2023-03-15 31750, 2023-11-29 current version
	OFFL	02.04.00 02.05.00 02.06.00	24655, 2022-07-17 28031, 2023-03-12 31705, 2023-11-26	28030, 2023-03-12 31704, 2023-11-26 current version
	RPRO	02.04.00	2818, 2018-04-30	24779, 2022-07-25
L2_NO2	NRTI	02.04.00 02.05.00 02.06.00	24697, 2022-07-20 28078, 2023-03-15 31751, 2023-11-29	28074, 2023-03-12 31750, 2023-11-29 current version
	OFFL	02.04.00 02.05.00 02.06.00	24655, 2022-07-17 28030, 2023-03-12 31705, 2023-11-26	28029, 2023-03-12 31704, 2023-11-26 current version
	RPRO	02.04.00	2818, 2018-04-30	24654, 2022-07-16
L2_HCHO	NRTI	02.04.01 02.05.00 02.06.01	24697, 2022-07-20 29878, 2023-07-20 31751, 2023-11-29	29861, 2023-07-19 31750, 2023-11-29 current version
	OFFL	02.04.01 02.05.00 02.06.01	24686, 2022-08-01 29818, 2023-07-16 31705, 2023-11-26	29817, 2023-07-16 31704, 2023-11-26 current version
	RPRO	02.04.01	2818, 2018-04-30	24865, 2022-07-31
L2_SO2 (incl. L2_SO2_LH from 02.05.00)	NRTI	02.04.01 02.05.00 02.06.01	24697, 2022-07-20 29878, 2023-07-20 31750, 2023-11-29	29861, 2023-07-19 31750, 2023-11-29 current version
	OFFL	02.04.01 02.05.00 02.06.01	3202, 2018-05-27 29818, 2023-07-16 31705, 2023-11-26	29817, 2023-07-16 31704, 2023-11-26 current version
	RPRO	02.04.01	2918, 2018-05-07	24898, 2022-08-03

Product ID	Stream	Version	In operation from (orbit #, date)	In operation until (orbit #, date)
L2_CO	NRTI	02.04.00 02.05.00 02.06.00	24697, 2022-07-20 28078, 2023-03-15 31750, 2023-11-29	28077, 2023-03-15 31750, 2023-11-29 current version
	RPRO	02.04.00	2818, 2018-04-30	24779, 2022-07-25
	OFFL	02.04.00 02.05.00 02.06.00	24655, 2022-07-17 28031, 2023-03-12 31705, 2023-11-26	28030, 2023-03-12 31704, 2023-11-26 current version
L2_CH4	RPRO OFFL	02.04.00 02.04.00 02.05.00 02.06.00	2818, 2018-04-30 24655, 2022-07-17 28031, 2023-03-12 31705, 2023-11-26	24779, 2022-07-25 28030, 2023-03-12 31704, 2023-11-26 current version
L2_CLOUD	NRTI	02.04.01 02.05.00 02.06.01	24697, 2022-07-20 29878, 2023-07-20 31751, 2023-11-29	29861, 2023-07-19 31750, 2023-11-29 current version
	OFFL	02.04.01 02.05.00 02.06.01	24655, 2022-07-17 29818, 2023-07-16 31705, 2023-11-26	29817, 2023-07-16 31704, 2023-11-26 current version
	RPRO	02.04.01	2818, 2018-04-30	24898, 2022-08-03
L2_AER_AI	NRTI	02.04.00 02.05.00 02.06.00	24697, 2022-07-20 28078, 2023-03-15 31750, 2023-11-29	28077, 2023-03-15 31750, 2023-11-29 current version
	OFFL	02.04.00 02.05.00 02.06.00	24655, 2022-07-17 28031, 2023-03-12 31705, 2023-11-26	28030, 2023-03-12 31704, 2023-11-26 current version
	RPRO	02.04.00	2818, 2018-04-30	24779, 2022-07-25
L2_AER_LH	NRTI	02.04.00 02.05.00 02.06.00	24697, 2022-07-20 28078, 2023-03-15 31750, 2023-11-29	28074, 2023-03-15 31750, 2023-11-29 current version
	OFFL	02.04.00 02.05.00 02.06.00	24655, 2022-07-17 28031, 2023-03-12 31705, 2023-11-26	28030, 2023-03-12 31704, 2023-11-26 current version
	RPRO	02.04.00	2818, 2018-04-30	24779, 2022-07-25

Table 1 – S5P TROPOMI data products and processor versions (NRTI for near real time, OFFL for off-line, and RPRO for reprocessed). Note: the operational phase (E2) of the S5P TROPOMI mission starts with orbit #2818.

Representative Quality Indicators

Representative values of key quality indicators (bias and dispersion vs. reference measurements, and special features) have been derived for the following S5P operational data products on the basis of the validation results reported in this document:

Product ID	Stream	Product	Bias	Dispersion	Special features
L2_O3	NRTI	O_3 column	0.5 %	2 %	Some increase in dispersion in the comparisons to ground-based measurements at SZA > 70°.
	OFFL	O_3 column	1.3 %	2 %	Some increase in dispersion in the comparisons to ground-based measurements at SZA > 70°.
L2_O3_TCL	OFFL RPRO	O_3 tropospheric column (CCD)	+19 %	26 %	Seasonal change of the bias at Atlantic sites coinciding with biomass burning season. Geographical imprints of sampling-related biases.
L2_O3_PR	NRTI	O_3 profile	5-10 %	10-30 %	Mean agreement better than 5 to 10 % in the troposphere and UTLS. Bias goes up to -15 % in the higher stratosphere (35-45 km), but with vertical oscillations. Dispersion of order of 30 % in the troposphere, and 10 to 20 % in the UTLS and upper stratosphere. Despite positive (troposphere) and negative (stratosphere) drifts at individual levels, no significant drift is detected for the vertically integrated profile.
	OFFL	O_3 profile	5-10 %	10-30 %	
	RPRO	O_3 profile	5-10 %	10-30 %	
L2_NO2	NRTI	NO_2 troposphere NO_2 stratosphere NO_2 total	-37 % -3 % 0±50 %	2.6 Pmolec/cm^2 0.3 Pmolec/cm^2 -	Bias and dispersion by column amount: Troposphere [<2 Pmolec/cm²] +13 % and [>15 Pmolec/cm²] -40 %. Total [6 Pmolec/cm²]: +5.9 % / -17.0 %.
	OFFL RPRO	NO_2 troposphere NO_2 stratosphere NO_2 total	-29 % -3 % -9 %	3.1 Pmolec/cm^2 0.3 Pmolec/cm^2 1.6 Pmolec/cm^2	
L2_HCHO	NRTI OFFL RPRO	HCHO, low HCHO, high	+32 % -30 %	9 Pmolec/cm² 29 Pmolec/cm^2	Bias and dispersion depend on column amount: [<2.5 Pmolec/cm²] positive bias, low dispersion, [>8 Pmolec/cm²] negative bias, high dispersion.
L2_SO2	NRTI	SO_2 column	0.2 DU	0.2 DU	Lack of validation stations in areas with high SO_2.
	OFFL	SO_2 column	0.2 DU	0.2 DU	
L2_SO2_LH	OFFL	SO2 layer height	-3 km	2.2 km	Temporal difference with IASI instruments. Significant eruptions and overlapping plumes show lower discrepancies < 1.5 km.
L2_CO	NRTI	CO column	+2 %	8 %	Along orbit stripes. 4% SZA and 2% seasonal dependence in bias, both within reference measurement uncertainty. Seasonal dependence can be reduced when taking into account the satellite averaging kernels.
	OFFL	CO column	+2 %	8 %	

Product ID	Stream	Product	Bias	Dispersion	Special features
L2_CH4	OFFL	CH4 column	+0.28 %	0.7 %	Along orbit stripes. Underestimation at low albedo. Remaining outliers with `qa_value` >0.5. Outlying CH_4 values observed along coastal or mountain regions – e.g., in Greenland.
L2_CLOUD	NRTI OFFL	CAL CTH (h) CAL CTH (l) CAL CMH (h) CAL CMH (l) CRB CH (h) CRB CH (l)	-36 % -14 % -12 % -16 % -19 % -21 %	1.9 km 0.6 km 1.7 km 0.5 km 1.9 km 0.5 km	Low clouds (l): CLOUDNET CTH<4km; high clouds (h): CLOUDNET CTH>4km. Discrepancy with FRESCO cloud height increases with decreasing cloud fraction.
L2_AER_AI	NRTI	aerosol index	-1.1 AI unit	0.1 AI unit	
	OFFL	aerosol index	-1.1 AI unit	0.1 AI unit	
L2_AER_LH	OFFL	aerosol layer height	50 hPa	100 hPa	Over ocean only. Larger bias and dispersion expected over land.

Table 2 – Representative quality indicators (bias, dispersion and special features) derived from the validation of the S5P TROPOMI data products listed in the **Table 1**, valid for all processor versions unless stated differently. CTH: cloud-top-height; CH: cloud height; CMH: cloud mean height; COT: cloud optical thickness.

Table of Contents

Document Information 3
Executive Summary 5
Processing Baseline Identification 12
Representative Quality Indicators 14
Table of Contents 16
1 Introduction 22
1.1 Background information on Sentinel-5 Precursor TROPOMI 22
1.2 Copernicus Atmospheric Mission Performance Cluster – Routine Operations Validation Service 22
1.3 Purpose, scope and outline of this document 23
2 S5P Data Quality Requirements 24
3 Validation Results: L1B_RA and L1B_IR 25
3.1 L1B products 25
3.2 Recommendations for data usage followed 25
3.3 Validation approach 25
3.3.1 Comparison to other satellites 26
3.3.2 Analysis with radiative transfer model calculations 26
3.3.3 Stability monitoring using Pseudo-Invariant Calibration Sites 27
3.4 Validation of L1B NRTI 27
3.5 Validation of L1B OFFL 27
4 Validation Results: L2_O3 30
4.1 L2_O3 products and requirements 30
4.2 Validation approach 30
4.2.1 Ground-based networks 30
4.2.2 Satellites 30
4.2.3 Field campaigns and modelling support 30
4.3 Validation of L2_O3 OFFL 31
4.3.1 Recommendations for data usage followed 31
4.3.2 Status of validation 31
4.3.3 Bias 32
4.3.4 Dispersion 34
4.3.5 Dependence on influence quantities 35
4.3.6 Long-term stability 36
4.3.7 Short term variability 37
4.3.8 Geographical patterns 37
4.3.9 Other features 37
4.4 Validation of L2_O3 NRTI 37
4.4.1 Recommendations for data usage followed 37
4.4.2 Status of validation 37
4.4.3 Bias 38

4.4.4 Dispersion ... 40
4.4.5 Dependence on influence quantities ... 41
4.4.6 Short term variability ... 42
4.4.7 Geographical patterns ... 42
4.4.8 Other features ... 42
4.5 Equivalence of NRTI and OFFL data, and comparison to other satellite datasets ... 42
5 Validation Results: L2_O3_TCL ... 44
5.1 L2_O3_TCL products and requirements ... 44
5.2 Validation approach ... 44
5.2.1 Ground-based networks ... 44
5.2.2 Satellites ... 44
5.2.3 Field campaigns and modelling support ... 45
5.3 Validation of L2_O3_TCL OFFL (CCD) ... 45
5.3.1 Recommendations for data usage followed ... 45
5.3.2 Status of validation ... 45
5.3.3 Bias ... 45
5.3.4 Dispersion ... 46
5.3.5 Dependence on influence quantities ... 46
5.3.6 Seasonal cycle and shorter term variability ... 46
5.3.7 Geographical patterns ... 47
5.3.8 Other features ... 47
6 Validation Results: L2_O3_PR ... 51
6.1 L2_O3_PR products and requirements ... 51
6.2 Validation approach ... 51
6.2.1 Ground-based networks ... 51
6.2.1.1 Balloon-borne ozonesonde ... 52
6.2.1.2 Differential absorption ozone lidars (DIAL) ... 52
6.2.2 Satellite intercomparisons ... 52
6.2.3 Analysis of information content ... 52
6.2.4 Analysis of daily global maps ... 53
6.2.5 Parameter correlation checks ... 53
6.2.6 Field campaigns and modelling support ... 53
6.3 Validation of L2_O3_PR v2.4-6 (RPRO/OFFL) ... 53
6.3.1 Recommendations for data usage followed ... 53
6.3.2 Status of validation ... 53
6.3.3 Vertical sensitivity, resolution and registration ... 59
6.3.4 Bias ... 59
6.3.5 Dispersion ... 60
6.3.6 Chi-square tests ... 60
6.3.7 Dependence on influence quantities ... 60
6.3.8 Temporal variability ... 61
6.3.9 Geographical patterns ... 61

6.3.10 Other features ... 61
7 Validation Results: L2_NO2 ... 64
7.1 L2_NO2 products and requirements ... 64
7.2 Validation approach ... 64
7.2.1 Ground-based monitoring networks ... 64
7.2.2 Satellites ... 66
7.2.3 Field campaigns and modelling support ... 66
7.3 Validation of L2_NO2 ... 67
7.3.1 Recommendations for data usage ... 67
7.3.2 Status of validation ... 67
7.3.3 Tropospheric NO_2 column ... 68
7.3.4 Stratospheric NO_2 column ... 75
7.3.5 Total NO_2 column ... 79
7.4 Equivalence of L2_NO2 NRTI and OFFL products ... 82
7.4.1 Tropospheric NO2 Column NRTI vs OFFL ... 82
7.4.2 Total NO_2 Column NRTI vs OFFL ... 83
7.5 Internal consistency of the NO_2 validation results ... 85
7.5.1 NO_2 absorption cross-sections ... 85
7.5.2 ZSL-DOAS and PGN with low pollution level (mountain-top, arctic) ... 85
7.5.3 Stations with multiple instruments (different geometries) ... 86
7.5.4 Consistency between MAX-DOAS and PGN network results ... 86
8 Validation Results: L2_HCHO ... 88
8.1 L2_ HCHO products and requirements ... 88
8.2 Validation approach ... 88
8.2.1 Ground-based monitoring networks ... 88
8.2.2 Satellites ... 90
8.2.3 Field campaigns and modelling support ... 90
8.3 Validation of L2_ HCHO ... 91
8.3.1 Recommendations for data usage ... 91
8.3.2 Status of validation ... 91
8.3.3 Bias ... 91
8.3.4 Dispersion ... 99
8.3.5 Dependence on influence quantities ... 101
8.3.6 Seasonal and short term variability ... 101
8.3.7 Geographical patterns ... 104
8.3.8 Impact of UPAS version changes ... 104
8.4 Equivalence of L2_HCHO NRTI and OFFL products ... 106
8.4.1 Bias ... 106
8.4.2 Dispersion ... 107
9 Validation Results: L2_SO2 ... 108
9.1 L2_SO2 products and requirements ... 108
9.2 Validation approach ... 108

9.2.1 Ground-based networks 108
9.2.2 Satellites 108
9.2.3 Field campaigns and modelling support 108
9.2.4 Test of the expectation of zero SO_2 SCDs (within detection limit) outside volcanic plumes and strongly polluted regions 108
9.3 Validation of L2_SO2 NRTI 109
9.3.1 Recommendations for data usage followed 109
9.3.2 Status of validation 109
9.3.3 Bias 113
9.3.4 Dispersion 113
9.3.5 Dependence on influence quantities 114
9.3.6 Short term variability 114
9.3.7 Geographical patterns 115
9.3.8 Other features 115
9.4 Equivalence of L2_SO2 NRTI and OFFL products 115
10 Validation Results: L2_SO2 Layer Height 116
10.1 Recommendations for data usage 116
10.2 Status of validation 116
10.2.1 Bias 120
10.2.2 Dispersion 121
10.2.3 Dependence on influence quantities 121
10.2.4 Short term variability 121
10.2.5 Geographical patterns 121
10.2.6 Other features 121
11 Validation Results: L2_CO 122
11.1 L2_CO products and requirements 122
11.2 Validation approach 122
11.2.1 Ground-based networks 122
11.2.2 Satellites 122
11.2.3 Field campaigns and modelling support 122
11.3 Validation of L2_CO OFFL 123
11.3.1 Recommendations for data usage 123
11.3.2 Status of validation 123
11.3.3 Bias 124
11.3.4 Dispersion 125
11.3.5 Dependence on influence quantities 125
11.3.6 Short term variability 126
11.3.7 Geographical patterns 126
11.3.8 Other features 127
11.4 Equivalence of L2_CO OFFL and NRTI products 127
11.5 Comparison of L2_CO OFFL standard and de-striped products 128
12 Validation Results: L2_CH4 129

12.1 L2_CH4 products and requirements 129
12.2 Validation approach 129
12.2.1 Ground-based networks 129
12.2.2 Satellites 129
12.2.3 Other TROPOMI XCH4 products 130
12.2.4 Field campaigns and modelling support 131
12.3 Validation of L2_CH4 OFFL 131
12.3.1 Recommendations for data usage followed 131
12.3.2 Status of validation 132
12.3.3 Bias 132
12.3.4 Dispersion 137
12.3.5 Dependence on influence quantities 137
12.3.6 Short term variability 137
12.3.7 Other features 137
12.4 Validation of L2_CH4 RPRO+OFFL sun-glint retrievals 139
13 Validation Results: L2_CLOUD 141
13.1 L2_CLOUD products and requirements 141
13.1.1 Differences between validation results before and after version 2.4.1 reprocessing 141
13.2 Validation approach 142
13.2.1 Overview of changes 142
13.2.2 Ground-based networks 142
13.2.3 Satellites 145
13.2.4 Alternative S5P cloud algorithms 145
13.2.5 Field campaigns and modelling support 145
13.3 Validation of L2_CLOUD OFFL 146
13.3.1 Recommendations for data usage 146
13.3.2 Status of validation 146
13.3.3 Quality screening numbers 147
13.3.4 Radiometric cloud fraction (L2_CLOUD CAL & L2_CLOUD CRB) 147
13.3.5 Cloud top height (L2_CLOUD CAL) and cloud height (L2_CLOUD CRB) 151
13.4 Comparison of L2_CLOUD NRTI and OFFL products 162
14 Validation Results: L2_AER_AI 163
14.1 L2_AER_AI products and requirements 163
14.2 Validation approach 163
14.2.1 Ground-based networks 163
14.2.2 Satellites 163
14.2.3 Field campaigns and modelling support 163
14.3 Validation of L2_AER_AI NRTI 164
14.3.1 Recommendations for data usage followed 164
14.3.2 Status of validation 164
14.3.3 Bias 167
14.3.4 Dispersion 167

14.3.5 Dependence on influence quantities ... 167
14.3.6 Short term variability ... 167
14.3.7 Geographical patterns ... 169
14.3.8 Other features ... 169
14.4 Equivalence of L2_AER_AI NRTI and OFFL products ... 169
15 Validation Results: L2_AER_LH ... 171
15.1 L2_AER_LH products and requirements ... 171
15.2 Validation approach ... 171
15.2.1 Ground-based networks ... 171
15.2.2 Satellites ... 172
15.2.3 Field campaigns and modelling support ... 172
15.3 Validation of L2_AER_LH ... 172
15.3.1 Recommendations for data usage followed ... 172
15.3.2 Status of validation ... 173
15.3.3 Version 1 data ... 179
15.3.4 Validation approach ... 183
15.3.5 Case studies ... 187
15.3.6 Bias ... 190
15.3.7 Dispersion ... 191
15.3.8 Dependence on influence quantities ... 191
15.3.9 Short term variability ... 192
15.3.10 Geographical patterns ... 192
15.3.11 Other features ... 193
16 References ... 196
16.1 Reference documents ... 196
16.2 Peer-reviewed articles ... 196
16.3 Electronic references ... 203
17 Acknowledgements ... 205
17.1 ATM-MPC S5P Routine Operations Validation Service ... 205
17.2 S5P validation facilities ... 206
17.3 Validation data ... 206
17.4 Agency support ... 208
18 Terms, definitions and abbreviated terms ... 209
18.1 Terms and definitions ... 209
18.2 Acronyms and abbreviations ... 210

1 Introduction

1.1 Background information on Sentinel-5 Precursor TROPOMI

TROPOspheric Monitoring Instrument (TROPOMI) [ER_TROPOMI] is the unique payload of the ESA/Copernicus Sentinel-5 Precursor mission (S5P) launched on October 13, 2017. The prime function of TROPOMI is to monitor the global distribution of atmospheric trace gases and aerosols for a better understanding of air quality, the ozone layer, atmospheric chemistry and transport, ultraviolet radiation, and climate change. The instrument is a nadir-viewing hyperspectral spectrometer measuring, in the ultraviolet-visible (270-495 nm), near infrared (675-775 nm) and shortwave infrared (2305-2385 nm), the solar radiation scattered by the Earth's atmosphere and reflected by the Earth's surface and by clouds, as well as solar spectral irradiance. Daily coverage at the high horizontal resolution of 7 x 3.5 km^2 before and 5.5 x 3.5 km^2 after the operations switch to smaller ground pixel size activated on the 6th of August 2019, is accomplished thanks to a Sun-synchronous polar orbit (equator crossing time of 13:30 local solar time) and a wide swath width of 2600 km across track. From the TROPOMI radiometric measurements of Earth's radiance and solar irradiance, on-ground data processors retrieve the atmospheric abundance of ozone (O_3), nitrogen dioxide (NO_2), formaldehyde (HCHO), sulphur dioxide (SO_2), carbon monoxide (CO), methane (CH_4), as well as cloud and aerosol properties.

The S5P mission is a key component of the space segment of the European Earth Observation programme Copernicus [ER_CoperESA]. As such, it has an operational and service-oriented vocation. With a 7-year nominal operation lifetime, the S5P mission aims at filling in the observational gap of key atmospheric composition data between, from one part, Envisat SCIAMACHY (operational in 2002-2012), EOS-Aura OMI (operational since 2004) and the EUMETSAT EPS MetOp GOME-2 series (initiated in 2006, with the latest MetOp-C launched in November 2018), and from the other part, the upcoming series of Copernicus Sentinel-4 and Sentinel-5 missions scheduled for 2025-2045.

1.2 Copernicus Atmospheric Mission Performance Cluster – Routine Operations Validation Service

Procured by an international consortium contracted by the European Space Agency (ESA), the Copernicus Atmospheric Mission Performance Cluster (ATM-MPC) provides an operational service-based response to the S5P mission requirements for quality control, calibration, validation and end-to-end system performance monitoring during the Routine Operations phase of the S5P mission.

In-flight calibration and characterisation of the TROPOMI instrument, long-term monitoring of the instrument sensor performance and ageing, and routine Quality Control (QC) of the operational Level-1 (radiometric) and Level-2 (geophysical) data products are coordinated by the Royal Dutch Meteorological Institute (KNMI), and documented on the *TROPOMI Portal for Instrument and Calibration* [ER_MPS] and the *TROPOMI Portal for Level-2 Quality Control* [ER_L2QC].

Geophysical validation of the operational Level-1 and Level-2 data products is coordinated by the Royal Belgian Institute for Space Aeronomy (BIRA-IASB), and documented on the Portal of the *TROPOMI Validation Data Analysis Facility* (VDAF) [ER_VDAF]. The TROPOMI routine operations validation service makes use of Fiducial Reference Measurements (FRM) and other correlative data of documented quality (ground-based and satellite measurements, dedicated field campaigns), to assess the overall quality, the compliance with mission requirements and the validity of uncertainty estimates of the TROPOMI data products. This service monitors validation results on a cyclic basis and updates every three months the present *Routine Operations Consolidated Validation Report* (ROCVR). It also contributes quality assessment support to the continuous evolution of the data processors.

1.3 Purpose, scope and outline of this document

The present document (DI-MPC-ROCVR / TD-VALREP) reports consolidated validation results for the S5P TROPOMI Level-1 and Level-2 operational data products. This report has been produced by the ATM-MPC Routine Operations Validation Service. It integrates validation results from the MPC Validation Data Analysis Facility (VDAF) consortium (**Table 23**) with support from other activities and dedicated field campaigns documented on the TROPOMI website [ER_TROPOMI], as well as ad hoc contributions from S5P Validation Team (S5PVT) AO projects [ER_S5PVT].

Updated with a trimestral frequency, S5P data quality information provided in this document supersedes that provided in previous versions. It complements S5P data quality information provided in the *Product Readme Files* (PRFs) attached to S5P data products released publicly. For details and for recommendations for data usage, data users are encouraged to read the PRF, *Product User Manual* (PUM) and *Algorithm Theoretical Basis Document* (ATBD) associated with the data products, all available on the Copernicus Sentinel Portal for S5P products and algorithms [ER_CoperATBD] and also on the TROPOMI Portal [ER_TROPOMI].

This ROCVR update #24 reports quality information for the latest versions of the S5P operational data products acquired from April 2018 to August 2024: L1B versions 01.00.00 to 02.01.00, versions 02.04.00 to 02.06.00 of the NL-L2 processor suite (O_3 profile, NO_2, CO and CH_4 columns, and AAI and ALH data), and version 02.04.01 to 02.06.01 of the UPAS processor suite (total and tropospheric O_3 columns, HCHO and SO_2 columns, and CLOUD). This document is structured as follows:

Document Information 3
Executive Summary 5
Representative Quality Indicators 14
Table of Contents 16
1 Introduction 22
2 S5P Data Quality Requirements 24
3 Validation Results: L1B_RA and L1B_IR 25
4 Validation Results: L2_O3 30
5 Validation Results: L2_O3_TCL 44
6 Validation Results: L2_O3_PR 51
7 Validation Results: L2_NO2 64
8 Validation Results: L2_HCHO 88
9 Validation Results: L2_SO2 108
10 Validation Results: L2_SO2 Layer Height 116
11 Validation Results: L2_CO 122
12 Validation Results: L2_CH4 129
13 Validation Results: L2_CLOUD 141
14 Validation Results: L2_AER_AI 163
15 Validation Results: L2_AER_LH 171
16 References 196
17 Acknowledgements 205
18 Terms, definitions and abbreviated terms 209

2 S5P Data Quality Requirements

Validation results can be interpreted to evaluate whether or not S5P Level 2 data products meet user requirements. Targets for key quality indicators of the S5P Level 2 data products have been formulated in the *S5P Geophysical Validation Requirements* document ([S5PVT-Req], Page 19) and the *S5P Cal/Val Plan for the Operational Phase* ([S5P-CSCOP], Page 14). Evolution of these requirements is supported by the Sentinel-5p Quality Working Group (QWG), who agreed (i) to adopt for tropospheric ozone column data the requirements expressed by the Climate Research Group (CRG) within ESA's Ozone_cci project, (ii) to revise requirements for SO_2 column data, and (iii) to provide maximum values of the estimates instead of ranges. Refined requirements were adopted for three different cases of SO_2 column regimes. Expressed in terms of measurement bias (estimate of the systematic measurement error) and uncertainty (measurement uncertainty, that is, dispersion of the quantity values being attributed to the measurand), these targets are reproduced hereafter in **Table 3**. Quality targets are typical of several known applications; nevertheless, it always remains the uttermost responsibility of any users to check the fitness of the S5P data for their own purpose, with respect to their own particular requirements.

Product ID	Level-2 Geophysical Quantity	Requirement: Vertical Resolution	Requirement: Uncertainty due to Systematic Effects	Requirement: Uncertainty due to Random Effects
L2_O3	Total O_3	total column	5%	2.5%
L2_O3_PR	O_3 profile (incl. troposphere)	6 km	30%	10%
L2_O3_TCL	O_3 tropospheric column	tropospheric column	25%	25%
L2_NO2	NO_2 tropospheric column	tropospheric column	50%	0.7 Pmolec.cm^{-2}
	NO_2 stratospheric column	stratospheric column	10%	0.5 Pmolec.cm^{-2}
L2_SO2	Total SO_2	total column	0.5 DU	1 DU
	Enhanced SO_2 (SCD <1.5 DU)	total column	0.5 DU	1 DU
	Enhanced SO_2 (SCD >1.5 DU)	total column	30%	30%
L2_SO2_LH	SO_2 layer height	total column	n/a	2 km
L2_HCHO	Total HCHO	total column	80%	12 Pmolec.cm^{-2}
L2_CO	Total CO	total column	15%	10%
L2_CH4	Total CH_4	total column	1.5%	1%
L2_CLOUD	Cloud Fraction	total column	20%	0.05
	Cloud Height (pressure)	total column	20%	0.5km (P<30hPa)
	Cloud albedo (optical thickness)	total column	20%	0.05 (10)
L2_AER_AI	Aerosol Absorbing Index	total column	1 AAI	0.1 AAI
L2_AER_ALH	Aerosol Layer Height	total column	100 hPa	50 hPa

Table 3 – Data quality targets for the operational Sentinel-5 Precursor TROPOMI Level 2 data products: vertical resolution and measurement uncertainty components associated with systematic and random effects, respectively (adapted by Sentinel-5p QWG from [S5PVT-Req] and [S5P-CSCOP]). SCD: slant column density, before the conversion to vertical column density using an air mass factor.

3 Validation Results: L1B_RA and L1B_IR

3.1 L1B products

This Section reports on the validation of the S5P TROPOMI L1B products Version 02.01.00 identified in **Table 4**. Current conclusions are based on a limited amount of version 02.01.00 data. The conclusions summarized hereafter need to be confirmed by longer time series and a larger amount of co-locations with correlative data in order to enable detection and quantification of potential patterns, dependences, seasonal cycles and longer-term features.

Table 4: – Identification of the S5P TROPOMI L1B products evaluated in this Section.

Product	Stream	Version	In operation from	In operation until
L1B_RA1/…/8	RPRO/OFFL	02.01.00	orbit 2818, 2018-04-30	Current version
L1B_IR_UVN/SIR	RPRO/OFFL	02.01.00	orbit 2818, 2018-04-30	Current version

Note: The operational phase (E2) of the S5P TROPOMI mission starts with orbit #02818.

3.2 Recommendations for data usage followed

An overview of the Sentinel-5p mission, the TROPOMI instrument and the algorithms for producing the L1b data products can be found in the Algorithm Theoretical Basis Document [ATBD]. Details of the data format are provided in the Input/Output Data Specification [IODS]. The metadata contained in the L1b data products are described in the Metadata Specification [MDS]. All these documents are available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

For Level 2 processing and related validation, the following additional notices apply:

- The L0-1b data processor annotates the data with quality assessment data in the fields `spectral_channel_quality`, `measurement_quality` and `ground_pixel_quality`. Level 2 developers are strongly encouraged to observe these quality fields in their retrievals and exclude flagged data as needed.
- All eight bands are processed individually in the L0-1b data processor. In case of missing data, for example in case of data dropouts during downlinks, this does not necessarily impact all bands (to the same extent). This means that a scanline can be missing for some bands, where it is not missing for other bands. When combining data from multiple bands, Level 2 algorithms should therefore always check and match the `delta_time` for these data and, in case of non-co-registered bands, the geolocation as well.
- For calculating reflectance from the radiance products, it is recommended to use the irradiance product with the sensing time close to the sensing time of the radiance product.

3.3 Validation approach

In-flight calibration and characterisation of the TROPOMI instrument, long-term monitoring of the instrument sensor performance and ageing, and routine Quality Control (QC) of the operational L1B data products are reported continuously on the *TROPOMI Portal for Instrument and Calibration* [ER_MPS]. These activities are out of scope of the present document.

Validation and verification of the radiance (L1B_RA) and irradiance (L1B_IR) Level-1b data products consist in comparison of spectral, radiometric and geolocation properties of the satellite data with those of external and independent datasets. Regular evaluation of Level-1b products is performed at DLR by experts independent from the assigned Level-1b algorithm developers and institute. Level-1b data are compared to various satellite instruments with similar characteristics that cover the different wavelength regions of TROPOMI (Suomi-SNPP OMPS-NP and VIIRS, OCO-2, GOSAT-2 TANSO-FTS-2), and to radiative transfer model calculations using reliable input parameters. Monitoring of stability is performed using comparisons of TROPOMI data over time over CEOS Pseudo-Invariant Calibration Sites (PICS). The approach follows guidance and incorporate findings from initiatives and collaborations such as CEOS WGCV (http://ceos.org), QA4EO (http://qa4eo.org), and GSICS (http://gsics.wmo.int).

3.3.1 Comparison to other satellites

Since S5P operates in tandem with NASA's Suomi-NPP platform (time difference ~3-5 min), a comparison to OMPS-NP spectra in the wavelength region 270-310nm and OMPSNM spectra in the interval 300-380 nm will be performed [Jaross (2017), Seftor et al. (2014), Wu et al. (2014)]. From VIIRS/S-NPP measurements in the VIS/NIR wavelength range (VIIRS bands M1, M2, M3, and M6) will be used [Cao et a. (2013), Xiong et al. (2020)].

In addition, for further comparisons in NIR data from OCO-2 (band B1 757-772nm) and TANSO-FTS-2/GOSAT-2 (band 1 757-775nm) can be utilized [Kataoka et a. (2017), Suto et al. (2021)]. The latter also allows for an evaluation of TROPOMI measurements in the 2.3μm spectral range (TANSO-FTS-2 band 3). For the comparison of TROPOMI to the other satellite sensors special attention will be paid to accounting for discrepancies regarding viewing geometry, spectral resolution, spatial resolution and observation times [Chander et al (2013)]. The impact of possible degradation effects of the individual sensors on the comparison will be investigated.

3.3.2 Analysis with radiative transfer model calculations

Concerning the evaluation of the TROPOMI products with radiative transfer model calculations, we will use the linearized pseudo-spherical vector discrete ordinate radiative transfer code VLIDORT [Spurr, 2006]. Simulations will be performed in the UV, VIS, and NIR wavelength range using geometric parameters and the spectral response function from TROPOMI. Atmospheric input parameters will be taken from the CAMS reanalysis dataset or the BASCOE assimilation system. Ozone profile shape information will be taken from ozonesonde flights. Only cloud- and aerosol-free pixels will be included and the sensitivity w.r.t. surface albedo and BRDF will be investigated in advance. The analysis will be performed for a selection of appropriate ground-based sites covering a sufficient number of latitudes and surface as well as atmospheric conditions [Tilstra et al. (2005), Van Soest et al. (2005), Liu et al. (2010), Cai et al. (2012), Wu et al. (2014), Tilstra et al. (2020)].

3.3.3 Stability monitoring using Pseudo-Invariant Calibration Sites

In addition to the comparative studies, we will use various terrestrial sites (e.g., Saharan PICS (USGS) or CEOS LANDNET sites such as Railroad Valley) which are suitable for monitoring the long-term radiometric stability of satellite sensors [Mishra et al. (2014), Sun et al. (2014), Uprety & Cao (2015), Coldewey-Egbers et al. (2018), Van Kempen et al. (2021)]. These sites fulfill a number of requirements regarding spatial uniformity and homogeneity, stability of spectral characteristics over time, and generally high reflectance to enhance the signal-to-noise ratio. Moreover, they are characterized by climatologically low aerosol loading, little rainfall, and practically no vegetation or human impact. Relative changes of the TROPOMI L1 products with respect to the beginning of the measurements will be analyzed in order to assess their temporal stability and to evaluate the degradation correction of the operational products [Ludewig et al. (2020)]. Findings and experiences from predecessor projects ESA GOME-Evolution and ESA FDR4ATMOS will be incorporated.

3.4 Validation of L1B NRTI

The near-real time L1b products are not distributed to users, and they are not validated separately. NRTI products use the same L01b data processor algorithms, and can only differ when the Calibration Key Data (CKD) used differs from OFFL. Currently no CKD is dynamically updated in OFFL, and hence no difference exists between NRTI and OFFL.

3.5 Validation of L1B OFFL

The validation of the wavelength assignment of the L1B_UVN version 2 products shows agreement of within 0.01 nm, which is within the pre-launch calibration accuracy. The reflectance in bands 1-3 is 1 % to 3 % lower than TOMS and the used ice radiance model. The radiance in bands 1-3 is up to 5 % smaller than OMPS radiance; above 320 nm this is a wavelength independent bias. Below 320 nm, the wavelength dependence seems to vary with the latitude. In band 1 around 280 nm the radiance deviates more than 10 % from OMPS values. The absolute radiometric calibration for UV radiance lacks accuracy and as a result may be updated in the future. In the spectrally overlapping regions of bands 2 and 3 there is a discrepancy of about 2 % in the L1b radiance signals. The radiance in band 6 was compared to model spectra in the continuum around the O2A band. The signal of TROPOMI is 1-2 % lower than the model. For bands 1 to 6 (UV, UVIS and NIR) degradation has been observed for the radiance. The degradation is the largest at short wavelengths. The decrease in radiance signal per 1000 orbits is between 0.31 % in band 1 and 0.02 % in band 6. The absolute and relative radiance radiometry of the SWIR bands were validated using reference stations in Railroad Valley and in the Saharan desert. Current validation results give upper limits of < 5 % for the absolute calibration and < 0.8 % for the relative calibration. The absolute irradiance calibration of TROPOMI has been compared to other published solar reference datasets. After an update to the calibration based on OMPS data, the UV and UVIS spectrometers agree within 0-5 % with the references. For extreme swath angles, the deviations are larger in the UV. For the NIR spectrometer the irradiance spectrum is approximately 1.5–3.5 % lower than the reference spectra. The SWIR spectrum is approximately 0.6 % lower than the closest reference spectrum.

The relative change of the reflectance derived from L1B V02.01.00 radiance and irradiance data for a small wavelength region in band 4 (436-454 nm) was analyzed using 18 Pseudo-Invariant Calibration Sites in the Saharan desert and the Arabian Peninsula. All S5P overpasses from May 2018 through August 2024 with a cloud fraction below 0.08 and a maximum distance of 8 km to the center of the site were investigated. **Figure 1** exemplarily shows the reflectance as a function of the orbit number for the Saharan sites Egypt-1, Niger-2, Libya-2, and Libya-4. The individual measurements were binned according to different viewing zenith angle sectors from 66° West to 66° East. These viewing zenith angle sectors were defined for each site separately based on the corresponding distribution of the

viewing angles. Since no BRDF correction was applied, a strong dependence of the reflectance on the viewing angle is seen. The reflectance is maximum for extreme off-nadir angles toward the western edge of the swath. In addition, a strong seasonal variation is observed. The solid black curve highlights nadir measurements (viewing zenith angle limited to the values closest to the nadir position). From these nadir pixels, the drift is computed. For all 18 sites the change in the reflectance per 1000 orbits is below 0.05 %.

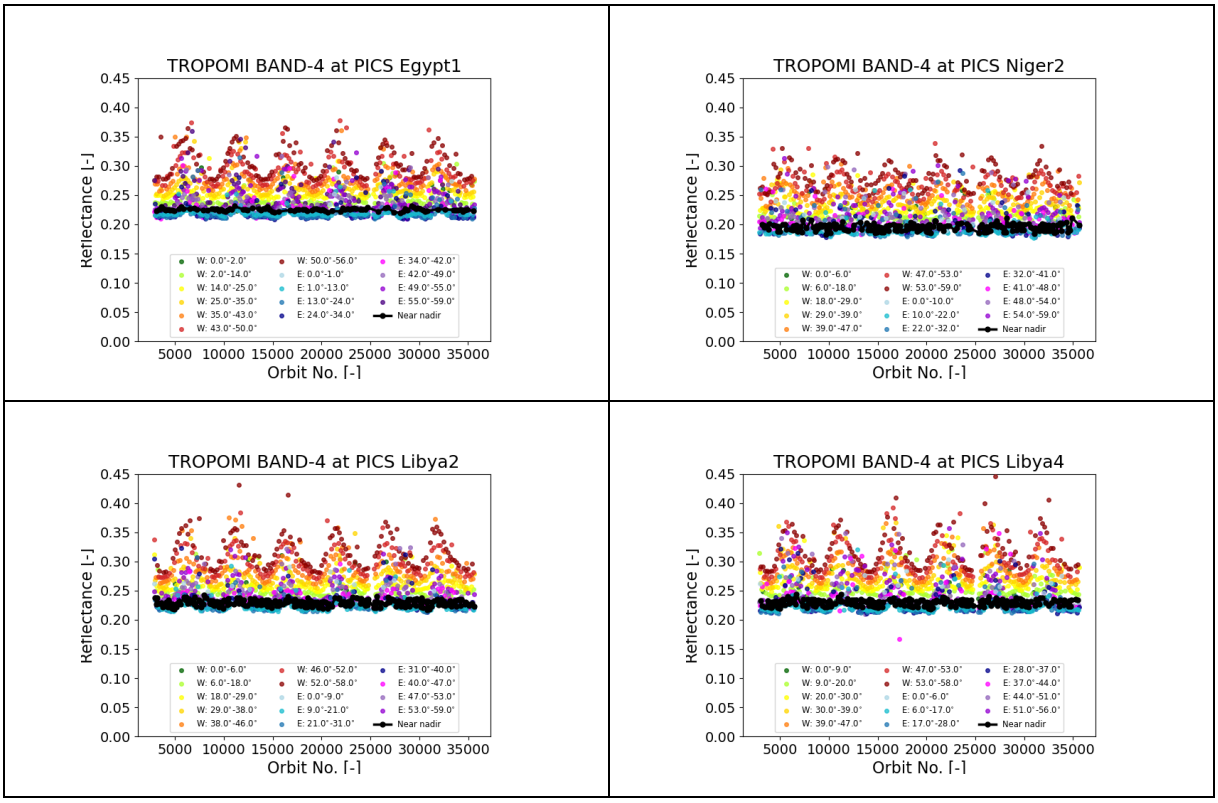


Figure 1: Reflectance derived from L1B v02.01.00 radiance and irradiance data for 436-454nm (band 4) as a function of the orbit number (May 2018 through August 2024) for PICS Egypt-1, Niger-2, Libya-2, and Libya-4. The individual measurements are binned for different viewing angle sectors from 66° West (green and red dots) to 66° East (blue and purple dots).

In addition, the TROPOMI irradiance for L1B_UVN version 02.01.00 in bands 1, 2, and 3 is compared to the corresponding irradiance from OMPS-NP (for TROPOMI band 1) and OMPS-NM (for TROPOMI bands 2 and 3) for the spectral intervals 270-300 nm, 300-330 nm, and 310-370 nm, respectively. Four spectra per year (01/06/2018, 01/09/2018, …, 01/06/2023) are extracted. L1B_UVN version 02.01.00 agrees well with the OMPS observations (see **Figure 2**). For TROPOMI bands 2 and 3, no wavelength dependence is observed. The ratio (TROPOMI vs OMPS) is very stable over the entire period.

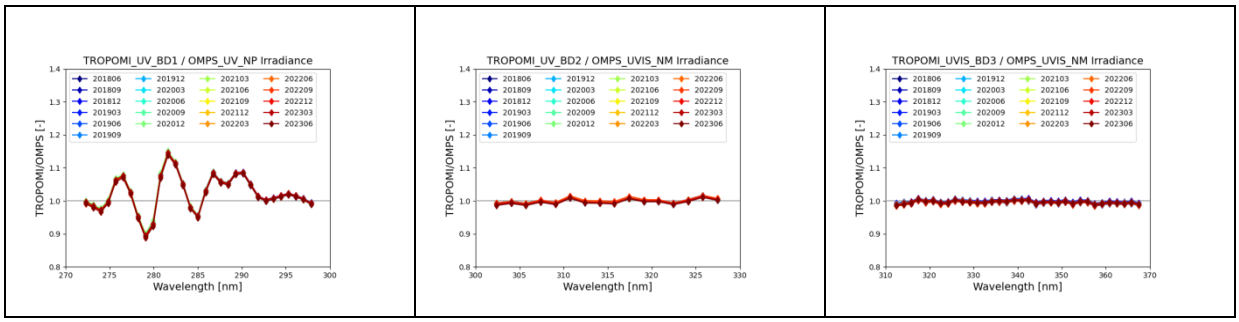


Figure 2: Ratio between TROPOMI L1B_UVN version 02.01.00 irradiance and OMPS irradiance as a function of wavelength for three spectral intervals. Left: 270-300nm, TROPOMI band 1 vs. OMPS-NP. Middle: 300-330nm, TROPOMI band 2 vs. OMPS-NM. Right: 310-370nm, TROPOMI band 3 vs. OMPS-NM.

4 Validation Results: L2_O3

4.1 L2_O3 products and requirements

This Section reports on the validation of the S5P TROPOMI L2_O3 product identified in **Table 1**. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. The NRTI and OFFL processors use different retrieval approaches, thus their respective validation is reported in separate subsections, and an additional section covers an analysis of their equivalence. The results presented here cover processor versions 02.04.01 (used for the full mission reprocessing, designated "RPRO" here) to 02.06.01 (the current operational processor). For validation results on other previous processor versions, the reader is referred to ROCVR version 17.1.0, available at https://mpc-vdaf.tropomi.eu/.

4.2 Validation approach

4.2.1 Ground-based networks

S5P TROPOMI L2_O3 total ozone column data are routinely compared to reference measurements acquired by instruments contributing to WMO's Global Atmosphere Watch (GAW): (1) Brewer (Kerr et al., 1981,1988) and (2) Dobson (Basher, 1982) UV spectrophotometers, and (3) NDACC Zenith Scattered Light (ZSL) DOAS UV-Visible spectrometers (Pommereau and Goutail, 1988, Hendrick et al., 2011). These are complemented with comparisons to direct-sun measurements obtained with the Pandora instruments of the Pandonia Global Network (PGN, a joint effort by NASA and ESA). Co-locations between S5P TROPOMI and direct-sun (DS) measurements are defined as "pixel contains station", with a maximum time difference of 3 hours. Note that direct-sun measurements obtained through the NDACC and WOUDC data archives are usually daily means of the individual measurements, while those from PGN are individual measurements (obtained at a rate of roughly eight per hour). To reduce co-location mismatch errors due to the significant difference in horizontal smoothing between S5P and ZSL-DOAS measurements, S5P O3 column values (from afternoon ground pixels at high resolution) are averaged over the footprint of the larger air mass to which the ground-based twilight zenith-sky measurement is sensitive. For more details about the validation methodology, see Lambert et al. (1997, 1999), Balis et al. (2007), Koukouli et al. (2015), Verhoelst et al. (2015), and Garane et al. (2019).

4.2.2 Satellites

S5P TROPOMI L2_O3 total ozone column data have also been compared to MetOp-A, MetOp-B, and MetOp-C GOME-2 ozone column data (versions GDP 4.8 for GOME-2A/B and GDP 4.9 for GOME-2C), to Suomi-NPP OMPS-nadir ozone column data, and to S5P ozone column data retrieved with the other S5P operational processor (NRTI vs. OFFL).

4.2.3 Field campaigns and modelling support

Since December 4, 2018, S5P L2_O3 NRTI total ozone data is monitored and assimilated operationally in the Copernicus Atmosphere Monitoring Service system (CAMS), which also assimilates ozone data from a list of other satellite instruments. See Inness et al. (2019) for further details. Specific checks are also carried out by CAMS to verify the effect of a particular event like e.g. a processor upgrade.

4.3 Validation of L2_O3 OFFL

4.3.1 Recommendations for data usage followed

Data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

In order to avoid misinterpretation of the data quality, it is recommended to use only those TROPOMI pixels associated with a qa_value above 0.5. This filter was applied here. As opposed to the implementation in the earliest versions of the processor, this filter no longer contains a threshold on SZA as this threshold was found to reject good retrievals at high SZA in key atmospheric conditions (polar ozone loss). More details are provided in the PRF.

4.3.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC VDAF and by S5P Validation Team (S5PVT) AO projects. This summary is based on coordinated operational validation activities carried out using the Automated Validation Server of the S5P MPC VDAF, the Multi-TASTE versatile multi-platform validation system operated at BIRA-IASB, and the ozone validation system operated at AUTH.

Current conclusions are valid for the S5P data obtained in the operational phase E2 of the mission, from May 2018 until August 2024, and on the reference data available at the time of this report: typically, until beginning of August 2024 for the PGN data, up to the end of June 2024 for the Dobson and Brewer data, and up to the end of August 2024 for the ZSL-DOAS SAOZ data. For the current report, Brewer and Dobson measurements were obtained through the World Ozone and UV Radiation Data Centre (WOUDC) in Toronto, the NDACC Data Host Facility, and WMO's Ozone Mapping Centre in Thessaloniki. If a station archives data both into WOUDC and NDACC HDF, the source with the most recent data is adopted. ZSL-DOAS measurements were collected through the SAOZ network Real-Time processing facility operated by CNRS LATMOS (LATMOS_RT) and the PGN data were obtained through EVDC. Over the period of this analysis, 300 to 2000 co-locations with respect to the reference data have been identified at about 40 Brewer and Dobson stations, at 30 PGN stations, and at 12 ZSL-DOAS SAOZ stations, sampling many latitudes from the Arctic to the Antarctic (**Figure 3**).

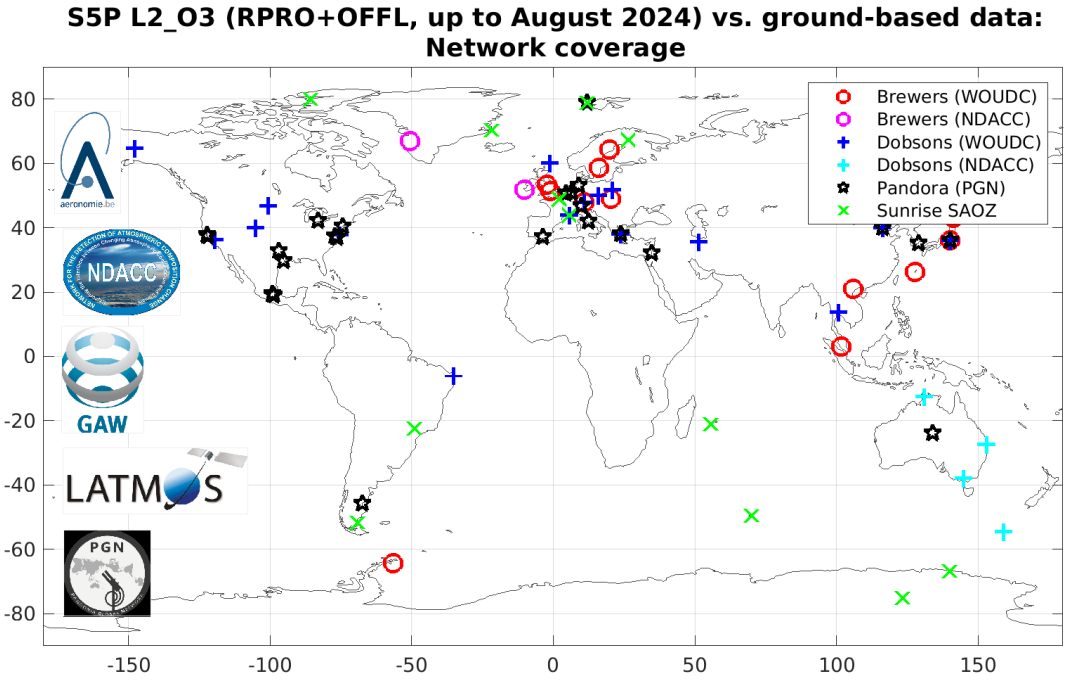


Figure 3: Geographical distribution of Brewer, Dobson, Pandora, and ZSL-DOAS ground-based stations for which co-locations with S5P L2_O3 OFFL ozone data have were used (period up to August 2024).

4.3.3 Bias

The systematic difference between S5p L2_O3 OFFL and reference ground-based data at individual stations rarely exceeds 3 %, as depicted in **Figure 4**. The median bias calculated over the entire ground-based networks is of the order of +1.3 %. Between 50˚S and 50˚N, the mean agreement with other satellite data derived with the same processor (GODFIT v4) is mostly within 1% as well (**Figure 5**). This median bias value falls well within the mission requirements (max. bias 5%).

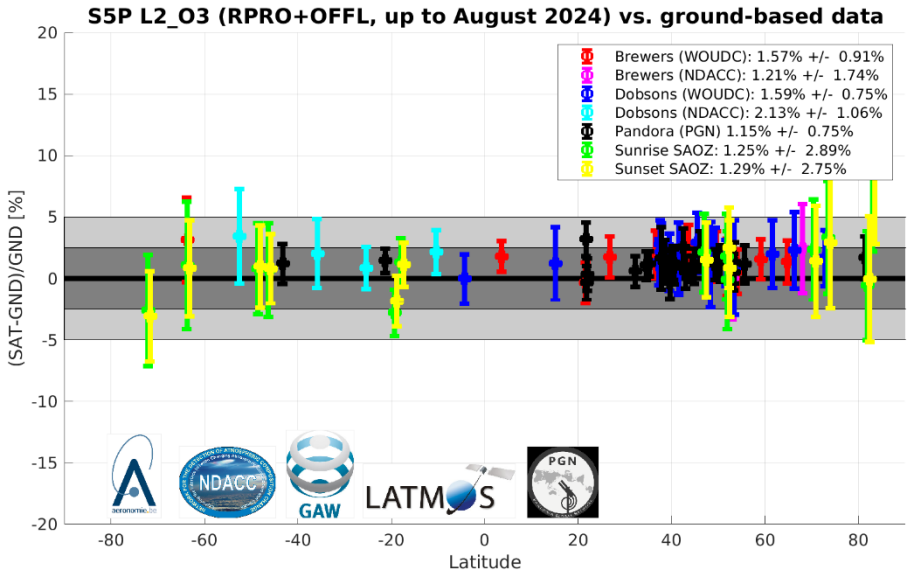


Figure 4: Meridian dependence of the median (the circular markers) and spread (±1 sigma, the error bars) of the percent relative difference between S5P TROPOMI L2_O3 (RPRO+OFFL up to August 2024) and ground-based (GND) ozone column data, represented at individual stations from the Antarctic to the Arctic and per measurement type. The values in the legend correspond to the median and spread of all median (per station) differences. For clarity, sunrise and sunset ZSL-DOAS results are represented separately (offset by -0.5˚ and +0.5˚ in latitude).

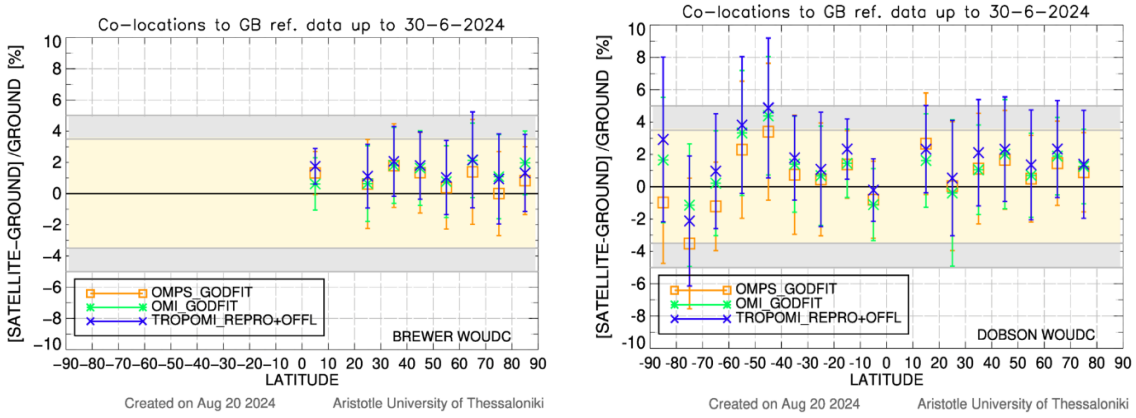


Figure 5: Comparison of the mean percentage differences between three satellite products (S5P L2_O3 OFFL, OMI GODFIT v4 and OMPS GODFIT v4) and ground-based total ozone data, versus latitude. The Brewer network comparisons are shown in the left-hand panel and the Dobson network comparisons are depicted in the right-hand panel, both datasets are downloaded from WOUDC. The period of data used for these plots is April 2018 –June 2024 (due to ground-based data availability). TROPOMI OFFL total ozone is reprocessed until 3 August 2022 and the operational OFFL product is used afterwards.

Time series of the evolution of the monthly mean difference for the comparison against Brewers in the Northern hemisphere are shown in **Figure 6**, revealing the excellent stability and the agreement between the NRTI and OFFL products. On 6 August 2019, the nominal ground pixel resolution of the TROPOMI measurement was reduced to 5.5 x 3.5 km², i.e. shorter by 1.5 km in the along-track direction, by reducing the integration time. This did not have an impact on the agreement between satellite and ground-based measurements. The slight drift seen in the last year of data is neither confirmed at individual high-quality sites, nor by the SAOZ or PGN comparisons and thus most likely related to the limited representativeness of the Brewer network in the most recent months.

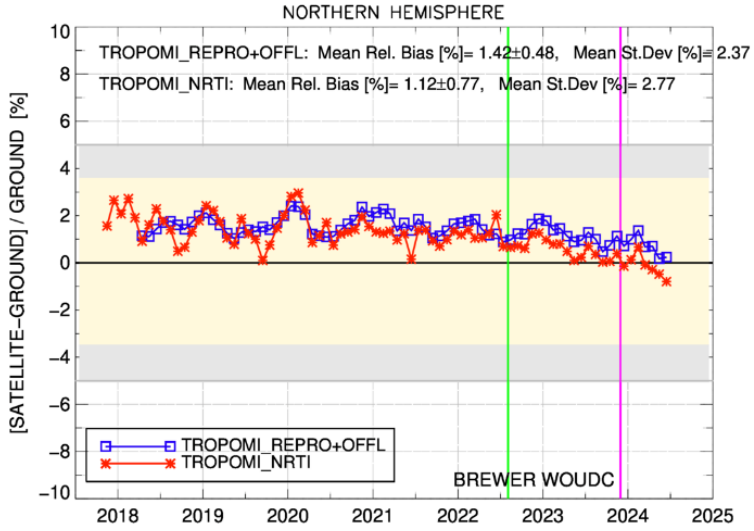


Figure 6: Time series of the monthly mean difference for the comparisons against Brewers in the Northern hemisphere for both the RPRO+OFFL product (blue) and the NRTI product (red). The green line denotes the switch to processor version 2.4.1, and the magenta line the switch to processor 2.6.1. For the OFFL data, earlier data are reprocessed with v2.4.1; for NRTI, these data correspond to earlier processors.

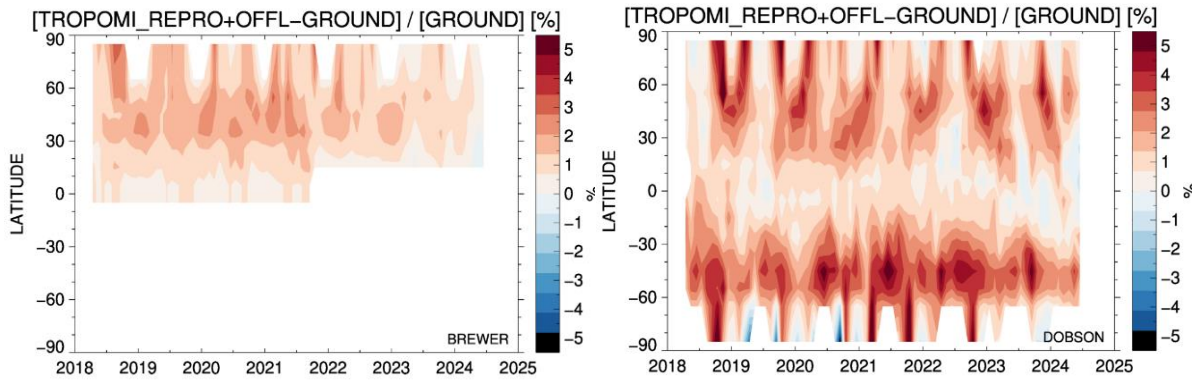


Figure 7: Time and latitude resolved analysis of the differences between RPRO+OFFL total ozone and Brewer measurements (left-hand panel) and Dobson measurements (right-hand panel). Strong seasonal features in the Dobson comparisons are partly due to the use of a fixed effective temperature in the Dobson measurement processing.

An analysis of the medium-term systematic differences as a function of latitude and time is shown in **Figure 7**. The overall slight overestimation by S5P is visible, and observed to be a little more pronounced at mid latitudes (best visible in the Brewer comparisons, i.e., the left-hand panel). The Dobson comparisons show more pronounced seasonal features at high latitudes, in particular for low-sun conditions, but these are (at least partly) attributable to the fixed effective temperature assumed in the Dobson data processing. A similar analysis (not shown here) for the historical sequence of OFFL processors (i.e., not reprocessed data) clearly reveals the increase in positive offset by approximately 1 % in July 2021 (related to an update in the L1b processor).

4.3.4 Dispersion

The ±1σ dispersion of the difference (between S5P and reference ground-based network data) around their median value rarely exceeds 3-4% for the comparisons with direct-sun instruments (cf. the error bars depicted in **Figure 4**). Combining random errors in both satellite and reference measurements with irreducible co-location mismatch effects, it is concluded that the random uncertainty on the S5P measurements falls within the mission requirements of max. 2.5 %.

Whether the uncertainties reported with the data (the so-called ex-ante uncertainties) are compatible with the observed dispersion can be tested by computing, per station, the χ_r^2 , defined as:

$$\chi^2{}_r = \frac{1}{N-1} \Sigma \frac{(SAT-GND - mean(SAT-GND))^2}{\sigma_{SAT}{}^2 + \sigma_{GND}{}^2}$$

where, σ_{SAT} and σ_{GND} refer to the ex-ante satellite and ground-based measurements respectively. The histogram of these χ_r^2 is shown in **Figure 8**. The histogram peaks close to one, with only a few values above 3. This suggests the ex-ante uncertainties to be realistic.

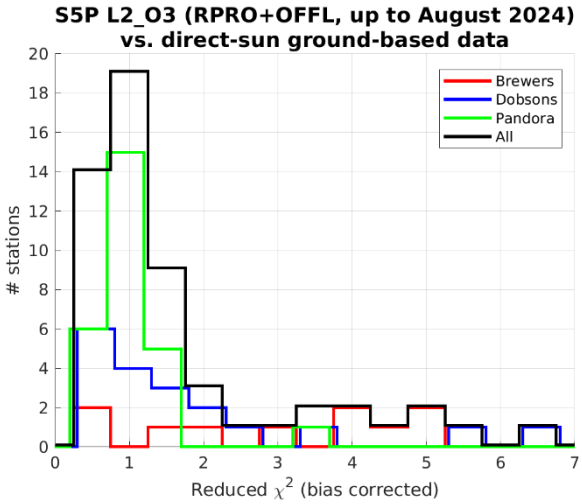


Figure 8: Histogram of the χ_r^2 values (one per reference instrument) for S5p OFFL total ozone versus different direct-sun ground-based instruments.

4.3.5 Dependence on influence quantities

The evaluation of potential dependence of the S5P bias and dispersion on the Solar Zenith Angle (SZA, evaluated up to 80°), surface albedo and cloud fraction (CF) of the TROPOMI measurement does not reveal any variation of the bias larger than 2-3 % over the range of those influence quantities (**Figure 9**).

The scatter of the data comparisons of about 2-3 % increases up to 5 % at large SZAs and at latitudes beyond 50°, which is expected knowing that random errors in both satellite and reference measurements as well as irreducible co-location mismatch errors increase at high latitude and low sun elevation.

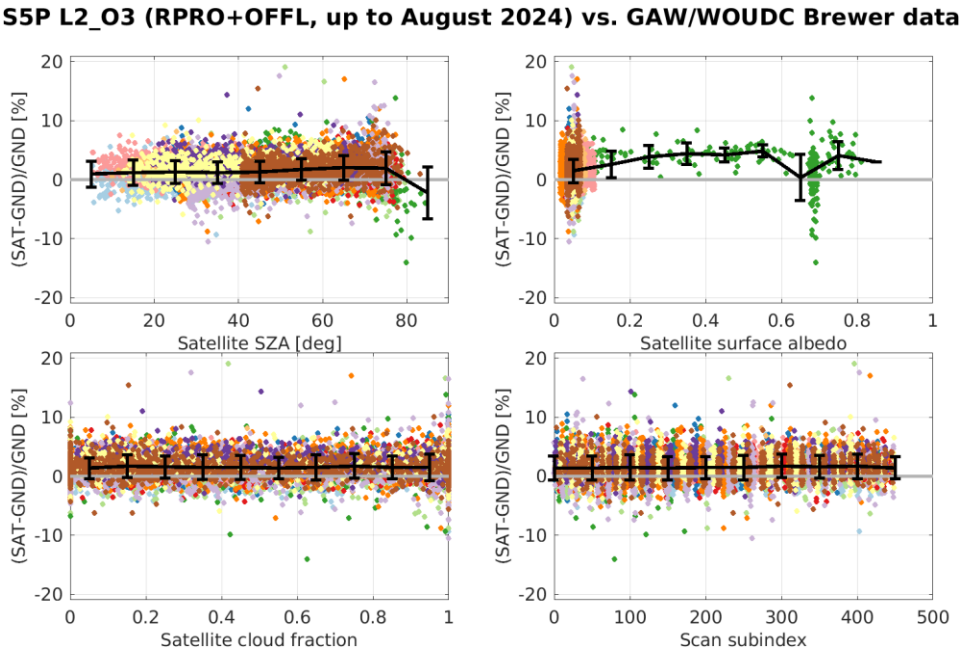


Figure 9: Dependence of the difference between S5P OFFL and ground-based Brewer total ozone data on the solar zenith angle (SZA), the surface albedo, the fractional cloud cover, and the scan sub-index of the satellite measurement. Black curve: mean and standard deviation over bins of 10 degrees of SZA, 0.1 of surface albedo, 0.05 of cloud fraction, and 50 pixels over across-track scan. Coverage of the high SZA and of the surface albedo is poor for these Brewer comparisons, but the SAOZ comparisons (not shown) which offer better coverage for these parameter ranges reveal no anomalies.

4.3.6 Long-term stability

Long-term stability was assessed by performing robust linear regression on the relative differences between the S5P and ground-based measurements at selected stations. This station selection was based on the following set of criteria:

- at least 5 years difference between the 1st and last measurement pairs,
- at least 250 measurement pairs in total,
- an inhomogeneity value (Sofieva et al., 2014) for the temporal distribution of the measurements below 0.25 for at least 4 years in the time series, ensuring an even distribution of measurement pairs over the better part of the full time series.

Figure 10 summarizes the results of this drift analysis in a pole-to-pole graph. For none of the networks, a significant drift was observed at a 1.5-2 %/decade sensitivity. Results at almost all selected stations comply with the GCOS 2022 Implementation Plan (IP) threshold value of at most 3 %/decade drift.

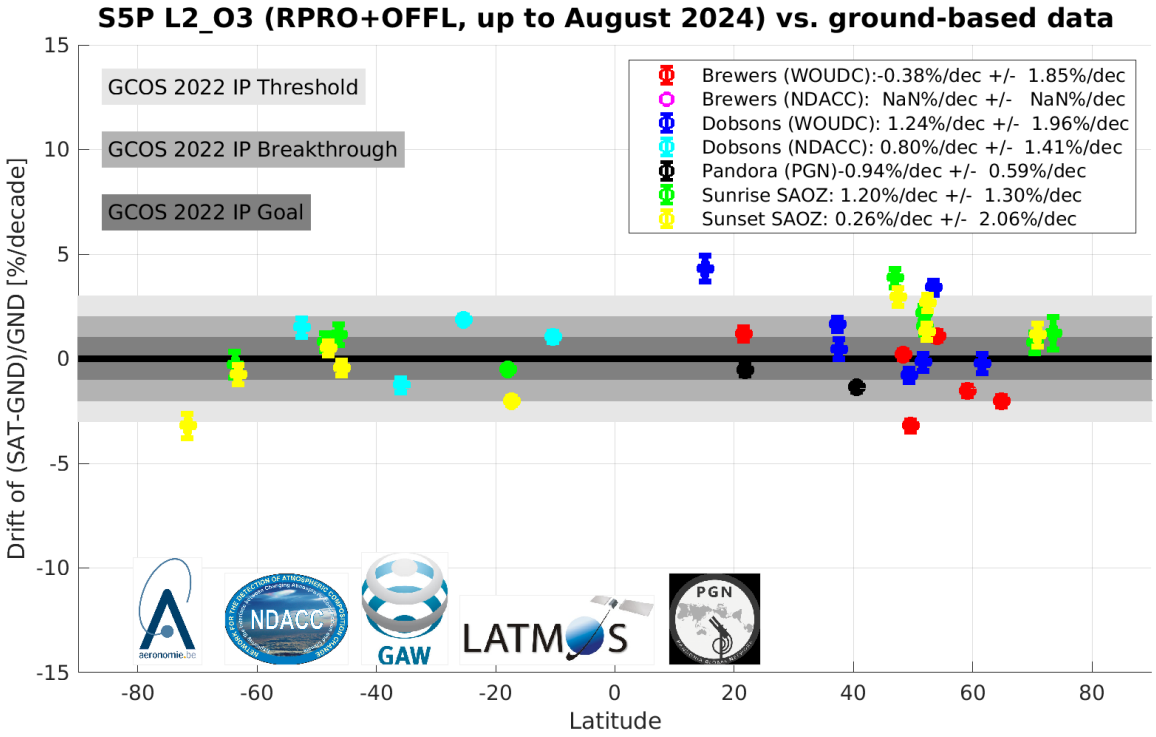


Figure 10: Drift of S5P OFFL versus the ground-based measurements, calculated as a robust linear regression on the relative differences at stations with a good (see main text for criteria) temporal sampling. Error bars denote the formal uncertainty on the slope in the linear regression. The uncertainty values in the legend correspond to the standard deviation of the station-to-station scatter across all latitudes.

4.3.7 Short term variability

Qualitatively, at all of the 50 ground-based reference stations, short scale temporal variations in the ozone column as captured by ground-based instruments are reproduced very similarly by S5P, as illustrated in **Figure 11**. The overall good agreement is corroborated by Pearson correlation coefficients always above 0.95.

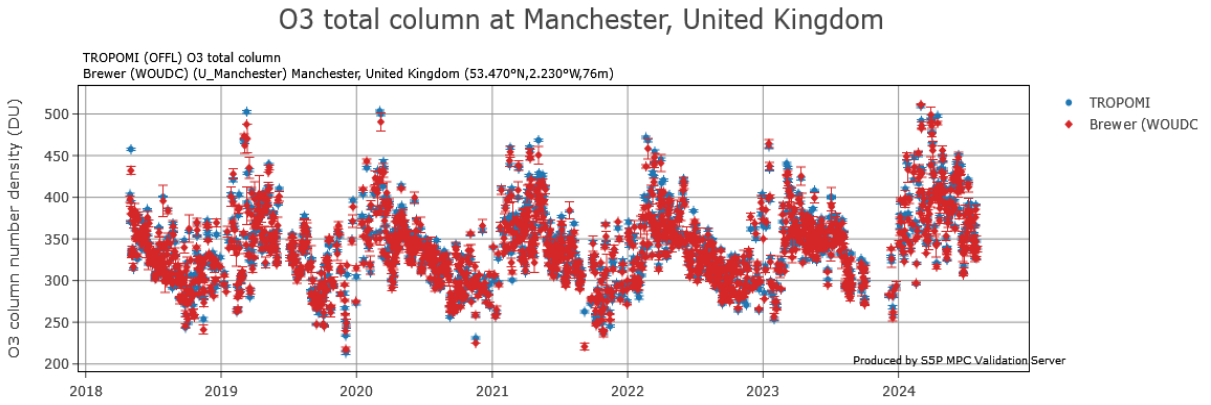


Figure 11: Time series of S5P TROPOMI OFFL and Brewer total ozone data at the station of Manchester in the United Kingdom (data courtesy J. Rimmer, University of Manchester).

4.3.8 Geographical patterns

None to report.

4.3.9 Other features

None to report.

4.4 Validation of L2_O3 NRTI

4.4.1 Recommendations for data usage followed

Data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

In order to avoid misinterpretation of the data quality, it is recommended to use only those TROPOMI pixels associated with a qa_value above 0.5. This filter was applied here.

4.4.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC VDAF and by S5P Validation Team (S5PVT) AO projects. This summary is based on coordinated operational validation activities carried out using the Automated Validation Server of the S5P MPC VDAF, the Multi-TASTE versatile multi-platform validation system operated at BIRA-IASB, and the ozone validation system operated at AUTH.

Current conclusions are valid for the S5P data obtained using processor versions 2.4.1 to 2.6.1, which cover the period from July 2022 until August 2024, and on the reference data available at the time of this report: typically, until the beginning of August 2024 for the PGN data, up to the end of June 2024 for the Dobson and Brewer data, and up to the end of August 2024 for the ZSL-DOAS SAOZ data. For the current report, Brewer and Dobson measurements were obtained through the World Ozone and UV Radiation Data Centre (WOUDC) in Toronto, the NDACC Data Host Facility, and WMO's Ozone Mapping Centre in Thessaloniki. If a station archives data into both WOUDC and NDACC HDF, the source with the most recent data is adopted. ZSL-DOAS measurements were collected through the SAOZ network Real-Time processing facility operated by CNRS LATMOS (LATMOS_RT) and the PGN data were obtained through EVDC. Over the period, with respect to the reference data available at the time of this analysis, of the order of 30 to 300 co-locations have been identified at about 20 Brewer and Dobson stations, at 24 PGN stations, and at 12 ZSL-DOAS SAOZ stations, sampling from the Arctic to the Antarctic (**Figure 12**).

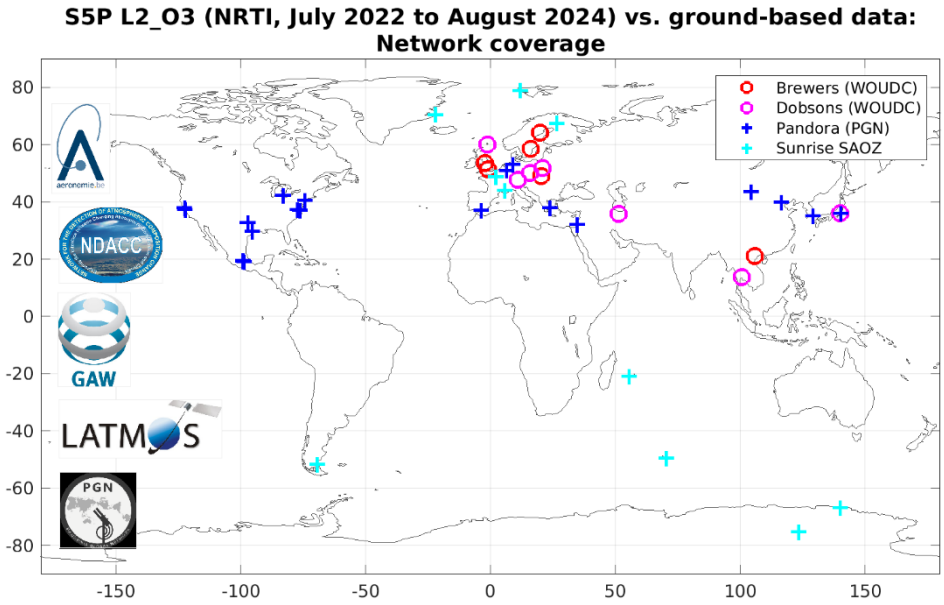


Figure 12: Geographical distribution of Brewer, Dobson, Pandora and ZSL-DOAS ground-based stations for which co-locations with S5P L2_O3 NRTI ozone data have been used (July 2022 until August 2024).

4.4.3 Bias

The systematic difference between S5p L2_O3 NRTI and reference ground-based data at individual stations rarely exceeds 2 %, as depicted in **Figure 13**. The median bias calculated over the entire ground-based networks is of the order of +0.5-1 %, S5P reporting higher values than the networks. Between 50˚S and 50˚N, the mean agreement with other satellite data is mostly within 1 % as well (**Figure 14**). This median bias value falls well within the mission requirements (max. bias 5 %). In addition, the agreement with the RPRO+OFFL product is within 1 %, as shown in **Figure 6**. On 6 August 2019, the nominal ground pixel resolution of the TROPOMI measurement was reduced to 5.5 x 3.5 km2, i.e. shorter by 1.5 km in the along-track direction, by reducing the integration time. This did not have an impact on the agreement between satellite and ground-based measurements.

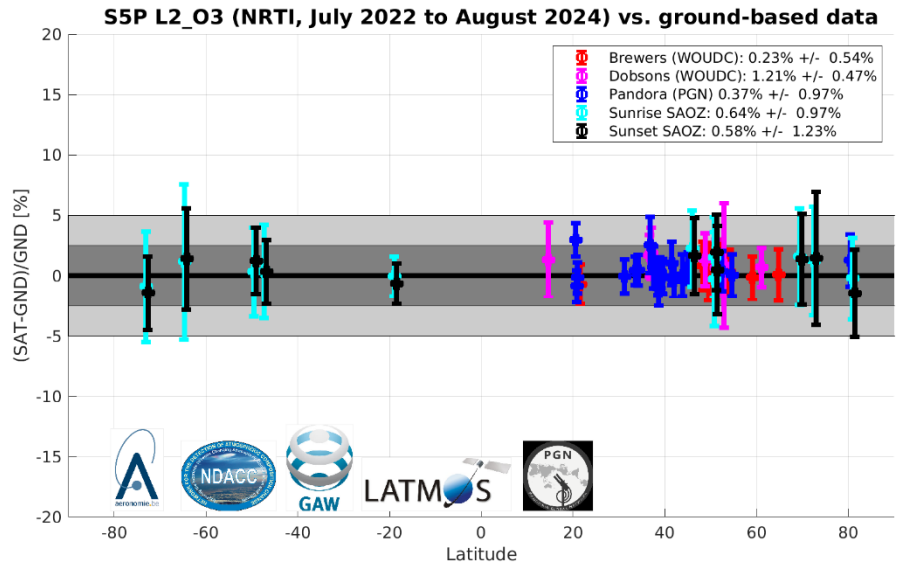


Figure 13: Meridian dependence of the median (the circular markers) and spread (±1 sigma, the error bars) of the percent relative difference between S5P TROPOMI L2_O3 (PDGS NRTI processors v2.4.1 and above, up to August 2024) and ground-based (GND) ozone column data, represented at individual stations from the Antarctic to the Arctic and per measurement type (Brewer, Dobson, Pandora, and ZSL-DOAS). The values in the legend correspond to the median and spread of all median (per station) differences. For clarity, sunrise and sunset ZSL-DOAS results are represented separately (offset by -0.5˚ and +0.5˚ in latitude).

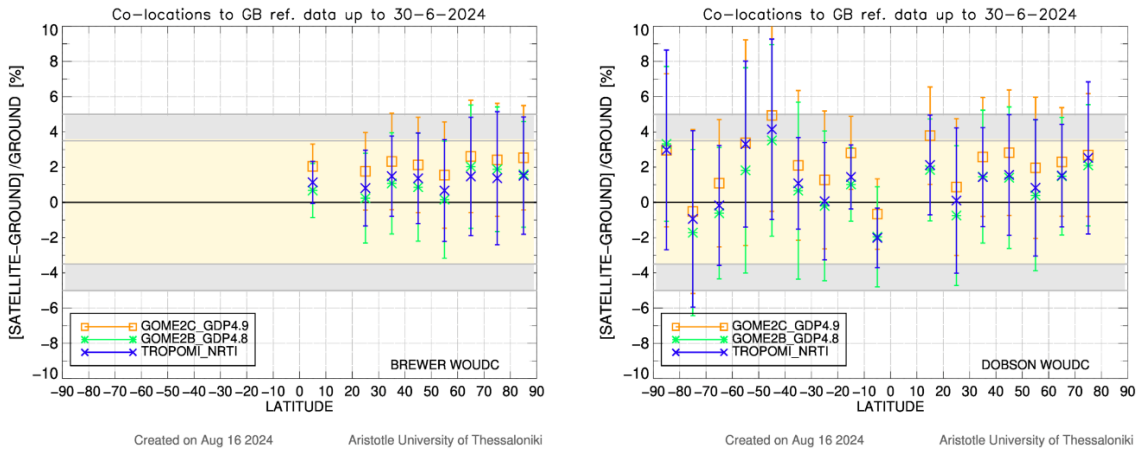


Figure 14: Comparison of the mean percentage differences between three satellite products (S5P TROPOMI L2_O3 NRTI, GOME-2B GDP 4.8 and GOME-2C GDP4.9) and ground-based total ozone data, versus latitude. The Brewer network comparisons are shown in the left-hand panel and the Dobson network comparisons are depicted in the right-hand panel, both datasets are downloaded from WOUDC. The period of data is July 2022 – June 2024.

An analysis of the medium-term systematic differences as a function of latitude and time is shown in **Figure 15**. The overall slight overestimation by S5P is visible, and a little more pronounced at mid latitudes (best visible in the Brewer comparisons, i.e., the left-hand panel). The Dobson comparisons show more pronounced seasonal features at high latitudes, in particular for low-sun conditions, which can be attributed to (1) the fixed effective temperature assumed in the Dobson data processing, and, (2) prior to the adoption of the GE_LER surface albedo product in July 2020, to poorer representation of difficult (e.g., snow and ice) surface albedo conditions. A similar analysis for the OFFL processor is shown in **Figure 7**.

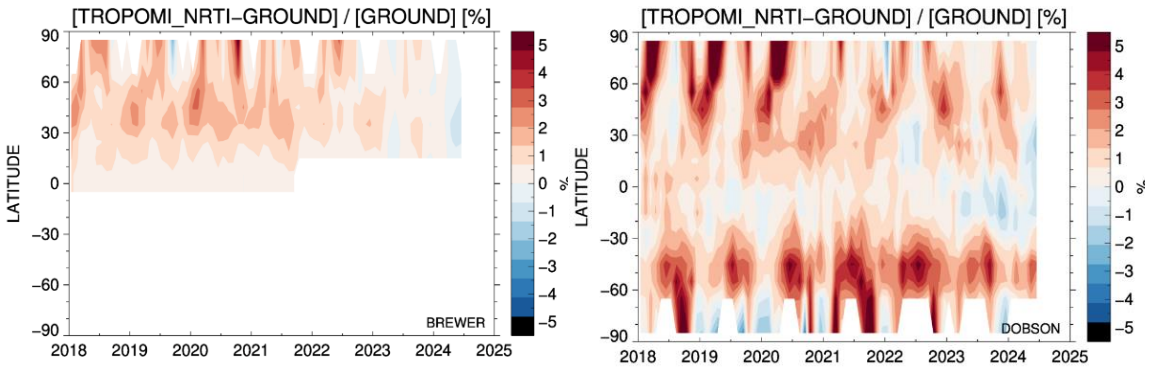


Figure 15: Time and latitude resolved analysis of the differences between NRTI total ozone and Brewer measurements (left-hand panel) and Dobson measurements (right-hand panel). Strong seasonal features in the Dobson comparisons are due (1) to the use of a fixed effective temperature in the Dobson measurement processing, and (2) due to the poorer representation of difficult surface albedo conditions before the adoption of the GE_LER surface albedo product in July 2020.

4.4.4 Dispersion

The ±1σ dispersion of the difference (between S5P and reference ground-based network data) around their median value rarely exceeds 3-4 % for the comparisons with direct-sun instruments (cf. the error bars depicted in **Figure 2**). Combining random errors in both satellite and reference measurements with irreducible co-location mismatch effects, it is concluded that the random uncertainty on the S5P measurements falls within the mission requirements of max. ±2.5 %. Whether the uncertainties reported with the data (the so-called ex-ante uncertainties) are compatible with the observed dispersion can be tested by computing, per station, the χ_r^2 , defined as:

$$\chi^2 = \Sigma \frac{(SAT - GND - mean(SAT - GND))^2}{\sigma_{SAT}{}^2 + \sigma_{GND}{}^2}$$

where σ_{SAT} and σ_{GND} refer to the ex-ante satellite and ground-based measurements respectively. The histogram of these χ_r^2 is shown in **Figure 16**. The histogram peaks around 0.5, which suggests the ex-ante uncertainties to be slightly too conservative as the peak should theoretically occur at one, given correct uncertainties on both data sets and no co-location mismatch errors.

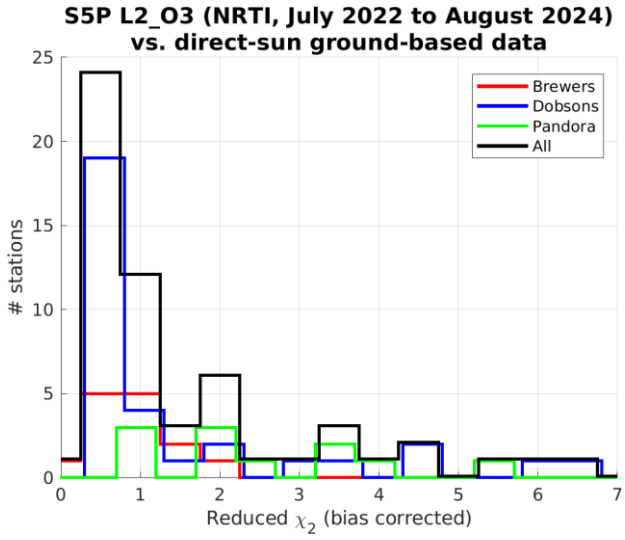


Figure 16: Histogram of the χ_r^2 values (one per reference instrument) for S5P NRTI total ozone versus different types of ground-based measurements.

4.4.5 Dependence on influence quantities

The evaluation of potential dependence of the S5P bias and dispersion on the Solar Zenith Angle (SZA, up to 80°), surface albedo, cloud fraction (CF) and scan sub-index of the TROPOMI measurement does not reveal any variation of the bias larger than 2-3 % over the range of these influence quantities (**Figure 17**). The scatter of the difference of about 2-4 % increases up to 7 % at large SZAs and at latitudes beyond 50°, as expected knowing that errors in both satellite and reference measurements as well as irreducible co-location mismatch effects increase at high latitudes and low sun elevation.

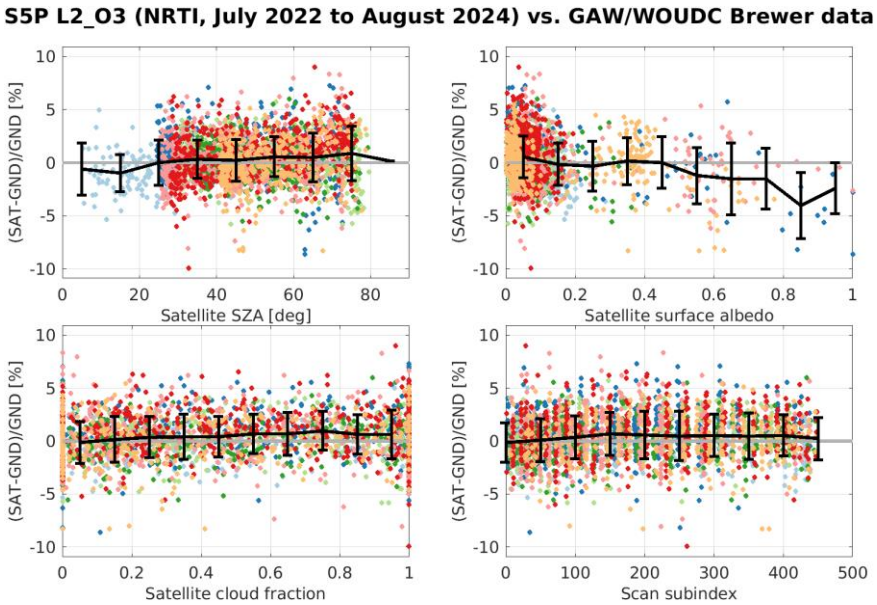


Figure 17: Dependence of the difference between S5P NRTI and ground-based Brewer total ozone data on the solar zenith angle (SZA), the surface albedo, the fractional cloud cover, and the scan sub-index of the satellite measurement. Black curve: mean and standard deviation over bins of 10 degrees of SZA, 0.1 of surface albedo, 0.05 of cloud fraction, and 50 pixels of the across-track scan. Coverage of the high SZA and of the surface albedo is poor for these Brewer comparisons, but the SAOZ comparisons (not shown) which offer better coverage for these parameter ranges reveal no anomalies.

4.4.6 Short term variability

Qualitatively, at all of the 50 ground-based monitoring stations, short scale temporal variations in the ozone column as captured by ground-based instruments are reproduced very similarly by S5P, as illustrated in **Figure 18**. The overall good agreement is confirmed by Pearson correlation coefficients always above 0.95.

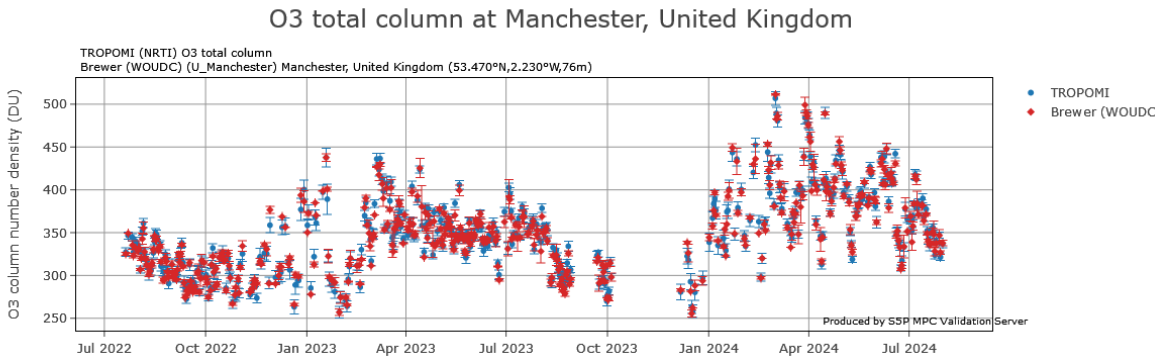


Figure 18: Time series of S5P TROPOMI NRTI (v2.4.1 and higher) and Brewer total ozone data at the station of Manchester in the United Kingdom (data courtesy J. Rimmer, University of Manchester).

4.4.7 Geographical patterns

None to report.

4.4.8 Other features

None to report.

4.5 Equivalence of NRTI and OFFL data, and comparison to other satellite datasets

While the previous sections demonstrate the quality of both the NRTI and OFFL data sets through their validation using ground-based reference data, a direct comparison (1) between both products and (2) between the S5P products and those from other satellite sensors, allows the identification of differences too subtle to be observed in the ground-based validation. To that end, this section covers direct comparisons between level-3 gridded (1° x 1° spatial resolution) TROPOMI NRTI and OFFL data, and GOME-2B (GDP4.8) total ozone column data. The GDP data processor is a DOAS-type processor, similar to the NRTI algorithm for S5P. The comparisons shown here are based on data sets filtered according to the official recommendations, i.e., the qa_value for the S5P data was required to be larger than 0.5.

The time range of the analysis is beginning of mission to December 2023. Both the OFFL and NRTI data sets used here are based on a full-mission reprocessing in version 2.4.1, extended with the more recent processor versions. Note that this NRTI full mission reprocessing is not publicly available as it was generated for internal quality control only.

Figure 19 reveals that the OFFL columns are on average about 1% larger than those retrieved with the NRTI processor at low and middle latitudes. At high latitudes, OFFL columns can be up to 3% smaller than those from the NRTI processor. Over the entire mission lifetime, there is a very minor positive trend in the differences at low latitudes. A consistent picture emerges from **Figure 20**, when considering that the GOME-2B DOAS retrievals are very similar to the S5P NRTI ozone product.

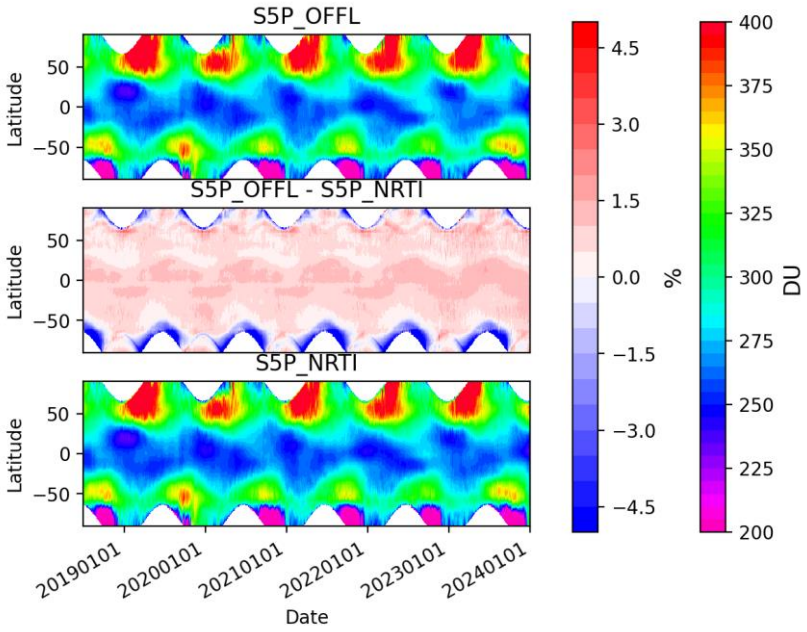


Figure 19: Time series of (the differences between) the zonal mean gridded (level-3) OFFL and NRTI total ozone column data sets.

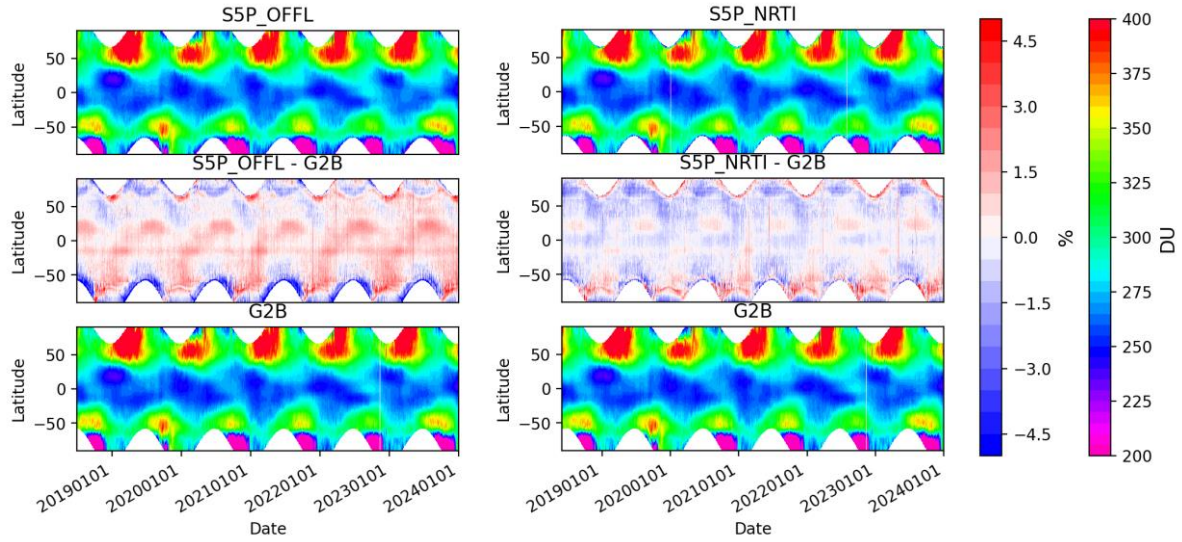


Figure 20: Zonal mean comparison between S5P TROPOMI and MetOp-B GOME 2 total ozone columns, for both OFFL (left-hand panel) and NRTI (right-hand panel).

5 Validation Results: L2_O3_TCL

5.1 L2_O3_TCL products and requirements

This Section reports on the validation of the S5P TROPOMI L2_O3_TCL RPRO and OFFL product identified in **Table 1**. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. The results presented here cover the offline processor versions 02.04.01, 02.05.00, and 02.06.01. Version 02.04.01 was also used for the full mission reprocessing (designated “RPRO” here). For validation results on processor versions prior to 02.04.01, the reader is referred to ROCVR version 18.1.0, available at https://mpc-vdaf.tropomi.eu/.

The S5P O3_TCL data files contain tropospheric ozone columns obtained by the Convective Cloud Differential algorithm (CCD). The CCD data are sampled daily and represent three-day averages of the ozone partial column between surface and 270 hPa (~10.5 km) under cloud-free conditions on a 0.5° latitude by 1° longitude grid between 20°S and 20°N. In contrast to most other S5P products in this document, it concerns a gridded data set, and, it covers about 2/3 of the actual vertical range of the tropical troposphere.

Variables related to a second tropospheric ozone algorithm, the Cloud Slicing Algorithm (CSA), are present in the data files but all corresponding entries are set to a fill value until public release of the CSA product. The CSA data are therefore not discussed in the following.

5.2 Validation approach

Routine validation of the S5P TROPOMI L2_O3_TCL tropospheric ozone data products entails both qualitative, visual inspections of daily maps of product variables and quantitative comparisons of these to independent reference measurements by ground-based and satellite instruments.

5.2.1 Ground-based networks

Reference measurements by ozonesondes launched at twelve stations of the ground-based SHADOZ network (ER_SHADOZ) are compared routinely to S5P data (see Hubert *et al.*, 2021). The SHADOZ data version used here is V06. The ozonesonde profile data are first quality controlled (Hubert *et al.*, 2016, 2021) and then integrated over the vertical range of the S5P CCD product (surface to 270 hPa) to obtain a comparable tropospheric column value. A reference measurement co-locates with a TROPOMI measurement provided that: (a) the SHADOZ station is located in the S5P CCD grid cell, and, (b) the ozonesonde was launched in the three-day satellite time window. Data that do not match these criteria are not used in the calculation of the quality indicators (e.g. **Figure 25**). If more than one reference tropospheric ozone column falls in a co-location window, then these are averaged prior to comparison. Such a double coincidence occurs very rarely in the considered data sample. Finally, it is important to note that the spatial and temporal sampling properties of satellite and reference data records are quite different, which adds mismatch uncertainties in the comparison results on top of the combined data uncertainties.

5.2.2 Satellites

S5P TROPOMI L2_O3_TCL tropospheric ozone column data are compared to Aura OMI, MetOp-B GOME-2 and MetOp-C GOME-2 tropospheric ozone column data using the GODFIT_v4 CCI algorithm developed within ESA’s Climate Change Initiative (CCI). It is based on the GODFIT total column data but the sampling was adapted to allow a more direct comparison to TROPOMI, i.e., 5 days averaging windows instead of monthly data and the tropospheric top pressure set to 270 hPa instead of 200 hPa. The horizontal resolution of the OMI, GOME-2B and GOME-2C data products was increased from 1.25° x 2.5° to 1° x 2°.

5.2.3 Field campaigns and modelling support

None for this report.

5.3 Validation of L2_O3_TCL OFFL (CCD)

5.3.1 Recommendations for data usage followed

Data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

In order to avoid misinterpretation of the data quality, we followed the recommendation to use only TROPOMI grid cells associated with a qa_value strictly above 0.7. This screening removes about 16 % of the S5P grid cells, usually between 15-20° latitude in the winter/spring hemisphere (see Sect. 5.3.8).

5.3.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC VDAF and by S5P Validation Team (S5PVT) AO projects. This summary is based on coordinated operational validation activities carried out using the Automated Validation Server of the S5P MPC VDAF and the Multi-TASTE versatile multi-platform validation system operated at BIRA-IASB.

Over the period 30 April 2018 to 20 August 2024, the ground-based validation analysis considers 2303 S5P OFFL CCD data products and 1586 ozonesonde flights at twelve stations across the tropics (**Figure 21**). S5P data averaged over the entire tropical region are also compared (**Figure 25**) to OMI (May 2018 – June 2023), GOME-2B (May 2018 – May 2023) and GOME-2C satellite data (July 2019 – June 2023).

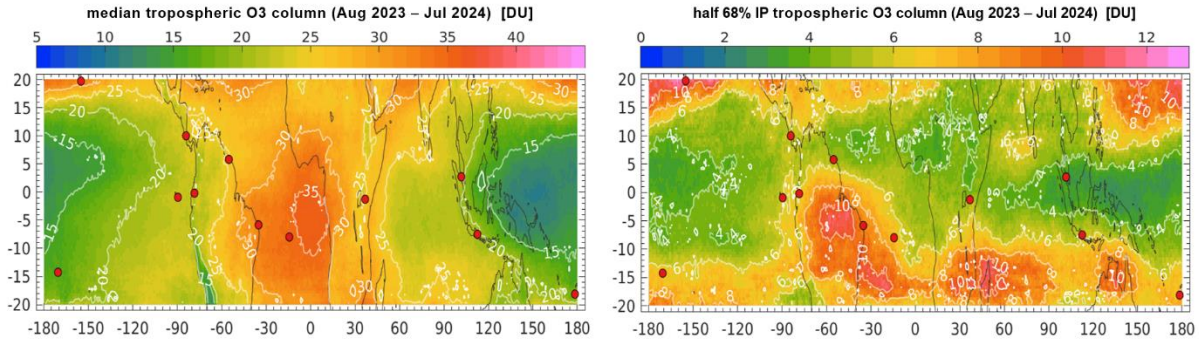


Figure 21: Median value (left) and half-width of 68% interpercentile (right) of S5P OFFL tropospheric ozone column data (CCD) over the last year of operations (August 2023 – July 2024). Red markers locate the twelve ground-based ozonesonde stations used in the validation analysis.

5.3.3 Bias

S5P tropospheric O_3 column values are on average larger than the ozonesonde values at all but one station (Watukosek, **Figure 23** and **Figure 25**). The mean bias over the network is +19 % or +3.4 DU (**Figure 25**, centre and bottom left). The inclusion of additional ozonesonde data especially at San Cristóbal and Quito, but also at Ascension Island and Heredia leads to a 3% or 0.5 DU reduction of the S5P bias estimates when compared to the previous issue (23.1.0, 17 July 2024) of this document. The mean bias is compliant with the mission requirement for a systematic uncertainty of maximum 25 %.

Difference time series between S5P and comparable satellite data (OMI, GOME-2B and GOME-2C) averaged over the 20°N – 20°S tropical belt are shown in **Figure 24**. The agreement with OMI is good, with a mean difference of +1.2 DU or +6 %. The larger mean difference of +3.5-3.8 DU or +19-21 % compared to the GOME sensors indicates a slight, general overestimation of TROPOMI which may (at least partly) be attributed to the different overpass times of MetOp-B/C (9:30 descending) and S5P (13:30 ascending) in combination with the diurnal cycle of tropospheric ozone.

5.3.4 Dispersion

At individual sites, the half 68 % interpercentile of the difference between S5P and ozonesonde data ranges between 17-35 % or 3.5-8.5 DU (**Figure 23** and **Figure 25**), and the network average is 26 % or 4.6 DU (**Figure 25**, centre and bottom right). Dispersion values at six SHADOZ stations are not compliant with the mission requirement for the random component of the uncertainty (<25%). However, four of these stations are located in an area with large natural percentage variability in the tropospheric O_3 field (due to high altitude location and/or proximity of Pacific low) and there is a considerable difference in spatio-temporal sampling between S5P and ozonesonde. In addition, the random component of the uncertainty of the ozonesonde measurement contributes about 5-10 % to the observed dispersion in the differences. Hence, the uncertainty of the S5P data is better than the 25 % observed dispersion in the comparisons to ozonesonde and therefore overall compliant with the mission requirement. Satellite-to-satellite comparisons exhibit a dispersion of about 3.1 DU or 16 % when averaged over the entire tropical belt. This is lower than the mission requirement and the average dispersion in comparison to the ground-based network (most likely due to the smaller difference in spatio-temporal sampling properties between satellite sensors).

5.3.5 Dependence on influence quantities

Nothing to report.

5.3.6 Seasonal cycle and shorter term variability

The difference between S5P and other satellite data records exhibit a clear seasonal structure (**Figure 24**). The phase varies with latitude but, generally, minima are noted around September-January and maxima around March-July. Peak-to-peak amplitude of the cycle lies around 1.5-2.5 DU, depending on latitude and reference instrument (Hubert *et al.*, 2021).

An annual cycle in the ground-based comparisons is found at about half of the ozonesonde sites (**Figure 26**). At Hilo (top left panel), a very clear maximum occurs in boreal winter-spring and a minimum in boreal summer-fall, with a peak-to-peak amplitude of about 18 DU (or 65 %). The source of this systematic seasonal effect at Hilo is not well understood. At Paramaribo (centre left), up to 10 DU (or 55 %) more positive biases are noted during July-November, coinciding with the biomass burning season. More elevated biases of 7-8 DU (or 25-40%) during a few months are seen at other sites around the Atlantic basin as well: Heredia (July-September), San Cristóbal (July-September), Natal (September-October), Ascension Island (October-November). Whether there is a causal relationship between these temporary increases in bias and more intense biomass burning around the Atlantic basin during these months is subject of further study. A second period with increases in S5P bias occurs between March and April, at Heredia, Natal and Ascension Island (and possibly San Cristóbal). Comparison statistics at Quito are limited, but current results on a seasonal cycle are consistent with those over San Cristóbal.

Co-located S5P and reference measurements correlate fairly well for stations with well-sampled comparison time series. Pearson's correlation coefficients range between 37 % (Watukosek, San Cristóbal) and 78 % (Natal) at individual stations, while the network average is 63 % (**Figure 25**, top left).

5.3.7 Geographical patterns

Annual median TROPOMI data (November 2022 – October 2023, **Figure 21**) capture the well-known South Atlantic ozone maximum associated with biomass burning, lightning and ozone precursors, as well as the well-known equatorial Pacific lows. Higher median levels in the 15°-20° tropical belts are a result of regular intrusions of ozone-rich air from higher latitudes. It shows the ability of S5P to observe the expected large-scale spatial patterns. At smaller scales, however, two sampling-related error patterns are noted.

The CCD algorithm requires an ample sampling of input total O3 column data to allow a robust estimate of a reference stratospheric O3 column. This requirement is not always fulfilled and, as a result, random errors of about 1 DU between neighbouring latitude bands are found in many S5P data products. The interplay between cloud coverage and S5P sampling imprints another random error pattern (up to 5 DU) that follows the progression of the S5P orbit. These errors are correlated in time and space and appear at small spatio-temporal scales.

Other known geophysical patterns and oscillations, such as the annual and semi-annual cycles, the biomass-burning season and the Madden-Julian Oscillation, are present in the S5P tropospheric O_3 data record as well. For an in-depth analysis, we refer to Hubert *et al.* (2021).

5.3.8 Other features

CCD data availability is much reduced poleward of ~15° latitude in the winter hemisphere (see time series at Hilo or Suva in **Figure 22**) since the algorithm requires a sufficient number of highly convective opaque clouds. Most of these are formed in or close to the Intertropical Convergence Zone (ITCZ) located mainly in the summer hemisphere. Suitable cloud conditions therefore occur less frequently in the winter-spring hemisphere.

On 6 August 2019, the nominal ground pixel resolution of the TROPOMI measurement was reduced to 5.5 x 3.5 km², i.e. shorter by 1.5 km in the along-track direction, by reducing the integration time. This did not have an impact on the agreement between satellite and ground-based measurements. Estimates of bias and dispersion before and after this change are consistent.

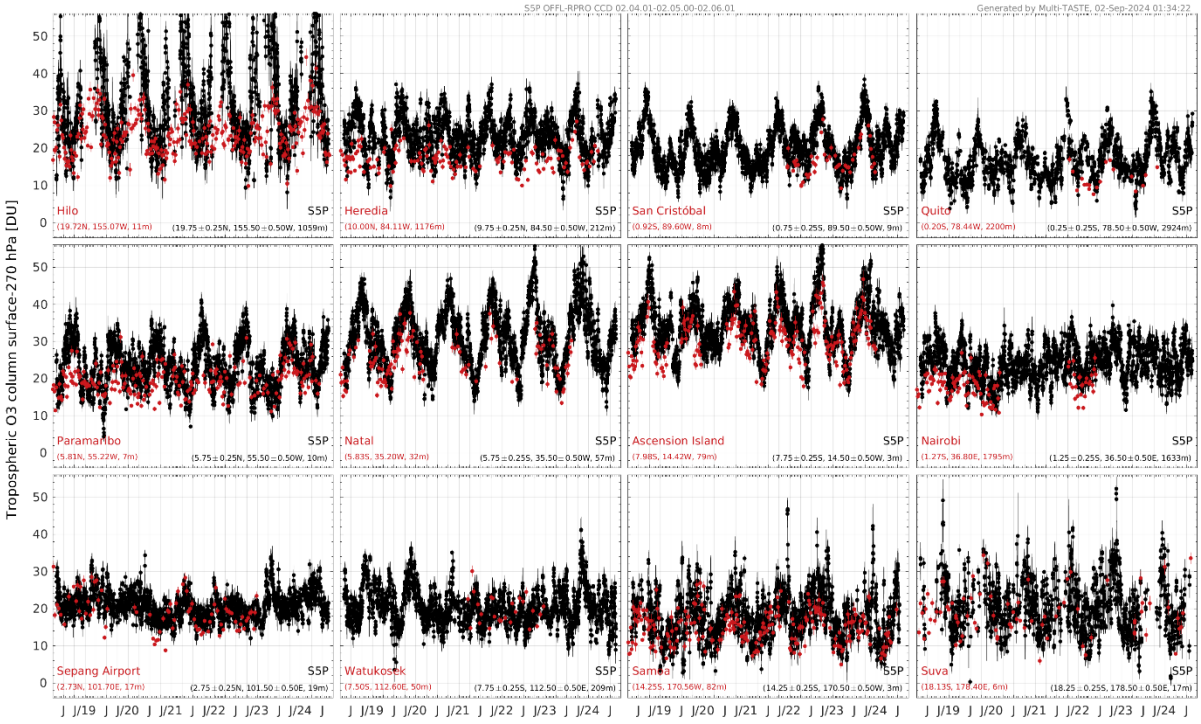


Figure 22: Time series of spatially co-located tropospheric O_3 column data by ozonesonde (red) and by S5P RPRO+OFFL v02.04.01-02.05.00-02.06.01 (black). All data were screened following recommendations by the data providers.

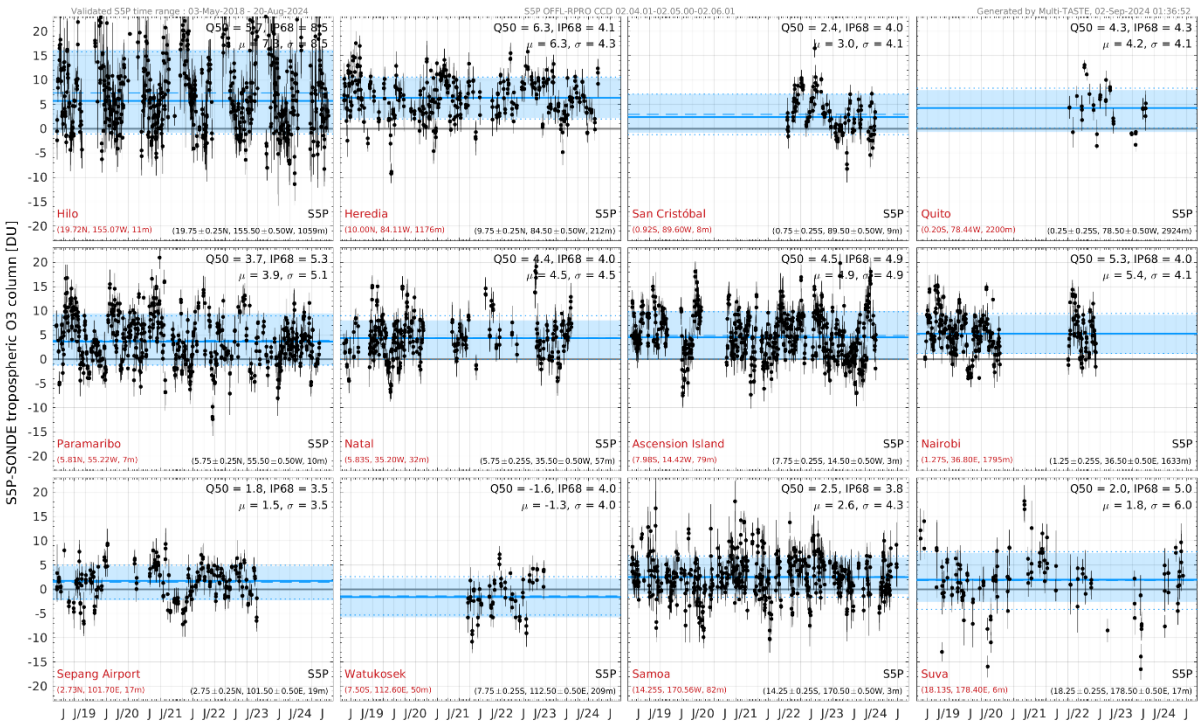


Figure 23: Time series of the absolute difference between spatially and temporally co-located S5P and ozonesonde tropospheric O_3 column data. The blue line and shaded area shows the median value and the range between the 16% and 84% percentiles. Positive values indicate a high bias of S5P w.r.t. the reference.

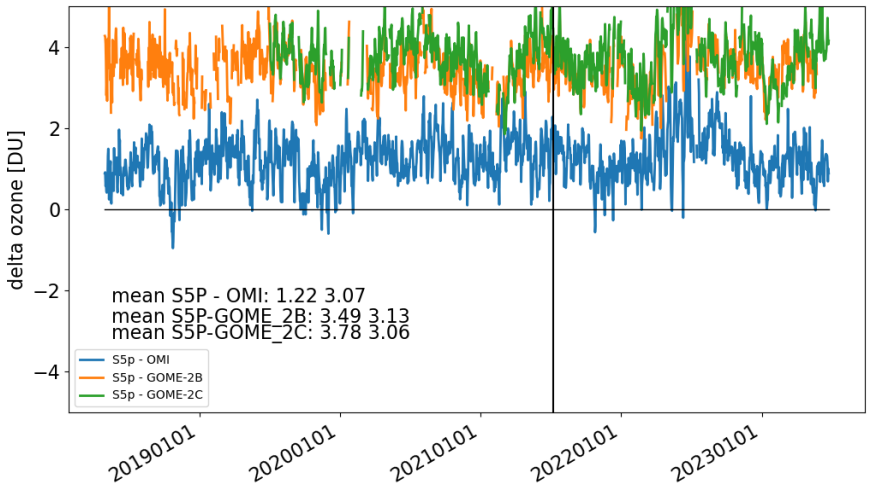


Figure 24: Difference time series of daily tropospheric O_3 column data averaged over the 20°S – 20°N tropical belt. S5P RPRO+OFFL CCD data are compared to OMI (blue), GOME-2B (orange), and GOME-2C (green) satellite data; positive values indicate a high bias of S5P w.r.t. the reference. The black vertical line indicates the transition from the RPRO to OFFL data files (see **Table 1**).

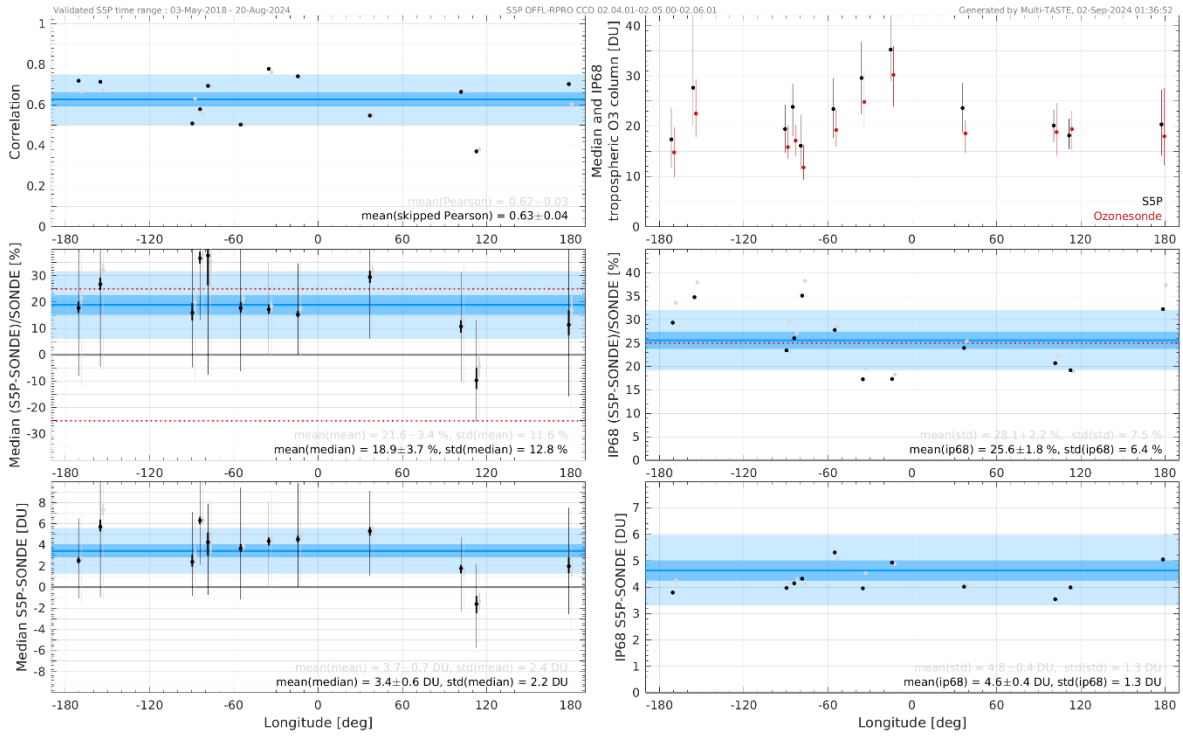


Figure 25: Overview of correlation (top left), median bias (middle & bottom left) and dispersion of the difference (middle & bottom right) of S5P tropospheric O_3 column data for each SHADOZ station (black markers). Black vertical bars represent the 68% interpercentile of the difference. The mean, standard error of the mean (1σ) and standard deviation (1σ) of the quality indicator across the network are shown as a horizontal blue line and shaded areas.

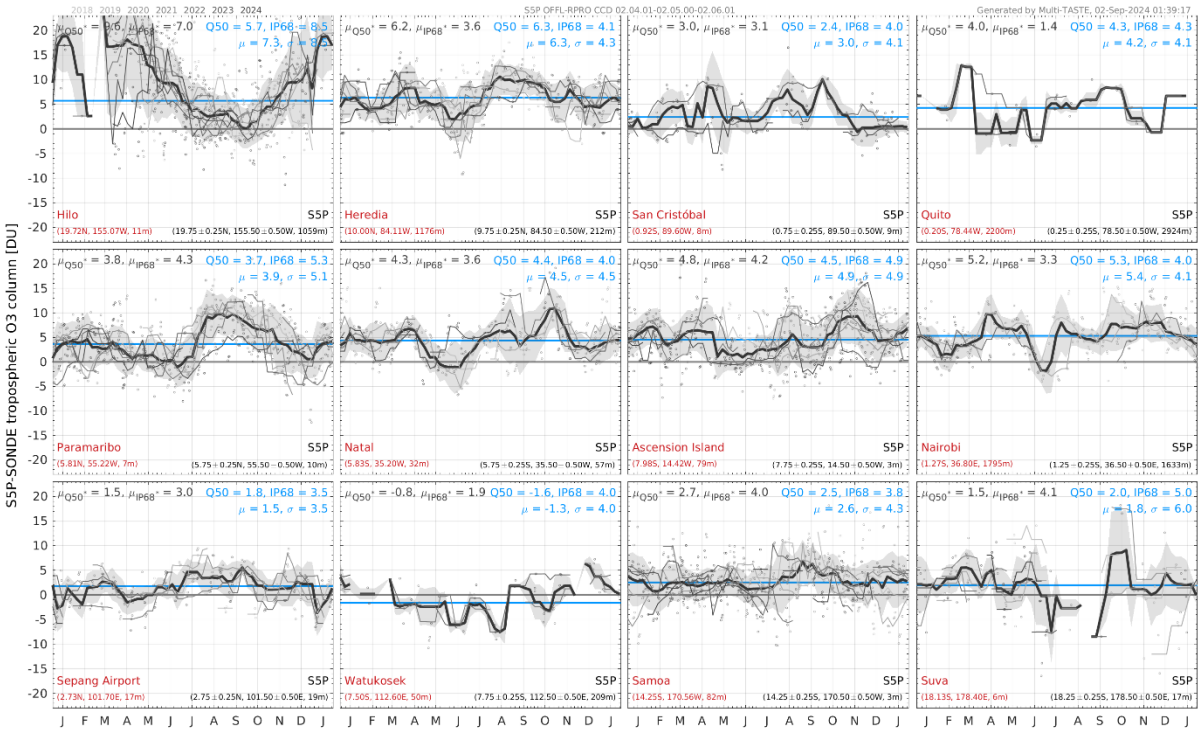


Figure 26: Annual cycle of the absolute difference between spatially and temporally co-located S5P and ozonesonde tropospheric O_3 column data. Individual comparison pairs are grey-shaded by year. Lines indicate the 29-day moving mean of the absolute difference for each year individually (thin grey) and all years combined (thin black). The blue line shows the median value over all comparison pairs. Positive values indicate a high bias of S5P w.r.t. the reference.

6 Validation Results: L2_O3_PR

6.1 L2_O3_PR products and requirements

This Section reports on the validation of the S5P TROPOMI L2_O3_PR data product identified in **Table 1**. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. These results are obtained by the MPC VDAF and by the S5P Validation Team (S5PVT) AO project CHEOPS-5p. The operational validation activities are carried out using the Automated Validation Server of the MPC VDAF and the Multi-TASTE versatile multi-platform validation system operated at BIRA-IASB. This section provides validation results for the reprocessed (RPRO) and offline (OFFL) PDGS processing streams (May 2018 to August 2024), which are available on ESA's data hub. Important dates and corresponding changes within this period are listed in **Table 5**. The near real-time (NRTI) stream differs hardly from the OFFL stream but follows a checkerboard pattern in its pixel selection (from v02.05), allowing rapid data processing and delivery. The operational S5P L2_O3_PR orbit data files contain, for each individual observation, the ozone number density on 33 pressure levels as retrieved by KNMI's operational algorithm, the integrated tropospheric and total ozone columns, and six integrated sub-columns (0-6, 6-12, 12-18, 18-24, 24-32, and 32-82 km). For the validation activities presented here, the station overpass files obtained from the PDGS processor in HARP format (v1.15) are considered. Detailed validation results can be found in Keppens, et al. (2024). This section has been updated with respect to the manuscript.

Table 5: Important changes in the operational TROPOMI O3_PR version 02.04.00 to 02.06.00 processing.

Date	Operational processing change
2018/04/30	RPRO start date using soft-calibration v2
2019/08/06	TROPOMI pixel resolution change
2022/07/25	RPRO end date
2022/07/25	OFFL start date using soft-calibration v1
2023/01/15	OFFL soft-calibration update to v2
2023/03/15	NRTI checkerboard pattern from v02.05
2023/11/29	NRTI and OFFL update to v02.06 (minor)

6.2 Validation approach

Validation of the S5P TROPOMI L2_O3_PR ozone profile data entails quantitative comparisons to independent reference measurements collected from ground-based monitoring networks, cross-validation with other satellite instruments, assessment of the retrieved information content based on the analysis of the associated averaging kernels, and visual inspections of daily maps of S5P ozone data and associated parameters.

6.2.1 Ground-based networks

S5P TROPOMI L2_O3_PR ozone profile data are compared to ground-based measurements acquired by instruments contributing to WMO's Global Atmosphere Watch (GAW), the Network for the Detection of Atmospheric Composition Change (NDACC), Southern Additional Ozonesonde programme (SHADOZ), and Tropospheric Ozone Lidar Network (TOLNET): (1) balloon-borne ozonesondes, (2) stratospheric differential absorption ozone lidars (DIAL), and (3) tropospheric DIAL. The ground-based data are collected through ESA's Validation Data Centre (EVDC) and the respective data host facilities of the ground-based networks.

6.2.1.1 Balloon-borne ozonesonde

Launched on board of small meteorological balloons, electrochemical ozonesondes measure the vertical distribution of atmospheric ozone partial pressure from the ground up to burst point, typically around 30 km. Their estimated bias is smaller than 5 %, and the precision remains within the order of 3 % (Smit et al., 2007). Caveats for using ozonesonde datasets include errors depending on instrument set up (buffer solution, pump efficiency correction...) and changes in ozonesonde type and measurement parameters with time. In the framework of the MPC, the VDAF-AVS performs automated data comparisons with ozonesonde datasets collected through the EVDC. These data originate from the NDACC Data Host facility, the SHADOZ archive, and World Ozone and UV Data Centre (WOUDC).

6.2.1.2 Differential absorption ozone lidars (DIAL)

Ozone differential absorption lidars (DIAL) can measure the vertical profile of ozone number density in the troposphere (500 m a.g.l. to 12-15 km) or in the stratosphere (8-10 km to 45-50 km). Ground-based systems perform network operation in the framework of the international Network for the Detection of Atmospheric Composition Change (NDACC) and the North American-based Tropospheric Ozone Lidar Network (TOLNet). For stratospheric measurements, effective vertical resolution typically degrades with altitude from a few hundred meters at 10 km to 3–5 km in the upper stratosphere, and the total uncertainty (systematic and random effects included) ranges from 4 % below 30 km to 10 % or more at 35 km and above (Leblanc et al., 2016). For tropospheric measurements, effective vertical resolution also degrades with altitude, from a few meters in the boundary layer to 2-3 km in the upper troposphere, and the total uncertainty ranges from 4 % (bottom) to 10-20 % (top). Between about 3 and 10 km, tropospheric ozone lidar measurements show a mean difference with ozonesonde of less than 2 % and a root-mean-square deviation below 3 %, which are well within the combined uncertainties of the two measurement techniques (Leblanc et al., 2018). The MPC VDAF and its AVS make use of DIAL ozone profile data available through EVDC, originating from the NDACC Data Host facility, and through the TOLNET data archive.

6.2.2 Satellite intercomparisons

Comparison to other satellite data extend ground-based validation to the global domain and increase the number of data comparisons. For stratospheric ozone, comparisons to limb and solar occultation sounders (MLS, OMPS-limb, ACE-FTS) are appropriate. For tropospheric ozone, comparisons can be made to OMI and OMPS-nadir, where the OMPS-nadir measurements have the best spatial and temporal co-registration with TROPOMI.

6.2.3 Analysis of information content

The information content of the S5P ozone profile data is assessed through algebraic analysis of the associated averaging kernel (AK) matrix generated by the same S5P processing algorithm. The row sums of the AK matrix indicate the vertical sensitivity of the S5P ozone profile retrieval (Rodgers, 2000). The trace of the AK kernel matrix gives the Degree of Freedom of the Signal (DFS), to be understood here as the amount of vertical sub-columns with independent ozone information from each other. The Full Width at Half Maximum (FWHM) of the AK corresponding to a given altitude gives an indication of the effective vertical resolution of the retrieved profile at this altitude. This effective resolution of the retrieved information is not the numerical resolution of the vertical grid used for the retrieval process, which is usually much higher than the true, physical resolution of the retrieved information. The true altitude registration of the retrieved profile information at a given altitude of calculation can be estimated as the barycentre or peak position of the associated AK at this calculation altitude.

6.2.4 Analysis of daily global maps

The MPC VDAF-AVS creates daily global maps of the six partial columns provided in the ozone profile product, together with the integrated total column. The latter is compared with the daily global map of the TROPOMI total column retrieval to assess their mutual consistency. Daily global maps easily allow identifying data gaps, retrieval artefacts, along-orbit striping, and other large-scale features that are not typically detected through comparison with respect to point-like ground-based data.

6.2.5 Parameter correlation checks

Using the in-house PyCAMA software, correlation checks are performed by KNMI on a broad selection of satellite data parameters within the orbit files. These checks provide a view on single-orbit features, correlations between retrievals of subsequent pixels, the appropriateness of the data flagging, etc. Relevant results can be found on the *TROPOMI Portal for Level-2 Quality Control* [ER_L2QC].

6.2.6 Field campaigns and modelling support

No specific campaigns and model-based studies have been foreseen to support the validation of S5P TROPOMI ozone profile data. However, ozone profile measurements acquired during campaigns not specific to TROPOMI validation are considered (e.g. the ozonesonde MATCH campaigns).

6.3 Validation of L2_O3_PR v2.4-6 (RPRO/OFFL)

6.3.1 Recommendations for data usage followed

Data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) of this data product, available online through the following link: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-5p/products-algorithms. In order to avoid misinterpretation of the data quality, we follow the PUM recommendation to use only TROPOMI ozone profile retrievals associated with a `qa_value` above 0.5.

6.3.2 Status of validation

Comparison results between ground-based reference measurements and coincident TROPOMI L2_O3_PR pixels (closest pixel on the same day with `qa_value` > 0.5) are obtained through the versatile Multi-TASTE validation system at BIRA-IASB, as part of both MPC and S5PVT CHEOPS-5p validation activities. Prior to their comparison to S5P ozone profile data, ground-based measurements – acquired at higher vertical resolution than S5P profile data – are convolved with the averaging kernels associated with the S5P retrievals to account for vertical smoothing differences (see e.g. Rodgers and Connor, 2003, Calisesi et al. 2005, Keppens et al., 2019). The S5P ozone subcolumns, on the other hand, are compared to vertically integrated ozonesonde measurements, again with preceding averaging kernel smoothing (**Figure 28**). The O3_PR retrieval's degrees of freedom per subcolumn are assessed as well, together with their dependence on influence quantities and their correlation with the retrieved subcolumn (**Figure 29**). In **Figure 30** and **Figure 31**, the difference of L2_O3_PR ozone number density profiles with respect to reference measurements is reported as a function of a selection of influence quantities (colour scales). Also included are the level-specific chi-square tests (von Clarmann, 2006; Keppens et al., 2019) and a selection of information content diagnostics: vertical sensitivity, altitude registration offset, and averaging kernel full width at half maximum (FWHM). The geographical distribution of the FRM stations depicted in **Figure 26** determines the domain of applicability of the validation results.

For the routine validation of the S5P/TROPOMI ozone profiles, the automated validation server (AVS, http://mpc-vdaf-server.tropomi.eu/o3-profile) deployed within the MPC VDAF facility collects S5P ozone profile data and correlative measurements to identify suitable co-locations, compare the co-located, data and produce S5P data quality indicators. The VDAF-AVS produces curtain plots (ozone number density as a function of altitude and time) of the satellite data at a selection of ground-based ozonesonde stations, together with curtain plots showing the difference between S5P and ground-based data. The VDAF-AVS also provides statistical estimates of the bias and dispersion of S5P data with respect to the ground-based measurements.

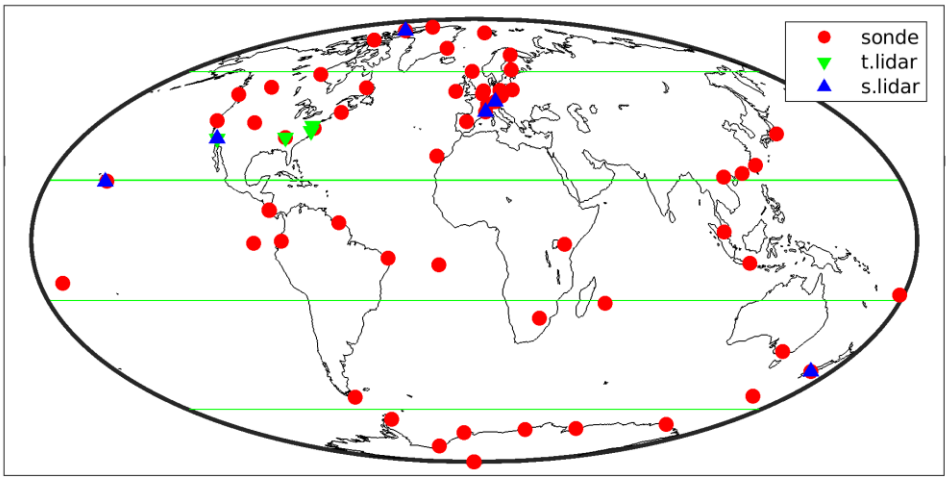


Figure 27: Geographical distribution of the ozonesonde and tropospheric (t.) and stratospheric (s.) lidar stations with which co-locations with S5P L2_O3_PR data have been identified and used in the data comparisons reported hereafter.

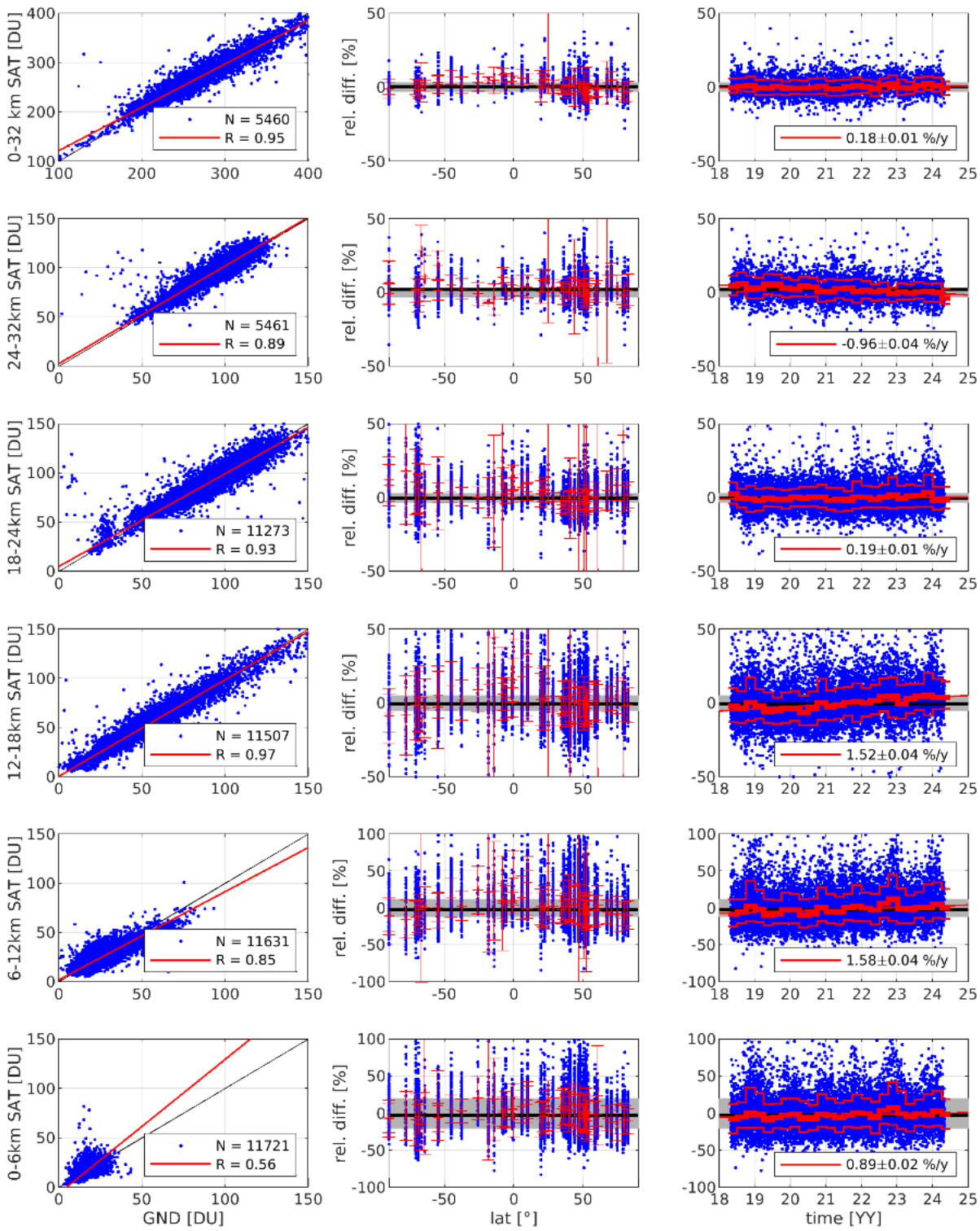


Figure 28: Correlation (R) between N lowest five ozone subcolumns as observed by TROPOMI and the coincident vertically integrated ozonesonde measurements (first column), and their overall sum (top row). Subsequent columns show the differences between satellite and ozonesonde subcolumns as a function of latitude, and time. 84, 50 (thicker), and 16 % quantiles are added in red (per station, per season, or for ten bins), together with the overall mean difference (black line), and the product requirements for each subcolumn (grey areas).

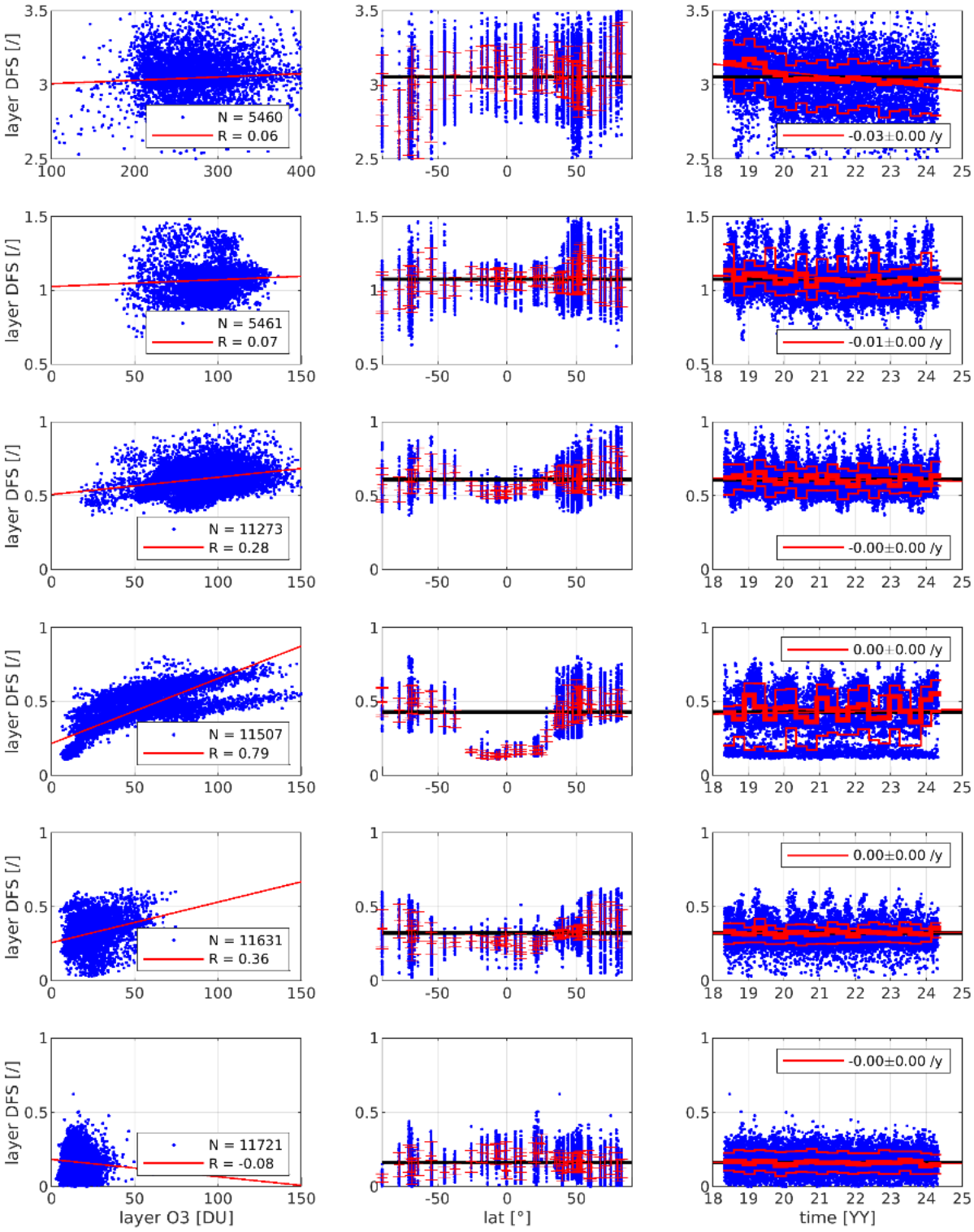


Figure 29: Correlation (R) between the six subcolumn DFS values and the six ozone subcolumns as observed by TROPOMI (first column) for the same N observations as in the previous figure. Subsequent columns show the layer-DFS as a function of latitude, and time. 84, 50 (thicker), and 16 % quantiles are added in red (per station, per season, or for ten bins) together with the overall layer mean (black line).

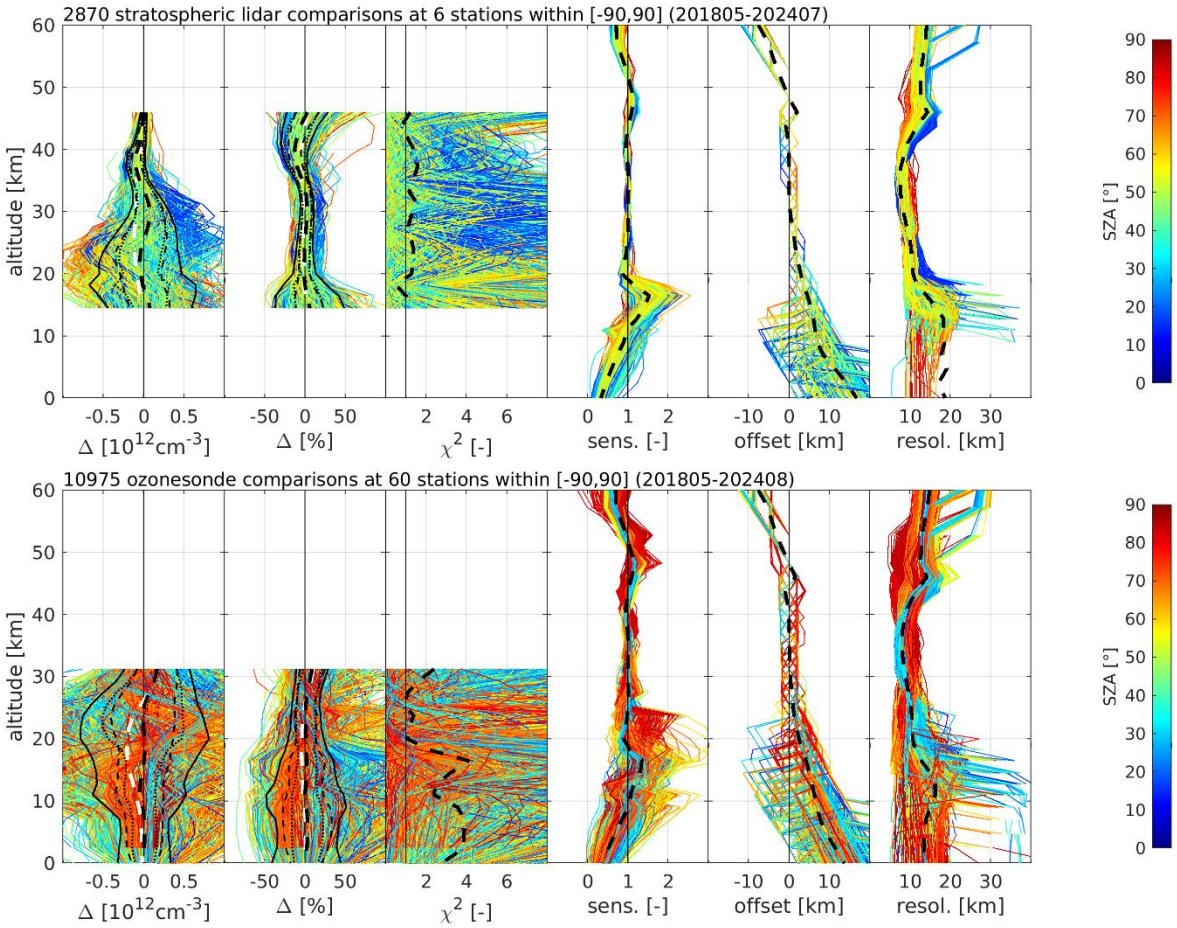


Figure 30: Comparison between S5P L2_O3_PR RPRO/OFFL ozone number density profile data and all co-located ground-based reference measurements, originating from stratospheric lidars (top) and ozonesondes (bottom). Every plate shows six graphs, respectively, from left to right: the difference and the percent relative difference between S5P and reference data, the chi-square profile, the vertical sensitivity, the altitude registration offset, and the averaging kernel FWHM associated with the S5P retrieval. The colour scale indicates the TROPOMI solar zenith angle. Black dashed lines show mean values (thick lines) and standard deviations (thin lines, around the mean), while white, dashed lines indicate the mean difference between the a-priori profile and the reference measurement. Dotted black lines indicate the total ex-ante (inductive) uncertainty of TROPOMI and the reference measurements combined (around the mean difference).

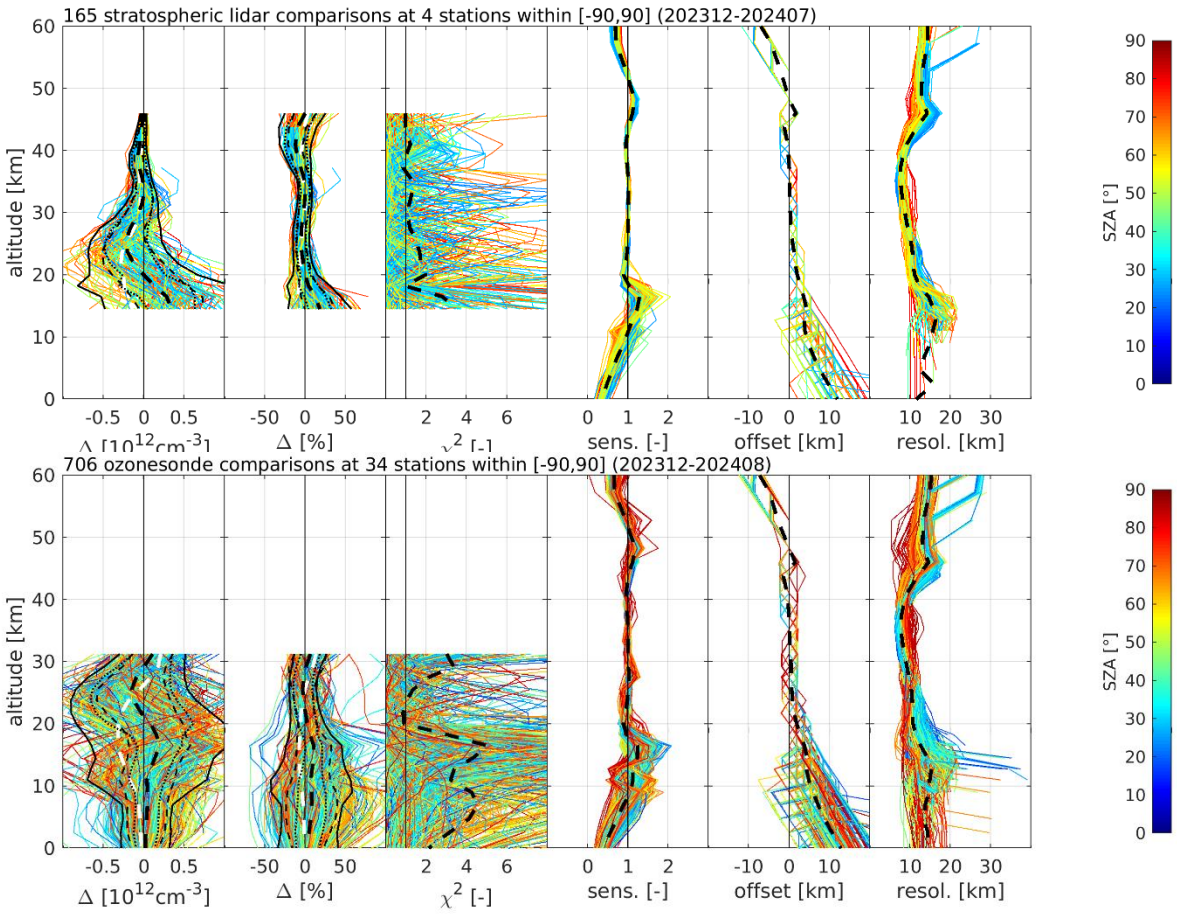


Figure 31: Comparison between S5P L2_O3_PR RPRO/OFFL ozone number density profile data and all co-located ground-based reference measurements, originating from stratospheric lidars (top) and ozonesondes (bottom), and for L2 v02.06 only. Every plate shows six graphs, respectively, from left to right: the difference and the percent relative difference between S5P and reference data, the chi-square profile, the vertical sensitivity, the altitude registration offset, and the averaging kernel FWHM associated with the S5P retrieval. The colour scale indicates the TROPOMI solar zenith angle. Black dashed lines show mean values (thick lines) and standard deviations (thin lines, around the mean), while white, dashed lines indicate the mean difference between the a-priori profile and the reference measurement. Dotted black lines indicate the total ex-ante (inductive) uncertainty of TROPOMI and the reference measurements combined (around the mean difference).

6.3.3 Vertical sensitivity, resolution and registration

The information content of the S5P ozone profile data is assessed through the analysis of the associated averaging kernels generated by the same L2_O3_PR data processor, as described in Section 6.2.3. Additionally, the ozone retrieval degrees of freedom (DFS) are assessed for each of the six sub-columns separately (**Figure 29**).

Overall, the retrieved information on ozone is distributed over five to six vertical subcolumns of independent information: about 5.5 on average, with values closer to six in the mid-latitudes and values just above five in the tropics and towards the poles. **Figure 30** shows that the vertical sensitivity of the S5P ozone profile data is nearly equal to unity at altitudes from about 20 km up to 50 km, meaning that the information in the retrieved product fully originates from the TROPOMI observation. It decreases rapidly at altitudes below 10 km and above 50 km. Around the tropopause between 10 and 20 km an oversensitivity (larger than one) can be observed, which is a rather typical compensating effect for the under-sensitivity below in nadir profile retrievals. This oversensitivity seems to be more pronounced for higher solar-zenith angles, and correlates with higher retrieved ozone concentrations, as can be observed from the plots in **Figure 29**. As a result, the retrieved ozone below about 24 km will also show a seasonal and meridian dependence (**Figure 28**, second and third columns). The vertical sensitivity usually drops below 0.5 towards the surface, meaning that most of the information in the lowest subcolumn originates from the prior profile, rather than from the spectral satellite measurement.

The altitude registration of the retrieved profile information usually is close to the nominal retrieval altitude in the 20-50 km altitude range, i.e., the offset is nearly zero. Yet the retrieval shows increasingly positive and negative offsets below and above the 20-50 km altitude range, respectively, which can reach 20 km towards the surface. This means most of the retrieved surface ozone information (ignoring the prior contribution) comes from the UTLS (upper troposphere to lower stratosphere), in agreement with the occurrence of a sensitivity peak at that altitude. The direction of the offset is always towards higher retrieval sensitivities, i.e., positive for the troposphere and negative for the highest altitudes. The effective vertical resolution of the TROPOMI ozone profile retrieval usually ranges within 10-15 km, with an optimum close to 7 km in the middle stratosphere (around 30-40 km).

6.3.4 Bias

Figure 30 contains all comparisons between TROPOMI ozone profiles and reference data, the latter AK-smoothed, and corresponding statistics. **Figure 30** focuses on comparisons for the latest data processor update (v2.6) only, showing slightly increased tropospheric biases (also see temporal variability below). Included are level-specific chi-square tests and information content diagnostics that facilitate the interpretation of the comparison results. The S5P ozone subcolumn comparisons in **Figure 28**, on the other hand, are only with respect to vertically integrated ozonesonde measurements, again with preceding averaging kernel smoothing. Compared to ozonesonde data, the S5P RPRO/OFFL data has a mean bias below 5-10 % in the troposphere (in black). This is slightly lower than the mean prior profile bias (in white) although the reduced sensitivity towards the surface indicates that a significant part of the retrieved information originates from the a-priori assumptions. In the stratosphere up to 35 km, stratospheric lidar data comparisons conclude to a mean bias of 5-10 % as well. The bias goes up to -15 % above (35-45 km), but with vertical oscillations (positive and negative), and hence exceeds the mean prior bias around 40 km. These oscillations of the bias may be due to a typically larger a-priori error in the mid and high stratosphere (above 20 %) in comparison with other retrievals.

6.3.5 Dispersion

S5P data comparisons with ozonesonde and stratospheric lidar data show a dispersion of order of 30 % in the troposphere, and 10 to 20 % in the UTLS and upper stratosphere.

6.3.6 Chi-square tests

Chi-square tests ($\chi^2 = (\Delta x)^T S_\Delta^{-1} \Delta x$) allow verifying whether the observed differences Δx between the satellite and reference profiles are consistent with the ex-ante (predicted) uncertainties on the difference S_Δ (von Clarmann, 2006; Keppens et al., 2019). The latter contains the satellite and reference covariances, and uncertainties that are due to sampling, smoothing, and retrieval differences. By application of vertical averaging kernel smoothing, however, retrieval differences and vertical sampling and smoothing differences are removed from the difference covariance S_Δ (Keppens et al., 2019). This means that for the results presented in **Figure 30** the difference covariance mainly contains the satellite and reference covariances. Horizontal and temporal sampling and smoothing differences are minimized to a nearly negligible level by application of strict collocation criteria (reference station within satellite pixel and a few hours measurement time difference only).

The chi-square plots in **Figure 30** (third graph in each plot) demonstrate that on average the observed differences confirm (χ^2 close to one) the ex-ante satellite and ground uncertainty estimates in the stratosphere, above about 20 km, despite the appearance of large outliers. However, around the tropopause and below (around 15-20 km and lower), the mean chi-square value increases up to about four, with especially high values for the tropics (low SZA) and Antarctic (high SZA and surface albedo). Here, the predicted (random) satellite uncertainty is smaller than what is observed (assuming correct reference uncertainties) by a factor of about two. This can also be seen in the difference plots, as the thin dashed lines representing dispersions of the difference are further away from the mean difference than the dotted lines representing the combined ex-ante uncertainties. Adding the smoothing difference error to the latter results in the full thin black lines, pointing at values that are typically 10-30 % higher. The largest smoothing errors (up to 50 %) occur in the UTLS and towards the surface (where kernel edge effects are also at play).

6.3.7 Dependence on influence quantities

Higher effective vertical resolutions (reduced kernel FWHM) can be observed for higher solar zenith angles (sideward atmospheric irradiation) in the troposphere and higher stratosphere (**Figure 30**), as can be expected for nadir (ozone) profilers and resulting in reduced vertical bias oscillations in the stratosphere (most clear from the lidar comparisons). On the other hand, the kernel width becomes ill-defined for the very broad averaging kernels that originate from tropospheric level retrievals at lower solar zenith angles. With only one retrieval degree of freedom up to about 18 km (lowest three subcolumns), the retrieved information is indeed very much vertically smeared, and a large part of the tropospheric ozone information comes from the prior profile. The low-SZA retrieval therefore goes hand in hand with increased vertical smoothing errors. An optical path length dependence is also observed for the sub-columns in **Figure 28**, which translates into a seasonal and meridian dependence of their bias.

The meridian dependence of the full profile bias with respect to ozonesondes is shown in **Figure 32** for five latitude bands, where increased tropospheric biases are observed for high solar zenith angles in the mid to high latitudes. As a result, increased biases are found for high-SZA observations above highly reflective scenes, like is the case for Antarctic (sea) ice. On the other hand, when the deviation from the prior profile becomes too strong, these observations are screened by the QA-value.

6.3.8 Temporal variability

For the TROPOMI RPRO/OFFL ozone profile data that are available now, comparison with the ozonesonde data reveal a positive drift for the lowest three subcolumns (0-18 km) and a negative drift of similar size for the 24-32 km subcolumn, while the drift is negligible for the subcolumn in between (18-24 km), as can be seen from **Figure 28**. The 6-12 km subcolumn shows the highest temporal difference change, with a positive drift close to 1.6 %/year, which is also caused by higher positive biases in 2022 and early 2023. Overall, no drift is found for the profile integrated from 0 to 32 km (upper row). More detailed, meridian drift assessments are shown in the right column of **Figure 32**. These plots show robust linear regression results for the temporal dependence of the TROPOMI bias with respect to ozonesonde measurements (again on the retrieval grid). The horizontal bars indicate two-sigma uncertainties on the drift from a bootstrapping technique with thousand samples. The significant positive and negative drifts that were observed for the subcolumns on the global scale are confirmed here for the tropics and mid-latitudes, with values up to 2-3 %/year below 20-25 km and minus 1-2 %/year above. No significant tropospheric drifts are detected towards the poles.

6.3.9 Geographical patterns

Global maps for March 16, 2024, are shown in **Figure 33**. Slight along-orbit striping can be observed, especially in the middle stratosphere (24-32 km sub-column). Global maps of the integrated L2_O3_PR ozone profile data and of the L2_O3 total column data (bottom row) look mutually consistent (same colour scale), with a slight overall negative bias for the integrated profile product with respect to the total column product (same processor version).

6.3.10 Other features

In the absence of clouds, data files sometimes contain negative surface albedo values. The TROPOMI ground pixels affected by this anomaly are usually located at the east and west edges of the across-orbit measurement swath. For now, these negative values are set to zero in the radiative transfer code and validation tools (as an influence quantity, also see above), but not in the ozone profile data distribution to users.

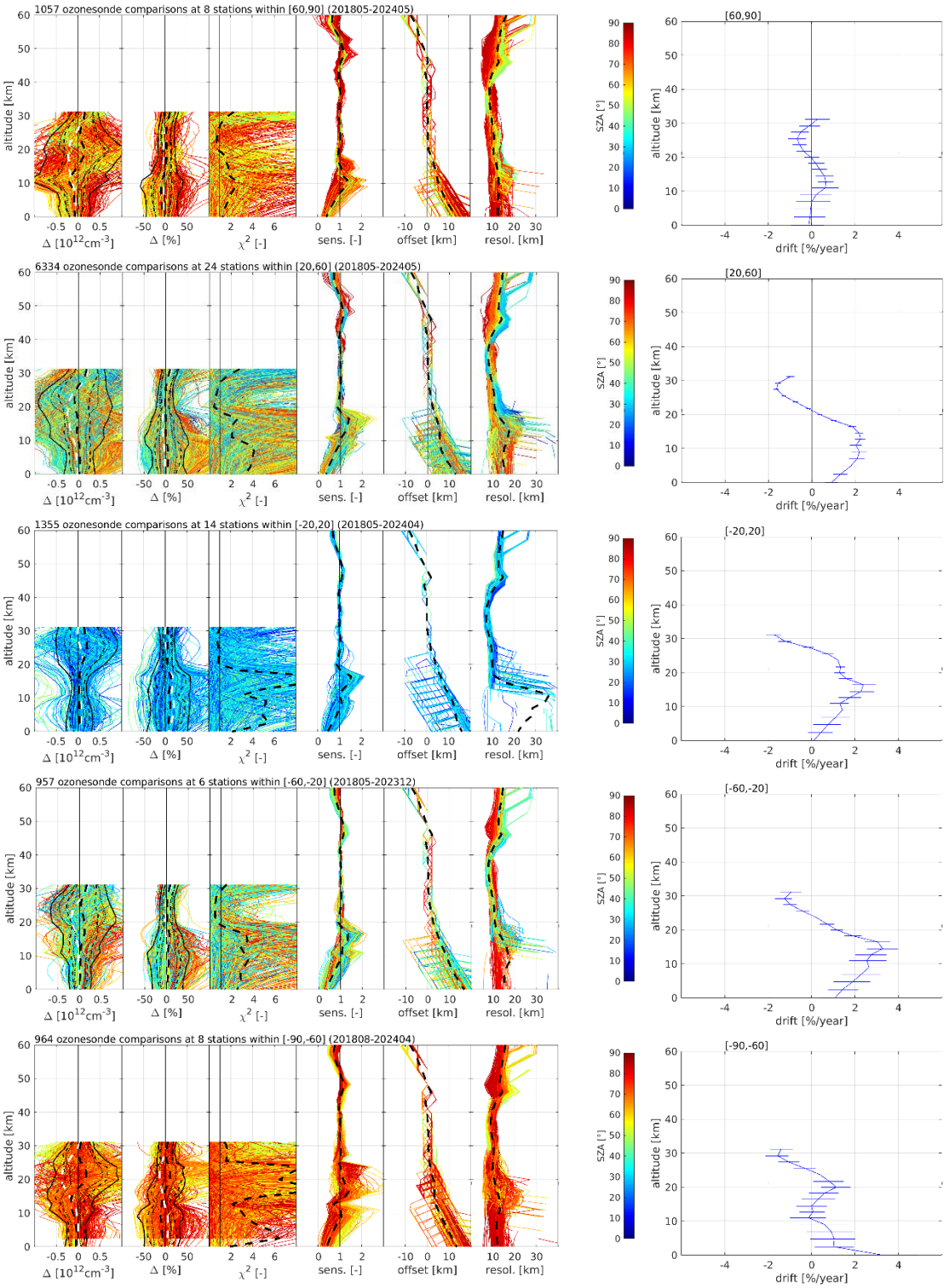


Figure 32: Comparison between S5P RPRO/OFFL ozone profiles (v2.4-6, May 2018 to May 2024) and co-located ozonesonde data in five latitude zones (Arctic to Antarctic, from top to bottom). Every row shows, from left to right: the difference and percent relative difference between S5P and ozonesonde data as a function of altitude, the chi-square, the vertical sensitivity, the altitude registration offset, the S5P averaging kernel FWHM, and the linear temporal drift. Colours indicate the SZA. Black dashed lines show mean values (thick lines) and standard deviations (thin lines, around the mean), while white, dashed lines indicate the mean difference between the a-priori profile and the reference measurement. Dotted black lines indicate the total ex-ante uncertainty of TROPOMI and the reference measurements combined (around the mean difference). Horizontal bars in the drift plots indicate 2-sigma uncertainty estimates on the drift using a bootstrapping technique with 1000 samples.

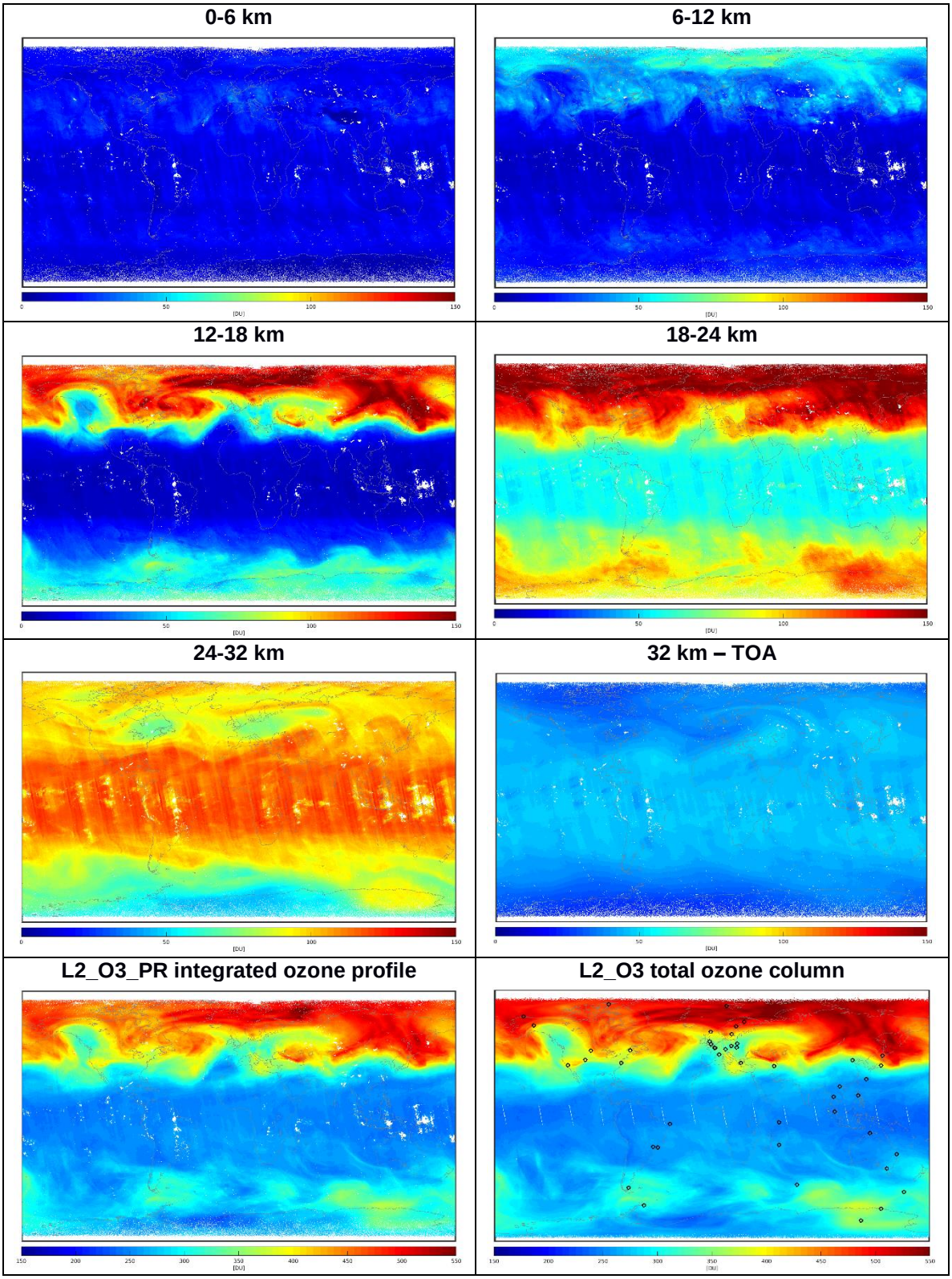


Figure 33: Upper six panels: daily global map for the six partial columns in the S5P L2_O3_PR OFFL v02.06.00 ozone profile product of March 16, 2024, as a function of altitude. The two lower maps show the total ozone column values obtained by integration of the L2_O3_PR profile data (left), and the map of L2_O3 total ozone column values for the same day (right) to check for their mutual consistency.

7 Validation Results: L2_NO2

7.1 L2_NO2 products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_NO2 data products as identified in **Table 1**: NO_2 tropospheric column, stratospheric NO_2 column, and NO_2 total column. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. The operational (E2) phase for the S5P TROPOMI mission started on 2018/04/30 with orbit 02818.

The latest processor version 02.06.00 was activated on 2023/11/26 (OFFL) and 2023/11/29 (NRTI), without changes to the algorithm. Data has been reprocessed to version 02.04.00 (RPRO) for the time span 2018/05/01 to 2022/07/16, using the TROPOMI DLER climatology. The NRTI data covers the full range of versions from 01.00.01 to 02.06.00 as there is no reprocessing necessary.

OFFL and NRTI products may differ because they are retrieved using ECWMF forecast meteorological data with a time difference as input for the CTM. Subsection 7.4 demonstrates evidence that NRTI and OFFL data do not differ significantly and that their respective validations yield similar conclusions.

7.2 Validation approach

7.2.1 Ground-based monitoring networks

Tropospheric NO2 – MAX-DOAS UV-Visible Spectrometers

S5P TROPOMI L2_NO2 tropospheric nitrogen dioxide column data are routinely compared to reference measurements acquired by MAX-DOAS (Multi-AXis Differential Optical Absorption Spectroscopy) UV-Visible spectrometers. Several of those instruments perform network operation in the context of the Network for the Detection of Atmospheric Composition Change (NDACC). MAX-DOAS tropospheric NO_2 column data have a maximum bias of 20 % and a precision better than 30 % at this set of stations. These estimates do not include the comparison errors, such as the spatial mismatch error (related to different field of view) and the difference in sensitivity coming from averaging kernels.

The validation with MAX-DOAS data from NItrogen Dioxide and FORmaldehyde VALidation (NIDFORVAL) has been included in a harmonized validation effort (Verhoelst et al., 2021). Since then, contact with NIDFORVAL PIs to extend their datasets and provision on ESA Atmospheric Validation Data Centre (EVDC) through conversion to fully GEOMS (Generic Earth Observation Metadata Standard) and HARP compatible data lead to the inclusion of new stations to the VDAF-AVS and several time-period updates.

Stratospheric NO_2 – ZSL-DOAS UV-Visible Spectrometers

S5P TROPOMI L2_NO2 stratospheric nitrogen dioxide column data are compared routinely to reference measurements acquired by Zenith-Scattered Light Differential Optical Absorption Spectroscopy (ZSL-DOAS) UV-Visible spectrometers (Pommereau and Goutail, 1988; Hendrick et al., 2011). The instruments perform network operations in the context of the Network for the Detection of Atmospheric Composition Change (NDACC). The ZSL-DOAS validation data (VDAF-AVS, 15 stations) have been obtained through the SAOZ near-real-time processing facility operated by the CNRS LATMOS (see **Figure 34**, red dots). They are complemented with measurements from 13 other NDACC affiliated ZSL-DOAS instruments (blue and green dots). The stations are located between 79°N and 75°S.

NDACC field intercomparison campaigns (Roscoe et al., 1999; Vandaele et al., 2005) conclude to an uncertainty of about 4-7% on the slant column density. Converting slant column into vertical column using a zenith-sky AMF, the uncertainty on the vertical column is estimated to be about 10-14 % for the latest data processing version (Yela et al., 2017; Bognar et al., 2019).

A limiting comparison factor comes from the temperature dependence of the NO_2 absorption cross-sections used in the DOAS retrieval of the slant column density. Most of the NDACC instruments use cross-sections at a single temperature of 220 K, which introduces a seasonal error of up to a few percent at middle and high latitudes.

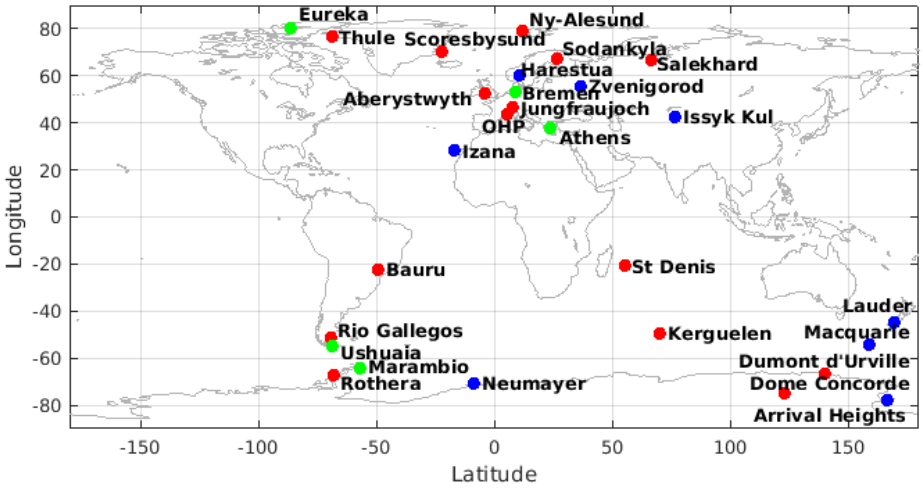


Figure 34: Geographical distribution of the NDACC ZSL-DOAS instruments routinely measuring stratospheric NO_2 and yielding co-locations with the current S5P L2_NO2 datasets. Stations marked with a red dot contribute fast delivery data coming from the LATMOS_RT facility. Blue and green dots depict the NDACC stations that contributes ZSL-DOAS data directly through the NDACC DHF and the AO project NIDFORVAL, respectively.

To account for effects of the photochemical diurnal cycle of stratospheric NO_2, the ZSL-DOAS measurements, which are obtained two times a day at twilight, are adjusted to the S5P overpass time using a model-based factor. This is calculated with the PSCBOX 1D stacked-box photochemical model (Errera and Fonteyn, 2001; Hendrick et al., 2004), initiated with daily fields from the SLIMCAT chemistry-transport model (CTM). The amplitude of the adjustment depends strongly on the effective SZA assigned to the ZSL-DOAS measurements, which is taken to be 89° here. The uncertainty related to this adjustment is in the order of 10%.

To reduce mismatch errors due to the significant horizontal smoothing differences between S5P and ZSL-DOAS measurements, S5P NO_2 values (from ground pixels at high resolution) are averaged over the air mass footprint where ground-based zenith-sky measurements are sensitive.

Additional confirmation is obtained by comparison with 3 mountain-top PGN (Pandonia Global Network) instruments where the measured signal corresponds rather more to the S5P L2_NO2 stratospheric column than the total column. The stations are Altzomoni (3985 m), Izaña (2360 m), and Mauna Loa (4169 m).

Stratospheric NO_2 – FTIR spectrometers

The ground-based FTIR instruments measure stratospheric NO_2 (e.g. Hendrick et al., 2012, Bognard et al., 2019, Garcia et al., 2021) with a precision of about 8-12 % and a systematic error of about 10 %. Within the AO project NIDFORVAL, the retrieval settings have been harmonized and applied for data of 26 FTIR stations for the S5P validation. We build the collocated pairs to be compared in several steps:

- When the collocation is not above the FTIR station, we use the line-of sight of the instrument (FTIR are direct sun measurements) instead and calculate the position along the line-of-sight corresponding to the altitude where the NO_2 FTIR averaging kernels show the maximum sensitivity (~30-35km). Then, S5P pixels are selected within 50 km of this position (about 150-200 pixels). Only pixels with `qa_value` > 0.5 are used. A collocation pair is only kept if at least 10 pixels can be averaged.
- The time coincidence criterion is set to ±1 hour of the satellite overpass time.
- The comparison methodology is the same as for HCHO validation using FTIR data (Vigouroux et al., 2020): (i) The FTIR a priori profile is substituted with the TROPOMI L2_NO2 one to get a corrected FTIR profile. (ii) The corrected profile is smoothed with the TROPOMI averaging kernel (Rodgers and Connor, 2003). In this process, since the TROPOMI averaging kernels are zero below the tropopause for the stratospheric NO_2, the tropospheric part of the FTIR profile is removed, and only stratospheric columns from both products are indeed compared. (iii) Both the individual manipulated FTIR columns and the individual S5P manipulated pixel columns are then averaged.
- Finally, the relative median bias at a single station is estimated by the median relative difference: MED[(SAT-REF)/REF]. Absolute-scale dispersion is estimated by the scaled Median Absolute Deviation: MAD = SF*MED[ABS(DIFF-MED(DIFF))]. The scaling factor SF of 1.4826 ensures that for a normal distribution, the MAD is equal to the standard deviation.

Total NO_2 – Pandora Direct-Sun UV-Visible Spectrometers

TROPOMI L2_NO2 nitrogen dioxide summed column data (troposphere + stratosphere) are routinely compared to reference measurements acquired by Pandora instruments. They perform network operation in the context of the Pandonia Global Network (PGN). Pandora total NO_2 data have maximum bias of 10-15 % and a precision of roughly 0.28 Pmolec/cm² (about 10 %). Although for most stations PGN v1.7 is replaced by v1.8, this is not yet everywhere the case, leading to a mix of versions at EVDC. PGN versions v1.7 and v1.8 have an additional uncertainty of about 10 % related to the temperature effect of the cross-section (see *Section 7.5.1*). Currently, 72 Pandora instruments are available for comparisons. But the PGN data stream to EVDC is temporarily halted due to technical issues, with last data from early December. Total column comparisons reported below therefore do not include the last months of S5P TROPOMI measurements.

The comparison criteria on the VDAF Automated Validation Server are: (i) TROPOMI L2_NO2 data with `qa_value` > 0.5; (ii) the TROPOMI ground pixel contains the Pandora station; (iii) Pandora negative values are excluded and only measurements with a flag not equal to 0 and 10 are used; (iv) Pandora measurements within ±30 min are averaged. If the Pandora instrument operates at an elevated station above low-lying tropospheric pollution, the Pandora measurement in absence of free troposphere NO_2 can also be representative of the stratospheric NO_2 column.

7.2.2 Satellites

TROPOMI L2_NO2 nitrogen dioxide column data are also compared to data from the Ozone Monitoring Instrument (OMI) retrieved with both the QA4ECV and the IUP-UB algorithm. OMI is on board the EOS-Aura satellite that was launched in July 2004.

7.2.3 Field campaigns and modelling support

None for this report.

7.3 Validation of L2_NO2

7.3.1 Recommendations for data usage

In order to avoid misinterpretation of the data quality, it is recommended to only use TROPOMI NO_2 measurements with `qa_value` ≥ 0.75. This removes cloudy scenes (cloud radiance fraction > 0.5), scenes covered by snow/ice, several other errors, and problematic retrievals. For stratospheric NO_2 retrievals and data comparisons, clouds are less of a problem, so that `qa_value` ≥ 0.5 can be used. Data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

7.3.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC VDAF and by S5PVT AO projects. Routine operations validation activities rely on the Automated Validation Server of the MPC VDAF, the Multi-TASTE versatile multi-platform validation system operated at BIRA-IASB, the validation tools of IUP-UB, and the HARP toolset (version 1.6). An in-depth discussion of the routine validation results up to March 2020 by Verhoelst et al. (2021) was published. Van Geffen et al. (2022) discussed the V2.2 improvements and Eskes et al. (2021) verified the reprocessed S5P-PAL data of V2.3.1.

In this section, we focus on the reprocessed v2.4.0 data (RPRO, from end of April 2018 to end of July 2022) and merge it with the OFFL v2.4.0/2.5.0/2.6.0 datasets available since 17/07/2022, 23/03/2023, and 26/11/2023, respectively. For the overlapping days in July 2022, the RPRO data will have preference to remove the impact of the spin-up period of the OFFL dataset.

Up-to-date validation results and consolidated validation reports are available through the MPC VDAF Portal. The PGN steadily extends its Pandora data provision on EVDC, making it accessible for the VDAF Automated Validation Server. There is regular contact with NIDFORVAL PIs to extend their MAX-DOAS datasets and their provision on EVDC (through conversion to fully GEOMS and HARP compatible data) and to do time-period updates. Moreover, contacts with new MAX-DOAS station PIs have been established that lead to new stations inclusions. Within NIDFORVAL, the FTIR PIs applied the harmonized settings, and submitted their data individually to the NIDFORVAL's PI. In summary, the FRM data streams cover the period from May 2018 to August 2024:

- MAX-DOAS tropospheric columns from the NIDFORVAL AO project from 31 stations. Vertical smoothing harmonization could be applied for 13 stations. MAX-DOAS tropospheric column data at the VDAF Automated Validation Server via EVDC are available from 8 stations (De Bilt, Cabauw (KNMI), Uccle, Xianghe (BIRA-IASB), Bremen, Athens (IUP-B), Mainz (MPIC); and Mohali (MPIC/IISER)). All stations on the server except Mohali are part of NDACC. At present, however, several of these stations are having/had instrumental problems and are temporarily down (Uccle, Xianghe, and Mainz).
- NDACC ZSL-DOAS stratospheric columns from 25 stations sample latitudes between 80°N (Eureka) and -75°S (Dome C). For the current report, covering the full mission reprocessing, so far only co-locations with the SAOZ network have been processed. A future version will include the other NDACC instruments.
- FTIR stratospheric columns from 26 stations between 80°N (Eureka) and -78°S (Arrival Heights).
- PGN total column direct-sun data are available from 66 stations (72 instruments) at the VDAF Automated Validation Server sample latitudes from 80°N to -46°S. Currently PGN is undergoing a version upgrade: at most sites but not yet all, PGN v1.7, is replaced by v1.8, leading in a few

cases to a mix of versions at the validation server. As PGN v1.7 is dissuaded for sites dominated by stratospheric NO_2, we excluded some stations for some plots.

7.3.3 Tropospheric NO_2 column

7.3.3.1 Bias

The OFFL NO_2 tropospheric column values are compared with NDACC MAX-DOAS data from 31 stations. The median bias over all stations and the full mission time is -1.5 Pmolec/cm^2 (approx. -28.7 %). A summary of the bias and spread for all stations is shown in **Figure 35** and **Figure 36**. The median bias for the subset of 8 MAX-DOAS stations in the VDAF-AVS is lower with -1.1 Pmolec/cm^2 (-18.8 %).

These results are within the mission requirement of a 50 % maximum bias. With a station-to-station dispersion (IP68/2 over all station medians) of 2 Pmolec/cm^2, for the 31 MAXDOAS stations ensemble, a single bias number for all stations has limited meaning. Three regimes can be identified for comparisons, as discussed in Verhoelst et al. (2021): (1) low tropospheric NO_2 values (median tropospheric column <2 Pmolec/cm^2), (2) polluted stations (from 3 to 14 Pmolec/cm^2), and (3) extremely polluted stations (>15 Pmolec/cm^2). The median bias in these regimes is about 0.2 Pmolec/cm^2 (12.6 %), -1.3 Pmolec/cm^2 (-28.6 %), and -8.1 Pmolec/cm^2 (-40.3 %), respectively.

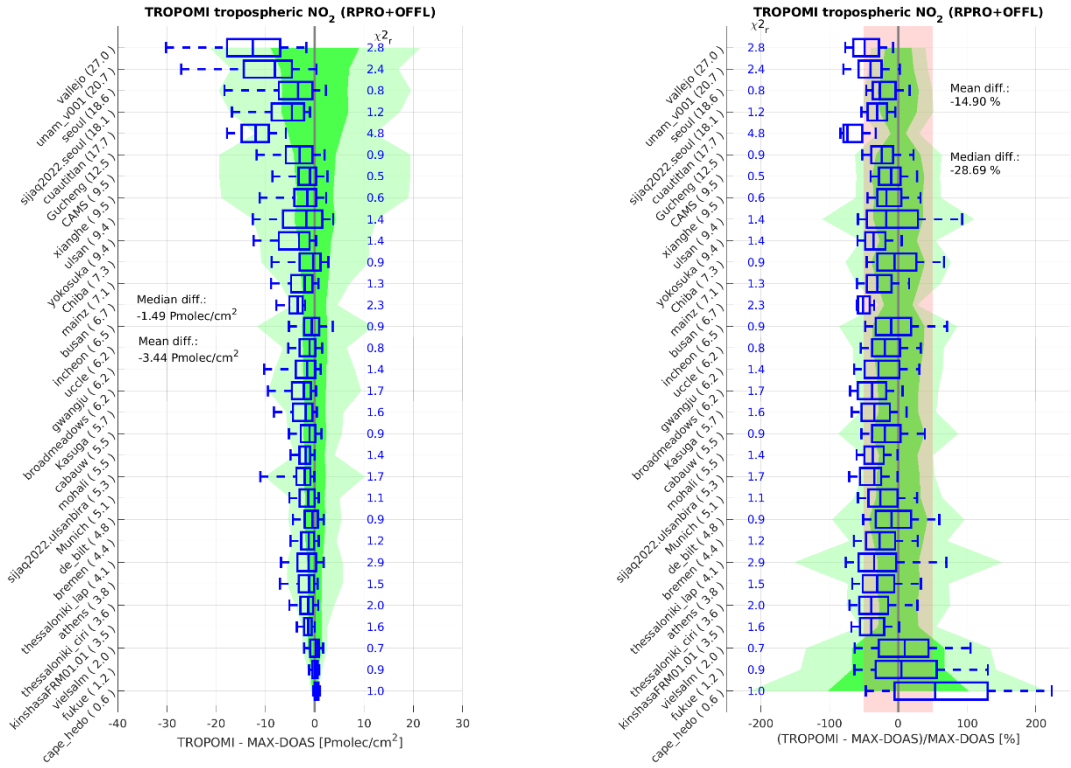
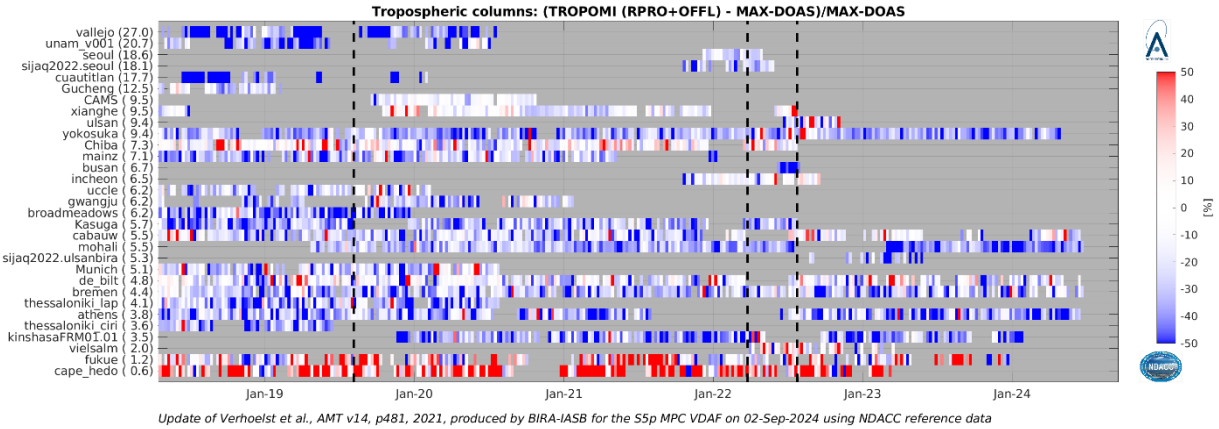


Figure 35: Box-and-whisker plots summarizing the bias and spread of the difference between S5P TROPOMI RPRO+OFFL v02.04/02.05/02.06 and MAX-DOAS NO_2 tropospheric columns (left: absolute, right: relative). Values between brackets in the labels denote the median tropospheric column at the station. The time frame is from May 2018 until end Augsut 2024 (see next figure). Stations are ordered by median tropospheric column. The median difference is represented by a vertical solid line inside the box, which marks the 25 and 75% quantiles. The whiskers cover the 9-91% range of the differences. The red shaded area represents the mission requirement of 50% for the uncertainty. Also indicated is the combined ex-ante uncertainty (green shaded areas, where dark green represents the median combined uncertainty and light green the 95% percentile), and the reduced chi^2 of the comparison for each station.



Update of Verhoelst et al., AMT v14, p481, 2021, produced by BIRA-IASB for the S5p MPC VDAF on 02-Sep-2024 using NDACC reference data

Figure 36: Time series – from May 2018 until end August 2024 – of the weekly averaged relative difference [%] between S5P RPRO+OFFL v02.04/02.05/02.06 and MAX-DOAS NO_2 tropospheric column data. Stations are ordered by median tropospheric column, in brackets.

A test case study of the effect of application of averaging kernels on the bias has been performed for ground-based MAX-DOAS stations providing low tropospheric profiles (with data up to end January 2024), which is the case at 18 stations (the BIRA-IASB, ChibaU, JAMSTEC, and UNAM stations and for the sites processed with FRM4DOAS for the GMAP 2021 and SIJAQ 2022 GEMS validation campaigns (ulsan, busan, seoul, sijaq2022.ulsanbira and sijaq2022.seoul)). In this case, the TROPOMI averaging kernel can be used to smooth the MAX-DOAS profiles and remove the profile contribution from the comparison, using the following formula:

$$VCD_{smoothed} = AK_{sat} * x_{MAXDOAS}$$

This analysis is illustrated in **Figure 37** where the absolute and relative differences are presented for original comparisons (grey boxes) and after smoothing of the ground-based data (black boxes). The median bias then generally decreases absolutely (for 12 cases) by about 10 to 30 % (see bottom of **Figure 37**). This reduction comes sometimes at the expense of a larger spread of the comparisons (e.g. Vallejo and Unam). In some cases, the final bias is slightly larger after smoothing (Vielsalm, Incheon, Xianghe). The large increase of 60% for Fukue and Cape Hedo can be attributed to the relative bias calculation from dividing with small tropospheric VCDs at very clean ground-based stations.

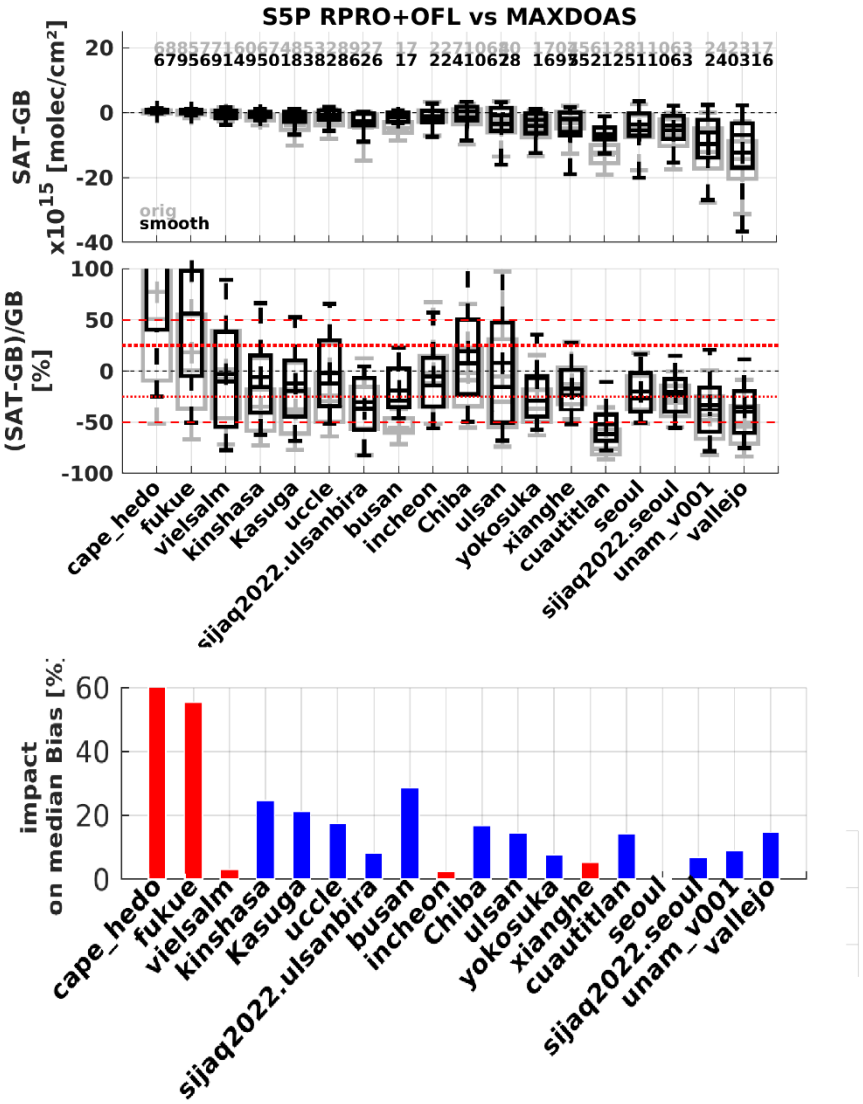


Figure 37: Box-and-whisker plots summarizing the bias and spread of the difference between S5P TROPOMI RPRO+OFFL v2.4-2.6 and MAX-DOAS NO_2 tropospheric columns (grey) and the smoothed MAX-DOAS NO_2 tropospheric columns (black). The third panel presents bar plot summarizing the change on the median bias between the original comparisons and the smoothed ones. Absolute values of the relative median biases are shown, where reductions are marked in blue bars and increases in red bars. The time frame is for all available data from May 2018 to end August 2024. The median difference is represented by a vertical solid line inside the box, which marks the 25 and 75% quantiles. The whiskers cover the 9-91% range of the differences and the crosses denotes the mean differences.

In summary, the tropospheric NO_2 bias depends on pollution level. On average, it is about -28 % for all stations, but it can be as high as -50 % for extreme pollution and +33 % in clean areas. Taking the sensitivities of the instruments into account by using the satellite averaging kernels with the low tropospheric MAX-DOAS profiles, the bias can be reduced by up to 20% absolutely.

7.3.3.2 Dispersion

The median dispersion in comparison to NIDFORVAL MAX-DOAS is 3.1 Pmolec/cm^2. The median IP68/2 dispersion for the different pollution levels as defined in the previous section is (1) 0.7 Pmolec/cm^2, (2) 2.9 Pmolec/cm^2, and (3) 5.9 Pmolec/cm^2, respectively. The uncertainty precision requirement of maximum 0.7 Pmolec/cm^2 is satisfied for the clean-station ensemble.

It must be noted that MAX-DOAS uncertainty sources and comparison errors also contribute to the dispersion. Moreover, systematic errors (e.g., seasonal cycle) can contribute. A part of the systematic error component can be removed by calculating the dispersion (IP68/2) around the OLS regression line instead of the dispersion between S5P and MAX-DOAS data (adapted from Schneider et al., 2006). The residual dispersion IP68/2 is 0.9 Pmolec/cm^2 at Mohali, between 1.2 and 1.7 Pmolec/cm^2 at the different European VDAF-AVS stations, and 2.3 Pmolec/cm^2 at Xianghe. There is a reasonably good correlation between TROPOMI and MAX-DOAS tropospheric columns with the Pearson R varying between 0.66 (Mohali) and 0.88 (Xianghe) with a mean of 0.78.

7.3.3.3 Dependence on influence quantities

Two key influence quantities for observations of tropospheric NO_2 are aerosol optical depth (AOD) and satellite cloud (radiance) fraction (CRF). The first one is retrieved within the MAX-DOAS NO_2 analysis. Please note that we expect a link between the two: (scattering) AOD is accompanied by increases in the “effective CRF”. When binning the MAX-DOAS comparison biases of each station by AOD from 0 to 2 in intervals of 0.5 and CRF from 0 to 0.5 in intervals of 0.05, a median bias increase towards larger bin values is found (**Figure 38**).

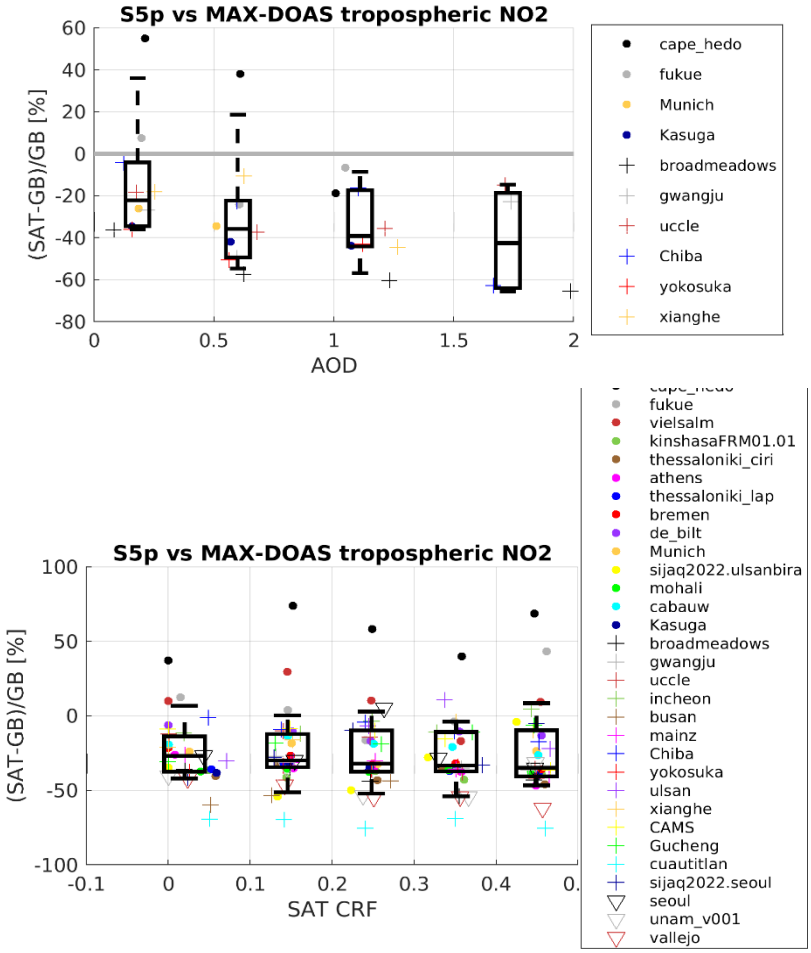


Figure 38: Relative differences between S5p TROPOMI RPRO+OFFL and MAX-DOAS NO_2 tropospheric columns as a function (upper panel) of MAX-DOAS AOD and (lower panel) S5p cloud radiance fraction (CRF). The time frame is from May 2018 until end of August 2024.

7.3.3.4 Seasonal and shorter term variability

Global zonal daily mean NO_2 tropospheric columns are shown in **Figure 39** for RPRO/OFFL versions 02.04.00 to 02.06.00. Overall, no obvious column changes or trends over mission lifetime are visible.

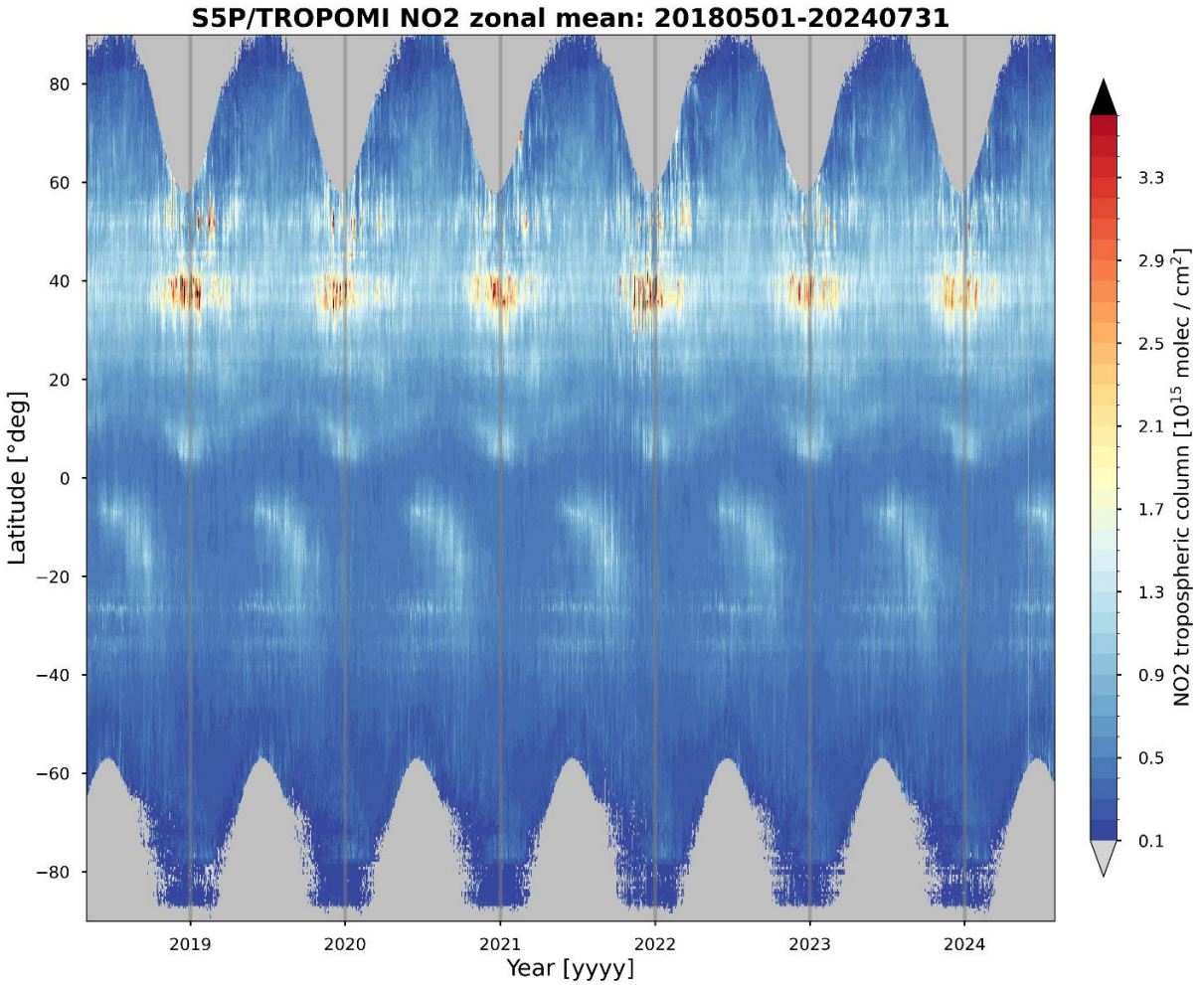


Figure 39: S5P TROPOMI NO_2 tropospheric columns [10^{15} molec/cm^2] as a function of day and latitude. The period covers six years from May 2018 to August 2024. RPRO/OFFL V02.04.00 and OFFL V02.05.00/02.06.00 data are used. Grid box size in latitude direction is 0.5°. The light grey vertical lines mark the beginning of each year.

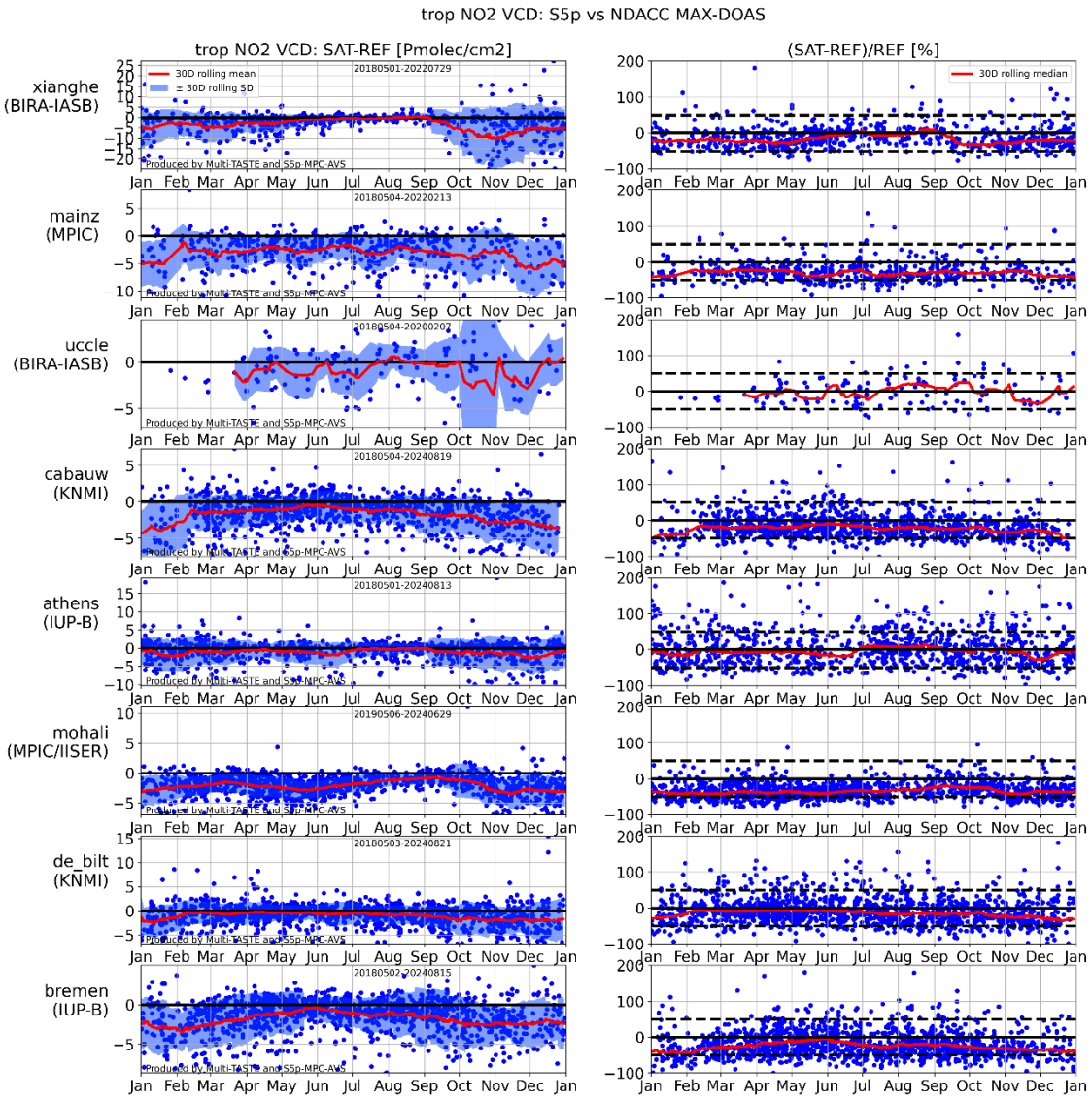


Figure 40: Seasonal cycle (with data mapped to one generic year) of the difference between S5P RPRO+OFFL and MAX-DOAS NO_2 tropospheric column data at eight stations. Difference (left) and relative difference (right). On the left column, the lowest 2.5 % data is not shown for visibility. Data was obtained from the VDAF Automated Validation Server on 2024/08/31.

Figure 40 presents the seasonal cycle of differences between S5P RPRO+OFFL and MAX-DOAS tropospheric NO_2 from the AVS. All comparison pairs are reported on a single year. TROPOMI measures lower values than MAX-DOAS in late fall and winter, when tropospheric NO_2 reaches its largest abundance. Over the entire year, the 30-day rolling median relative difference is within the mission requirements for the bias.

7.3.3.5 Geographical patterns

In general, erroneous geographical patterns or artefacts cannot be detected in L2_NO2, as shown in the maps for MAR/APR/MAY 2024 over central Europe (**Figure 40**) in comparison to 2019. A reduction in columns can be seen over Europe and China. Interesting is the slight increase in ship emissions around West Africa and Malaysia and the reduction in the Red Sea due to regional conflict in the Gulf of Aden.

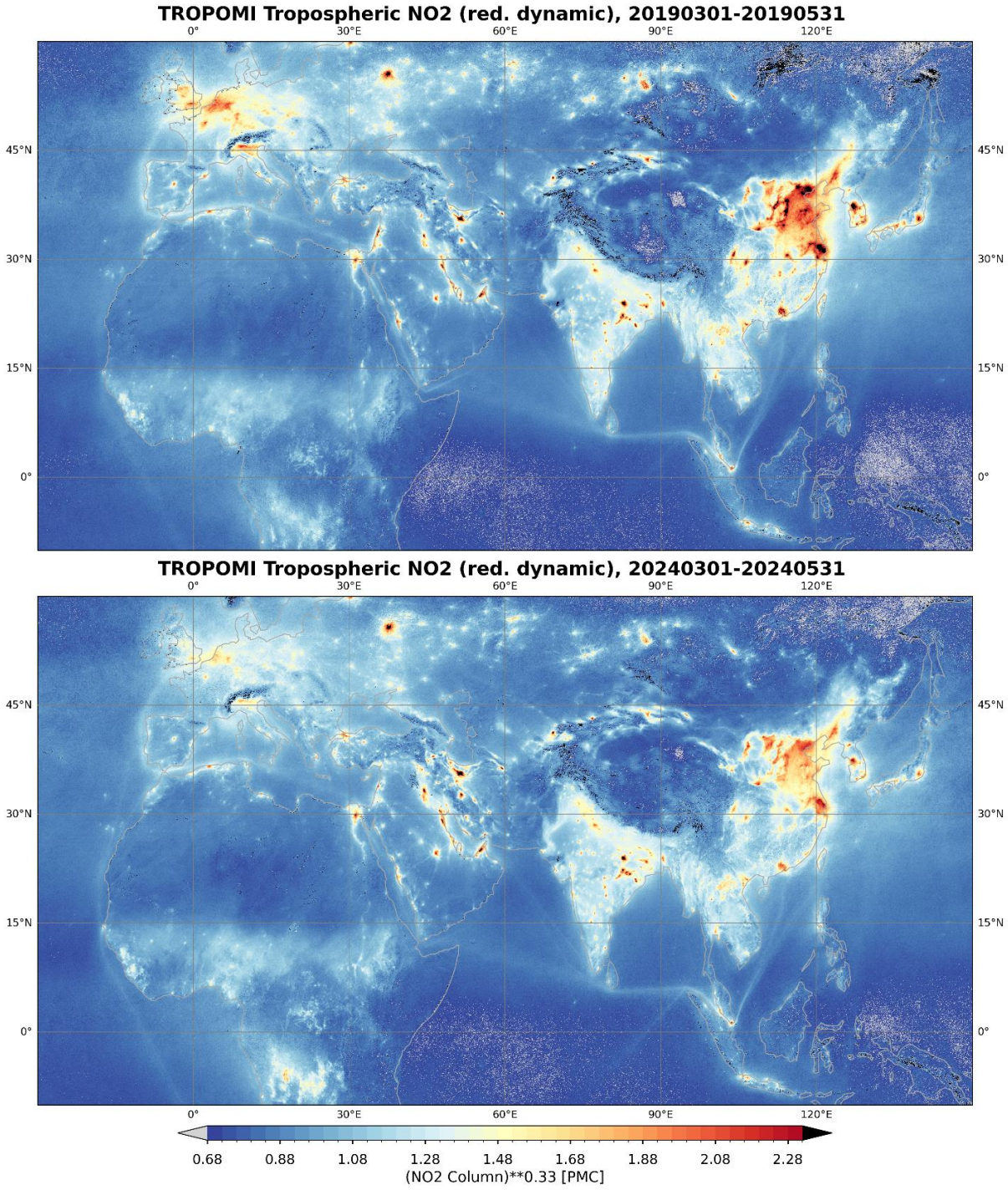


Figure 41: S5P RPRO/OFFL tropospheric NO_2 [-10°S, 60°N, -30°W, 150°E]. The data is binned on a grid of 0.05° latitude/longitude for Mar/Apr/May 2019 and 2024. A qa_value > 0.75 and a cloud fraction CF<0.2 are used to reduce the amount of data and exclude cloudy scenes. 1 PMC stands for 1 Pmolec/cm^2. The third root of values was calculated to better cover the dynamics.

7.3.3.6 Other features

Ordinary linear regression (OLS) of S5P vs MAX-DOAS yields fairly good correlation coefficients (0.71-0.66, see before), but low slopes. When $S5P=a*MXD+b$, a varies between 0.4 (Bremen) to 0.6 (Athens, Xianghe). It is known, however, that this approach is only correct in the limit that all random errors are in S5P data. For the MAX-DOAS vs S5P OLS (i.e., assuming the opposite limit that all random errors are in MAX-DOAS), still with *S5P = a*MXD + b*, one obtains slopes closer to unity, varying between 0.74 (Mainz) and 1.0 (De Bilt).

7.3.4 Stratospheric NO_2 column

7.3.4.1 Bias

RPRO+OFFL stratospheric NO_2 column values are generally lower than the photochemical corrected ground-based ZSL-DOAS values by approximately -0.10 Pmolec/cm2 (-3 %), with a station-to-station scatter of the mean bias of similar magnitude (**Figure 42**). For the current report, the validation results are based on comparisons with the automated SAOZ stations using a photochemical correction and sampling the latitude range from 80°N (Eureka) to -75°S (Dome C). A time series for each station is given in **Figure 43**. The remainder of the NDACC instruments will be included in a next version of this report. These stratospheric column results are within the mission requirement of 10% maximum bias (equivalent to 0.2-0.4 Pmolec/cm^2, depending on latitude and season).

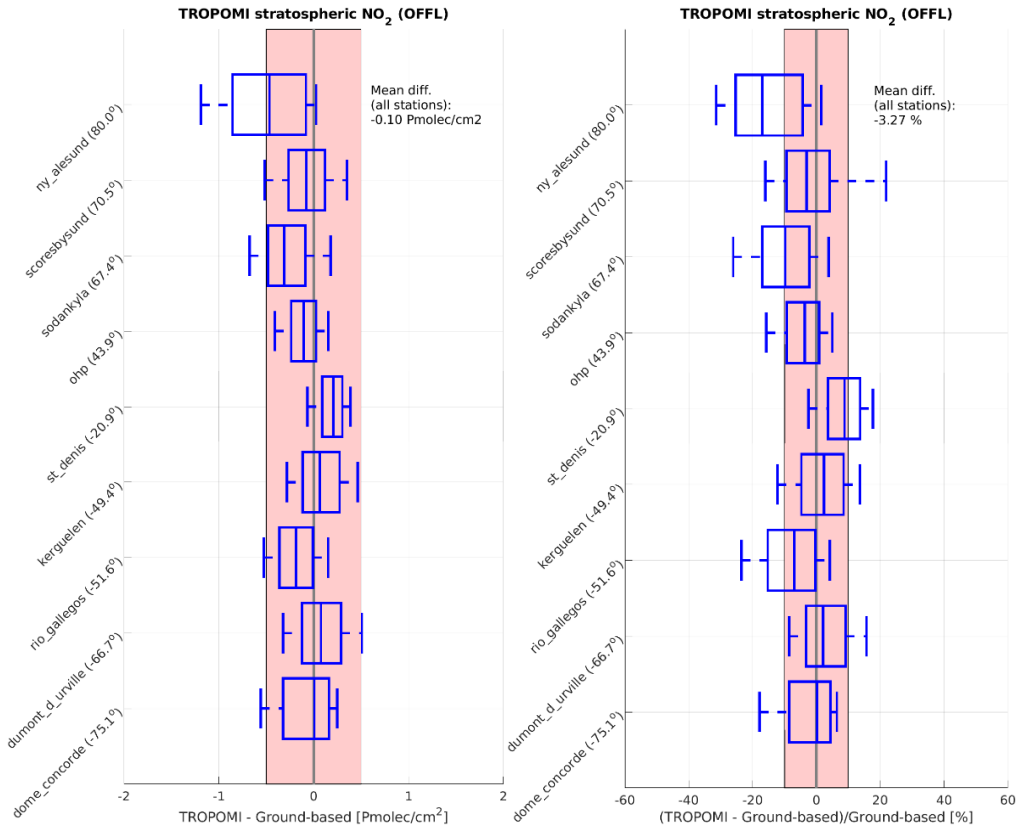


Figure 42: Pole to pole bias and spread (absolute differences in the left-hand panel, relative differences in the righ-hand panel) of S5P TROPOMI RPRO+OFFL and NDACC ZSL-DOAS NO2 stratospheric columns. For the current report, only the SAOZ data sets covering the full mission lifetime are shown. The median difference per station is represented by a vertical solid line inside the box, which marks the 25 and 75% quantiles and the whiskers denote the 9-91% range. The red shaded area represents either the mission requirement of 0.5 Pmolec/cm^2 for the random part of the uncertainty (left-hand panel) or that of 10% on the systematic uncertainty (right-hand panel). Station name and latitude of the station are added on the left. The time frame is May 2018 to August 2024.

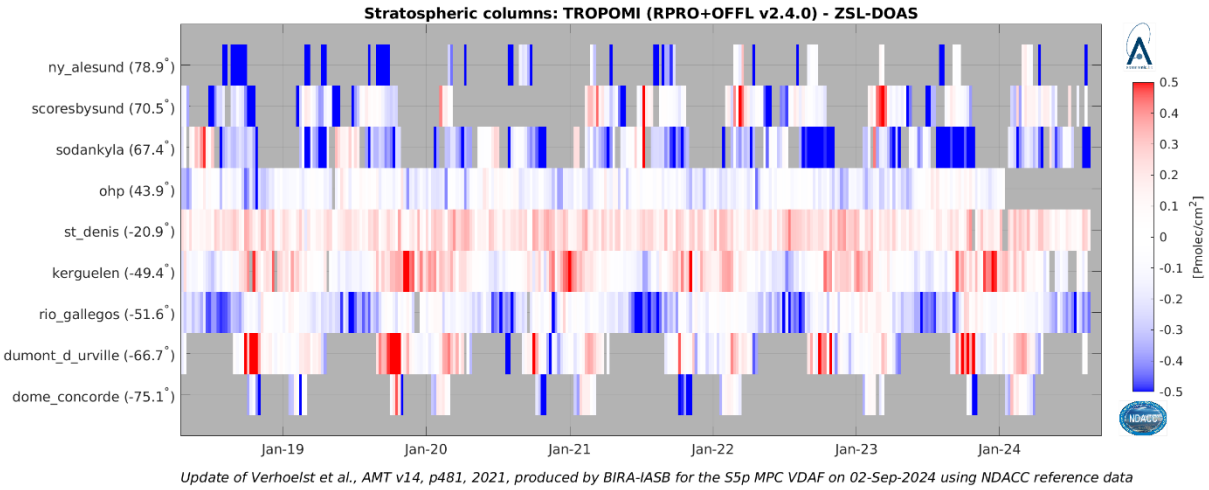


Figure 43: Time series – from August 2018 until August 2024 – of S5P RPRO+OFFL L2_NO2 and NDACC ZSL-DOAS NO_2 stratospheric column differences, weekly averaged [Pmolec/cm^2]. For the current report, only the SAOZ data sets covering the full mission lifetime are shown.

The median bias between S5P and FTIR NO_2 stratospheric columns is +4.7 % for the combined 26 stations (see **Figure 44**). Most of the sites show biases that are within the mission requirement of 10% maximum bias. Larger biases are observed at high latitude and tropical stations (9-13%, see **Figure 42**). The median positive bias of TROPOMI compared to FTIR is opposite to the negative one obtained with the ZSL-DOAS network (-5 %).

FTIR and ZSL-DOAS measurements were compared at sites where both techniques are available (Eureka, Sodankyla, Bremen, Izaña, and Maïdo) to check if the different TROPOMI biases are due to the ground-based data themselves or the difference in the validation methodologies used for FTIR and ZSL-DOAS.
ZLS-DOAS has indeed a positive median bias compared to FTIR of about 11% (individual biases from 0 to 16%), giving confidence that the difference in the TROPOMI biases of +3 and -6% obtained with FTIR and ZSL-DOAS, respectively, is not due to the different validation methodologies. The station-to-station 1σ scatter is 5.3%, which is similar to the ZSL-DOAS network and shows a good consistency of the FTIR network as well.

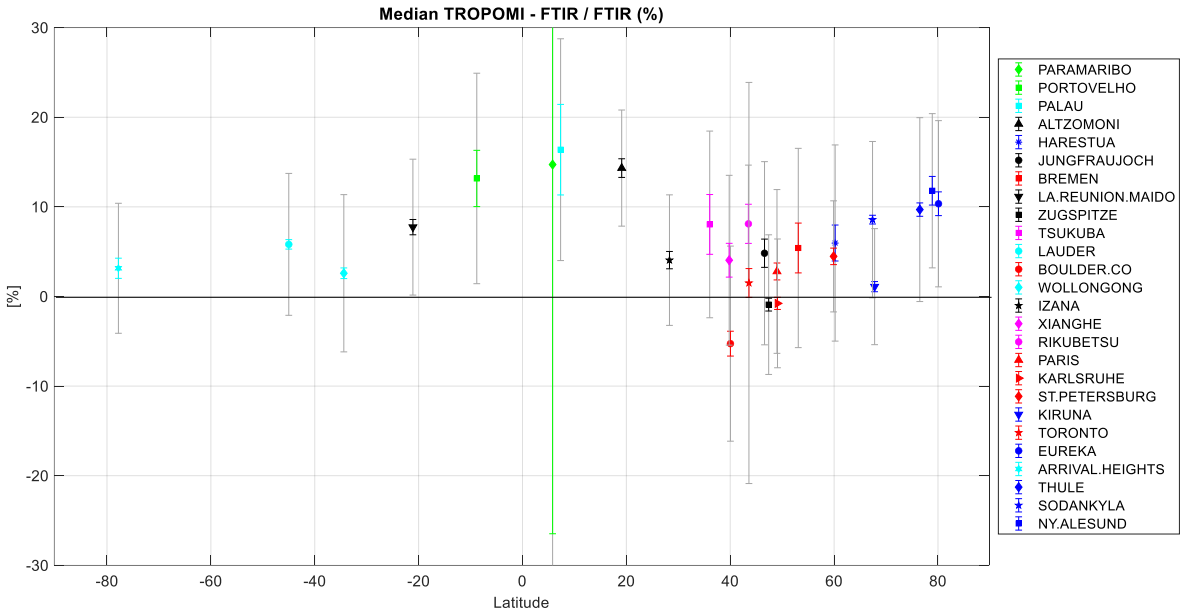


Figure 44: Median bias at each NDACC FTIR station as a function of latitude, calculated as the median of the percentage difference between S5P L2_NO2 and the FTIR NO_2 stratospheric column measurement. The grey bars are the scaled MAD (see Sect. 6.2.1), and the coloured bars are the ±2σ error on the bias.

7.3.4.2 Dispersion

From ZSL-DOAS comparisons, the ±1σ dispersion of the difference between stratospheric column and reference data around their median value rarely exceeds 0.3 Pmolec/cm² at stations without tropospheric pollution (cf. the box plots in **Figure 42**). When combining random errors in the satellite and reference measurements with irreducible collocation mismatch effects, it can be concluded that the random uncertainty on the S5P stratospheric column measurements falls within mission requirements of maximum 0.5 Pmolec/cm². The mean Pearson-R is 0.96 ± 0.1.

Similar conclusions are reached from FTIR comparisons. The scaled MAD (equivalent to 1σ dispersion) of the differences is 0.28 Pmolec/cm² for all data together, with a Pearson correlation coefficient of 0.94. At individual stations, it never exceeds 0.5 Pmolec/cm², except at Toronto (0.6) and Paramaribo (0.8). Note that at Paramaribo, only 7 collocations occur, which can explain the larger scatter and bias observed there. The robust correlation coefficient is 0.97, as shown in the scatter plot (**Figure 45**).

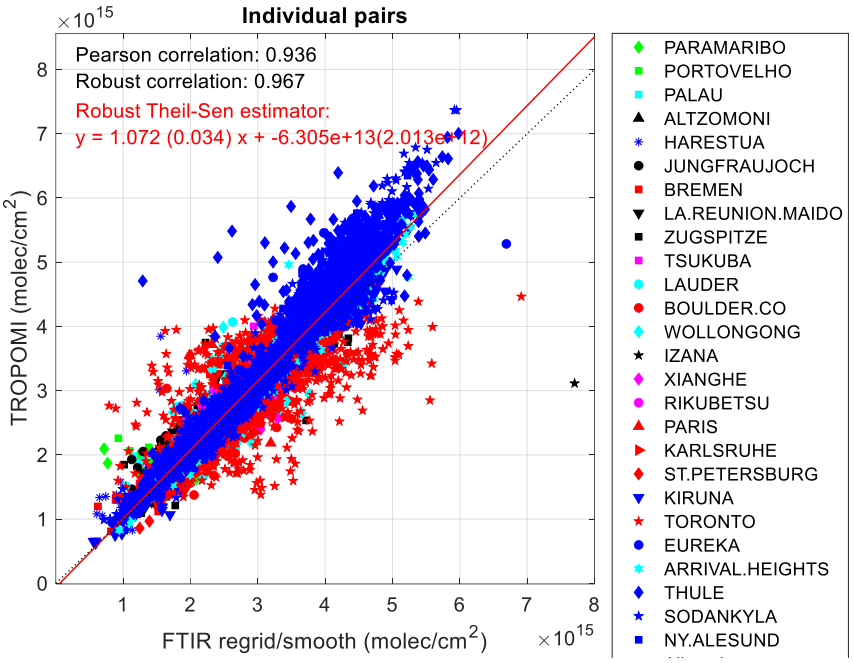


Figure 45: Scatter plot of co-located S5P TROPOMI and NDACC FTIR stratospheric NO_2 column data.

7.3.4.3 Dependence on influence quantities

The evaluation of potential dependences of the S5P stratospheric column on Solar Zenith Angle (SZA), cloud fraction (CF) and surface albedo of the S5P measurement does not reveal bias variations much larger than 0.2 Pmolec/cm^2 over the range of the influence quantities (**Figure 46**).

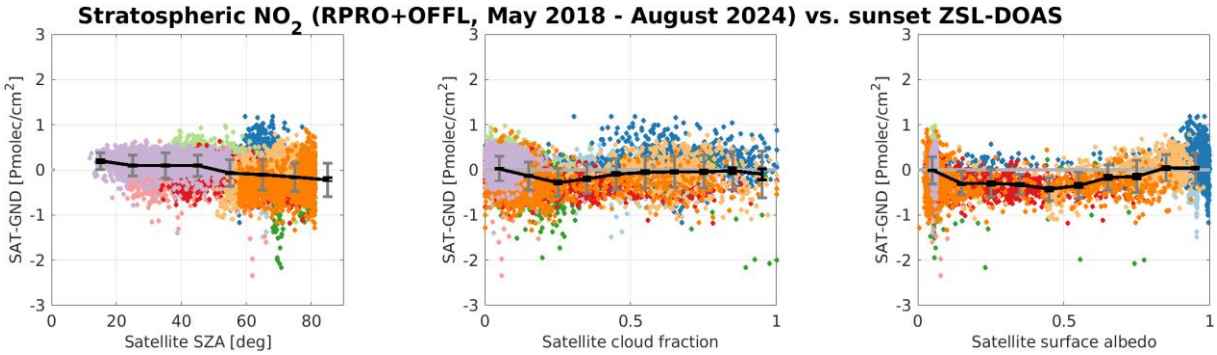


Figure 46: Difference between S5P L2_NO2 RPRO+OFFL and ground-based SAOZ stratospheric NO2 columns as a function of the satellite solar zenith angle (SZA), satellite cloud fraction, and satellite surface albedo. Mean and standard deviation are calculated over bin widths of 10° degrees in SZA, 0.1 in CF, and 0.1 in surface albedo (solid black line and grey bars). Co-locations cover the period from May 2018 to August 2024.

7.3.4.4 Seasonal cycle and shorter term variability

TROPOMI and ground-based ZSL-DOAS instruments both capture the short-term variabilities (at daily and monthly scales) of the NO_2 stratospheric column, as illustrated for the NDACC station of Kerguelen Island in **Figure 47**. The ground-based SAOZ data acquired at twilight were adjusted to account for the photochemical diurnal variation between twilight and the early afternoon S5P overpass time.

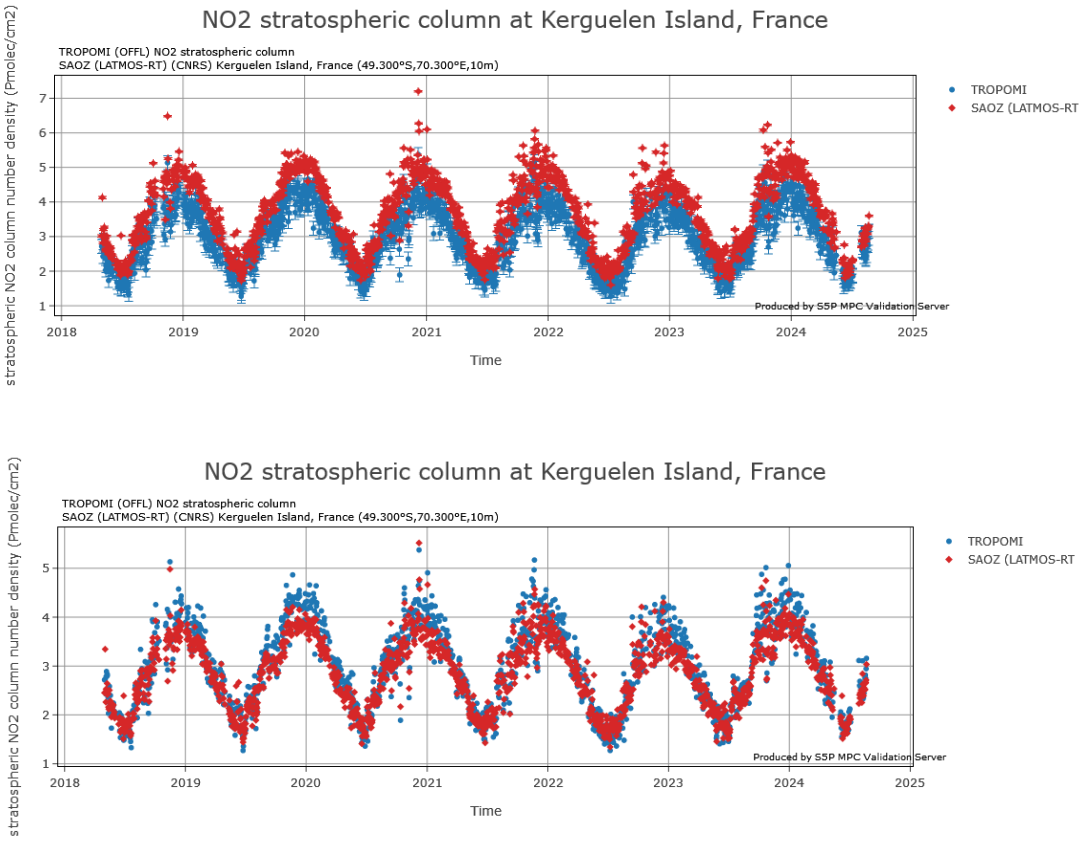


Figure 47: Time series of S5P OFFL L2_NO2 stratospheric NO2 column data (blue dots) co-located with ground-based SAOZ twilight measurements (red dots) at sunset performed by LATMOS at the NDACC southern mid-latitude station of Kerguelen Island. In the upper plot, the photochemical correction is deactivated to offset the two time-series and to better see the day-to-day variability. The time frame is July 2018 to August 2024.

7.3.4.5 Geographical patterns

None to report.

7.3.4.6 Other features

None to report.

7.3.5 Total NO_2 column

7.3.5.1 Bias

Based on measurements from 70 Pandora stations between 80.05°N and -45.8°S available at VDAF-AVS (**Figure 48**), the median bias is -8.6 % (-0.6 Pmolec/cm^2) with a station-to-station 1-σ scatter of 17 % and a Pearson correlation coefficient of 0.7. The results are within the 30% accuracy requirement, which is the average of the tropospheric and stratospheric bias maxima. Dividing the stations by pollution level at 6 Pmolec/cm^2, the bias is +5.9 % for the 28 lower polluted stations and -17.0 % for the 42 higher polluted stations.

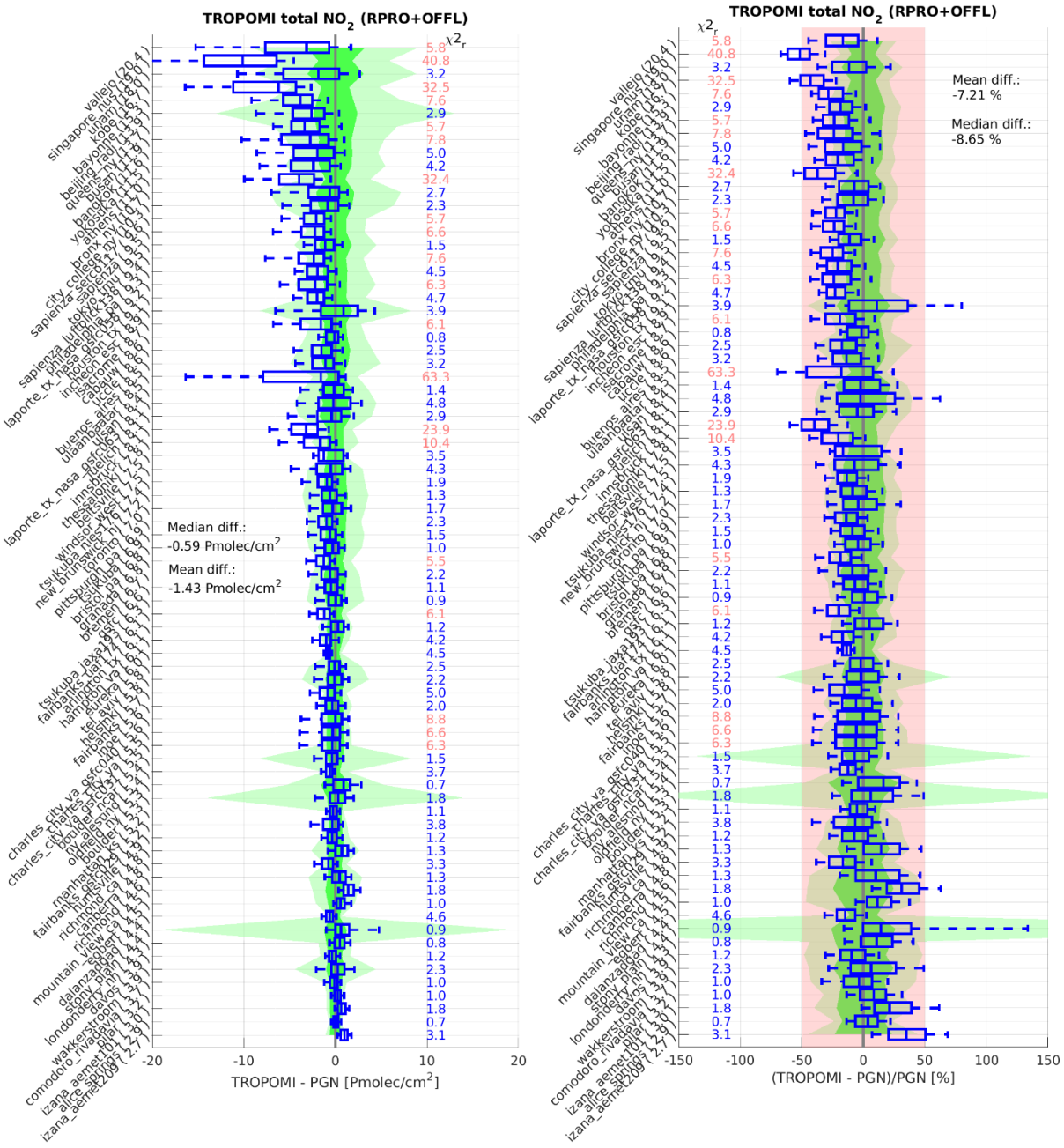


Figure 48: Box-and-whisker plots summarizing the bias and spread [Pmolec/cm²] (left) and relative bias and spread [%] (right) between S5P TROPOMI RPRO+OFFL and PGN Pandora NO_2 total column data. Conventions of the boxplots are identical to **Figure 35**. Stations are ordered by median total column. The time frame is May 2018 to August 2024. Poor reduced chi²s (i.e. those above 5) are marked in red. The large deviations at Singapore could be due to much localized pollution but require further investigation.

On 6 August 2019, there was a change in the TROPOMI ground pixel size but this affects neither the bias nor the dispersion in the comparisons.

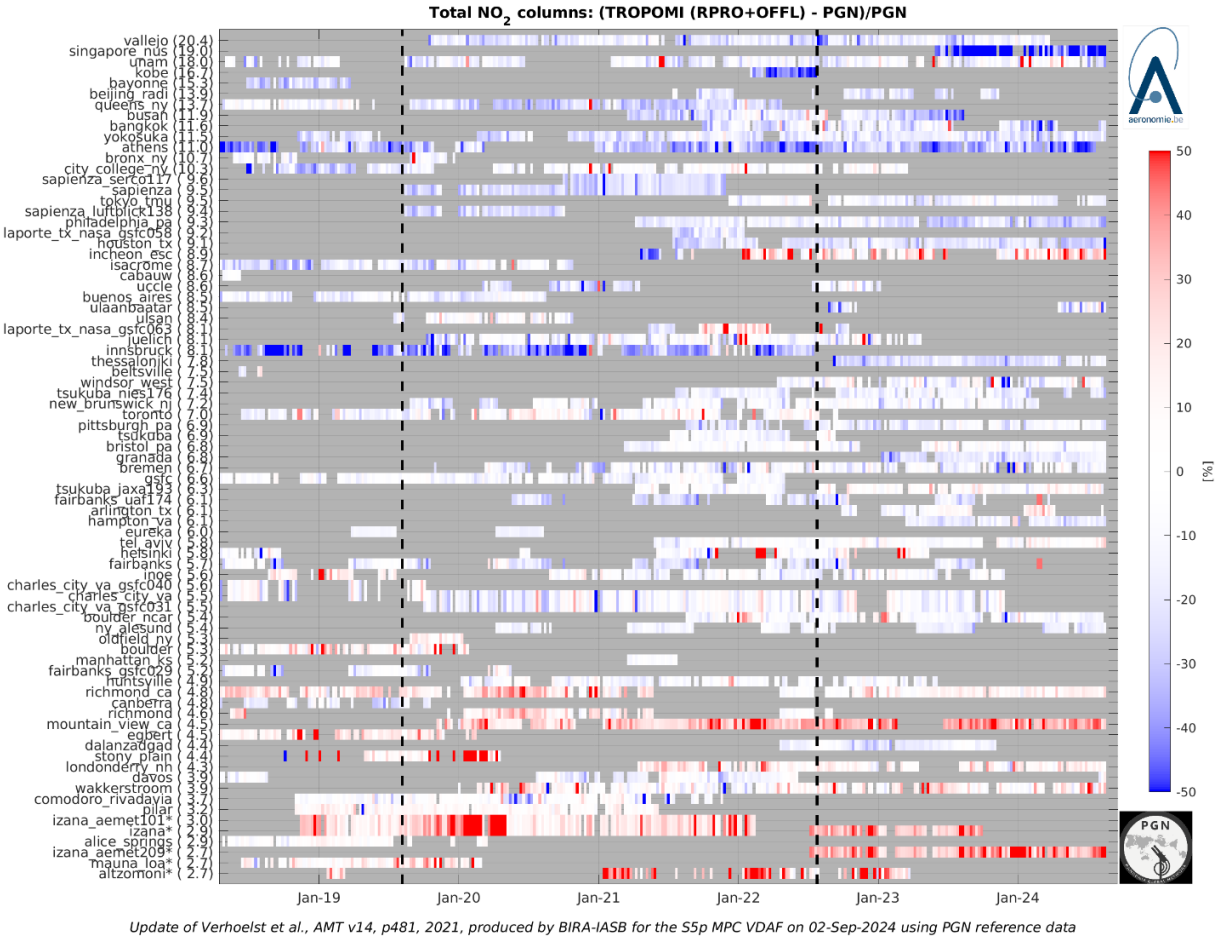


Figure 49: Time series – from May 2018 until August 2024 – of S5P RPRO+OFFL and PGN Pandora NO_2 total column differences, weekly averaged [%]. The black dashed lines indicate, respectively, the activation of the finer horizontal resolution and the switch between the RPRO and operational OFFL data. Stations are ordered by median total column, in brackets.

7.3.5.2 Dispersion

The overall dispersion from reprocessed V2.4.0 and OFFL V2.4.0-2.6.0 data is 1.6 Pmolec/cm^2 for the 72 Pandora instruments at AVS. The dispersion of the S5P and PGN Pandora differences depends strongly on the station. Small dispersions (IP68/2) are observed at Eureka, Ny-Ålesund, Alice Springs, Mauna Loa, Pilar, Izaña, Dalanzadgad, Comodoro Rivadavia, Davos (0.3 - 0.6 Pmolec/cm^2) that are within the mission precision requirement, and higher values elsewhere (e.g., 2 - 5 Pmolec/cm^2 at New York Bronx, Sapienza Rome, New York City College, Unam). The per-station median of the IP68/2 is 1.5 Pmolec/cm^2. The mean Pearson-R is 0.69 and varies from relatively low (e.g., 0.42 at Fairbanks) to high (0.91 at Yokosuka).

7.4 Equivalence of L2_NO2 NRTI and OFFL products

This section shows evidence that the L2_NO2 NRTI and OFFL products do not differ significantly and that their respective validations yield similar conclusions. We show the differences between the two datasets for the tropospheric and total column data.

7.4.1 Tropospheric NO2 Column NRTI vs OFFL

To demonstrate the closeness of the NRTI and OFFL L2_NO2 products at MAX-DOAS stations, L2_NO2 NRTI and L2_NO2 OFFL, each co-located with MAX-DOAS, were obtained from the validation server, and the subset of pixels, common to both NRTI and OFFL, was determined. Statistical results for Bremen and Mainz are summarized in **Table 6**.

Table 6 Statistics on the comparison of the common subset of L2_NO2 NRTI, L2_NO2 RPRO+OFFL and co-located MAX-DOAS, for the stations Bremen and De Bilt (*: unit of Pmolec/cm2).

Bremen: 294 common co-locations Orbits range from 24758 (2022-07-24) to 35441 (2024-08-15).			
	NRTI-OFFL	NRTI-MXD	OFFL-MXD
Mean(diff) ±sem*	-0.16	-1.44±0.14	-1.29±0.14
Median(diff)*	-0.09	-0.81	-0.74
Std(diff)*	0.5	2.4	2.4
1/2 IP68(diff)*	0.3	2.0	1.8
Pearson R	0.97	0.65	0.67
Slope	0.96	0.37	0.39
De Bilt: 313 common co-locations Orbits range from 24701 (2022-07-20) to 35527 (2024-08-21).			
	NRTI-OFFL	NRTI-MXD	OFFL-MXD
Mean(diff) ±sem*	-0.15	-1.33±0.15	-1.18±0.15
Median(diff)*	-0.13	-0.92	-0.80
Std(diff)*	0.4	2.7	2.6
1/2 IP68(diff)*	0.3	1.9	1.8
Pearson R	0.99	0.67	0.69
Slope	1.00	0.53	0.54

The mean difference between NRTI and OFFL is of the same order or smaller as the standard error on the mean difference of NRTI-MAX-DOAS and OFFL-MAX-DOAS. Therefore, the bias difference between NRTI and OFFL is statistically not significant. Also, the difference dispersion between NRTI and OFFL is small compared to the difference dispersion between either NRTI or OFFL on the one hand and MAX-DOAS on the other hand. The good match between NRTI and OFFL is also demonstrated by the high Pearson R values and the near unity slope of the linear regression.

Daily tropospheric columns from NRTI and OFFL V2.6.0 data streams are binned to 0.1° grid cells and differentiated (Figure 50) to inspect the spatial difference distribution. Data correlates well with a Pearson correlation coefficient of 0.99 and a global relative mean difference of 0.13 % (2024/08/15). Differences above $\pm$0.5 Pmolec/cm^2 are rarely found. NRTI columns tend to be lower in the Northern hemisphere (North America, Europe, China) above 35°N.

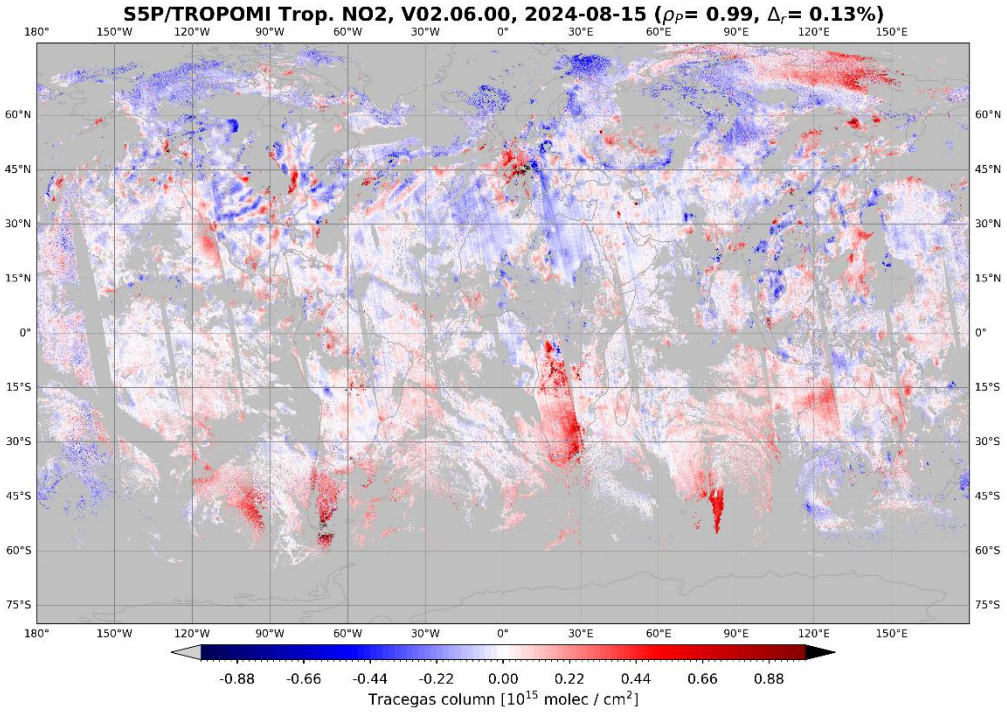


Figure 50: Difference between S5P NRTI and OFFL V2.6.0 daily tropospheric NO_2 column data on 2024/08/15. The difference is calculated for columns > 0.5 Pmolec/cm^2, cloud fractions < 0.2 and a `qa_value` > 0.75. Data was binned on 0.1° x 0.1° grid cells.

7.4.2 Total NO_2 Column NRTI vs OFFL

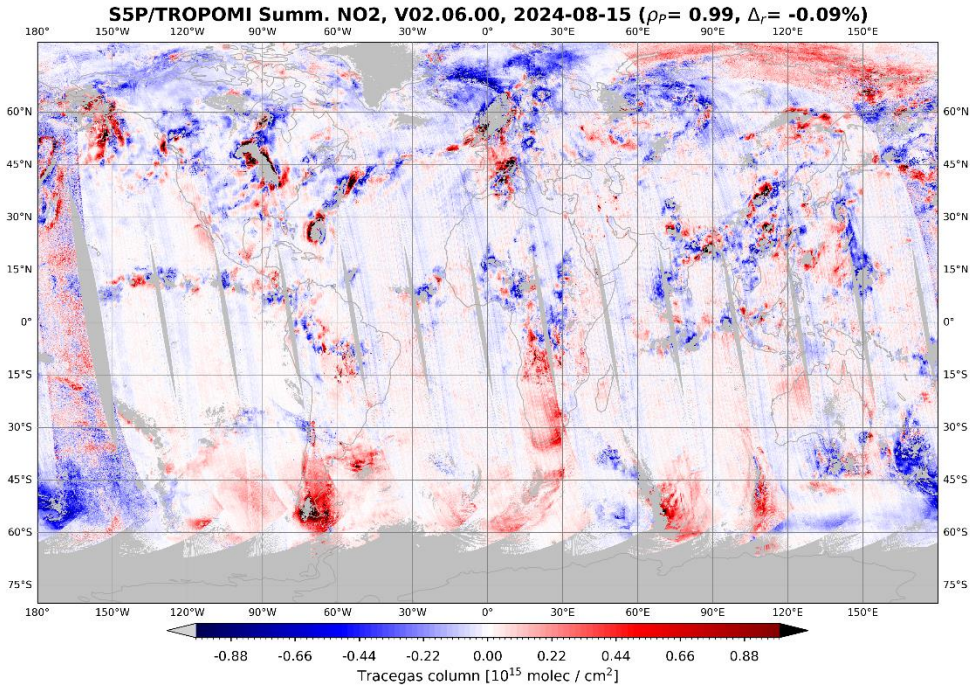


Figure 51: Difference between S5P NRTI and OFFL V2.6.0 daily total NO_2 column data for 2024/05/25. The difference is only calculated for columns > 0.5 Pmolec/cm^2, cloud fractions < 0.6 and a `qa_value` > 0.5. Data was binned on 0.1° x 0.1° grid cells.

The global relative total column difference is in the range of ±0.2 Pmolec/cm² (2024/08/15). On average, NRTI summed NO_2 columns are very similar (-0.1 %), with a Pearson correlation coefficient of 0.99. The comparison as previously analysed for the stratospheric and tropospheric columns reveals that most of the features seen in the sub-columns compensate each other in the total columns, indicating a different troposphere-stratosphere separation (**Figure 51**).

7.5 Internal consistency of the NO_2 validation results

This section focuses on the internal consistency checks of the NO_2 validation results. This relies on analysis of potential source of inconsistencies (such as impact of different NO_2 cross-sections) or synergistic analysis of several instruments types.

7.5.1 NO_2 absorption cross-sections

A potential source of inconsistencies between the different data products lies in the NO_2 absorption cross sections (and their dependence on temperature) that are used in the DOAS retrieval of the slant column density (SCD). The TROPOMI NO2 retrieval explicitly accounts for these temperature effects by using the co-located temperature profiles in troposphere and stratosphere from the ECMWF meteorological analyses. An overview of the different NO_2 cross sections choices made for each instrument is provided in **Table 7**, reproduced from Verhoelst et al. (2021). For a detailed discussion we refer to this work. The main conclusions are:

- A small (few percent) seasonal cycle in the stratospheric column comparisons can be expected, due to the seasonal variation in stratospheric temperature not being accounted for in the ZSL-DOAS data processing.
- PGN columns may either overestimate by up to 10% when the column is mostly stratospheric or underestimate by a similar order of magnitude when large tropospheric amounts are present, due to the use of a fixed effective temperature of 254.4 K. This is addressed with variable effective temperatures in the v1.8 processing that is currently being implemented across the network.
- The MAX-DOAS data may be biased in either direction by a few percent when tropospheric and/or stratospheric temperatures differ strongly from the 298 K and 220 K default temperatures.

Table 7 - NO2 cross section source and temperature for the different instruments (Verhoelst et al., 2021).

Instrument	reference	temperature	comments
S5p TROPOMI	Vandaele et al. (1998)	220K	With temperature correction in AMF (Zara et al., 2017)
ZSL-DOAS	Vandaele et al. (1998)	220K	
ZSL-DOAS	Harder et al. (1997)	227K	NIWA instruments
MAX-DOAS	Vandaele et al. (1996)	298K	tropospheric retrieval only
MAX-DOAS	Vandaele et al. (1998)	298K and 220K	Orthogonalized following Peters et al. (2017)
PGN	Vandaele et al. (1998)	254.4K	PGN processor v1.7

7.5.2 ZSL-DOAS and PGN with low pollution level (mountain-top, arctic)

Three of the PGN direct-sun instruments are located near the summit of a volcanic peak: Altzomoni (3985m a.m.s.l) in the State of Mexico, Izaña (2360m a.m.s.l.) on Mount Teide on the island of Tenerife, and Mauna Loa (4169m a.m.s.l.) on the island of Hawaii. At these high-altitude stations, the total column measured by the ground-based direct-sun instrument misses most of the tropospheric (potentially polluted) part and as such becomes representative of the TROPOMI stratospheric column (with a minor free tropospheric column part). These stations have been used in Verhoelst et al. (2021) for the stratospheric comparison. As for the zenith-sky data, a minor negative median difference (TROPOMI-GB) of the order of -0.2 Pmolec/cm2 was detected.

At Arctic stations (Eureka, Ny-Ålesund), where the tropospheric contribution to the column is expected to be small, both ZSL-DOAS and PGN instruments are located. All instruments follow the temporal evolution of S5P NO_2 rather well. However, there is a clear negative bias (TROPOMI lower than PGN) of about -0.8 Pmolec/cm^2 or -15% for the two PGN instruments, and a much smaller to no negative bias for the ZSL-DOAS instruments (no bias at Eureka, -0.4 Pmolec/cm^2 at Ny-Ålesund when considering the same time period). Also here, it is expected that the upcoming PGN release will reduce this discrepancy.

7.5.3 Stations with multiple instruments (different geometries)

There are a number of stations that have several instruments, covering different viewing geometries, and thus allowing an investigation of the internal consistency of the validation results. Direct-sun and MAX-DOAS instruments measure total and tropospheric NO_2 columns, respectively. By subtracting the TROPOMI stratospheric VCD from the direct-sun total columns, an estimation of the tropospheric NO_2 can be obtained and compared to MAX-DOAS results (Pinardi et al., 2020). Past comparisons (e.g., Figure 4 of Pinardi et al., 2020) point to a good consistency, with high correlations and biases of 10 to 15%. Current stations with these two types of instruments are Athens, Thessaloniki, Xianghe, Uccle, Yokosuka and Unam.

Preliminary comparisons have been done, but there is insufficient data for consolidated conclusions so far, due to impact of the seasonality, short overlapping periods and the need to consider differences in horizontal and vertical NO_2 representatively.

7.5.4 Consistency between MAX-DOAS and PGN network results

Another way to explore the consistency of the validation results is to compare the results at the network level, by comparing TROPOMI tropospheric columns to the MAX-DOAS values on one hand (cf. Section 6.3.3) and on the other hand to the calculated PGN-trop values (PGN total columns minus collocated TROPOMI stratospheric columns). This is an interesting comparison, even considering the different sensitivity (vertical kernels) of MAX-DOAS and PGN-trop.

Figure 51 represents the box-whisker plots for these PGN-trop comparisons. Like for the MAX-DOAS comparisons, a small (in absolute terms) positive bias is seen for clean stations and a negative bias for more polluted sites. The transition between these regimes appears to occur at slightly larger columns in the PGN-trop comparisons (near 3 Pmolec/cm^2 instead of near 2 Pmolec/cm^2). Some stations hosting both a MAX-DOAS and a Pandora instrument present some apparent inconsistencies. For instance, the Unam PGN-trop median value (15.6 Pmolec/cm^2) is quite below the MAX-DOAS one (between 19.9 and 21.1 Pmolec/cm^2, depending on the viewing directions), and the Athens PGN-trop median value (7.8 Pmolec/cm^2) is higher than the MAXDOAS one (3.7 Pmolec/cm^2). In the latter case however the 2 instruments are not located at the same station in Athens, with the MAX-DOAS on top of a hill, missing the urban boundary layer contribution. As discussed in the previous subsection, more analysis should be done (keeping only common time periods, exploring the line of sight influence and the impact of the clouds, on the comparisons, in addition to comparing the GB datasets themselves before any TROPOMI collocation.

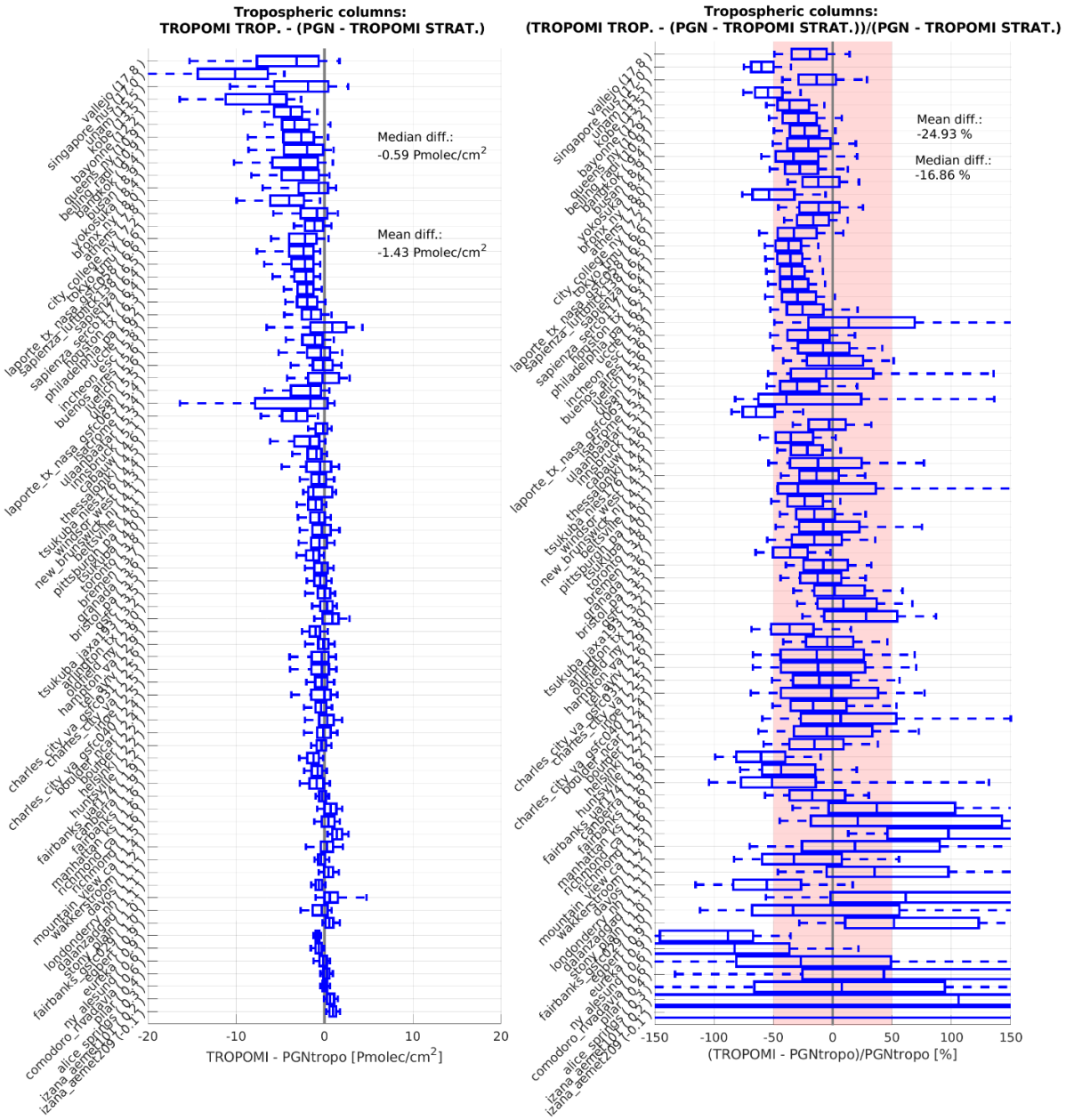


Figure 52: Box-and-whisker plots summarizing the bias and spread of the difference between S5P TROPOMI RPRO+OFFL v02.04/02.08 and PGN-trop NO2 tropospheric columns (i.e., PGN total columns minus the S5P stratospheric columns). Conventions of the boxplots are identical to **Figure 35**. Stations are ordered by median tropospheric column. The period is May 2018 to August 2024.

8 Validation Results: L2_HCHO

8.1 L2_ HCHO products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_HCHO product identified in **Table 1**, the HCHO total column. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. As the NRTI and OFFL processors are producing very similar data products, mainly the validation of the L2_HCHO OFFL product is reported hereafter. Subsection 8.4 shows evidence that NRTI and OFFL data do not differ significantly and that their respective validations yield similar conclusions.

The operational (E2) phase for the S5P TROPOMI mission starts with orbit #02818 on April 30, 2018. L2_HCHO NRTI and OFFL product version 02.04.01 was released July 17, 2022 with reprocessed data to version 02.04.01 covering the period of E2 until end of July 2022. The current version 02.06.01 was activated on 2023-11-26 (OFFL) and 2023-11-29 (NRTI). Although there was no change in the HCHO processor itself with version 02.06.01, the CLOUD product (which is used as input) received a significant update in that version.

8.2 Validation approach

8.2.1 Ground-based monitoring networks

S5P L2_HCHO data are routinely validated through comparisons with respect to ground-based measurements acquired by NDACC FTIR and MAX-DOAS UV-visible instruments performing network operation in the framework of NDACC. For S5P validation purposes, those measurements are collected either automatically through EVDC or manually through S5PVT AO projects with faster data delivery (e.g., CESAR AO ID 28596 and NIDFORVAL AO ID 208607).

8.2.1.1 Fourier Transform Infrared Spectrometers

TROPOMI L2_HCHO formaldehyde column data are compared to reference measurements acquired at NDACC FTIR stations. FTIR measurements have a median systematic uncertainty of 13% and a median random uncertainty of 0.3 Pmolec/cm^2 (Vigouroux et al., 2018). The vertical sensitivity of FTIR is similar to that of S5P HCHO, with lower sensitivity close to the surface.

The comparison methodology is described in Vigouroux et al. (2020). Here we only give a brief outline:

- S5P pixels are selected within 20 km of the FTIR station (about 30-40 pixels). Only pixels with a `qa_value` > 0.5 are used. A collocation pair is only kept if at least 10 pixels can be averaged.
- The time coincidence criterion is set to ±3 hours of the satellite overpass time.
- The following data manipulations are performed: (i) The FTIR a priori profile is substituted with the TROPOMI L2_HCHO one to get a corrected FTIR profile. (ii) The corrected profile is smoothed with the TROPOMI averaging kernel (Rodgers and Connor, 2003). (iii) Scaling is applied to take into account altitude differences between pixel level and station altitude. (iv) Both the individual manipulated FTIR columns and the individual S5P manipulated pixel columns are then averaged.
- The relative median bias at a single station is estimated by the median relative difference: MED[(SAT-REF)/REF]. Absolute-scale dispersion is estimated by the scaled median absolute deviation from the median (MAD): 1.4826*MED[ABS(DIFF-MED(DIFF))]. The scaling factor of 1.4826 ensures that for a normal distribution, the MAD is equal to the standard deviation.

8.2.1.2 **Pandora Direct-S*un* UV-Visible Spectrometers**

TROPOMI L2_HCHO column data are routinely compared to reference measurements acquired by Pandora instruments. They perform network operation in the context of the Pandonia Global Network (PGN).

8.2.1.3 MAX-DOAS UV-Visible Spectrometers

TROPOMI L2_HCHO formaldehyde column data are routinely compared to reference measurements acquired by MAX-DOAS UV-Visible spectrometers. MAX-DOAS HCHO column data have a maximum bias 20% with a precision better than 30%. The MAX-DOAS vertical sensitivity differs from the S5P HCHO sensitivity. While MAX-DOAS has a higher sensitivity close to the surface and a lower sensitivity at higher altitudes, the reverse is true for S5P HCHO. Currently two channels are used to acquire MAX-DOAS data and perform the comparisons, each with their own comparison methodology:

- VDAF Automated Validation server: The S5p pixels are kept for `qa_value` ≥ 0.5. It covers the MAX-DOAS measurement location, and is within ±0.5 h of the MAX-DOAS measurement. All MAX-DOAS measurements within ±0.5 h of the satellite overpass are averaged. No a priori substitution or averaging kernel is applied.
- NIDFORVAL AO project: S5P pixels with a `qa_value` ≥ 0.5 are kept. The average of S5P pixels within 20 km radius is compared with the average of MAX-DOAS measurements within ±3 h of satellite overpass. Note that these are the same co-location criteria as for the FTIR comparisons. For stations that also deliver an averaging kernel, a priori substitution followed by averaging kernel smoothing (Rodgers and Connor, 2003) is optionally applied. Relative bias and absolute-scale dispersion are calculated as for the validation based on FTIR data.

8.2.1.4 Mutual consistency of the FTIR and MAX-DOAS ground-based data

The Xianghe station (China, 39.75° N, 116.96°E) is one of the few stations where both FTIR and MAX-DOAS instruments measure in parallel with a current overlap of about 2 years. In addition, there are also direct-sun measurements from the BIRA MAXDOAS (a bit more than 1-year overlap), plans to install a Pandora and a second MAXDOAS instrument was operated by USTC. It is thus an excellent candidate to test the consistency of the two techniques in a polluted station. Summary results (**Figure 53**) show very consistent results between direct-sun and FTIR data, a bit more spread for the direct-sun vs MAX-DOAS data (with larger MAX-DOAS values in winter time) and good regression but with large negative intercept for the FTIR vs MAX-DOAS original columns.

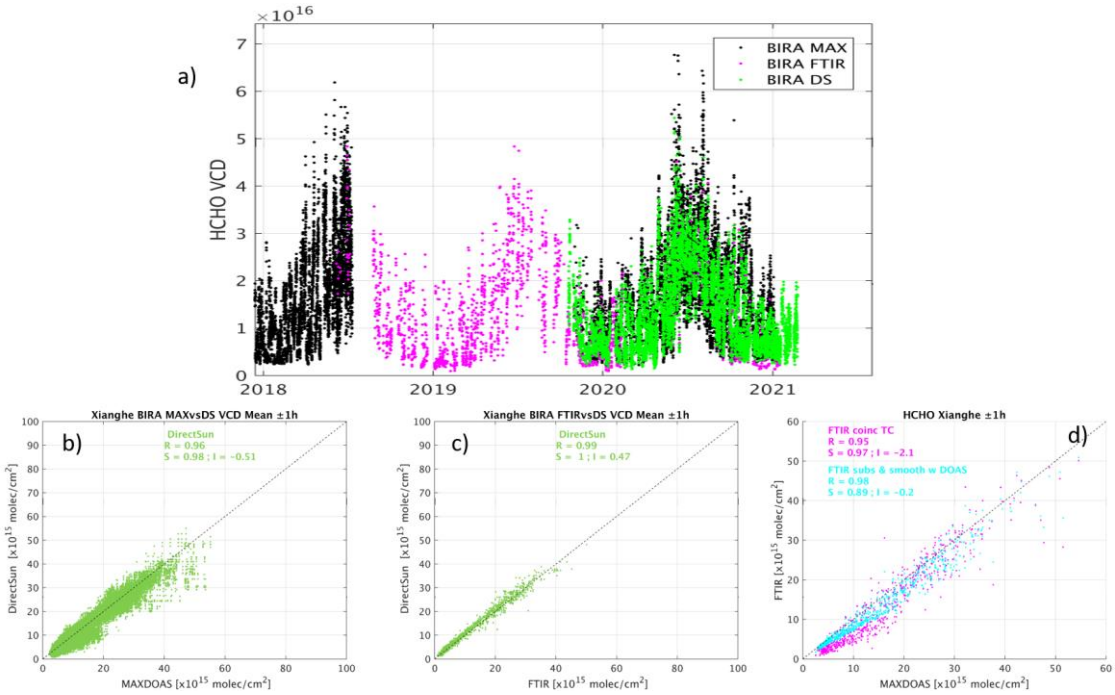


Figure 53: a) Time-series of the MAX-DOAS, FTIR and direct-sun HCHO columns in Xianghe from BIRA-IASB instruments; Scatter-plot between the b) direct-sun and the MAX-DOAS, c) direct-sun and FTIR and d) FTIR and MAX-DOAS for the raw comparisons and when taking into account a-priori profile substitution and smoothing.

When taking into account the MAXDOAS and FTIR sensitivities and using the Rodgers and Connor (2003) methodology with a priori substitution and smoothing, the initial MAX-DOAS vs. FTIR bias of 27 % is reduced to 15 %. A regression analysis shows a reduced slope, but also a much reduced intercept value between the two instruments (from $y = 0.97 x - 2.1*10^{15}$ to $y= 0.89 x - 0.2*10^{15}$). More information can be found in Pinardi et al. (MAXDOAS workshop meeting 2021).

8.2.2 Satellites

TROPOMI L2_HCHO formaldehyde column data are also compared to similar data from the EOS-Aura Ozone Monitoring Instrument (OMI) using the QA4ECV L2 product (http://doi.org/10.18758/71021031).

8.2.3 Field campaigns and modelling support

Nothing to report.

8.3 Validation of L2_ HCHO

8.3.1 Recommendations for data usage

In order to avoid misinterpretation of the data quality, only those TROPOMI pixels associated with a `qa_value` > 0.5 (no error flag, cloud radiance fraction at 340 nm < 0.5, SZA equal to or below 70°, surface albedo smaller than or equal to 0.2, no snow/ice warning, air mass factor > 0.1) have been used as recommended. For further details, including how to apply the averaging kernel and a priori profile in comparisons, data users are encouraged to read the Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, which are available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

8.3.2 Status of validation

This section presents a summary of the validation results obtained with the Validation Data Analysis Facility (VDAF) of the S5P Mission Performance Centre (MPC) and by the S5P Validation Team (S5PVT) AO projects CESAR and NIDFORVAL. Up-to-date validation results and consolidated validation reports are available through the MPC VDAF Portal at http://mpc-vdaf.tropomi.eu.

The status of the FRM data streams is as follows:

- Comparisons with UV-Vis MAX-DOAS. Six MAX-DOAS stations contribute data to the VDAF Automated Validation Server with a temporal coverage of collocations from May 2018 to August 2024.
- Data from eight UV-Vis MAX-DOAS stations are available through the NIDFORVAL AO project. It covers the period from May 2018 to February 2022, dependant on the station.
- Comparisons with NDACC FTIR follow the methodology of Vigouroux et al. (2020). The current number of stations is 29. Note that only 10 sites among the 29 provided data after the last May 2023 update, up to August 2024.
- The FTIR dataset covers the period May 2018 to August 2024 at VDAF (18 stations).
- Comparisons with PGN Pandora from 34 stations (36 instruments) that are available at MPC VDAF. The dataset covers the period May 2018 to August 2024.

8.3.3 Bias

The following results, using ground-based FTIR, MAX-DOAS, Pandora, and OMI satellite data, show that the TROPOMI HCHO bias is usually well within the 40% mission requirements and always within the 80% upper limit. Monthly mean HCHO differences are shown in **Figure 55** with respect to the 3 different reference measurements.

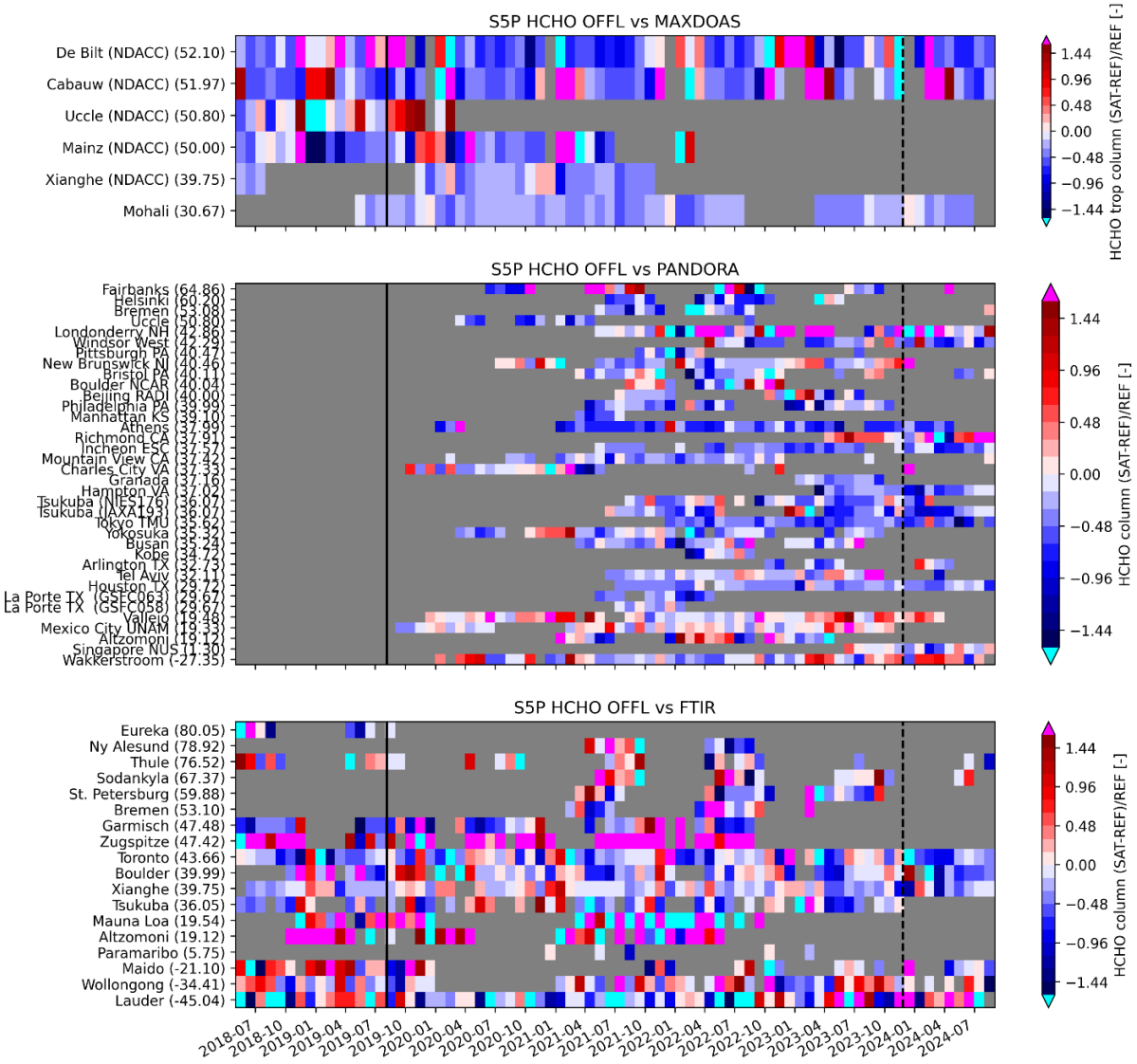


Figure 54: Time series of monthly mean relative differences [%] of S5P OFFL/RPRO V02.04.01-02.06.01 HCHO versus MAXDOAS, PANDORA, and FTIR. Data was obtained from the VDAF Automated Validation Server on 2024/09/02. Indicated are the pixel size switch (2019/08/16) and the update to v2.3.1 (2023/11/26).

8.3.3.1 Fourier Transform Infrared Spectrometers

TROPOMI v02.04.00/02.05.00 shows a positive bias of +30±4 % for clean stations (HCHO <2.5 Pmolec/cm^2) and a negative bias of -30±1 % for high emission stations (>8 Pmolec/cm^2) in comparison to correlative data from 29 NDACC FTIR stations, covering the period from May 2018 to August 2024, as illustrated in **Figure 55**. Details about the applied methodology are described in Vigouroux et al. (2020). Using the robust Theil-Sen estimator to derive slope and intercept of TROPOMI vs FTIR, a constant positive bias of 1.12±0.03 Pmolec/cm^2 and a proportional bias of 0.62±0.02 is obtained (updated from Vigouroux et al. (2020), Fig. 4).

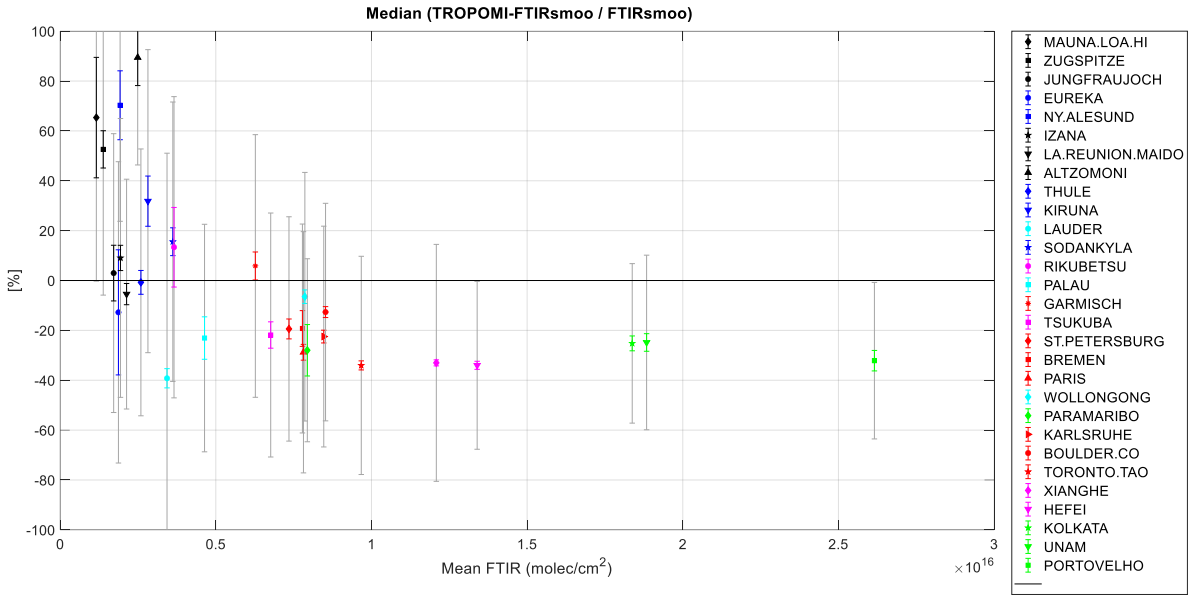


Figure 55: Percent bias between S5P L2_HCHO and NDACC FTIR HCHO column data at each station as a function of the mean FTIR total column (10^{16} molec/cm^2). The grey bars are the systematic uncertainty on the difference, and the coloured error bars are the 2-σ errors on the bias. The temporal range is May 2018 to August 2024 (updated from Vigouroux et al. (2020), Fig. 3).

8.3.3.2 PGN Pandora

The median bias is -33 % (-3.2 Pmolec/cm^2) with a Pearson correlation coefficient 0.32. This is based on measurements from 36 stations (VDAF-AVS, 2024/08/26, #collocations 11516),

8.3.3.3 MAX-DOAS UV-visible Spectrometers

Figure 56 shows difference time series S5P-MAXDOAS for the six stations from the AVS, with indication of the 30-day rolling monthly mean or median. Biases, if they occur, are mostly negative, and within the 80% requirements. Pixel size switch has no obvious impact on the comparison. A bias peak occurs at Mohali around the update to version v2.6.1 (2023/11/26), but this is not necessarily causally related since a similar peak occurs end 2022 at Mohali.

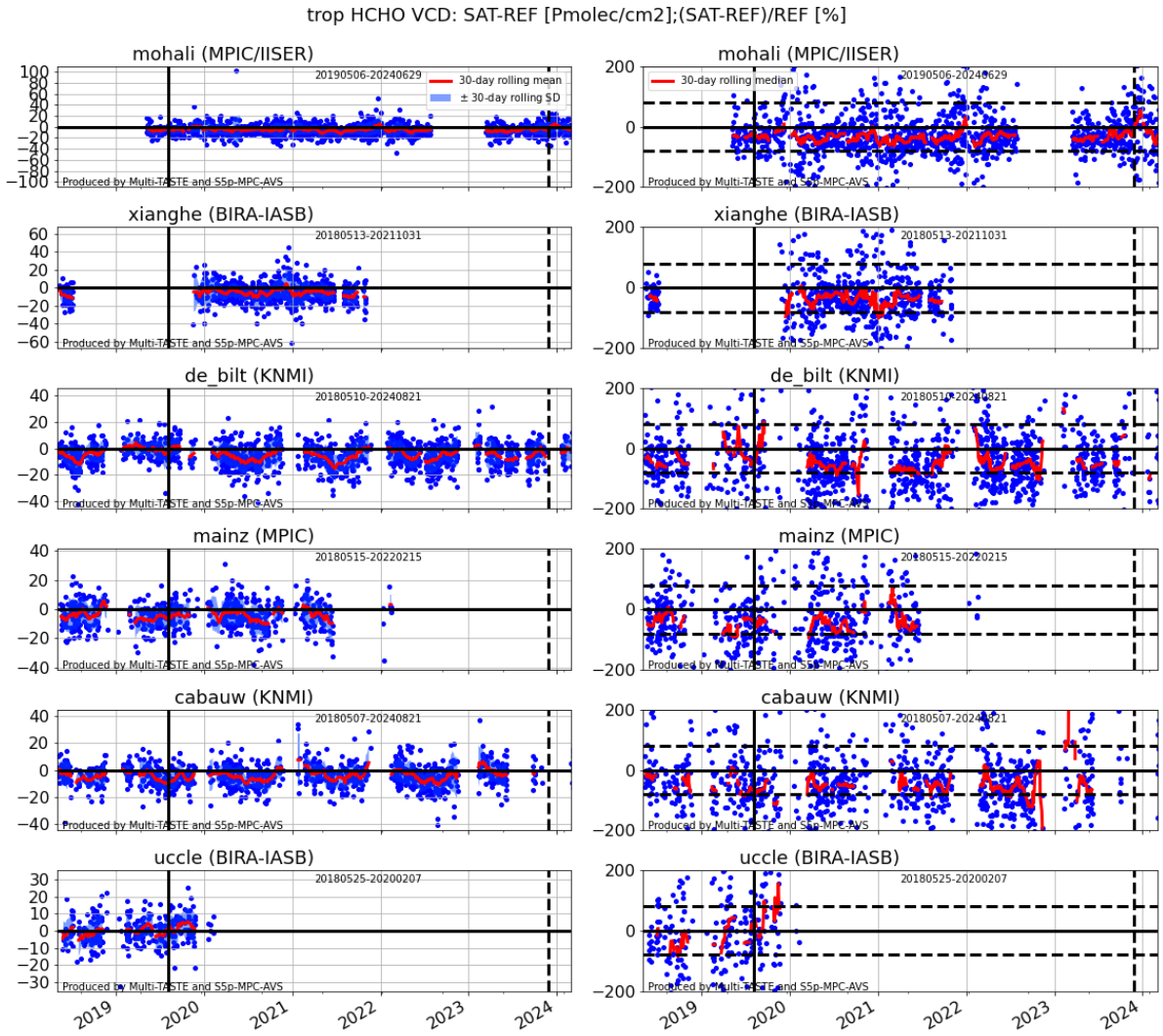


Figure 56: Time series of the difference [Pmolec/cm^2] and relative difference [%] between S5P RPRO+OFFL and MAX-DOAS HCHO column data at six stations. Data was obtained from the VDAF Automated Validation Server on 2024/09/02. Full line: pixel size switch at 2019/08/06. Dashed line: update to v2.6.1 at 2023/11/26.

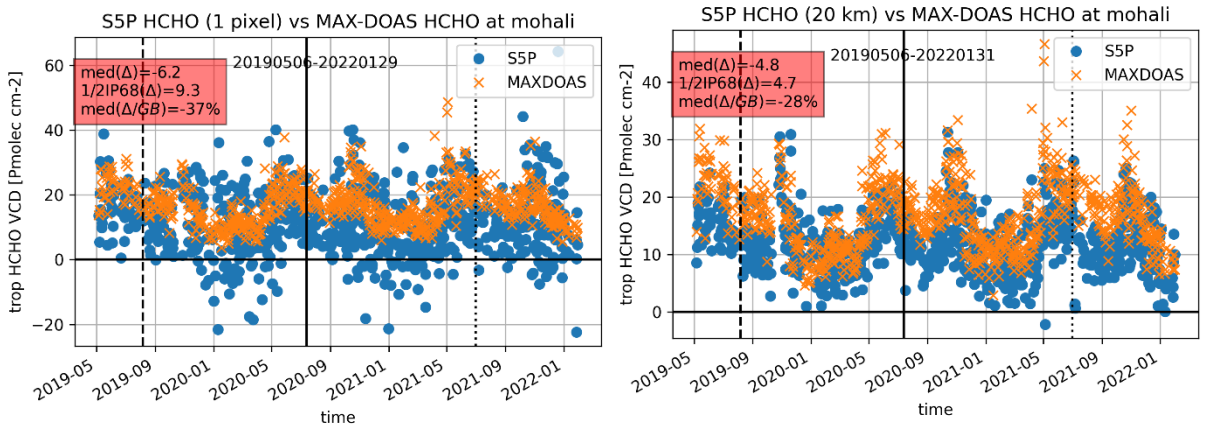


Figure 57: Time series of HCHO columns [Pmolec/cm^2] from S5P RPRO+OFFL and MAX-DOAS data at the Mohali station. (Left) Single S5P pixel, ±0.5h MAX-DOAS averaging, (Right) 20 km area average for ≥10 pixels, ±3h averaging. Data was processed on 2022/02/28.

The VDAF comparisons are done for a single pixel of S5P versus 0.5 h means of MAX-DOAS. This leads to larger scatter. An area averaged 20km column of S5P data in comparison to 3 h MAX-DOAS means are better suited to detect seasonal cycles. An example is shown in **Figure 56** for the station Mohali, where bias, dispersion, and negative values are reduced for the area averaged comparisons.

In the NIDFORVAL AO project, eight stations are used that provide data in the *GEOMS* format as required in the project. **Figure 58** (upper panel) shows that the median bias varies between -5% to -59%, with a median for all stations of -37%. Among these 8 stations, 3 are providing profiles and averaging kernels (AK) which allow to take the difference in a-priori profiles and vertical resolution of the instruments into account (Rodgers and Connor, 2003), as done with FTIR data. This is particularly important for the MAX-DOAS instruments due to different shape of AK compared to TROPOMI as shown in **Figure 59**. For these three stations, the biases improve when using the averaging kernel (**Figure 58**, lower panel).

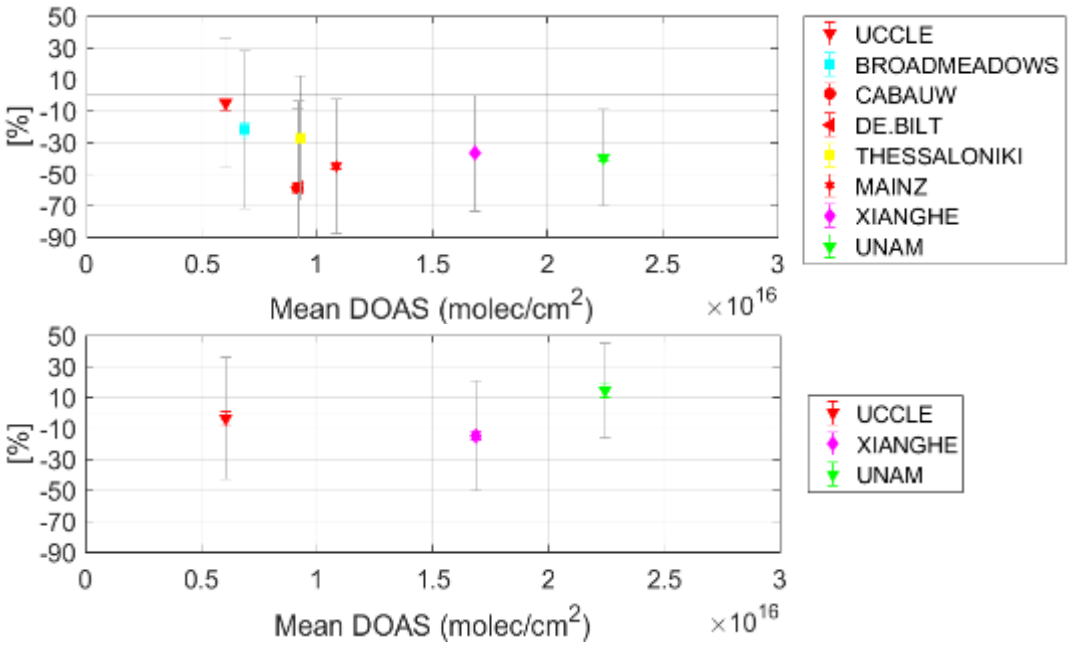


Figure 58: Bias at each station (%) as a function of the mean DOAS total columns (10^{16} molec/cm^2). The grey bars are the systematic uncertainty on the differences, and the coloured error bars are the 2-σ error on the bias. Top panel: normal DOAS data without any modification. Bottom panel: DOAS data after Rodgers and Connor (2003) is applied (a priori substitution and smoothing with the TROPOMI averaging kernels).

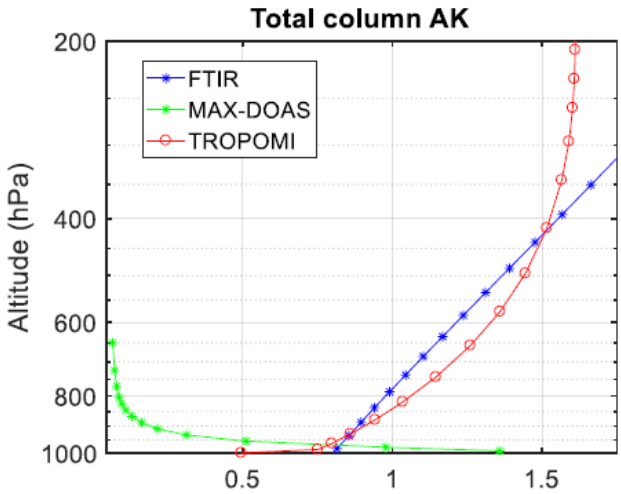


Figure 59: Typical total column averaging kernels for S5P TROPOMI and the two ground-based instrument types: FTIR (blue), MAX-DOAS (green), and TROPOMI (red). This illustrates the problem of the vertical smoothing difference error in these comparisons, as the instruments "see" different parts of the column.

Further separating the biases for low (< 2.5 Pmolec/cm^2) and high (> 8 Pmolec/cm^2) HCHO columns as done for FTIR comparisons, we get for the smoothed DOAS bias of +27±25 % and -10±2 %, respectively. The large uncertainty on the bias for low-HCHO levels is due to the small number of data involved, because the DOAS stations used in this study are not situated in a clean environment (see x-axis in **Figure 58**, > 5 Pmolec/cm^2).

Based on measurements from the 6 MAXDOAS stations available at VDAF-AVS and only using reprocessed V02.04.01 and OFFL v02.05.00 / 02.06.01 data until August 2024 (5455 collocations), the median bias is -41 % (-4.8 Pmolec/cm^2) with a Pearson correlation coefficient of 0.3.

8.3.3.4 Consolidation of FTIR and MAX-DOAS validation results on bias

The FTIR data and the MAX-DOAS data (from two streams: NIDFORVAL and VDAF server) used for the S5P HCHO validation are different in scope (FTIR network having more stations and covering a wider range of HCHO values), harmonization (FTIR network being the more harmonized one), vertical sensitivity (FTIR vertical sensitivity being closer to that of TROPOMI) and uncertainty (FTIR having the smaller systematic error and random error uncertainty). Thus, differences between FTIR and MAX-DOAS validation results can at least be partly attributed to the above-cited factors.

The FTIR network covers very low per-station mean HCHO column values (down to 1.2 Pmolec/cm^2), which is not the case for the MAX-DOAS network (lower bound is at Uccle, which is moderately polluted). To check the consistency of validation results of both networks one should therefore consider a common range of HCHO levels. De Smedt et al. (2021, paper in preparation), using a larger set of NIDFORVAL MAX-DOAS stations, found that in the HCHO column range of 3-6 Pmolec/cm^2, TROPOMI columns do not have a significant bias towards the MAX-DOAS stations, in agreement with the results for FTIR (Vigouroux et al., 2020). Note that the NIDFORVAL MAX-DOAS network does not include mean levels below 3 Pmolec/cm^2.

However, one should consider that the results of De Smedt et al. (2021) are taken for unsmoothed MAX-DOAS columns. In addition, the previous section makes clear that, unexpectedly, the agreement between unsmoothed MAX-DOAS and smoothed FTIR is in fact better than the agreement between smoothed MAX-DOAS and smoothed FTIR. Note however that (i) there are only 3 smoothed MAX-DOAS in use, and (ii) the result is strongly driven by the MAX-DOAS station Unam with a large bias change upon smoothing. More profile MAX-DOAS data is therefore needed to draw strong conclusions. As the FTIR data have the broadest scope, include more stations and are more harmonized, FTIR validation results are provided as representative quality indicator.

8.3.3.5 OMI QA4ECV comparisons

The TROPOMI HCHO algorithm was designed in parallel with the QA4ECV OMI algorithm in order to create a consistent time series of early afternoon observations. The QA4ECV OMI HCHO dataset is now exceeding 17 years (2005-2021), including four years of overlap with TROPOMI, allowing for a meaningful comparison at different scales. As presented in the TROPOMI HCHO ATBD, all retrieval settings have been chosen as similar as possible for the two L2 products, as well as the auxiliary datasets with the important exception of the cloud products (De Smedt et al., 2018; 2021).

While the QA4ECV OMI product is based on the O2–O2 absorption feature around 477 nm, and considers a fixed cloud albedo of 0.8 (version 2.0), the TROPOMI product uses the S5P operational cloud product in CRB (Cloud as Reflecting Boundary) mode (OCRA/ROCINN-CRB). The S5P ROCINN algorithm is based on the O2 A-band around 760 nm and simultaneously retrieves the cloud-top height and cloud albedo. Systematic differences between the cloud parameters will result in differences in the air mass factors, influencing the comparisons. To get around this difference between OMI and TROPOMI, it is advised to replace the cloud-corrected AMFs by clear-sky AMFs (no cloud correction applied). Both types of AMFs are provided in both L2 products.

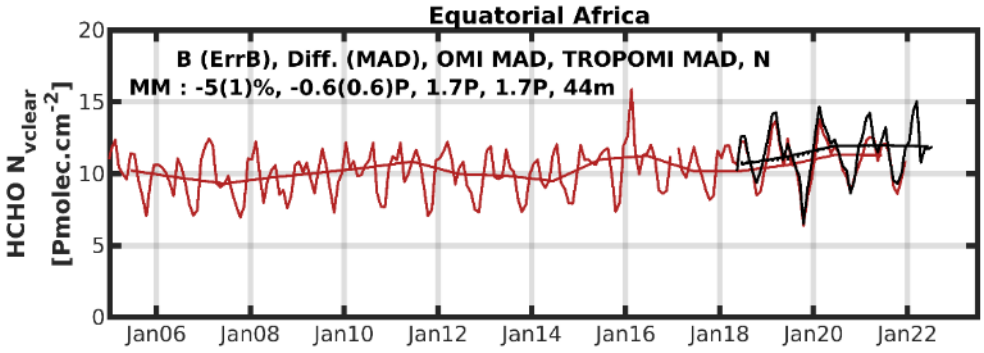


Figure 60: Example of monthly and yearly averaged HCHO columns (Nv_clear) retrieved from OMI (October 2004–December 2021, in red) and TROPOMI collection 3 (2018–March 2023, in black) in a large region of equatorial Africa. Absolute and relative biases between OMI and TROPOMI HCHO monthly averaged columns are given at the top, as are the median deviations of the OMI and TROPOMI averaged columns.

We calculated averaged columns in 35 regions covering a broad range of emission levels and observation conditions. As the regions are large, many observations are included (on average 500/day for OMI, 12500/day for TROPOMI). To obtain daily and monthly comparison pairs, we keep coincident days of observations. An example of the OMI and TROPOMI time series is given in **Figure 60**.

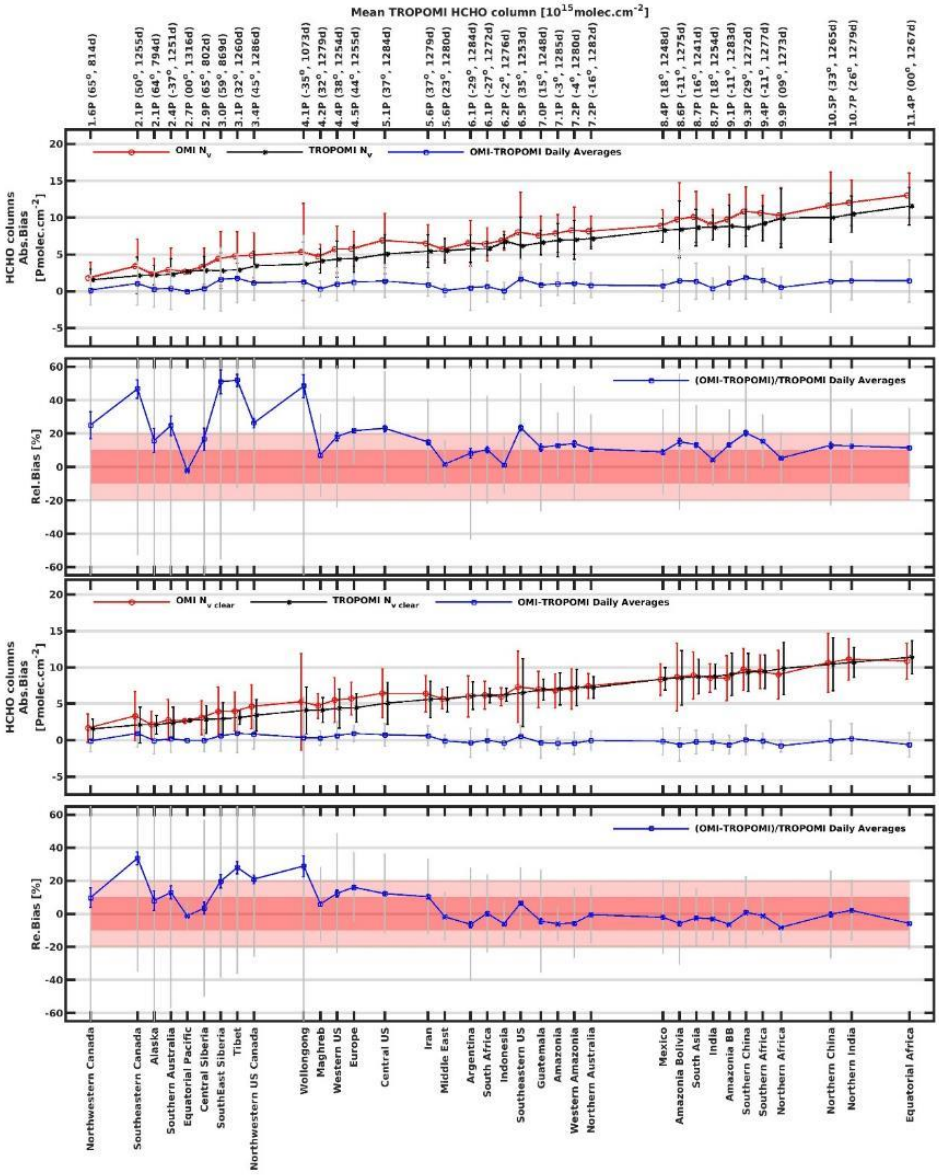


Figure 61: Absolute and relative biases between OMI and TROPOMI HCHO daily averaged tropospheric columns using cloud corrected AMF (N_v, two upper panels) or clear sky AMF (N_v_clear, two bottom panels) for the large regions represented on **Figure 55**. Regions are sorted as a function of the median TROPOMI HCHO column. Values of the averaged HCHO columns are provided on the top axis, as well as the numbers of common days taken for the comparison and the latitude of the region. The median OMI (red) and TROPOMI (black) columns are plotted together with the absolute differences (in blue). Error bars represent the median deviations of the columns, or the median absolute deviations of the differences (in grey). Pink areas indicate 10 % and 20 % bias.

Figure 61 presents the mean bias between OMI and TROPOMI HCHO tropospheric columns for the 35 regions. Numbers are provided for daily averaged columns applying a cloud correction (upper panels) or not (lower panels). As discussed in De Smedt et al. (2021), biases up to 30% related to the cloud correction are observed over Tropical regions where the clouds are the highest in altitude (Africa, South America, South Asia), and a smaller but systematic effect, up to 15%, is observed over mid-latitude polluted regions such as China, India, US or Europe. We note that the differences between N_v and N_v_clear are mainly significant for the OMI HCHO columns.

It has been reported that the cloud pressures retrieved from TROPOMI and from OMI present a bias (OMI clouds are higher in altitude, Compernolle et al., 2020). This translates into OMI cloud-corrected air mass factors generally smaller than TROPOMI AMFs by 5 to 30%, depending on the cloud altitude, and therefore in a positive bias of the OMI HCHO VCD compared to the TROPOMI product. It is therefore important to keep in mind that the use of different cloud products may introduce inconsistencies, which may be resolved by using clear HCHO VCDs (N_v_clear).

When comparing N_v_clear, the biases are strongly reduced below 10% in all regions where the HCHO levels are larger than 5 Pmolec/cm^2, and the TROPOMI columns are found to be slightly larger than OMI on average (-3±1.2%). In mid-Northern-latitudes/moderate emissions (2-5 Pmolec/cm^2) regions such as Europe, Central and Western US, North Western Canada, Siberia or Tibet, OMI columns present a remaining bias of about 15±3%, while in the regions of Canada and Alaska, a larger bias of about +30±7% remains. The mean columns are lower over those regions, and differences in sampling (pixel size and OMI row degradation) start playing a bigger role.

8.3.4 Dispersion

The dispersion is evaluated for several scenarios: single pixel comparisons with MAX-DOAS (from VDAF-AVS), 20-km radius pixel average comparisons with MAX-DOAS (from NIDFORVAL), and 20-km radius pixel average comparisons with FTIR, for all stations and for clean stations only. The dispersion difference obtained for the 20-km radius pixel averages is also recalculated to a theoretical single-pixel value by multiplying with $\sqrt{}$(#pixels). However, one should take into account that this formula assumes that random error is uncorrelated and only originates from the satellite.

8.3.4.1 Fourier Transform Infrared Spectrometers

The median absolute deviation (MAD) remains close to the mission requirement of 12 Pmolec/cm^2, using data from the latest update of the NIDFORVAL project (29 FTIR stations covering the period May 2018 – August 2024). In this work, we do not use a single TROPOMI pixel (as in the MPC Automated Validation Server) but an average of about 30 TROPOMI pixels (20 km around the station). The MAD for the 29 stations taken together is 3.04 Pmolec/cm^2, corresponding to a theoretical single-pixel dispersion of 3.04×$\sqrt{}$(#pixels) = 19 Pmolec/cm^2, slightly above the 12 Pmolec/cm^2 requirement. The MAD is slightly higher than with the previous TROPOMI versions (2.85 Pmolec/cm2), but this is due to the fact that we measure the dispersion with the median, therefore reducing the effect of outliers. If we compare the standard deviations instead, we see a nice reduction from 4.3 to 3.9 Pmolec/cm^2 for V02.04.01-02.06.01, which is due to the fact that obvious outliers are filtered in this version.

However, to evaluate the TROPOMI precision, it is more relevant to compare the MAD obtained at clean stations only because MAD is less sensitive to the additional collocation error in regions far from emissions. At clean conditions, the TROPOMI precision is much better than the pre-launch requirements: 1.52 Pmolec/cm^2 for the 35-pixels-average, corresponding to a theoretical single pixel precision of 9 Pmolec/cm^2. These updated results are very similar to results in Vigouroux et al. (2020). We note that, while the pre-launch requirements are reached, and while the provided TROPOMI random uncertainty agree with the estimated single-pixel precision, the MAD of the 20km-averaged-pixels at clean stations is about 1.6 to 2 times larger than the random uncertainty budget provided in the TROPOMI files, pointing to a possibly too optimistic TROPOMI random uncertainty budget (Vigouroux et al., 2020).

8.3.4.2 PGN Pandora

The dispersion of 9.6 Pmolec/cm^2 is similar to the other instruments and within mission requirements.

8.3.4.3 MAX-DOAS UV-visible Spectrometers

The MAD of the difference of S5P (single pixel) with respect to MAX-DOAS ranges from 8 Pmolec/cm^2 at Uccle to 10 Pmolec/cm^2 at Xianghe. This is within the mission requirement of precision of 12 Pmolec/cm^2. Using the 20km-averaged-pixels within NIDFORVAL (~about 42 pixels here), as done for FTIR and not applying vertical smoothing (Rodgers and Connor, 2003), we obtain a MAD of 3.0 and 2.9 Pmolec/cm^2 at Cabauw and De Bilt, respectively. This corresponds to a single pixel precision of 19-20 Pmolec/cm^2, which is twice larger than the pre-launch requirement of precision. However, if we look at the cleanest DOAS station Uccle to avoid larger collocation errors, the MAD is 2.6 Pmolec/cm^2 for the 34-pixels average comparisons, leading to a single pixel precision of 14 Pmolec/cm^2. If we apply the Rodgers and Connor (2003) technique, the MAD between TROPOMI and the $DOAS_{smoo}$ data is reduced at all the five stations, except at Uccle where the smoothing has little effect (**Figure 58**), leading to a single pixel precision of 15 Pmolec/cm^2 there.

In summary, the dispersion of the difference (8 to 10 Pmolec/cm^2, mainly polluted stations) is already within the dispersion requirement of 12 Pmolec/cm^2 for single-pixel comparisons with MAX-DOAS. Even lower dispersions (2 to 4 Pmolec/cm^2) are obtained when using 20-km averaged pixels (NIDFORVAL FTIR, NIDFORVAL MAX-DOAS), as the random error is reduced. However, recalculating to a theoretical single pixel dispersion by multiplying with √(#pixels), the dispersion requirement is now slightly (FTIR) and strongly (MAX-DOAS) higher than the dispersion requirement. But comparison error and FRM random error in the case of MAX-DOAS make an important contribution at polluted stations. The theoretical single pixel dispersion (~8 Pmolec/cm^2) at clean FTIR stations is within the dispersion requirement, but 1.6 times larger than the random uncertainty budget provided in the TROPOMI files.

8.3.4.4 Consolidation of FTIR and MAX-DOAS validation results on dispersion

The mission requirement for S5P HCHO on dispersion is 12 Pmolec/cm^2. The number is valid at a single-pixel level. The single-pixel comparisons with MAX-DOAS from the VDAF validation server (mainly polluted stations) have a single-pixel dispersion of the difference of 8 to 10 Pmolec/cm^2. This is only an upper bound to the S5P HCHO dispersion, as there are also contributions from MAX-DOAS random error and from comparison error. Nonetheless, the result is already within the dispersion requirement of 12 Pmolec/cm^2, confirming that at single-pixel level, the dispersion requirement is met.

Even lower dispersions (2 to 4 Pmolec/cm^2) are obtained when using 20-km averaged pixels (NIDFORVAL FTIR, NIDFORVAL MAX-DOAS), as the random error is reduced. However, recalculating to a theoretical single pixel dispersion by multiplying with √(#pixels), the dispersion requirement is now slightly (FTIR) and strongly (MAX-DOAS) higher than the dispersion requirement. But comparison error and (in the case of MAX-DOAS) FRM random error make an important contribution at polluted stations. It is therefore more appropriate to focus on clean FTIR stations. The theoretical single pixel dispersion (~8 Pmolec/cm^2) at clean FTIR stations is within the dispersion requirement, but 1.6 times larger than the random uncertainty budget provided in the TROPOMI files. We can thus conclude that both at single-pixel level and for 20-km averaged pixel areas, the dispersion requirement is met.

8.3.4.5 OMI QA4ECV Data Record

For individual pixels, the standard deviation of individual OMI and TROPOMI observations in remote regions with no local emissions is about 7 and 5 Pmolec/cm^2, respectively. When averaging data over large regions, the dispersion due to random uncertainties is greatly reduced compared to individual observations. As summarized in **Table 8**, the median absolute deviations of the monthly averaged columns are equivalent for OMI and TROPOMI (1.8 Pmolec/cm^2), while the median deviations of their differences are significantly lower (0.5 Pmolec/cm^2). This indicates that at this spatiotemporal resolution, the natural variability dominates the dispersion of the averaged observations. Looking at the daily averaged columns, the TROPOMI median deviation is lower than for OMI (2.2/2.7), but still larger than the median deviation of their differences (1.5). Note that these estimates still include the natural variability of the columns themselves. If an in the remote Equatorial Pacific is considered, the observations represent constant background values and the seasonal variability is further reduced. In such conditions, the dispersion of the OMI daily observations is 3.5 Pmolec/cm^2, while only 1 Pmolec/cm^2 for TROPOMI.

Low dispersion is related to the large number of observations included in the averages. The frequent occurrence of extreme outliers advocates the use of the median difference as a quality indicator instead of the mean difference.

Table 8: Median absolute deviation of the OMI and TROPOMI daily and monthly averaged columns (N_{v_clear}), in large regions and in 20km-radius area. Median absolute deviations (MAD) of differences between OMI and TROPOMI columns are also given in the last column.

Dispersion	OMI MAD [10^{15} molec.cm^{-2}]	TROPOMI MAD [10^{15} molec.cm^{-2}]	OMI-TROPOMI MAD [10^{15} molec.cm^{-2}]
Monthly Regional	1.8	1.8	0.5
Daily Regional	2.7	2.2	1.6
Monthly 20km	3.3	2.5	2.4
Daily 20km	7.8	3.7	7.1
Daily 20km in the Equatorial Pacific	3.5	1.0	3.7

8.3.5 Dependence on influence quantities

None to report.

8.3.6 Seasonal and short term variability

Fourier Transform Infrared Spectrometers

The seasonal variability captured by TROPOMI is similar to the one reported by FTIR. As an illustration, **Figure 62** shows as an example the HCHO time series at Boulder. The Pearson correlation between the TROPOMI and FTIR monthly means for all 29 stations is 0.92.

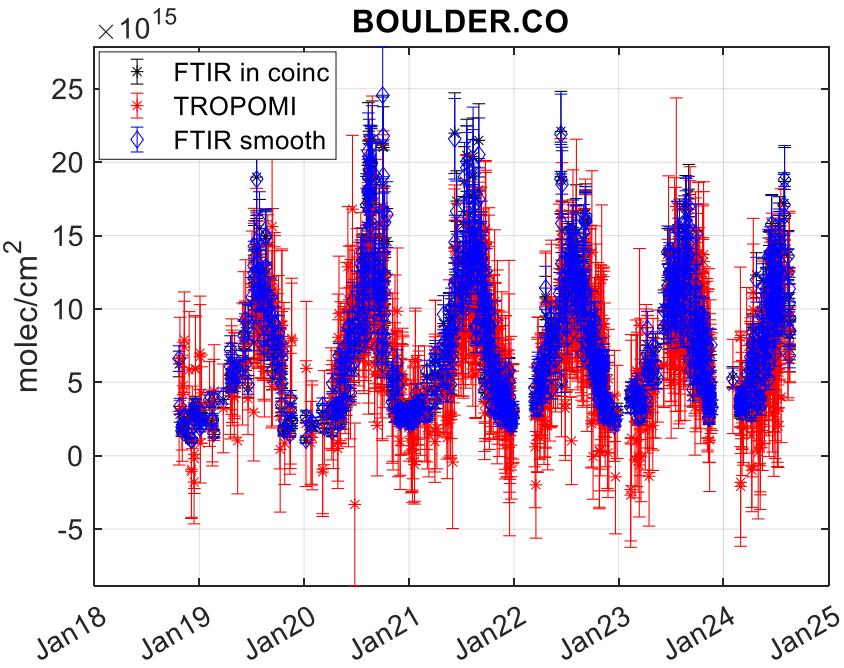


Figure 62: TROPOMI and FTIR HCHO time series at Boulder. Period is May 2018 to August 2024.

MAX-DOAS UV-Visible Spectrometers

The comparisons of TROPOMI and NIDFORVAL MAX-DOAS HCHO data show a monthly mean correlation of 0.86 for 8 stations, and 0.88 after smoothing for 3 stations providing averaging kernels. But it varies strongly from 0.96 (Xianghe) to 0.55 (UNAM, Mexico City). **Figure 63** shows two stations with measurements in 2022, Mohali and Chiba. The Chiba station is not in GEOMS format, but the CHIBA group format is foreseen to be added in the validation server in the near future.

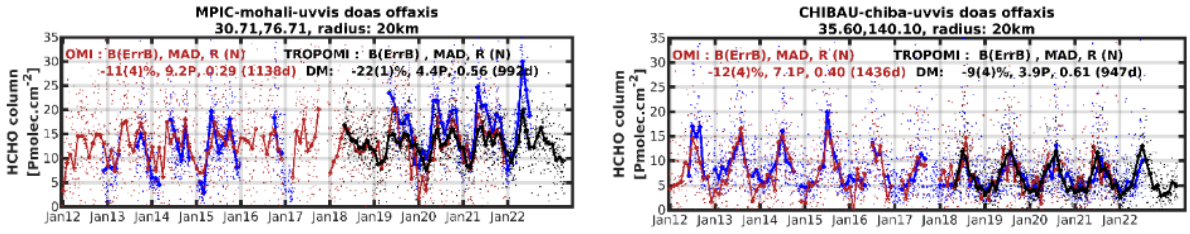


Figure 63: *OMI, TROPOMI and MAX-DOAS HCHO time series at Mohali and Chiba. Time frame is Jan 2012 to May 2023.*

OMI QA4ECV comparisons

Day to day correlation between OMI and TROPOMI is very high above emission regions (see also Figure 61). **Figure 64** shows the global distribution of the OMI and TROPOMI HCHO VCD (*Nv_clear*) over the year 2019. Furthermore, **Figure 65** presents the time variation of the monthly averaged OMI and TROPOMI HCHO columns (*Nv* and *Nv_clear*) in the regions outlined on the TROPOMI map, as well as the mean absolute bias between OMI and TROPOMI columns (third row). We do not observe a change in time in the bias between OMI and TROPOMI. As already mentioned, the main source of bias lies in the different cloud parameters (*Nv*). When looking at the *Nv_clear* bias for which the cloud correction is discarded, the main remaining dependency lies in the latitude.

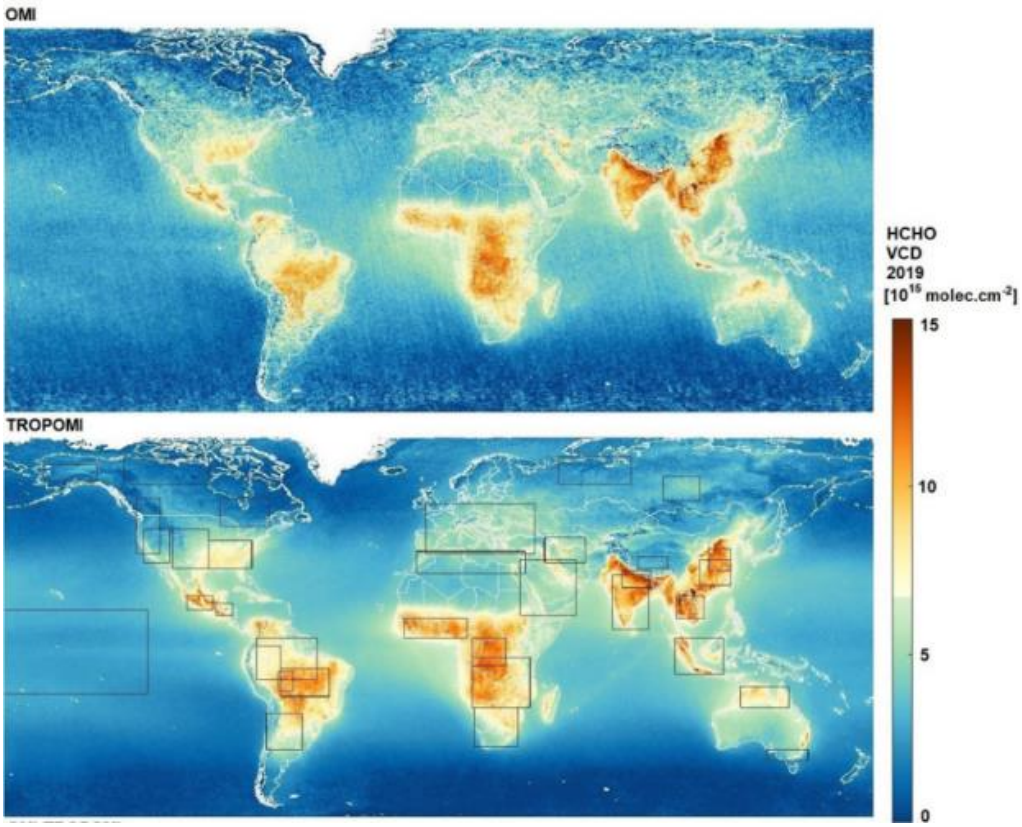


Figure 64: Average HCHO tropospheric column (N_{v_clear}) retrieved from OMI (upper panel) and S5P TROPOMI (lower panel) in 2019. Limits of the regions selected for the comparisons are identified on the TROPOMI map. The same grid is used for both dataset (0.05°). Data are filtered using the recommended product quality flags.

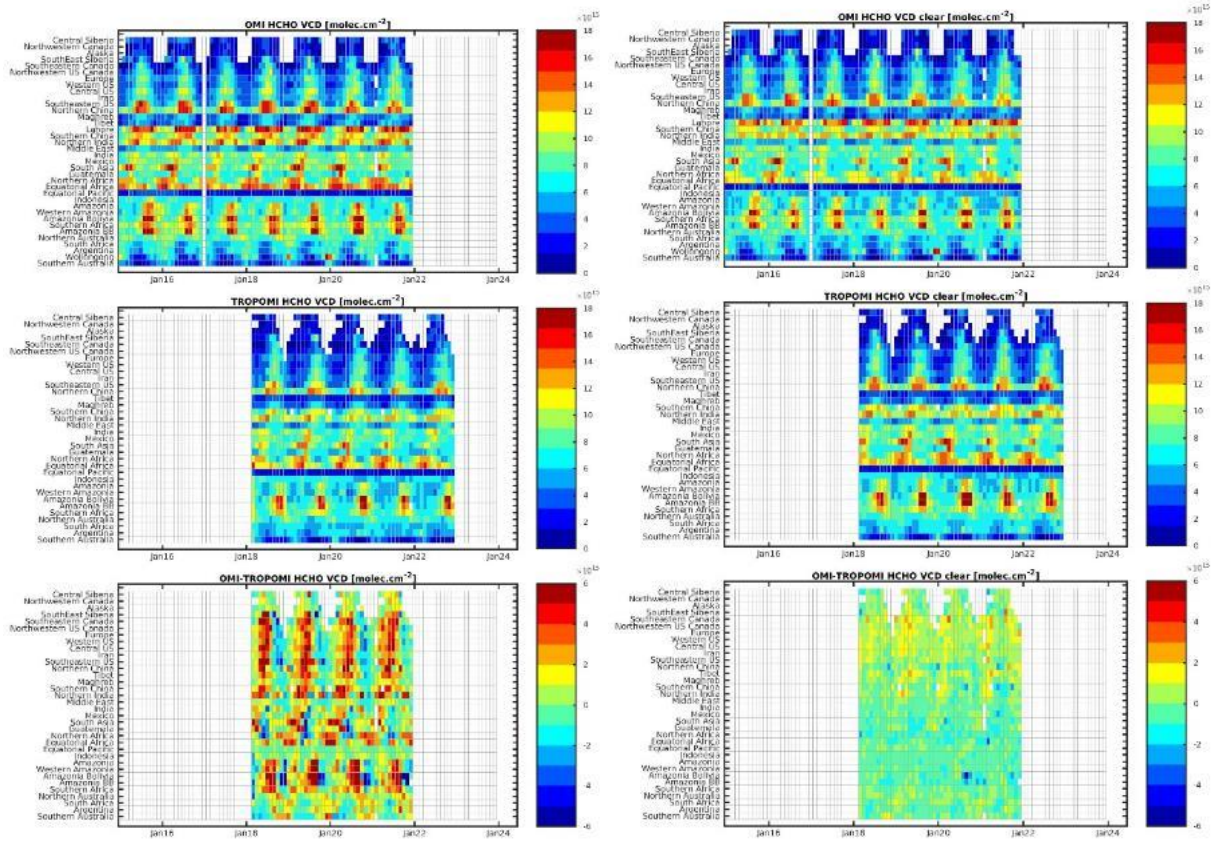


Figure 65: Time series of the regional monthly averaged OMI (first row), TROPOMI collection 3 (second row) HCHO VCD (*N_v*, left panels) and VCD clear (*N_v_clear*, right panels), and the mean absolute bias between OMI and TROPOMI columns (third row).

8.3.7 Geographical patterns

Figure 64 shows the comparison of OMI and TROPOMI HCHO columns (N_{v_clear}) averaged over one full year (2019). We observe a very good overall agreement. Differences range from 2 Pmolec/cm2 over Tropics to -2 Pmolec/cm2 over mid-latitude regions. The gain in TROPOMI precision can be observed at the global scale, mainly at larger latitudes where the OMI sampling is most affected in 2019.

8.3.8 Impact of UPAS version changes

Figure 66 shows the HCHO changes from 2018 to 2022. Mean HCHO vertical columns and their standard deviation in the remote Equatorial Pacific Ocean are plotted as a function of day and across-track position. The background values are stable in time. The zonal mean HCHO column from 2021 until February 2024 in **Figure 66** shows no impact from the processor changes between V02.04.01 and V02.06.01.

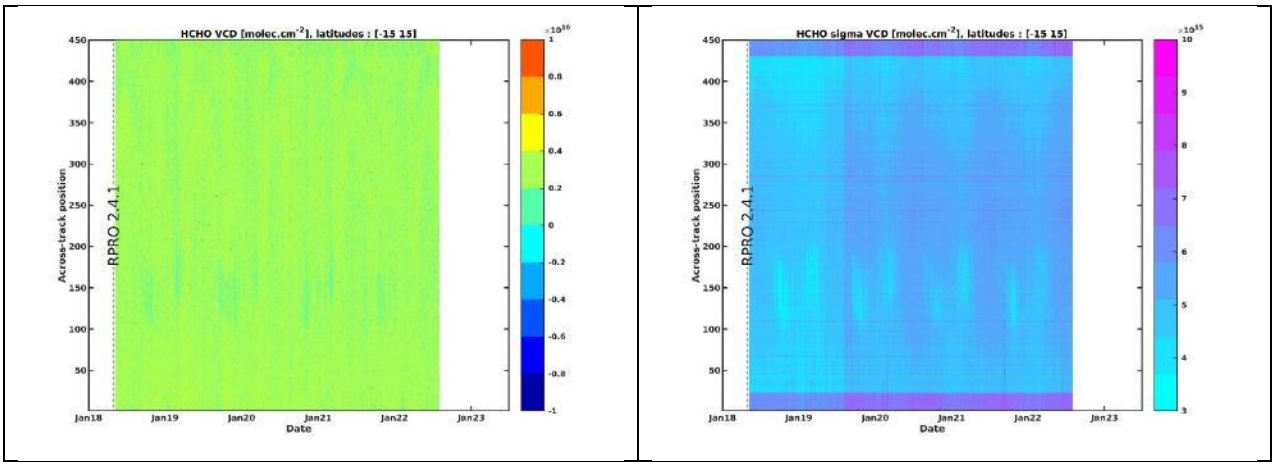


Figure 66: S5P TROPOMI HCHO column data (left) and their standard deviation (right) as a function of day (x-axis) and across-track position (y-axis). Data are averaged in the reference sector (the remote Equatorial Pacific). Results are shown for the reprocessed data collection 3.

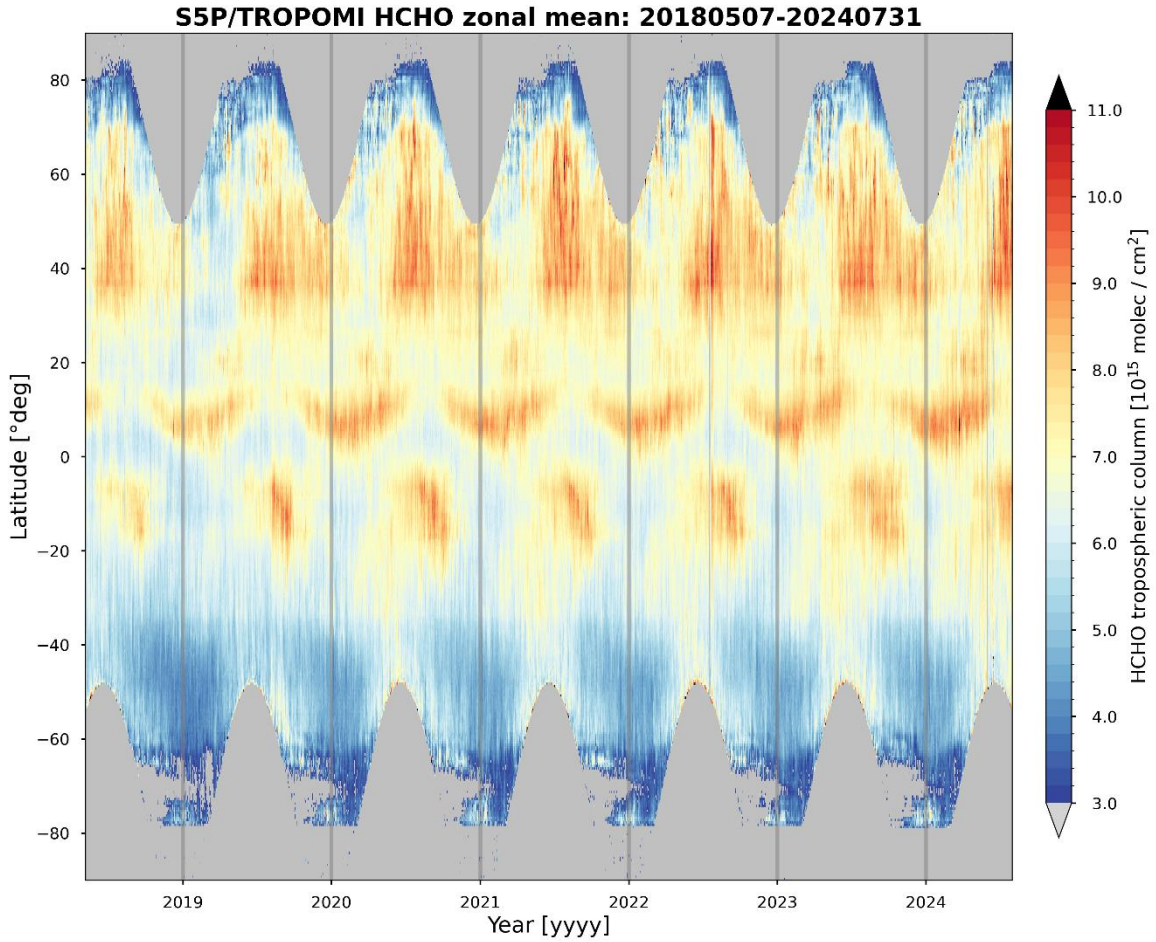


Figure 67: S5P TROPOMI HCHO column data [10^{15} molec/cm^2] as a function of day and latitude. The period is January 2019 to August 2024 (V02.06.01). Grid box size in latitude direction is 0.5°. The grey vertical lines mark the processor version changes, the black lines the beginning of each year.

8.4 Equivalence of L2_HCHO NRTI and OFFL products

We demonstrate the closeness of L2_HCHO NRTI and OFFL products at the MAX-DOAS stations De Bilt, Cabauw, and Mohali. NRTI (V02.04.01-02.06.01) and OFFL (RPRO and OFFL V02.04.01-02.06.01) L2_HCHO results, each co-located with MAX-DOAS, were obtained from the VDAF Automated Validation Server. A subset of pixels, common to both NRTI and OFFL, was chosen and differences between NRTI, OFFL, and MAX-DOAS determined. The statistical results are summarized in **Table 9.**

Table 9 - Statistics on the comparison of the common subset of L2_HCHO NRTI, L2_HCHO RPRO+OFFL and co-located MAX-DOAS for the stations Cabauw, De Bilt, and Mohali (*: unit of Pmolec cm^{-2}). The Automated Validation Server was consulted on 2024/08/21.

Cabauw: 277 common co-locations Orbits range from 24744 (2022-07-23) to 35527 (2024-08-21).			
	NRTI vs OFFL	NRTI vs MXD	OFFL vs MXD
Mean(diff)±sem*	-0.01	-3.61±0.55	-3.60±0.55
Median(diff)*	-0.03	-3.30	-3.14
Std(diff)*	0.7	9.1	9.1
1/2 IP68(diff)*	0.3	7.8	7.9
Pearson R	1.00	0.20	0.20
Slope	1.00	0.29	0.29
De Bilt: 338 common co-locations Orbits range from 24744 (2022-07-23) to 35527 (2024-08-21).			
	NRTI vs OFFL	NRTI vs MXD	OFFL vs MXD
Mean(diff)±sem*	-0.04	-4.74±0.51	-4.69±0.51
Median(diff)*	-0.00	-5.05	-4.80
Std(diff)*	0.7	9.4	9.4
1/2 IP68(diff)*	0.4	8.9	8.8
Pearson R	1.00	0.20	0.21
Slope	0.99	0.28	0.29
Mohali: 396 common co-locations Orbits range from 24728 (2022-07-22) to 34772 (2024-06-29).			
	NRTI-OFFL	NRTI-MXD	OFFL-MXD
Mean(diff)±sem*	0.01	-5.59±0.50	-5.60±0.51
Median(diff)*	0.18	-6.42	-6.43
Std(diff)*	2.1	10.0	10.1
1/2 IP68(diff)*	1.3	9.2	9.2
Pearson R	0.98	0.34	0.34
Slope	0.97	0.51	0.51

8.4.1 Bias

At the MAX-DOAS stations, the bias (both mean and median difference) of L2_HCHO NRTI vs. L2_HCHO OFFL is smaller than that of either L2_HCHO NRTI or L2_HCHO OFFL with respect to MAX-DOAS (see **Table 10**). More importantly, the bias of NRTI vs. OFFL is smaller than the standard error on the mean difference of either NRTI or OFFL with respect to MAX-DOAS. The bias differences between NRTI and OFFL are therefore statistically not significant. Similar conclusions are found using FTIR data (Vigouroux et al., 2020).

An example of the spatial distribution of daily differences is shown in **Figure 68** (2024/08/15) for V02.06.01 data. The relative differences Δ_r are 0.7 % with a high Pearson correlation coefficient ρ_P of 0.97. Differences are usually within ±1 Pmolec/cm^2. An across track striping is visible for the eastern part of the orbits.

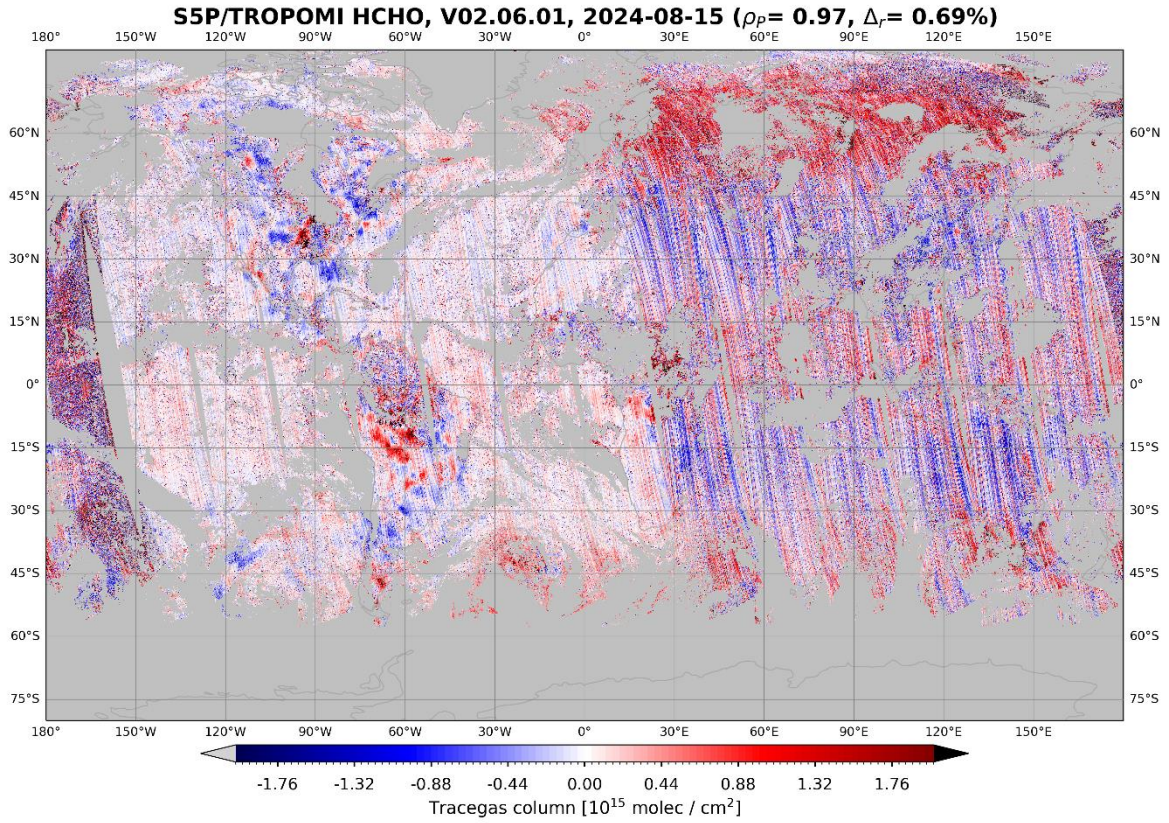


Figure 68: Difference between HCHO (V02.06.01) NRTI and OFFL column data for 2024/08/15 (binned to 0.1°x0.1° resolution). The difference is only calculated for columns above 0.5 Pmolec/cm^2 and a `qa_value` above 0.5. PMC stands for Pmolec/cm^2.

8.4.2 Dispersion

Both standard deviation and ½ 68% interpercentile (1/2IP68) of the NRTI/OFFL differences are much smaller than those between either NRTI/MAX-DOAS or OFFL/MAX-DOAS, indicating a much smaller dispersion between NRTI and OFFL. This is also confirmed by the near-unity NRTI/OFFL Pearson R correlation coefficient and slope (e.g. **Figure 68**). These are much lower for both NRTI and OFFL vs MAX-DOAS.

9 Validation Results: L2_SO2

9.1 L2_SO2 products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_SO2 product identified in **Table 1**: the sulphur dioxide total column. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. The NRTI and OFFL processors producing very similar data products, only validation of the L2_SO2 NRTI product is reported hereafter. Subsection 9.4 demonstrates evidence that NRTI and OFFL data do not differ significantly and that their respective validations yield similar conclusions.

9.2 Validation approach

9.2.1 Ground-based networks

Boundary layer pollution (SO_2 total)

S5P TROPOMI L2_SO2 sulphur dioxide column data are compared to ground-based MAX-DOAS UV-visible observations. However, currently the number of available stations in strongly polluted regions is still very rare. Outside strongly polluted regions, the SO_2 column is below the detection limit of both the MAX-DOAS and satellite measurements. For the validation of the S5P TROPOMI L2_SO2 sulphur dioxide column data MAX-DOAS and Pandora measurements at Xianghe (China), Greater Noida (India), Mohali (India), Basra (Iraq), Mexico City and Wakkerstrom (South Africa) are included in this report.

Volcanic plumes (SO2 enhanced)

S5P TROPOMI L2_SO2 sulphur dioxide column data are compared to MAX-DOAS UV-visible measurements collected from the Network for Observation of Volcanic and Atmospheric Change (NOVAC) [ER_NOVAC]. Because of the strong SO_2 concentration gradients in volcanic plumes, the comparison is not performed using the SO_2 columns but rather using the derived SO_2 fluxes.

9.2.2 Satellites

S5P TROPOMI L2_SO2 sulphur dioxide column data are compared to similar data from EOS-Aura OMI and Suomi-NPP OMPS.

9.2.3 Field campaigns and modelling support

S5P TROPOMI L2_SO2 sulphur dioxide column data are compared to car MAX-DOAS measurements performed in Lahore.

9.2.4 Test of the expectation of zero SO_2 SCDs (within detection limit) outside volcanic plumes and strongly polluted regions

Outside strongly polluted regions and volcanic plumes, the atmospheric SO_2 concentrations are very low and the corresponding SO_2 columns are below the detection limit of S5P TROPOMI. Thus S5P TROPOMI measurements outside strongly polluted regions and volcanic plumes are used to check the consistency of the S5P TROPOMI L2_SO2 sulphur dioxide column data with the assumption of SO_2 slant column densities (SCD) of zero. From this test, also the spread of the S5P TROPOMI L2_SO2 sulphur dioxide column data is quantified.

9.3 Validation of L2_SO2 NRTI

9.3.1 Recommendations for data usage followed

The quality of the observations depends on many factors which are taken into account in the definition of the `qa_value`. While it is a handy way of filtering observations of less quality, the "quality assurance value" should also be considered with caution, as it is a compromise to take into account several aspects, such as: processing errors, presence of clouds or snow/ice, observations affected by sun glint, South Atlantic Anomaly, possible contamination by volcanic SO_2, absence of background correction, and important variables out of range (importantly the AMF).

The `qa_value` is a continuous variable, ranging from 0 (error) to 1 (all is well). In order to avoid misinterpretation of the data quality, it is recommended at the current stage to only use those TROPOMI pixels associated with a `qa_value` above 0.5.

For further details, data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms

9.3.2 Status of validation

So far the validation of the S5P TROPOMI L2_SO2 sulphur dioxide column data is mainly based on satellite to satellite comparisons (**Figure 69**, **Figure 70**), for which good agreement is found with OMI and OMPS measurements. Validation for polluted regions using ground based MAX-DOAS data is limited to three stations in polluted regions (Xianghe, Mexico City, and Wakkerstroom, South Africa, see **Figure 71** to **Figure 75**). In general good agreement was found. However, it should be noted that for these comparisons the SO_2 columns were mostly close to or below the detection limit of S5P TROPOMI.

S5P TROPOMI L2_SO2 sulphur dioxide column data were also compared to ground based MAX-DOAS measurements from the NOVAC network. However, the SO_2 columns were not compared directly, because of the strong gradients across volcanic plumes. Instead, the derived SO_2 fluxes were compared, for which good agreement was found.

Outside strongly polluted regions and volcanic plumes, the atmospheric SO_2 SCDs were found to be consistent with the assumption of zero within the measurement uncertainties.

From these comparisons, (details are shown below) the following conclusions are drawn:

- over polluted regions the requirements are fulfilled;
- over volcanic plumes the bias requirement is fulfilled, while the random requirement can be exceeded occasionally, which is not seen as a substantial restriction of the data quality;
- from the time series of averaged SO_2 SCDs (and their errors and standard deviations) it is concluded that the requirements are fulfilled. The bias and spread are typically below 0.2 DU.

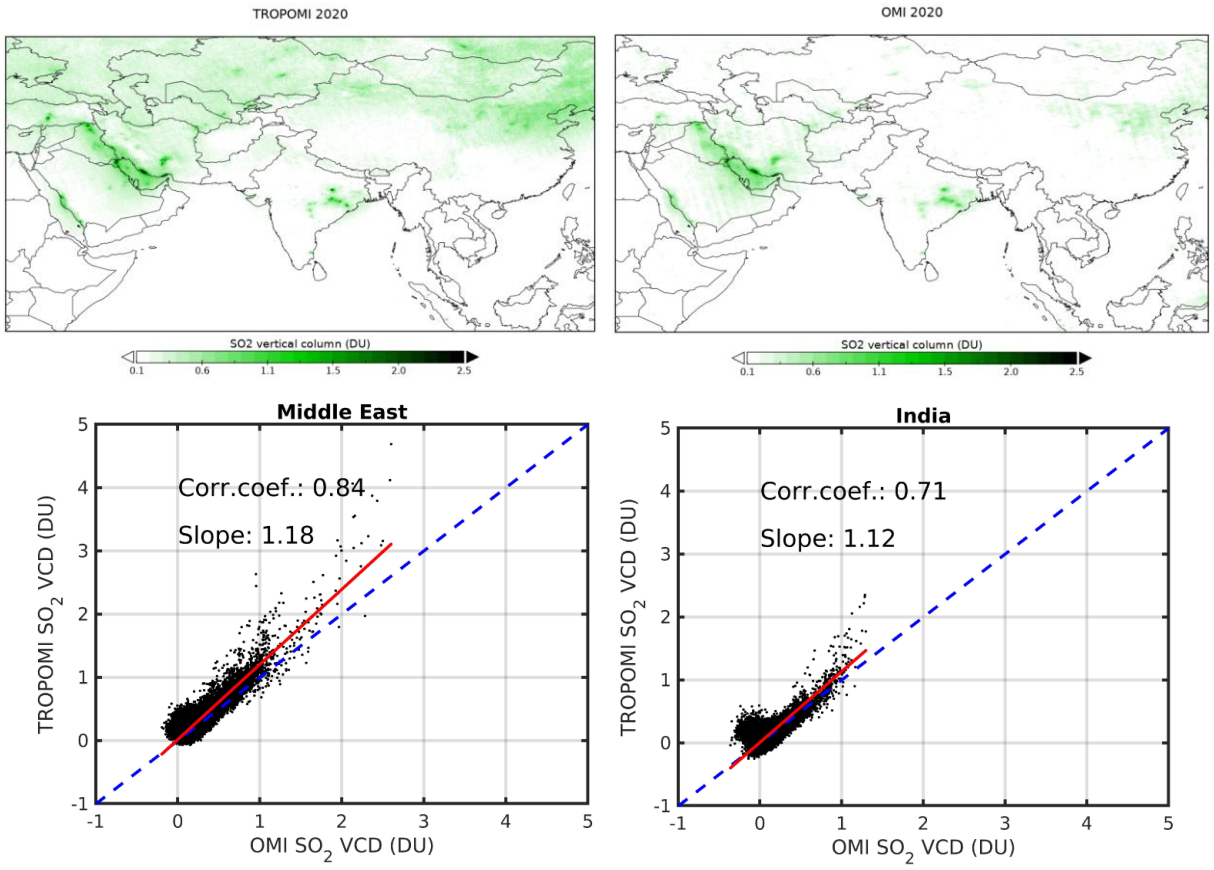


Figure 69: Top: Comparison of the average distribution (01 Jan 2020 – 31 Dec 2020) of the SO_2 VCDs derived from TROPOMI and OMI over regions with strong air pollution. Except high latitudes, both data sets show very good agreement. Bottom: Correlation plots TROPOMI versus OMI over the Middle East and India. Note that a fixed AMF of 0.4 was used for both retrievals to exclude the effect of different profile assumptions. Courtesy of Nicolas Theys, BIRA-IASB.

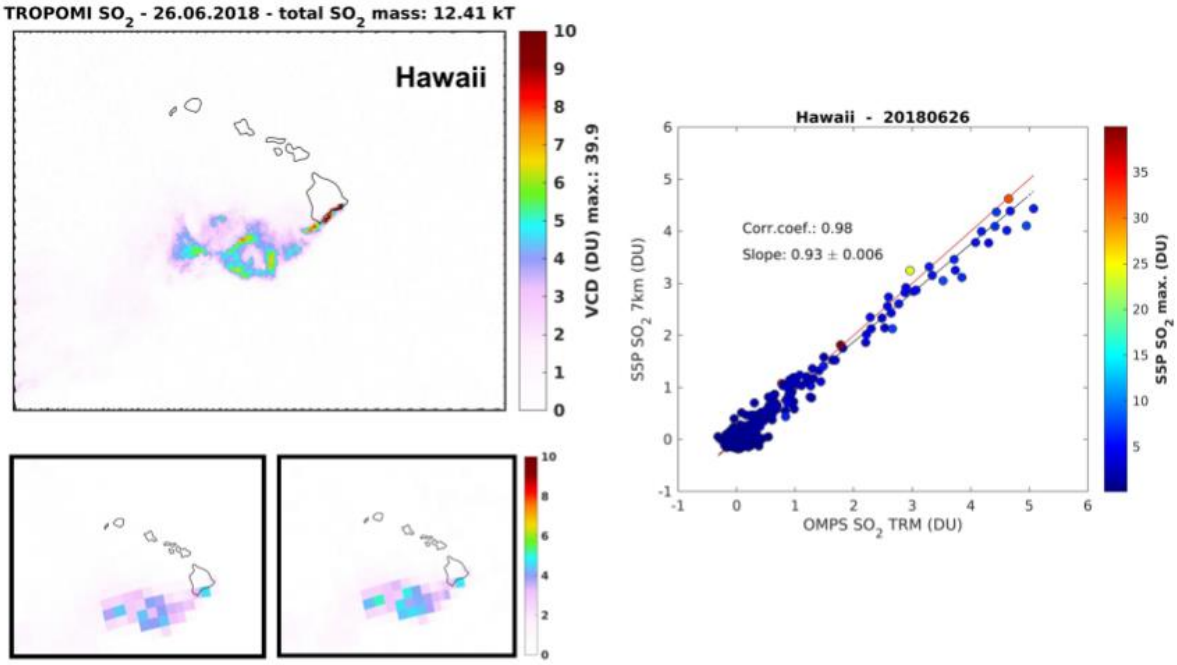


Figure 70: Comparison of TROPOMI and OMPS measurements of the volcanic plume of Kilauea on 26 June 2018. The large figure shows the original TROPOMI data. The two small figures show the spatially degraded TROPOMI data and the OMPS data. The figure right shows the correlation plot of the degraded TROPOMI data versus the collocated OMPS data. Courtesy of C. Li and N. Krotkov, NASA/GSFC.

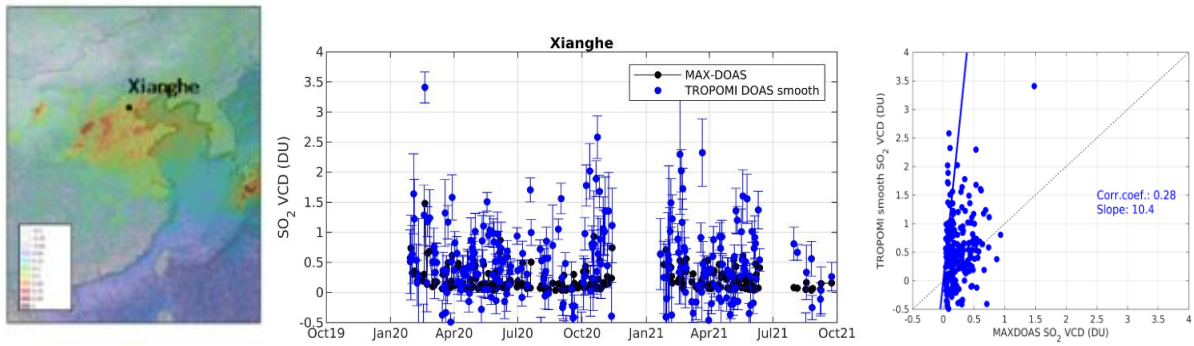
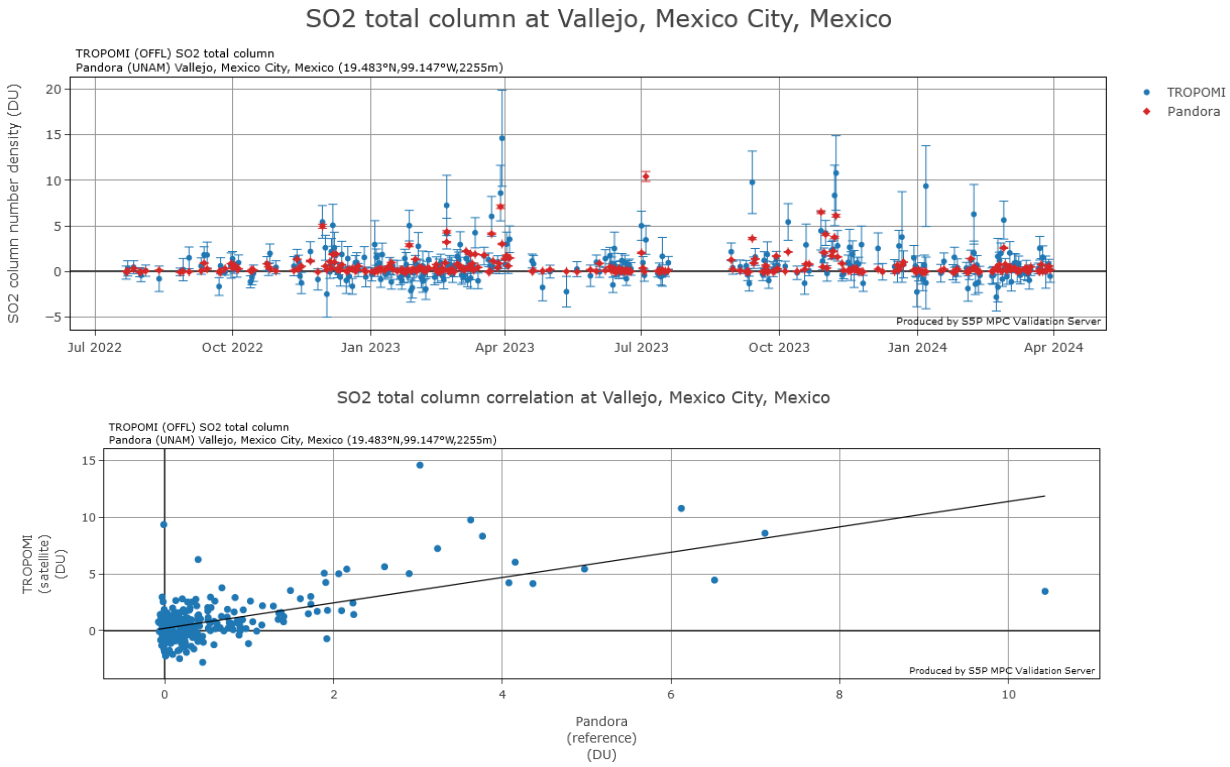
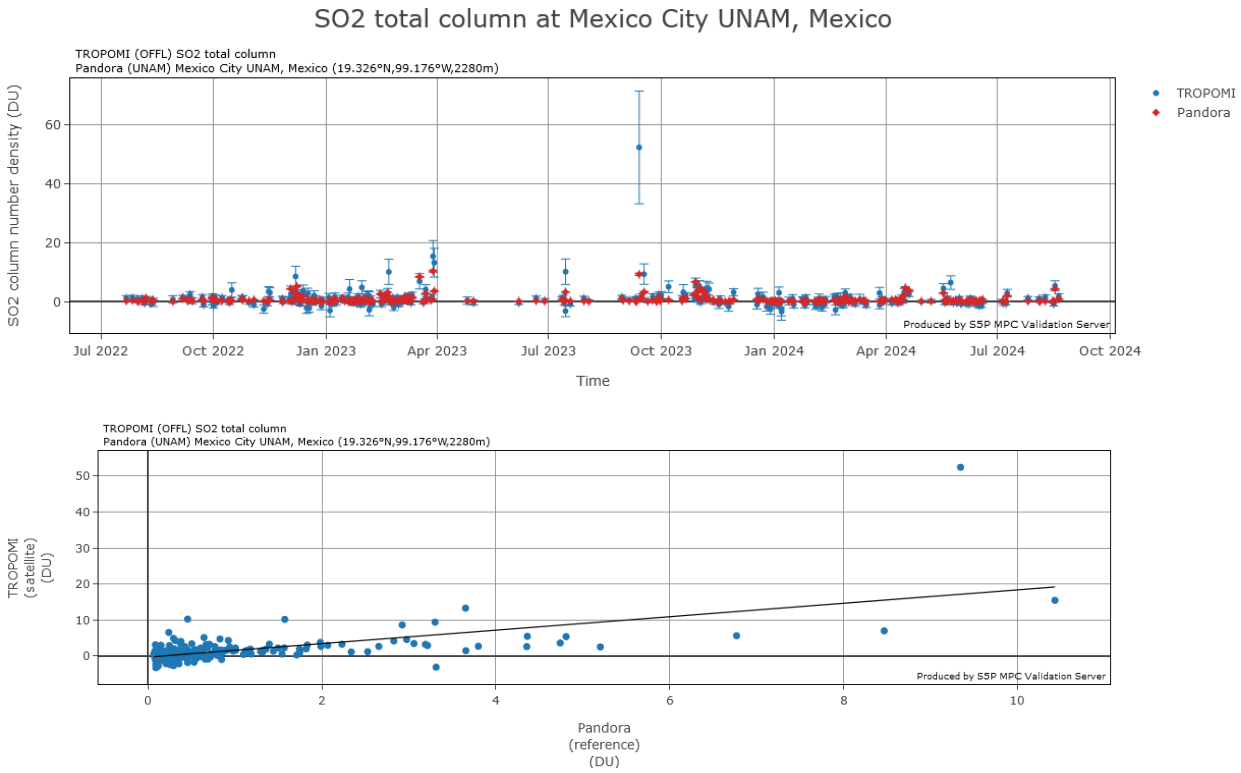


Figure 71: Comparison of TROPOMI SO_2 VCDs to MAX-DOAS measurements (daily means) at Xianghe (China). The following selection criteria were applied: distance < 25 km, CF< 0.2, AMF > 0.2, time window +/- 1h around overpass. Courtesy of N. Theys (BIRA-IASB).



R= 0.63, slope: 1.12

Figure 72: Comparison of TROPOMI SO_2 VCDs to Pandora measurements (daily means) at Vallejo (Mexico City) from July 2020 to April 2024. The reference data was measured by Universidad Nacional Autonoma de Mexico and the calibration and processing was conducted within the framework of the Pandonia Global Network. Data are preliminary and are obtained from the Sentinel-5P MPC Validation Server: https://mpc-vdaf-server.tropomi.eu/so2/so2-nrti-pandora/.



R= 0.65, slope: 1.87

Figure 73: Comparison of TROPOMI SO_2 VCDs to Pandora measurements (daily means) at Observatorio Atmosferico Altzomoni, UNAM, Mexico (3985m altitude) from December 2019 to August 2024. The reference data was measured by Universidad Nacional Autonoma de Mexico and the calibration and processing was conducted within the framework of the Pandonia Global Network. Data are preliminary and are obtained from the Sentinel-5P MPC Validation Server: https://mpc-vdaf-server.tropomi.eu/so2/so2-nrti-pandora/.

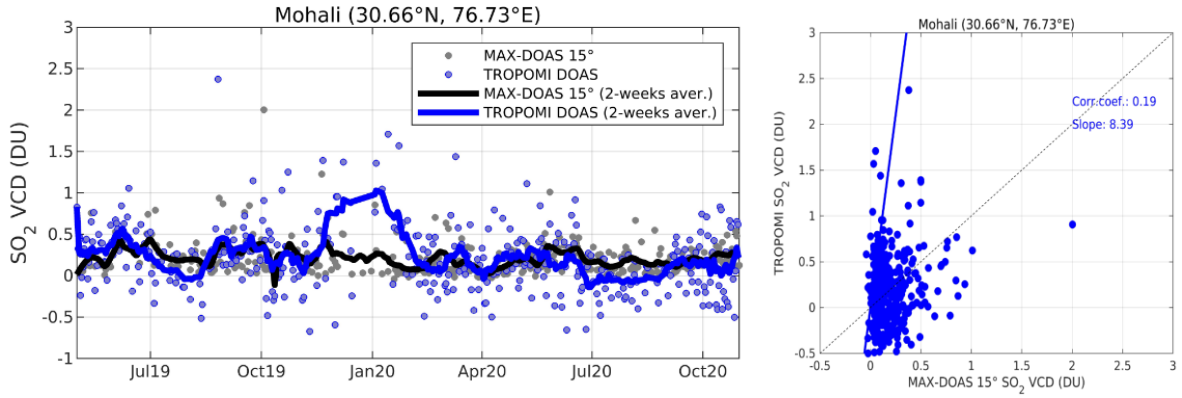
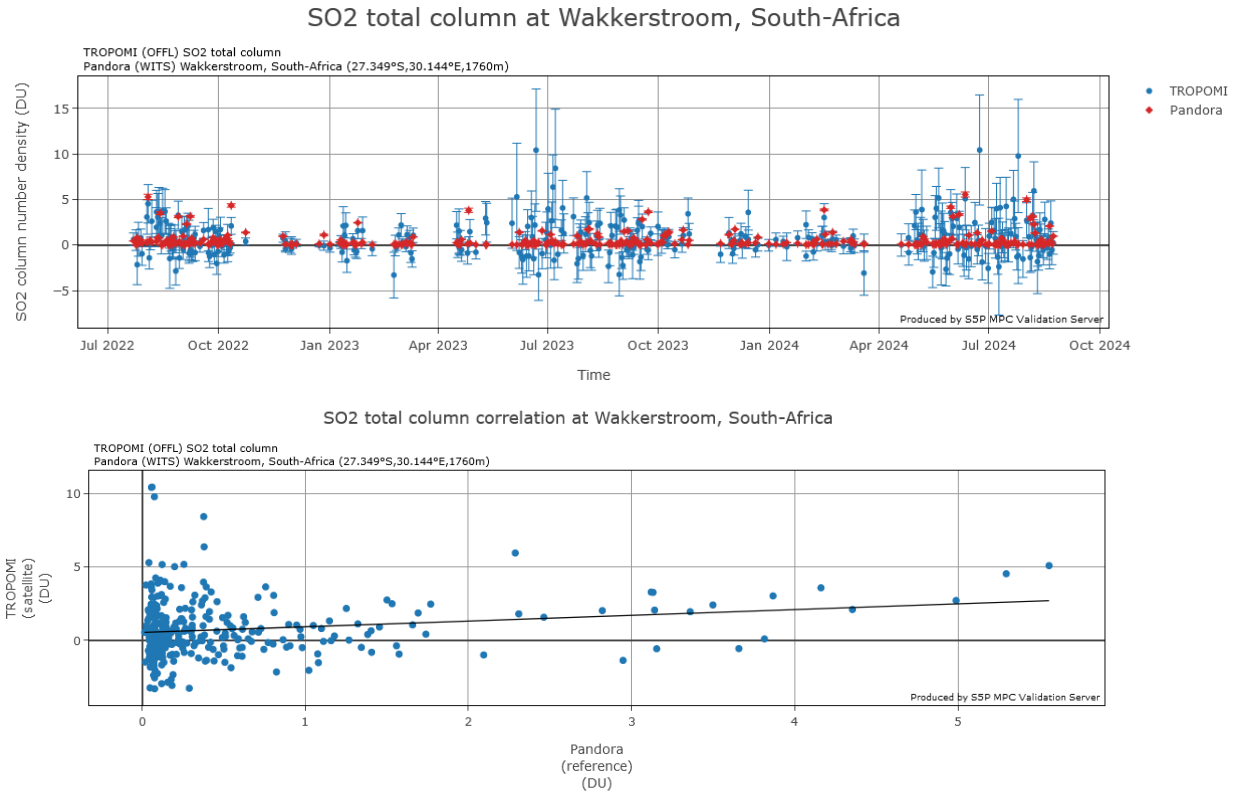


Figure 74: Comparison of TROPOMI SO_2 VCDs to MAX-DOAS measurements (daily means) at Mohali (India) from June 2019 to November 2020. The following selection criteria were applied: distance < 25km, CF<0.2, AMF>0.2, time window +/- 1h around overpass. Courtesy of N. Theys (BIRA-IASB), data provided by, Vinod Kumar, Vinayak Sinha, Sebastian Donner, Steffen Dörner, Thomas Wagner.



R= 0.17, slope: 0.39

Figure 75: Comparison of TROPOMI SO_2 VCDs to Pandora measurements (daily means) at Wakkerstroom from December 2019 to August 2024. The reference data was processed within the framework of the Pandonia Global Network. Data are preliminary and are obtained from the Sentinel-5P MPC Validation Server: https://mpc-vdaf-server.tropomi.eu/so2/so2-nrti-pandora/.

9.3.3 Bias

The bias is well within requirements for boundary pollution. From the time series of averaged SO_2 SCDs it is estimated that the bias is within 0.2 DU. For volcanic plumes, very good agreement with other satellite observations is found (<10%), but due to the lack of validation by ground-based measurements, the true bias might be larger in some cases.

9.3.4 Dispersion

The dispersion is well within requirements for observations of the boundary pollution. For observations of dense volcanic plumes, the dispersion is probably within the requirements, but due to the lack of validation by ground-based measurements, the true dispersion is at the moment difficult to quantify. From the time series of the standard deviation of the SO_2 SCDs it is estimated that the dispersion is within 0.2 DU.

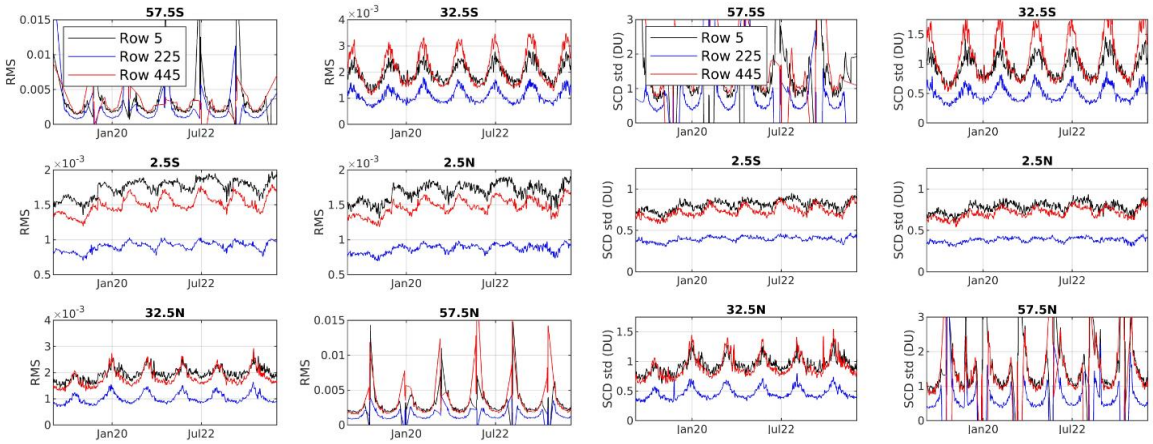


Figure 76: Temporal evolution of the measurement error (left) and the standard deviation (right) for selected 5° latitude bands and 3 detector rows from December 2018 to August 2024. Good qualitative agreement between both quantities is found indicating that the random uncertainty is well characterized by the measurement error. Larger errors (and standard deviations are found at the edges of the detector and towards high latitudes. The jump in the RMS and standard deviations in low and mid-latitudes in August 2019 are caused by the reduction of the ground pixel size. Courtesy of N. Theys (BIRA-IASB).

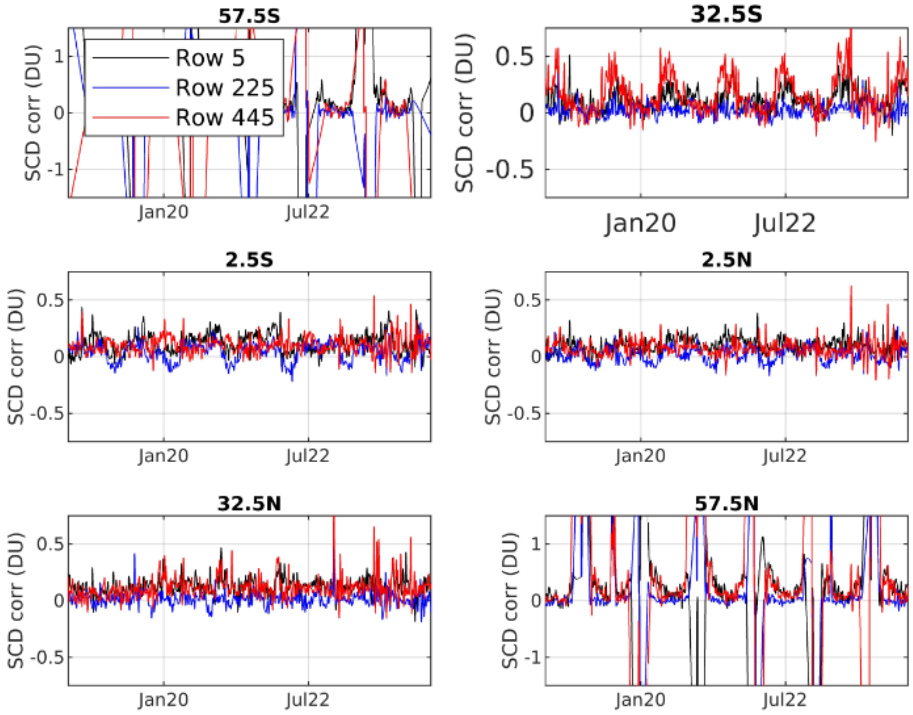


Figure 77: Temporal evolution of the averaged SO_2 SCD for selected 5° latitude bands and 3 detector rows from December 2018 to August 2024. The values are close to zero and show relatively small day-to-day variations. The larger variations in August 2019 are caused by strong volcanic eruptions. Courtesy Nicolas Theys (BIRA-IASB).

9.3.5 Dependence on influence quantities

Slightly larger bias and dispersion are found towards higher SZA.

9.3.6 Short term variability

The short term variability can be estimated from the time series of averaged SO_2 SCDs (outside periods with strong volcanic eruptions). It is estimated to below about 0.1 DU.

9.3.7 Geographical patterns

Slightly larger bias and dispersion are found at higher latitudes, likely as an effect of high solar zenith angles.

9.3.8 Other features

None to report.

9.4 Equivalence of L2_SO2 NRTI and OFFL products

The NRT and offline SO_2 products are very similar, as illustrated by the comparison of the SO_2 SCDs of the three data versions hereafter. Thus, the validation activities performed for the OFFL data product (see above) are also representative for the NRTI data product. This finding is illustrated in **Figure 78**.

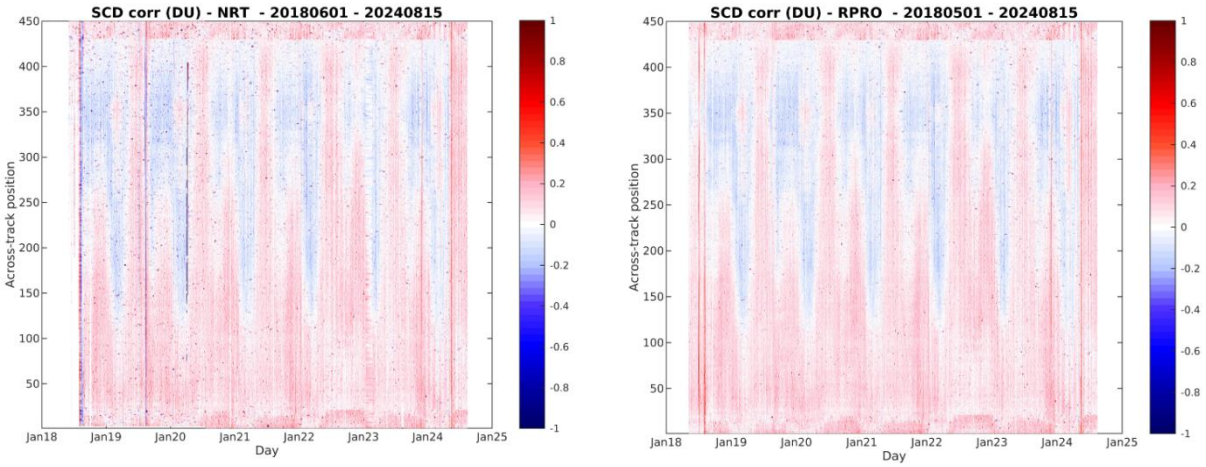


Figure 78: Comparison of the NRT (left) and offline (right) SO_2 data products. Shown are the time series of background corrected SO_2 SCDs for all 450 detector rows from June 2018 to August 2024 Most of the short time features (vertical coloured lines) are caused by individual strong volcanic eruptions. For the NRT product, a strong spike occurs in early April 2020, which is caused by a downlink issue that corrupted an irradiance file. Courtesy of Nicolas Theys, BIRA-IASB.

10 Validation Results: L2_SO2 Layer Height

10.1 Recommendations for data usage

In the case of the SO_2_layer_height product the quality value is described in the qa_value_layer_height_parameter. In order to avoid misinterpretation of the SO_2 plume height data quality, it is recommended at the current stage to only use those TROPOMI pixels associated with a `qa_value_layer_height` above 0.5. A further quality flag that verifies if the layer height retrieval has been performed with the required precision, specified as sulfurdioxide_layer_height_flag, is incorporated. The flag values are static (0, 1, 2, 4, 8, 16, 32, 64, 128) and retrievals are considered successful for values lower than 32. More specifically, the flag is: 0 for successful SO_2 layer height retrieval, 1 for exceeded input data range, 2 for layer height exceeding height of 30 km (out of training range), 4 for high aerosol absorption, 8 for negative aerosol absorption, 16 for retrieved VCD below 15 DU (below training range), 32 for retrieved LH below 0 km, 64 for layer height not triggered and 128 for missing input data. Here, the sulfurdioxide_layer_height_flag is set to lower than 16 to disregard the noisy observations and pixels with equivalent SO_2 values higher than 20 D.U. are only considered, corresponding to strong volcanic eruptions and robust SO_2 plumes.

For further details, data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

10.2 Status of validation

The validation of the S5P TROPOMI L2_SO2 sulphur dioxide plume height is based on satellite-to-satellite comparisons (Figure 78, Figure 80), during large enough volcanic eruptions, and with airmass trajectories retrieved from the HYSPLIT model (Stein et al., 2015). For this purpose, the Metop[B/C] IASI "SO2_vertical_level" product, i.e., the retrieved vertical level of the SO_2 plume, for the morning (AM) observations was utilized. IASI provides near global coverage twice a day, with a local overpass times at 9.30 AM and PM. The AM observations were selected due to the temporal proximity to the TROPOMI overpass time (at 13.30 local time). The validation shown in this report covers eruptions for the September 2023 to May 2024 period, where the pixels with corresponding SO_2 levels higher than 20 D.U. led to an activation of the SO_2 Plume Height algorithm. Details about the eruptions included in this validation process (e.g., start and end date of the acquired SO_2 layer height, name of the volcano and location) can be found in Table 11.

Table 10. Volcanic eruptions details used so far for the validation of S5P/TROPOMI SO_2 layer height.

Date start	Date end	Volcano	Country	Latitude	Longitude
19/09/2023	26/09/2023	Popocatepetl	Mexico	98.622°W	19.023°N
25/09/2023	03/10/2023	Shishaldin	United States	54.756°N	163.97°W
31/10/2023	1/11/2023	Klyuchevskoy	Russia	56.056°N	160.642°E
20/11/2023	22/11/2023	Ulawun	Papua New Guinea	5.05°S	151.33°E
02/03/2024	08/05/2024	Fernandina	Ecuador	0.37°S	91.55°W
16/03/2024	08/05/2024	Reykjanes	Iceland	63.817°N	22.717°W
16/04/2024	21/05/2024	Ruang	Indonesia	2.3°N	125.37°E

Figure 79 shows the Ruang volcano eruption SO_2 layer height on April 18,2024, as sensed by the three instruments, IASI on both MetOp-B & MetOp-C and TROPOMI). As expected, IASI/MetOp-B and MetOp-C show a similar plume distribution as TROPOMI, which reported a slight northwestern dispersion of the plume due to the ~3h difference between the overpass of the satellites. All three instruments sense the SO2 plume around 115°E-125°E and 0°N to 10°N.

A quantitative representation of the SO_2 plume height of the Ruang volcano eruption, sensed by the three instruments, is given in **Figure 80**. TROPOMI SO_2 layer height ranges from 2.5 km to 20 km, whereas IASI/Metop-[B&C] SO_2 plume height ranges from 2 km to 18 km. The mean TROPOMI SO_2 layer height is approximately 14.2±2.3 km the mean IASI/Metop-[B&C] is 14±3.1 km. The two instruments show a slightly different SO_2 layer height distribution likely attributed to the different overpass times. IASI instruments overpass time was closer to the volcano eruption time, which was approximately at 20:45pm on April 17, 2024. Reports from the Volcanology and Geological Hazard Mitigation Center of Indonesia (PVMBG, https://vsi.esdm.go.id/, last accessed on August 29, 2023) indicate that plumes rose up to 20km and drifted northwest by 21:20 pm of the same day. During the IASI overpass (~9.30 am), on the 18th of April, significant levels of SO_2 are observed in the 2-8km range and in the 16-18km range, whereas during the TROPOMI overpass (~13.30 pm) the SO_2 plume is mainly observed in the 8-20km range. Hence, TROPOMI cannot seemingly sense the lower altitude plumes. Overall, the two instruments report similar mean SO_2 layer height levels with discrepancies of approximately 12 %.

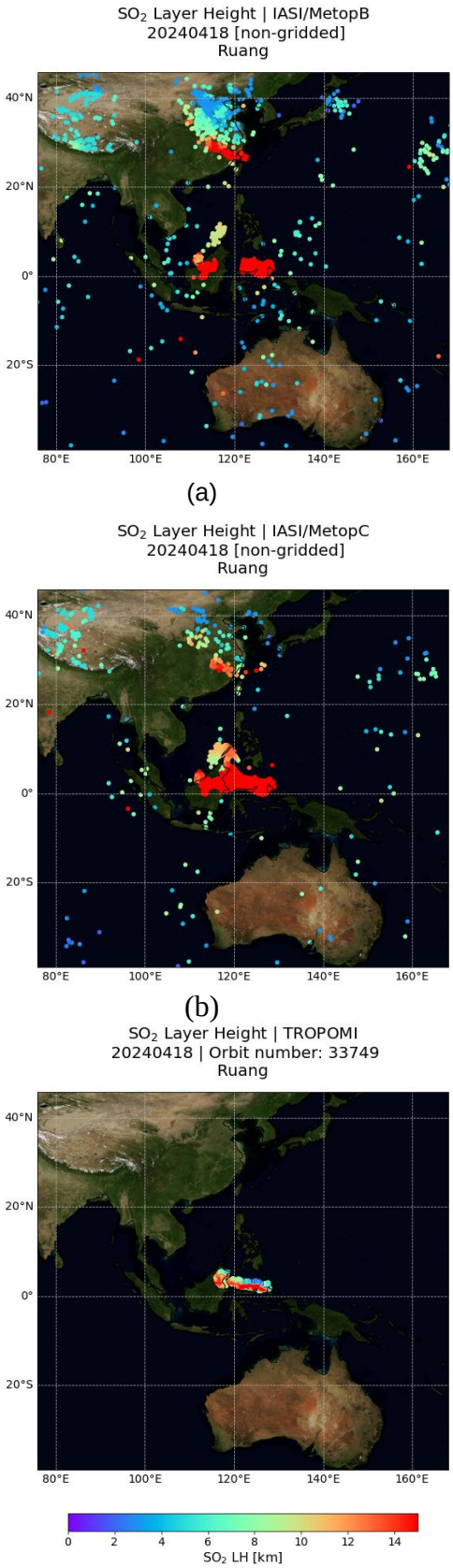


Figure 79. SO_2 layer height [km] during the Ruang eruption on April 18, 2024, as sensed by (a) IASI/MetOp-B; (b) IASI/MetOp-C; (c) S5P TROPOMI.

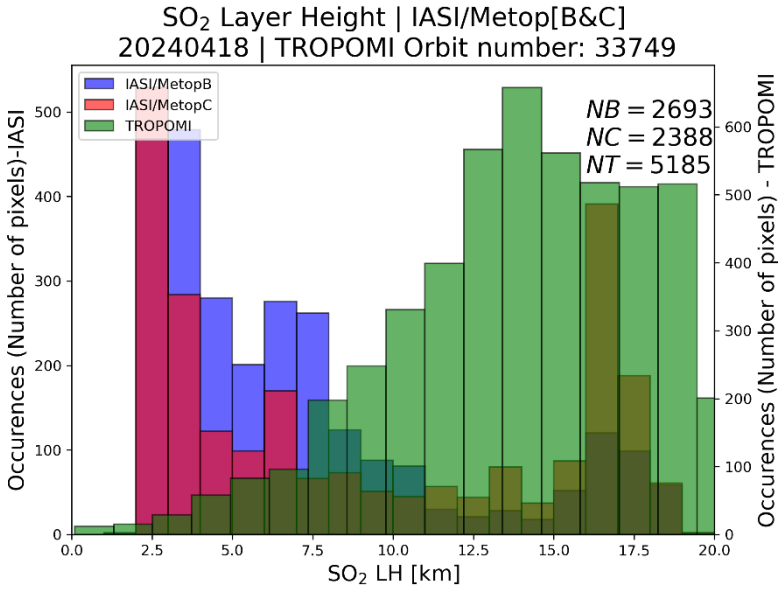
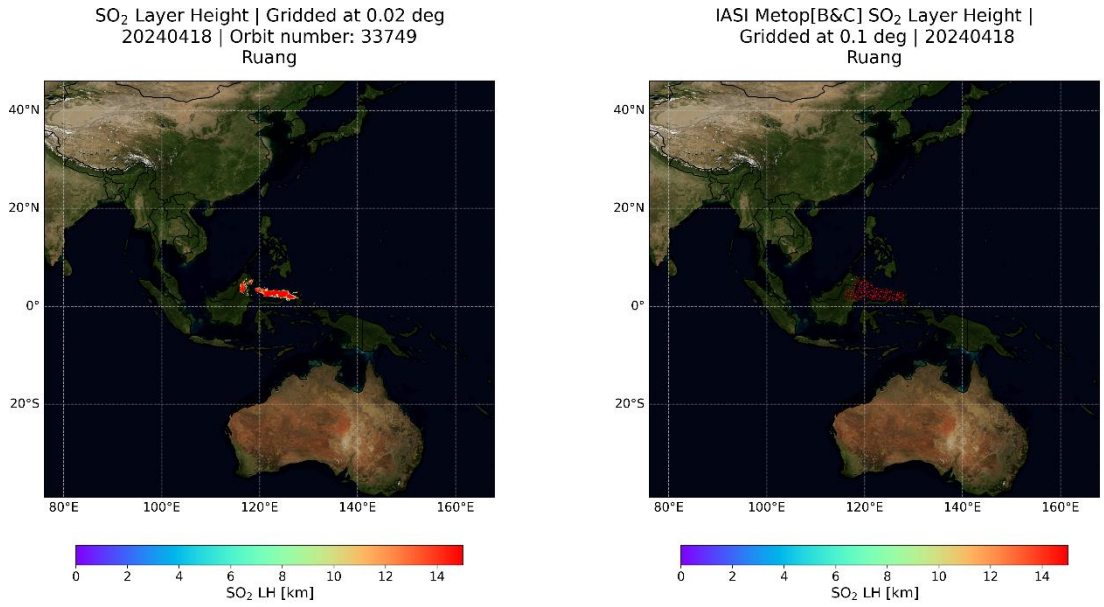


Figure 80. Histogram of the SO_2 layer height [in km] of the Ruang volcano eruption on April 18, 2024, as sensed by IASI/Metop-B [blue], IASI/Metop-C [red] and S5P/TROPOMI [green]. The left y-axis denotes the number of pixels of valid SO_2 layer height retrievals for the IASI instruments, whereas the right y-axis denotes the valid SO_2 layer height retrievals of the TROPOMI instruments. Only the retrievals corresponding to the TROPOMI lat/lon extents were selected for all instruments.

To get a more realistic comparison close to collocated observations, IASI/Metop-B and Metop-C observations are merged and regridded to the TROPOMI grid. **Figure 81** shows the TROPOMI SO_2 layer height [left], and the combined and regridded IASI/Metop[B&C] SO_2 plume height observations [right] for the Ruang volcano eruption on the 18th of April 2024. The SO_2 plume is located in the same area, at around 115°E-130°E and 2°N-8°N. TROPOMI mean SO_2 layer height is found approximately at 14km, whereas IASI/Metop[B&C] at 16km.



(a) (b)

Figure 81. Gridded SO_2 layer height [in km] during the Ruang volcano eruption on April 18, 2024, as sensed by (a) S5P/TROPOMI; (b) IASI/MetOp-[B&C].

An alternative source for the validation of the TROPOMI SO2_Layer_Height product is the forward and backward airmass trajectories of the HYSPLIT model. With the forward trajectories, predefined start points from the volcano at specific altitudes at the eruption time can be used to indicate the height of airmasses and the dispersion of the SO_2 plume. With the backward trajectories, defined endpoints in specific locations and altitudes within the SO_2 plume can be used to backtrack the origin and the altitude of the airmass. Here, the forward trajectories of the HYSPLIT model are compared with the TROPOMI SO2_Layer_Height product for the Ruang volcano eruption on the 18th of April, as shown in Figure 81. Reports from the local Volcanology and Geological Hazard Mitigation Center of Indonesia (PVMBG) indicate that on April 18 lower altitudes plumes (up to 7km) drifted in a southeastern direction. Higher altitude plumes (up to 15km) drifted northwest and plumes up to 20km drifted in a western-northwestern direction. Hence, forward trajectories starting from the volcano with predefined heights of 0.7km, 7km, 14km and 21km were selected. Overall, the forward trajectories show a good agreement with TROPOMI observations and the local reports of the plume dispersion. The lower-altitude forward trajectories (1-7km) show a southeastern dispersion, where TROPOMI observations report a mean plume height of 4km. The trajectory starting at 14km disperse westwards, agreeing with TROPOMI, where a mean plume height of 13km is reported. Finally, the forward trajectory, starting at 21km, demonstrates a west-northwest dispersion where the TROPOMI SO_2 plume is reportedly at 20km. The HYSPLIT model airmass trajectories will be utilized to bridge the gap between the overpasses of S5P/TROPOMI and MetOp satellites in order to obtain temporally collocated observations and more realistic comparisons between the sensors. The revert in the overpass will be applied to all eruptions and discussed in following Validation Reports.

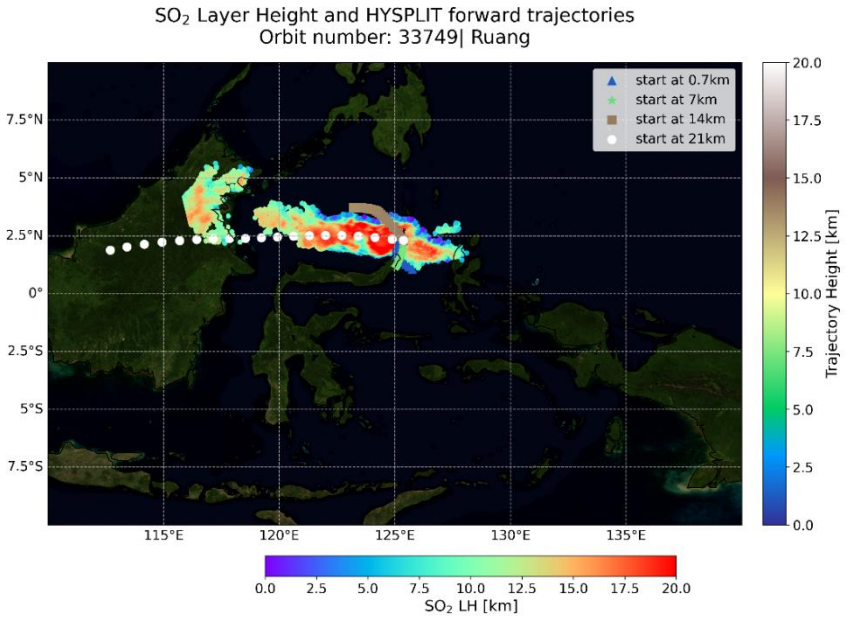


Figure 82: SO2_Layer_Height sensed by TROPOMI compared with the HYSPLIT air mass forward trajectories over Ruang. Forward trajectories start at the eruption time (20:45 local time) and the selected heights of 0.7 km (blue), 7 km (green), 14 km (brown), and 21 km (white).

10.2.1 Bias

The same gridding process was followed for all the studied volcanoes included in this validation report during the September 2023 - May 2024 period. Comparisons with IASI were conducted between each single TROPOMI orbit for which the algorithm was activated (i.e., SO_2 VCD higher than 20 DU).

Figure 81 shows the scatter plot between the daily mean TROPOMI SO_2 layer height observations (y-axis) and combined IASI/Metop-[B&C] observations, regridded in the TROPOMI grid. Different colours identify different TROPOMI orbits. Overall, TROPOMI shows lower SO_2 layer height retrievals compared to the combined IASI/Metop[B&C] SO_2 plume height observations by ~35 %, equivalent to a mean SO_2 layer height difference of 2.5 km. In more significant volcano eruptions with robust and overlapping SO_2 plumes between the involved instruments, such as Shishaldin and Ulawun, and Ruang, the discrepancies are lower (~12 %) and equivalent to approximately 1.5 km. This indicates a good agreement between the different sensors, bearing in mind the difference in the overpass time over the volcanoes. The HYSPLIT method is not applied yet for the direct comparison between TROPOMI and IASI as shown in **Figure 83** below. Thus for these comparisons it should be noted that both measurements are not made at exactly the same time.

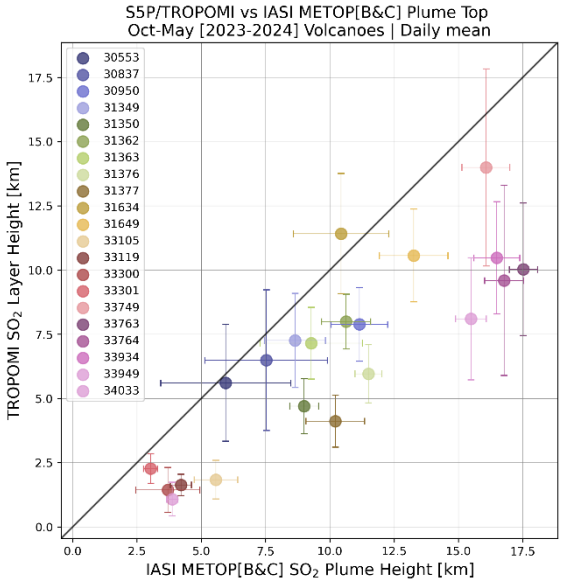


Figure 83. Scatter plot between the daily mean TROPOMI SO_2 layer height observations (y-axis) and combined IASI/Metop-[B&C] observations retrieved during volcanic eruptions of the September 2023 - May 2024 period.

10.2.2 Dispersion

Not applicable to the L2 SO2 Plume Height geophysical parameter.

10.2.3 Dependence on influence quantities

The S5P SO_2 Plume Height algorithm is activated only when a large enough SO_2 load is retrieved for a specific pixel, currently set at 20 D.U.

10.2.4 Short term variability

Not applicable to the L2 SO_2 Plume Height geophysical parameter.

10.2.5 Geographical patterns

Not applicable to the L2 SO_2 Plume Height geophysical parameter.

10.2.6 Other features

None to report.

11 Validation Results: L2_CO

11.1 L2_CO products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_CO product identified in **Table 1**: the carbon monoxide total column. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. Comparison results of L2 CO OFFL standard and destriped products are discussed separately and only the destriped product is mentioned in the quality indicators in **Table 2**.

11.2 Validation approach

11.2.1 Ground-based networks

S5P TROPOMI L2_CO carbon monoxide column data are routinely compared to reference measurements obtained from FTIR spectrometers performing network operation in the context of the Network for the Detection of Atmospheric Composition Change (NDACC, http://ndacc.org), the Total Carbon Column Observing Network (TCCON, https://tccondata.org) and the Collaborative Carbon Column Observing Network (COCCON[1], https://www.imk-asf.kit.edu/english/COCCON.php). **Figure 84** displays the geographical distribution of the NDACC and TCCON stations. Near-infrared TCCON measurements provide CO column averaged (xCO) data with typical uncertainty values of 2% for the bias and 1% for the precision. The COCCON measurements are calibrated to TCCON and show similar performance as TCCON. Solar infrared NDACC measurements provide CO total column data with a typical total uncertainty of 3 %.

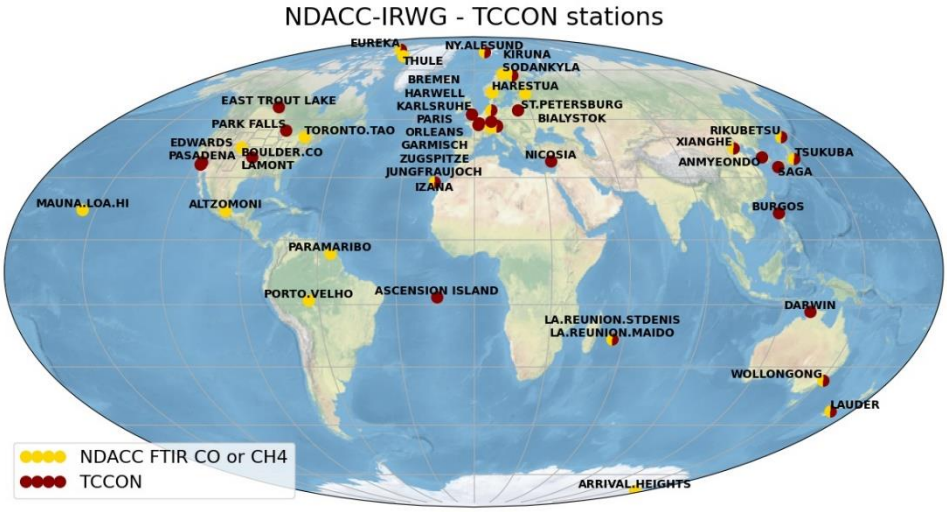


Figure 84: Geographical distribution of NDACC and TCCON FTIR stations measuring atmospheric carbon monoxide column data. Some stations contribute to the two networks.

11.2.2 Satellites

None for this report.

11.2.3 Field campaigns and modelling support

None for this report.

[1] Comparisons with COCCON will be included in the next ROCVR.

11.3 Validation of L2_CO OFFL

11.3.1 Recommendations for data usage

The Product Readme File (PRF) recommends the use of S5P data with a `qa_value` above 0.5 and the validation results reported hereafter are obtained by filtering the pixels using this recommendation.

For further details, data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this product: https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

11.3.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC and only uses the NL-L2 processor version 02.04 or higher. Validation results and consolidated validation reports for previous processors are available through the MPC VDAF Portal at http://mpc-vdaf.tropomi.eu.

Current conclusions are based on the amount of reference measurements available at the time of this analysis, yielding comparison pairs from April 2018 through July 2024. Routine validation is done using the Automated Validation Server of the MPC VDAF, the CO validation system operated at BIRA-IASB, and the HARP toolset.

TROPOMI observations co-located with the TCCON and NDACC measurements are found by selecting all filtered TROPOMI pixels within a radius of 50 km around each station and with a maximal time difference of 1h for TCCON and 3h for NDACC observations. The 1-hour interval can be justified by noting that TCCON instruments acquire only one type of spectra, while NDACC instruments are set up measure different types of spectra, making the number of available CO observations sparser. To reduce the influence of the two priors in the comparison, the S5P CO prior is substituted in the ground-based (NDACC and TCCON) measurement (Rodgers 2003). The validation procedure for the NDACC and TCCON based comparisons includes an adaptation of the TROPOMI CO column to the altitude of the ground-based FTIR instrument.

Since August 6, 2019 (orbit 9388), S5P measures with increased spatial resolution from 7km to 5.5km along track. This change in operations did not change the performance of the CO NRTI and OFFL product. TCCON released a new data version (GGG2020) on April 26, 2022 and this version has been used for the validation of S5P products in this report. Due to an anomaly in the TCCON CO prior at highly polluted sites, we see an increase in the bias for those sites (e.g. Caltech). As of 1 April 2024, TCCON changed its prior from GEOS FP-IT to GEOS IT. The result of this prior change is minimal for most gases, except for CO where a significant change is observed especially for urban sites.

Statistics are presented for two different selection rules on co-located pixels: a closest[2] pixel selection is used for the standard and destriped CO column and for the standard CO column, to reduce the noise caused by the striping pattern, a group of at least 5 and maximum 10 closest pixels is averaged and compared to a reference measurement.

[2] (in distance)

11.3.3 Bias

The systematic difference between S5P L2_CO daily mean data and correlative ground-based measurements is on an average 2% with respect to NDACC data and -2.5 % with respect to TCCON data see **Table 11**. At some stations higher values are seen (Toronto, Altzomoni) and these are likely due geographical colocation issues (mountain, city). For high latitudinal stations (>70deg) an increased positive bias is observed with values reaching 5% at Antarctica and 3% (Thule, Ny Alesund) to 7% (Eureka) for Arctic stations All individual bias estimates fall well within the mission requirements (15%). **Figure 85** does not reveal any significant degradation in bias with time.

Table 11 NDACC network statistics comparing the OFFL closest destriped CO columns against the standard CO columns with averaging of S5P pixels. Both comparison methods produce similar results.

	S5P-NDACC OFFL destriped					S5P-NDACC OFFL (pixel averaging)					
station	#	rel. std	correlation	rel diff bias(%)	std rel diff (%)	#	rel. std	correlation	rel diff bias(%)	std rel diff (%)	lat
EUREKA	591	0.7	0.91	6.73	6.54	583	0.8	0.96	7.39	4.06	80
NY.ALESUND	149	0.9	0.9	3.01	7.85	149	1	0.96	4.17	4.55	78.9
THULE	4924	0.9	0.93	3.02	6.02	4870	0.9	0.96	3.52	4.41	76.5
KIRUNA	957	0.8	0.9	0.94	7.26	951	0.9	0.96	0.78	4.18	67.8
HARESTUA	192	0.9	0.79	1.36	8.93	192	0.9	0.96	1.27	4.13	60.2
ST.PETERSBURG	1572	0.9	0.91	2.98	6.18	1573	0.9	0.93	2.83	5.09	59.9
BREMEN	384	0.9	0.88	0.77	5.52	365	1	0.93	0.44	4.09	53.1
GARMISCH	2801	0.8	0.78	6.42	9.09	2755	0.9	0.86	5.02	6.5	47.5
ZUGSPITZE	3555	0.9	0.77	0.18	8.03	3500	1	0.89	-0.63	5.24	47.4
JUNGFRAUJOCH	970	0.8	0.76	1.24	8.7	966	1	0.88	2.01	5.34	46.5
TORONTO.TAO	2551	0.9	0.86	6.71	9.33	2502	1	0.94	5.1	5.81	43.6
RIKUBETSU	110	0.8	0.82	-0.18	7.04	110	1	0.88	0.72	5.2	43.5
BOULDER.CO	5769	1	0.86	-0.89	7.19	5742	1	0.88	-0.67	6.56	40
XIANGHE	2887	1.2	0.94	2.65	7.9	2890	1.1	0.88	4.59	10.83	39.7
TSUKUBA	433	0.9	0.87	-0.24	5.98	432	1	0.86	0.73	6.19	36
IZANA	1459	1	0.74	0.14	6.84	1446	1	0.78	1.61	5.92	28.3
MAUNA.LOA.HI	333	0.8	0.78	-3.97	12.24	331	1	0.95	0.37	4.6	19.5
ALTZOMONI	645	0.8	0.76	3.81	11.48	639	0.9	0.65	5.26	11.93	19.1
PARAMARIBO	67	1.1	0.77	-3.45	6.86	67	1.1	0.71	-2.4	7.77	5.8
PORTO.VELHO	172	1	0.96	-0.51	9.79	168	1.1	0.93	-1.38	10.45	-8.8
LA.REUNION.MAIDO	1483	0.9	0.9	3.53	9.72	1473	0.9	0.95	4.19	6.3	-21.1
WOLLONGONG	3294	0.9	0.81	-0.1	14.89	3257	0.5	0.5	2.36	37.54	-34.4
LAUDER	4346	0.9	0.9	1.32	7.86	4310	0.9	0.95	1.19	5.12	-45
ARRIVAL.HEIGHTS	606	0.7	0.83	5.35	13.85	604	0.8	0.95	6.34	6.47	-77.8
		0.89	0.85	1.7	8.55		0.95	0.88	2.28	7.43	

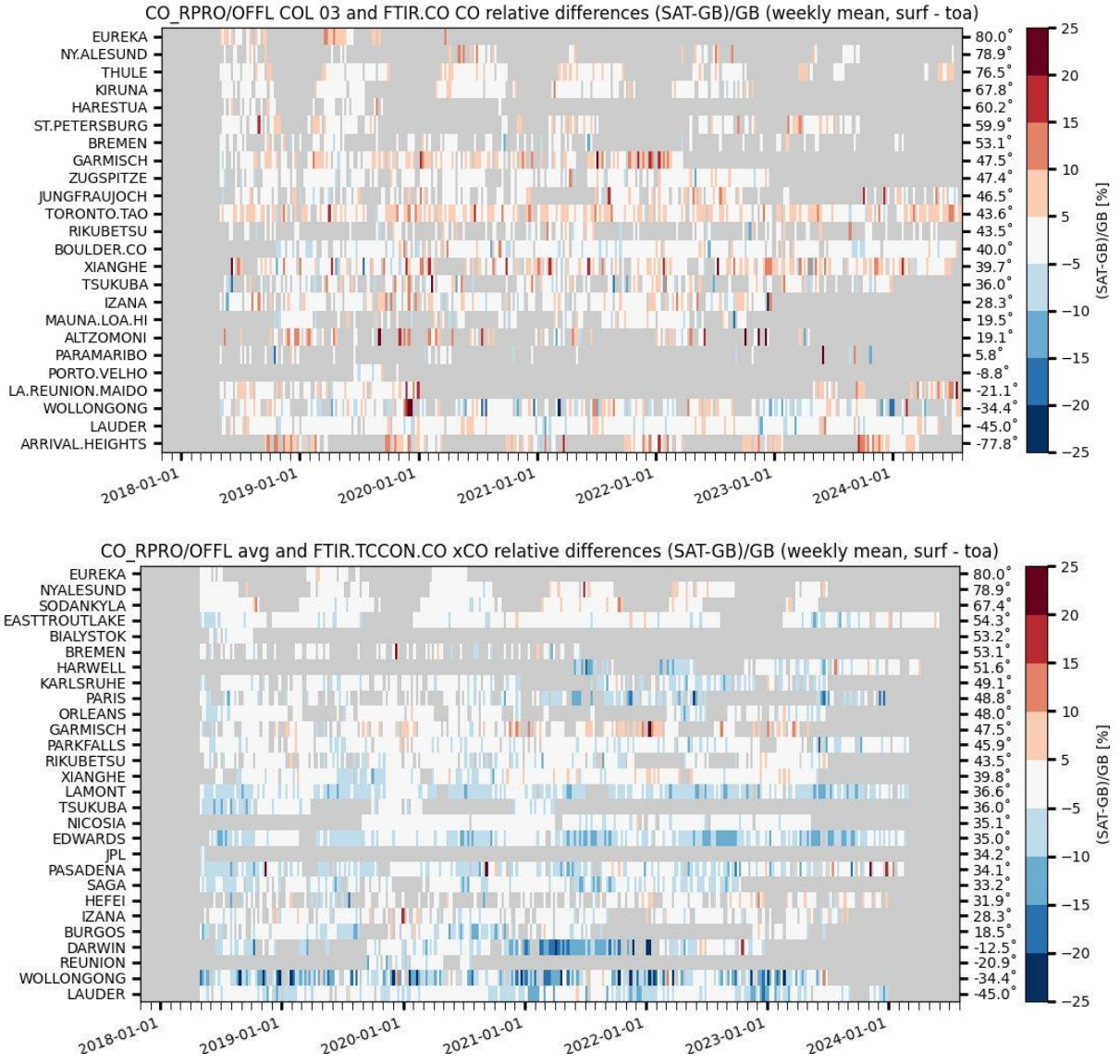


Figure 85: Relative bias between S5P L2_CO OFFL and ground-based CO column data at NDACC (top), TCCON (bottom) FTIR stations. Over the April 2018 – July 2024 period the plots do not show a clear meridian dependence or temporal change in the weekly averaged biases. Colour bar gradients are chosen such that they match with the measurement's uncertainties.

11.3.4 Dispersion

The one-σ dispersion of the relative mean bias around its mean is of the order of 8% for NDACC and 5% for TCCON. The individual values for the different NDACC stations are indicated in **Table 11.** This dispersion can be considered as an upper boundary of the random uncertainty of the satellite data. The dispersion on the differences when averaging pixels in the comparison is reduced with 1% for NDACC and 0.85 % for TCCON compared to the destriped CO product. This suggests that some random uncertainty remains present in the destriped product.

11.3.5 Dependence on influence quantities

For the evaluation of potential dependence of the S5P bias and spread on the Solar Zenith Angle (SZA), the comparisons for the stations in the northern hemisphere are grouped and the correlation plot in **Figure 87** shows an increase of the relative bias with the solar zenith angle of about 4% between 20deg and 80deg (this estimate uses the Theil-Sen). When using a rolling mean through the time series (based

on daily medians, with a 21-day window), a slight seasonal cycle is observed of approximately 2 % for the high latitude arctic sites with a lower bias in autumn and a higher positive bias in spring. These estimates fall within the reported reference measurement uncertainties and are therefore not considered significant.

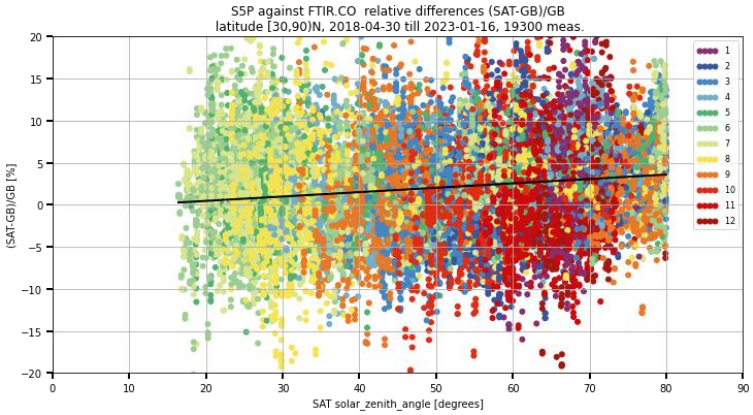


Figure 86 Relative differences versus solar zenith angle for the comparison with pixel averaging for all stations in the northern hemisphere (latitudes in [30,90]). The black line is a Theil-Senn slope estimate of approximately 0.05%/10deg.

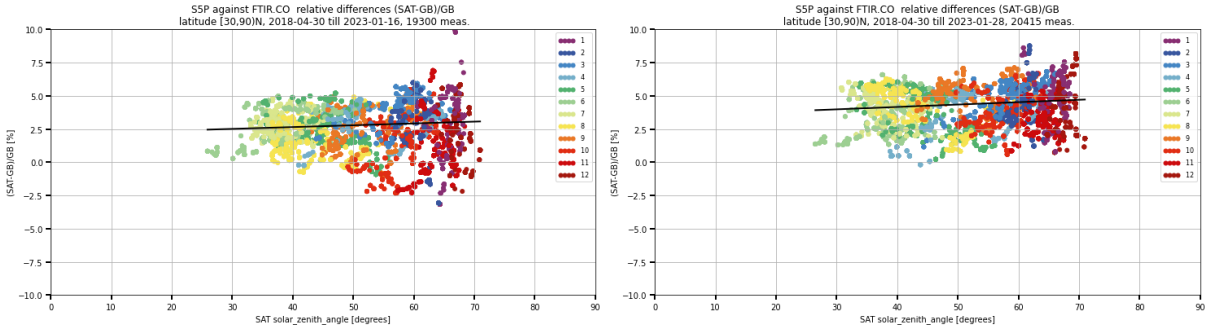


Figure 87 Left: similar to **Figure 86**, but with a rolling median (21-day window) applied to reduce the scatter and to show seasonal dependences of the relative differences. The right plot shows the comparisons where the NDACC profiles are smoothed with the TROPOMI averaging kernel, which reduces the seasonal dependence.

11.3.6 Short term variability

For all the NDACC and TCCON stations, short scale temporal variations in the CO column as captured by ground-based instruments are reproduced very similarly by S5P L2_CO OFFL. This overall good agreement is confirmed by individual Pearson correlation coefficients well above 0.75 for all sites and on average reaching 0.89 for the destriped product and above 0.9 for the comparison with pixel averaging (**Table 11**).

11.3.7 Geographical patterns

Individual S5P L2_CO column data show stripes of erroneous CO values below 10 % in the flight direction, and is probably associated to calibration issues of TROPOMI, see the CO PRF. This data quality issue leads to the provision of a “corrected” CO column in the L2 data files. The effect of the destriping method is discussed in Section 3.

11.3.8 Other features

NRTI granules from one S5P orbit have overlapping pixels. In order to avoid duplicated pixels in the validation statistics, pixels from the first 12 (before July 3 2019) or 16 (after July 3 2019) scanlines have been filtered.

11.4 Equivalence of L2_CO OFFL and NRTI products

The L2_CO NRTI processor uses the same settings as the OFFL processor. **Table 12** confirms that the statistical quality indicators for both OFFL and NRTI products are very similar.

Table 12 NDACC network statistics comparing the OFFL data from collection 3 with the NRT data (since July 21 2022). Both products OFFL and NRT behave very similar.

	S5P-NDACC OFFL C3					S5P-NDACC NRT C3					
station	#	rel. std	correlation	rel diff bias(%)	std rel diff (%)	#	rel. std	correlation	rel diff bias(%)	std rel diff (%)	lat
NY.ALESUND	36	1.1	0.96	5.04	5.56	34	1.1	0.96	4.93	5.71	78.9
THULE	1023	0.9	0.96	4.23	4.51	1012	1	0.96	4.21	4.55	76.5
KIRUNA	52	0.7	0.55	0.49	4.7	52	0.6	0.55	0.56	4.79	67.8
ST.PETERSBURG	399	0.9	0.94	3.59	5.08	400	0.9	0.94	3.69	5.18	59.9
BREMEN	52	0.8	0.87	1.03	4.07	50	0.8	0.87	1.01	4.16	53.1
ZUGSPITZE	346	0.9	0.75	-0.07	5.78	333	0.9	0.75	-0.08	5.82	47.4
JUNGFRAUJOCH	253	0.9	0.82	2.91	6.71	250	0.9	0.82	2.95	6.65	46.5
TORONTO.TAO	889	0.9	0.94	4.68	6.41	864	0.9	0.94	4.77	6.49	43.6
RIKUBETSU	31	1	0.85	1.7	6.06	31	1	0.84	1.53	6.21	43.5
BOULDER.CO	2465	1	0.89	-0.48	5.66	2418	1	0.89	-0.55	5.7	40
XIANGHE	1241	1.2	0.88	5.69	10.54	1225	1.1	0.88	5.84	10.54	39.7
TSUKUBA	132	0.9	0.86	0.44	5.49	133	0.9	0.86	0.38	5.52	36
IZANA	142	1.9	0.27	0.91	9.5	142	1.9	0.27	0.93	9.46	28.3
MAUNA.LOA.HI	15	0.4	0.62	-0.75	6.52	15	0.4	0.67	-0.22	6.46	19.5
ALTZOMONI	38	1	0.2	6.94	17.48	39	1	0.25	5.94	16.18	19.1
PARAMARIBO	7	1.4	0.67	0.24	15.43	7	1.4	0.67	0.32	15.46	5.8
LA.REUNION.MAIDO	388	0.7	0.88	6.04	7.19	387	0.7	0.87	5.88	7.36	-21.1
WOLLONGONG	1071	1	0.82	-0.02	10.68	1075	1	0.81	-0.01	10.82	-34.4
LAUDER	1054	1	0.96	0.62	5.21	1032	0.9	0.96	0.58	5.26	-45
ARRIVAL.HEIGHTS	180	0.8	0.95	7.48	6.54	179	0.8	0.95	7.69	6.54	-77.8
		0.97	0.78	2.54	7.46		0.97	0.79	2.52	7.44	

11.5 Comparison of L2_CO OFFL standard and de-striped products

A comparison between the standard and de-striped column using the “closest” pixel comparison shows that the de-striped product has reduced dispersion in the relative differences and slightly higher correlation (see also the results on the VDAF server). Relative bias do not show significant changes.

Table 13 NDACC network statistics comparing the OFFL destriped CO columns against the standard CO columns available in the S5P L2 (closest pixel comparison). The dispersion on relative differences is reduced with approximately 1 % in the destriped product.

	S5P-NDACC OFFL destriped					S5P-NDACC OFFL standard product					
station	#	rel. std	correlation	rel diff bias(%)	std rel diff (%)	#	rel. std	correlation	rel diff bias(%)	std rel diff (%)	lat
EUREKA	591	0.7	0.91	6.73	6.54	591	0.7	0.91	7.01	6.77	80
NY.ALESUND	149	0.9	0.9	3.01	7.85	149	0.9	0.88	3.27	8.48	78.9
THULE	4924	0.9	0.93	3.02	6.02	4927	0.9	0.91	2.94	6.96	76.5
KIRUNA	957	0.8	0.9	0.94	7.26	957	0.8	0.88	1.32	8.16	67.8
HARESTUA	192	0.9	0.79	1.36	8.93	192	0.8	0.75	1.16	10.15	60.2
ST.PETERSBURG	1572	0.9	0.91	2.98	6.18	1573	0.8	0.89	2.88	7.04	59.9
BREMEN	384	0.9	0.88	0.77	5.52	368	0.9	0.87	0.77	5.98	53.1
GARMISCH	2801	0.8	0.78	6.42	9.09	2801	0.8	0.75	7.21	10.07	47.5
ZUGSPITZE	3555	0.9	0.77	0.18	8.03	3555	0.9	0.73	0.77	9.16	47.4
JUNGFRAUJOCH	970	0.8	0.76	1.24	8.7	966	0.8	0.71	0.84	9.95	46.5
TORONTO.TAO	2551	0.9	0.86	6.71	9.33	2553	0.9	0.84	6.64	10.48	43.6
RIKUBETSU	110	0.8	0.82	-0.18	7.04	110	0.8	0.79	0.7	8.02	43.5
BOULDER.CO	5769	1	0.86	-0.89	7.19	5764	1	0.8	-0.36	8.97	40
XIANGHE	2887	1.2	0.94	2.65	7.9	2894	1.1	0.93	3.07	8.55	39.7
TSUKUBA	433	0.9	0.87	-0.24	5.98	432	0.9	0.86	-0.33	6.4	36
IZANA	1459	1	0.74	0.14	6.84	1459	0.9	0.68	0.17	8.15	28.3
MAUNA.LOA.HI	333	0.8	0.78	-3.97	12.24	333	0.7	0.73	-3.53	14.24	19.5
ALTZOMONI	645	0.8	0.76	3.81	11.48	644	0.8	0.72	4.48	12.85	19.1
PARAMARIBO	67	1.1	0.77	-3.45	6.86	67	1.1	0.72	-3.69	7.59	5.8
PORTO.VELHO	172	1	0.96	-0.51	9.79	172	1	0.95	-0.38	10.05	-8.8
LA.REUNION.MAIDO	1483	0.9	0.9	3.53	9.72	1477	0.9	0.88	3.57	11.22	-21.1
WOLLONGONG	3294	0.9	0.81	-0.1	14.89	3299	0.8	0.79	-0.61	16.37	-34.4
LAUDER	4346	0.9	0.9	1.32	7.86	4346	0.8	0.85	1.02	9.89	-45
ARRIVAL.HEIGHTS	606	0.7	0.83	5.35	13.85	607	0.7	0.81	4.85	15.63	-77.8
		0.89	0.85	1.7	8.55		0.86	0.82	1.82	9.63	

12 Validation Results: L2_CH4

12.1 L2_CH4 products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_CH4 product identified in **Table 1**: the methane total column. Validation results are discussed with respect to the product quality targets outlined in **Table 3**.

12.2 Validation approach

12.2.1 Ground-based networks

S5P TROPOMI L2_CH4 methane column data are routinely compared to reference measurements obtained from FTIR spectrometers performing network operation in the context of the Network for the Detection of Atmospheric Composition Change (NDACC, http://ndacc.org), the Total Carbon Column Observing Network (TCCON, https://tccondata.org) and the Collaborative Carbon Column Observing Network (COCCON[3], https://www.imk-asf.kit.edu/english/COCCON.php). **Figure 84** displays the geographical distribution of the NDACC and TCCON stations. Near-infrared TCCON measurements provide calibrated methane column averaged (xCH_4) data with typical uncertainty values of 0.5% for the precision and 0.2% for the accuracy. The COCCON measurements are calibrated to TCCON and show similar performance as TCCON. Solar infrared NDACC measurements provide CH_4 total column data with a lower accuracy (typically 3%) and precision (1.5%). The required accuracy (<1.5%) and precision (<1%) for S5P implies that we mainly focus on the validation with TCCON and COCCON measurements.

12.2.2 Satellites

XCH4 measurements by the Thermal and Near Infrared Sensor for Carbon Observation Fourier transform spectrometer (TANSO-FTS) on board the Greenhouse gases Observing SATellite (GOSAT) satellite are used for the validation of the TROPOMI XCH4 data. GOSAT was launched in 2009, and it performs three-point observations in a cross-track swath of 790 km with 10.5 km resolution on the ground at nadir, which results in global coverage approximately every 3 days.

We use the GOSAT proxy XCH4 data product produced at SRON in the context of the ESA GreenHouse Gas Climate Change Initiative (GHG CCI) project (Buchwitz et al., 2019, 2017). This XCH4 product is retrieved using the RemoTeC/proxy retrieval algorithm. The proxy approach (Frankenberg et al., 2005) infers a CO2 and CH4 total column from observations at 1.6 μm, ignoring any atmospheric scattering in the retrieval.

To compare TROPOMI and GOSAT XCH4, we compute daily mean XCH4 in a 2x2-degree grid, and then estimate the averaged bias and its standard deviation.

[3] Comparisons with COCCON will be included in the next ROCVR.

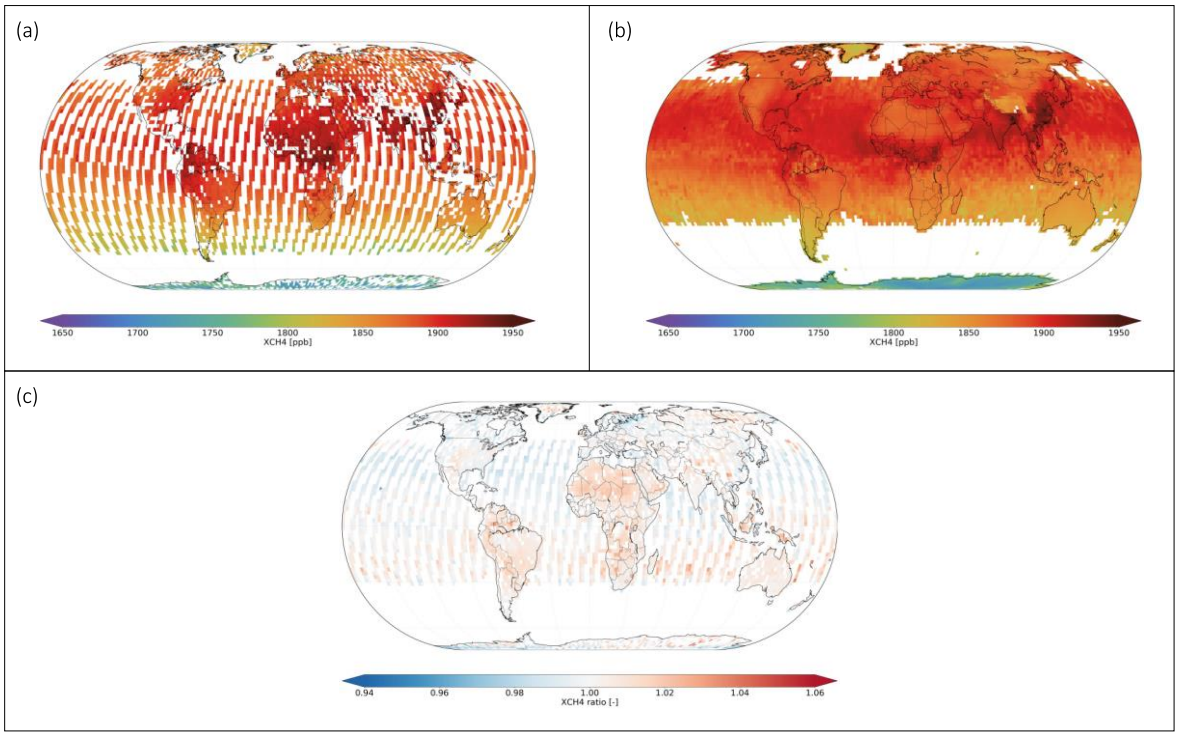


Figure 88: Global distribution of XCH4 measured by (a) GOSAT and (b) TROPOMI and (c) the ratio of GOSAT to TROPOMI XCH4. Daily collocations are averaged to a 2x2 degree grid for the year 2022.

Figure 88 shows XCH_4 measured by GOSAT, TROPOMI, and the ratio of both. The comparison yields a mean bias TROPOMI-GOSAT of -7.7 ppb ± 18.3 ppb (−0.41 ± 0.97 %) and a Pearson's correlation coefficient of 0.83.

12.2.3 Other TROPOMI XCH4 products

Besides the operational TROPOMI XCH4 product, there is a scientific product using the weighting function modified differential optical absorption spectroscopy (WFM-DOAS) method to retrieve CO and CH4 (Schneising et al., 2019). Here we compare the pre-operational TROPOMI XCH4 product to the WFMD product.

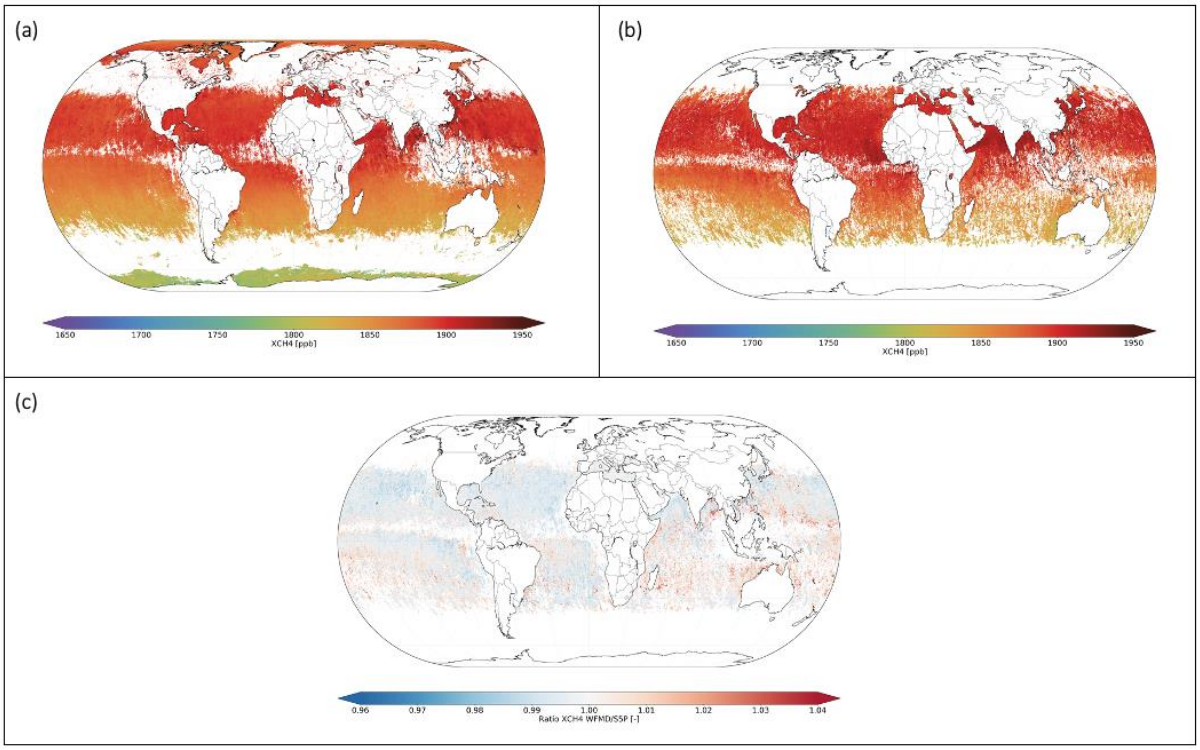


Figure 89: Global distribution of XCH4 measured by TROPOMI retrieved with (a) WFMD-DOAS retrieval algorithm, (b) the pre-operational retrieval algorithm and (c) the ratio of both. Daily collocations are averaged to a 0.1 x 0.1-degree grid for the year 2022.

To compare the two TROPOMI XCH4 products, we compute yearly averaged bias and its standard deviation. **Figure 90** shows XCH4 retrieved over 2022 by WFMD-DOAS algorithm and by the pre-operational algorithm, together with their ratio. The WFMD-DOAS product covers areas over the ocean further away than the sun-glint limit and it also does not perform a strict cloud filtering as the operational retrieval. Over the ocean, the comparison yields a mean bias of -3.0 ppb ± 16.4 ppb (-0.16 ± 0.9 %) and a Pearson's correlation coefficient of 0.81.

12.2.4 Field campaigns and modelling support

None for this report.

12.3 Validation of L2_CH4 OFFL

12.3.1 Recommendations for data usage followed

The Product Readme File (PRF) recommends the use of only S5P data with a `qa_value` above 0.5 and the validation results hereafter are obtained by filtering the pixels using this recommendation.

The S5P L2 data contains two xCH_4 column values: the standard retrieved product and a bias corrected product. Both products are validated separately, but only the bias corrected is mentioned in the quality indicators in **Table 2**.

For further details, data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product: https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

12.3.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC and only uses the NL-L2 processor version 02.04 or higher. Validation results and consolidated validation reports for previous processors are available through the MPC VDAF Portal.

TROPOMI observations co-located with the ground-based FTIR measurements are found by selecting all filtered TROPOMI pixels within a radius of 100 km around each station and with a maximal time difference of 1h for TCCON and COCCON and 3h for NDACC observations. The 1-hour interval can be justified by noting that TCCON instruments acquire only one type of spectra, while NDACC instruments are required to measure different types of spectra, making the number of CH_4 observations sparser. In the comparison, the a priori in the TCCON and NDACC retrievals have been substituted with the S5P CH_4 a priori (Rodgers 2003). For NDACC data the method described in Rodgers (2003) is followed one step further and the FTIR CH_4 concentration profile (with the S5P prior substituted) is additionally smoothed with the S5P column averaging kernel. The validation procedure for the NDACC and TCCON based comparisons includes an adaptation of the TROPOMI CH_4 column to the altitude of the ground-based FTIR instrument.

Current conclusions are based on the S5P and reference measurements available at the time of this analysis, which yield comparison pairs from April 2018 through July 2024. Routine validation is done using the Automated Validation Server of the MPC VDAF, the CH_4 validation system operated at BIRA-IASB, and the HARP toolset.

Since August 6 2019 (orbit 9388) S5P measures with increased spatial resolution from 7km to 5.5km along track. This change in operations did not change the performance of the methane OFFL product. TCCON released a new data version (GGG2020) on April 26, 2022. This data has been used for the validation of S5P products and shown in this report. As of 1 April 2024, TCCON changed its prior from GEOS FP-IT to GEOS IT. The result of this prior change is minimal for CH_4.

Statistics are presented for the selection rules on co-located pixels where to reduce the noise a group of at least 5 and maximum 10 closest pixels is averaged and compared to a reference measurement.

12.3.3 Bias

The systematic difference (the mean of all relative differences) between S5P L2_CH4 and TCCON dry air methane column data is on an average -0.34 % (standard) and +0.22 % (bias corrected), well within the mission requirements. Comparisons against NDACC in **Table 14** estimate the relative difference of -0.85 % for the standard S5P methane product and -0.06 % for the bias corrected product, both estimates are below the reported systematic uncertainties. The NDACC and TCCON reference data are obtained from different spectral regions and explains the difference in bias estimates.

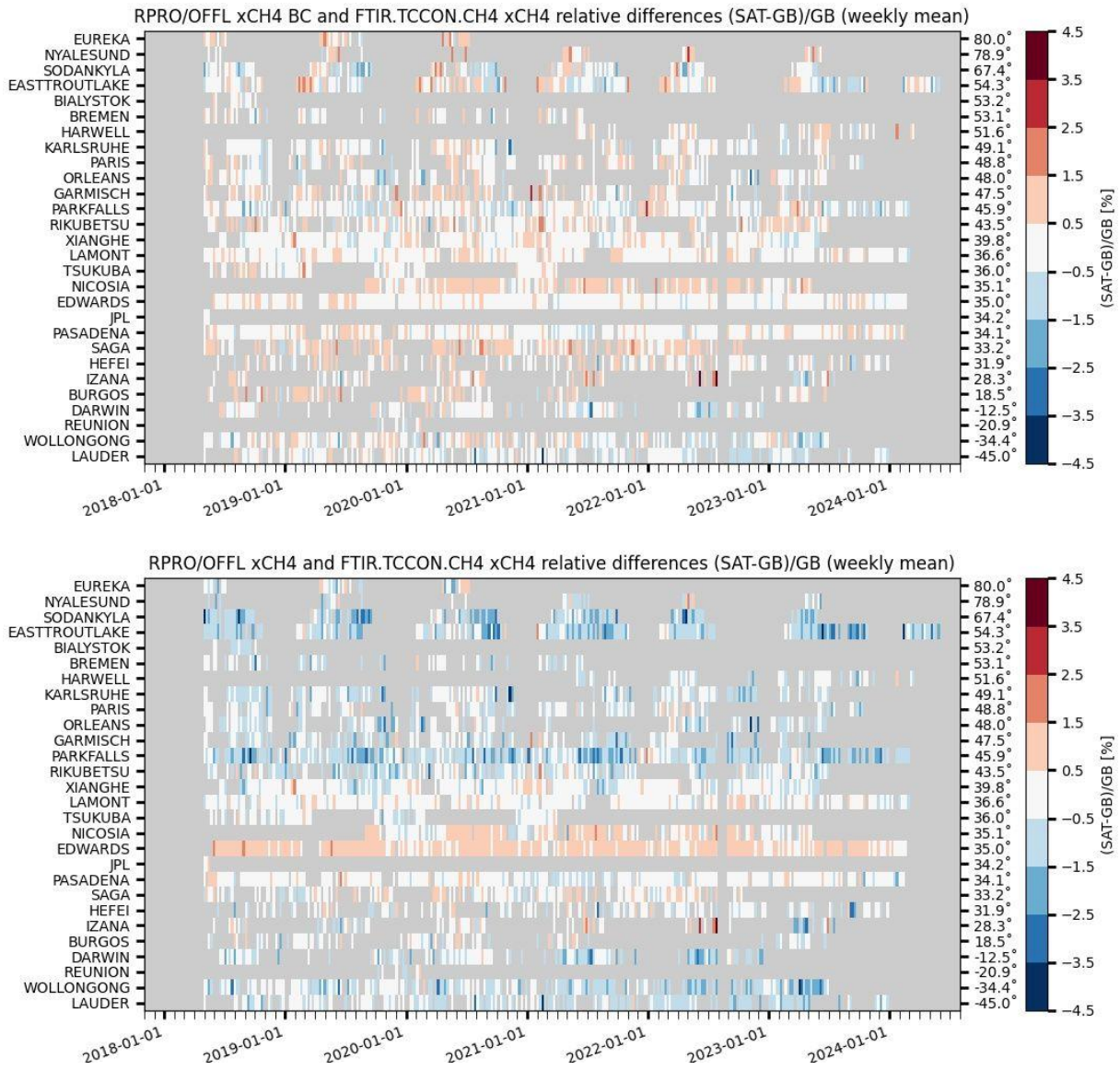


Figure 91: Mosaic plots of relative biases between S5P L2_CH4 RPRO+OFFL and ground-based CH_4 TCCON column data for the bias corrected (top) and standard (bottom) methane products. Over the April 2018 – July 2024 time period the plots do not show a clear meridian dependence or temporal change in the weekly averaged biases. Colour bar gradients are chosen such that they match with the measurement's uncertainties.

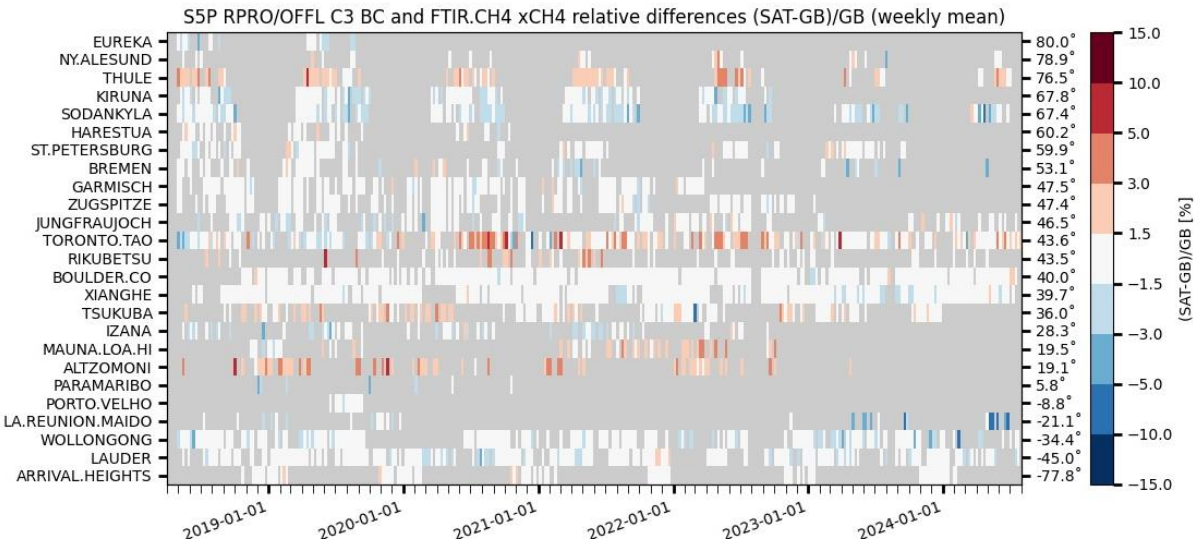


Figure 92 Mosaic plots of relative biases between S5P L2_CH4 RPRO+OFFL and ground-based CH_4 NDACC column data for the bias corrected methane products. Over the April 2018 – July 2024 time period the plots do not show a clear meridian dependence or temporal change in the weekly averaged biases.

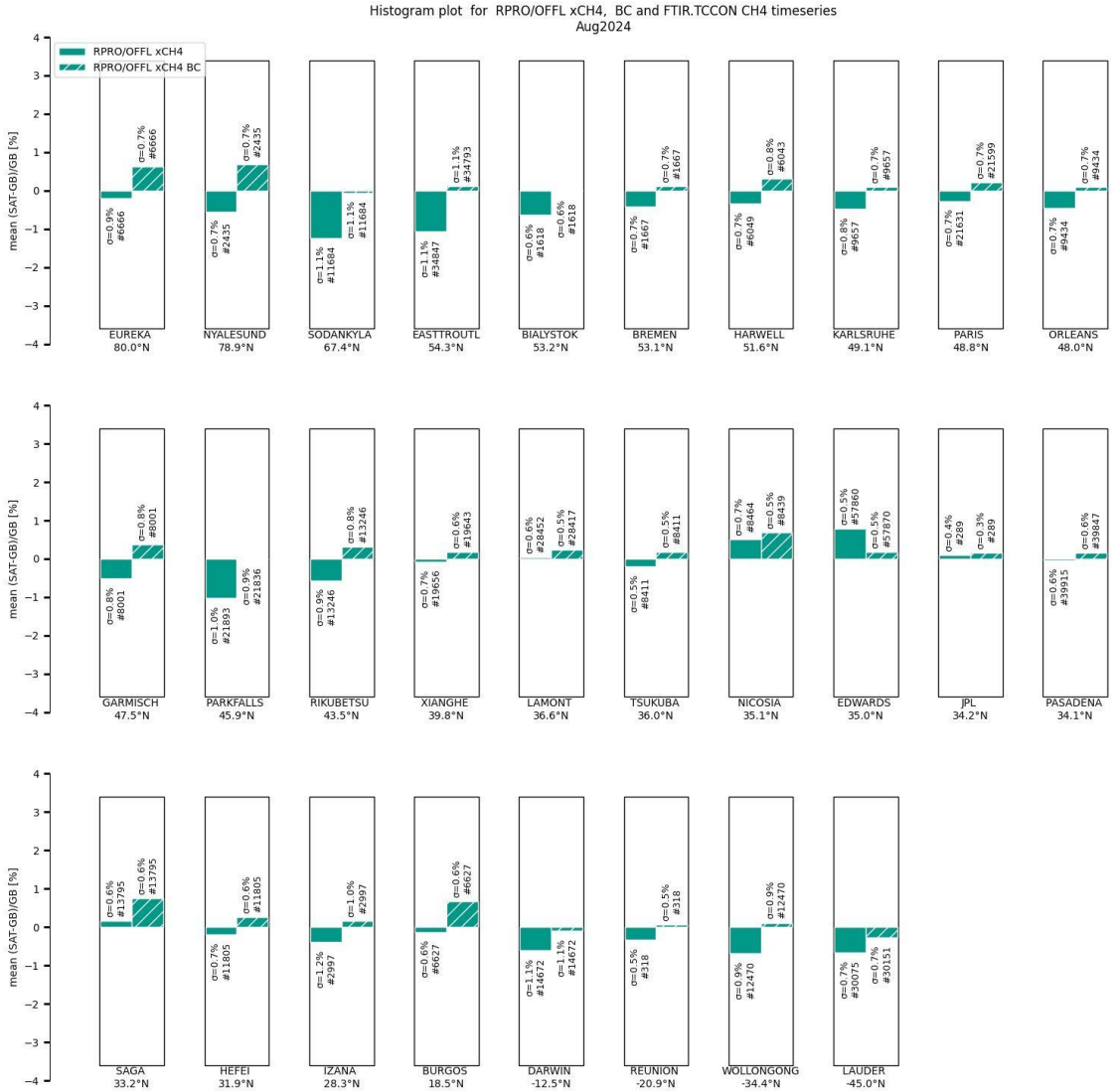


Figure 93: Chart of relative mean difference between S5P L2_CH4 and FTIR CH_4 column data at 24 TCCON stations within the time range April 2018 until July 2024. The stations are sorted with decreasing latitude. The relative mean difference of the standard and bias-corrected xCH4 products never exceeds the mission requirements (bias below 1.5 %).

	S5P-NDACC xCH4 smooth 100km 3hr					S5P-NDACC xCH4 smooth 100km bc 3hr					
station	#	rel. NDACC std	correlation	rel diff bias(%)	std rel diff (%)	#	rel. NDACC std	correlation	rel diff bias(%)	std rel diff (%)	lat
EUREKA	304	1.5	0.3	-1.95	1.3	304	1.6	0.57	-1.06	1.06	80
NY.ALESUND	104	1.5	0.48	-0.61	1.59	104	1.5	0.63	0.61	1.4	78.9
THULE	964	0.8	0.62	0.75	1.33	964	0.8	0.63	2.06	1.32	76.5
KIRUNA	707	0.8	0.7	-2.28	1	707	0.8	0.71	-1.15	0.99	67.8
SODANKYLA	7192	1.1	0.66	-2.37	1.34	7175	1.1	0.68	-1.18	1.31	67.4
HARESTUA	183	1.2	0.43	-0.29	1.11	183	1.2	0.55	0.88	1	60.2
ST.PETERSBURG	1238	1	0.64	-0.87	1.12	1238	1	0.71	0.07	1.01	59.9
BREMEN	255	1.9	0.65	-0.76	1.71	255	1.8	0.68	-0.22	1.66	53.1
GARMISCH	1406	0.8	0.65	-0.9	0.9	1406	0.9	0.68	-0.01	0.83	47.5
ZUGSPITZE	1238	1.2	0.6	-1.11	1.11	1238	1.2	0.64	-0.22	1.08	47.4
JUNGFRAUJOCH	666	1	0.56	-1.09	1.22	666	1	0.57	-0.43	1.22	46.5
TORONTO.TAO	1378	1.8	0.18	0.24	2.77	1382	1.8	0.16	0.87	2.82	43.6
RIKUBETSU	57	1.3	0.7	0.22	1.51	57	1.4	0.71	1.07	1.51	43.5
BOULDER.CO	4017	1.1	0.66	-0.5	0.98	4017	1.1	0.67	-0.3	0.96	40
XIANGHE	2966	1.1	0.83	-0.71	0.91	2964	1.1	0.85	-0.47	0.86	39.7
TSUKUBA	313	1.4	0.75	0.79	1.17	313	1.4	0.75	1.22	1.17	36
IZANA	184	0.7	0.64	-1.7	1.14	184	0.7	0.66	-1.2	1.09	28.3
MAUNA.LOA.HI	74	1	-0.03	0.13	1.9	74	0.9	0.06	0.89	1.83	19.5
ALTZOMONI	440	1.2	0.78	2.47	1.02	440	1.2	0.79	2.69	1	19.1
PARAMARIBO	10	1.2	-0.44	-3.15	1.52	10	0.9	-0.4	-2.23	1.74	5.8
PORTO.VELHO	87	0.9	0.21	-1.8	0.97	87	1.1	0.28	-0.86	0.88	-8.8
LA.REUNION.MAIDO	254	0.9	-0.21	-3.68	3.14	254	0.9	0.04	-2.56	2.75	-21.1
WOLLONGONG	2270	1.1	0.66	-1.5	1.15	2271	1	0.71	-0.68	1.1	-34.4
LAUDER	2660	1.1	0.72	-1.2	1.08	2658	1.2	0.71	-0.82	1.1	-45
ARRIVAL.HEIGHTS	256	1.1	0.74	-0.97	0.89	256	1	0.78	0.32	0.84	-77.8
		1.14	0.5	-0.91	1.36		1.13	0.55	-0.11	1.3	

Table 14 – Overview of statistical quality indicators for the co-located S5P and NDACC time series. The relative mean difference of the corrected xCH4 product slightly exceeds the mission requirements (bias below 1.5 %) only at a few NDACC stations (i.e. Thule, Maido and Altzomoni).

12.3.4 Dispersion

The 1σ spread of the relative difference (between the S5P and the TCCON methane column data) around the mean value is of the order of 0.76 % for standard products and 0.71 % for bias corrected product. This dispersion can be considered as an upper boundary of the random uncertainty of the satellite data. The values for the individual stations are mostly below the mission requirements (precision <1%) for both the bias corrected and standard products. Except for four stations (Sodankylä, East Trout Lake, Izaña, and Darwin) where the value is around the limit.

Because NDACC measurements are reported with a higher random uncertainty, the NDACC estimate for the precision of 1.5 % therefore exceeds the mission requirement and should be ignored.

12.3.5 Dependence on influence quantities

At this stage, the evaluation of potential dependence of the S5P bias and spread on the Solar Zenith Angle (SZA) is hard to evaluate: at high latitude stations e.g., Sodankylä and Kiruna, the bias during spring and autumn (both seasons have high SZA) changes sign.

The relative differences show a dependence on the surface albedo, which is corrected in the bias corrected product. The relative difference of the bias corrected product shows a remaining weak dependence in low albedo case.

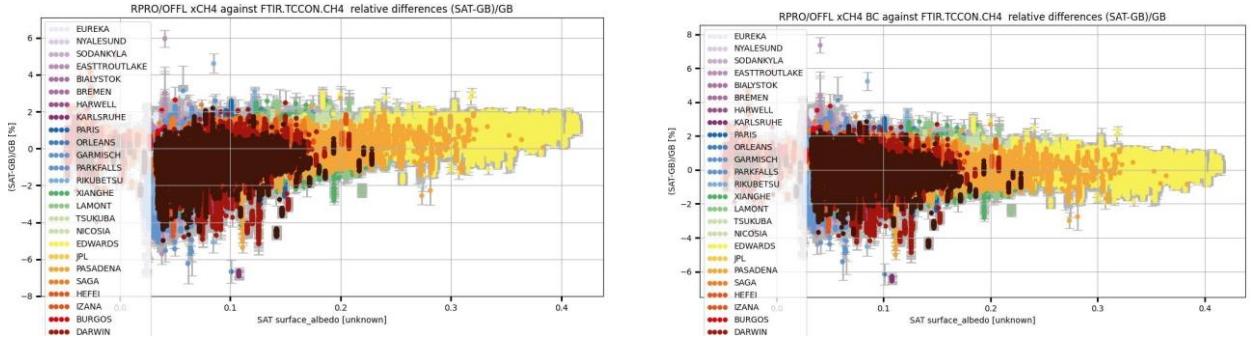


Figure 94: Dependence of the S5P-TCCON relative difference on surface albedo. The left column shows the standard S5P product and the right column the bias corrected S5P product. The bias correction removes the surface albedo dependence of the standard S5P product.

12.3.6 Short term variability

For all the NDACC and TCCON stations, short scale temporal variations in the CH_4 column as captured by ground-based instruments are reproduced very similarly by S5P L2_CH4 OFFL. The individual Pearson correlation coefficients are on average 0.75 for standard product and 0.78 for bias corrected product for all TCCON sites.

12.3.7 Other features

Filtering on `qa_value` > 0.5 does not remove all pixels considered bad. Some pixels with too low and too high methane concentrations are still present.

Outlying methane values are observed along coastline regions or mountain regions, see for example Greenland in **Figure 96**.

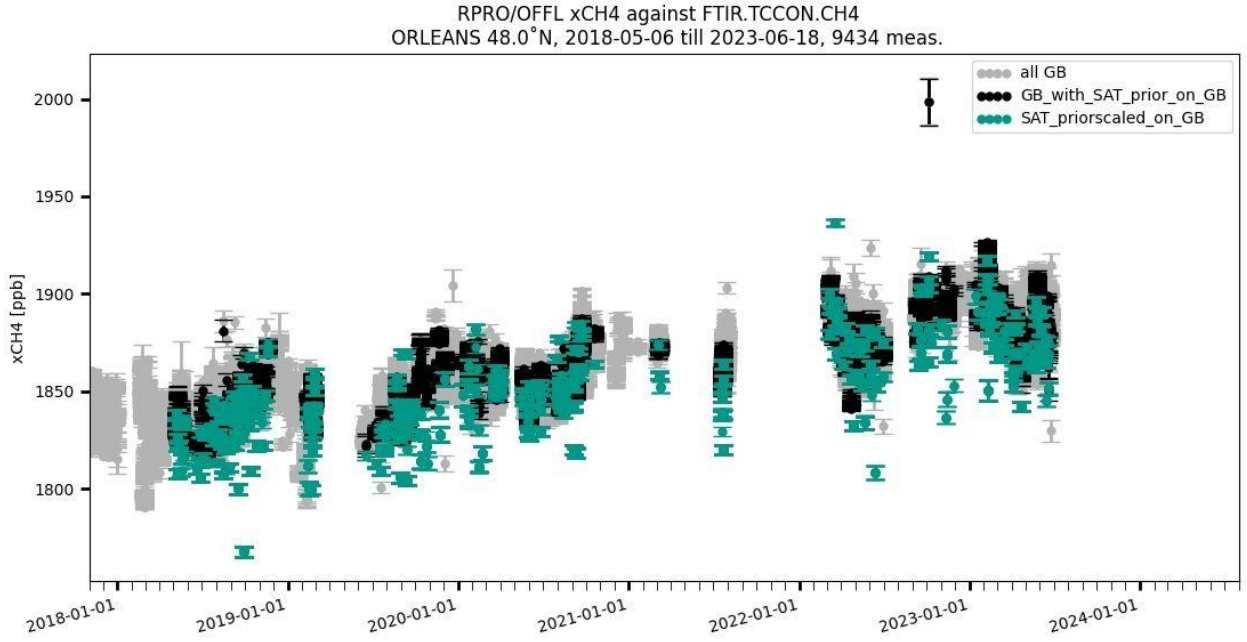


Figure 95: S5P L2_CH4 XCH_4 time series over Orleans where low values of XCH_4 are observed for several days.

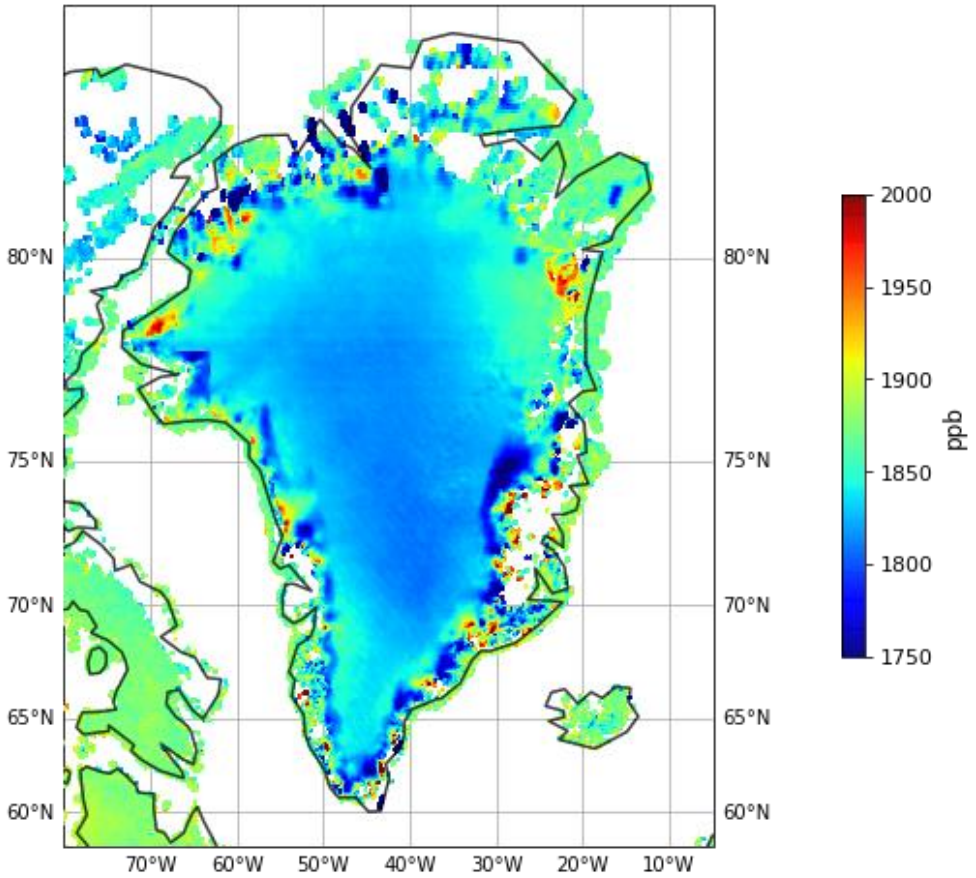


Figure 96: Map showing XCH_4 concentrations above Greenland and parts of North America for three years averaged data with pixels with qa_value > 0.5. The XCH_4 pixels in Greenland shows outliers along the mountain and coastline regions. Similar features are observed at the Antarctic.

12.4 Validation of L2_CH4 RPRO+OFFL sun-glint retrievals

Ocean measurements performed over sun-glint geometry improve significantly the TROPOMI XCH_4 product coverage. Since sun-glint geometry depends on the position of the sun relative to the satellite, these measurements have a seasonal cycle and different parts of the ocean are covered by the sun-glint geometry at different times of the year. It covers approximately an area of 30 degrees extension in latitude (see **Figure 97** – right). In November-February it covers mostly the Northern hemisphere, and April-September it covers mostly the Southern hemisphere. The typical annual coverage is shown in **Figure 97** – left, which corresponds to the year 2022.

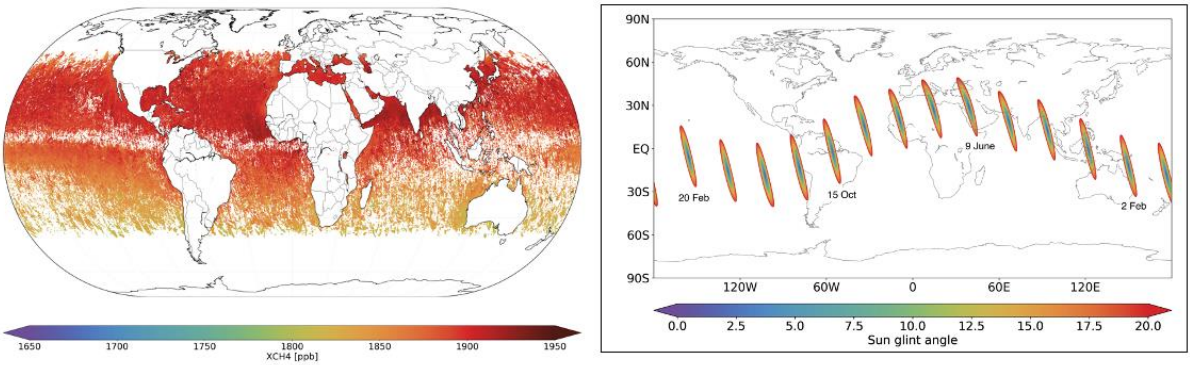


Figure 97: (left) Annual coverage of XCH_4 retrievals over the ocean for the year 2022, and (right) seasonal cycle of the sun-glint geometry.

To evaluate the sun-glint retrievals, the scientific L2 product for the period between April 2018 and September 2023 is used. A similar validation approach is used as for the standard methane L2 data, except that an additional filtering is enabled to keep only sun-glint pixels over ocean. Only the bias-corrected methane values in the scientific L2 products are considered.

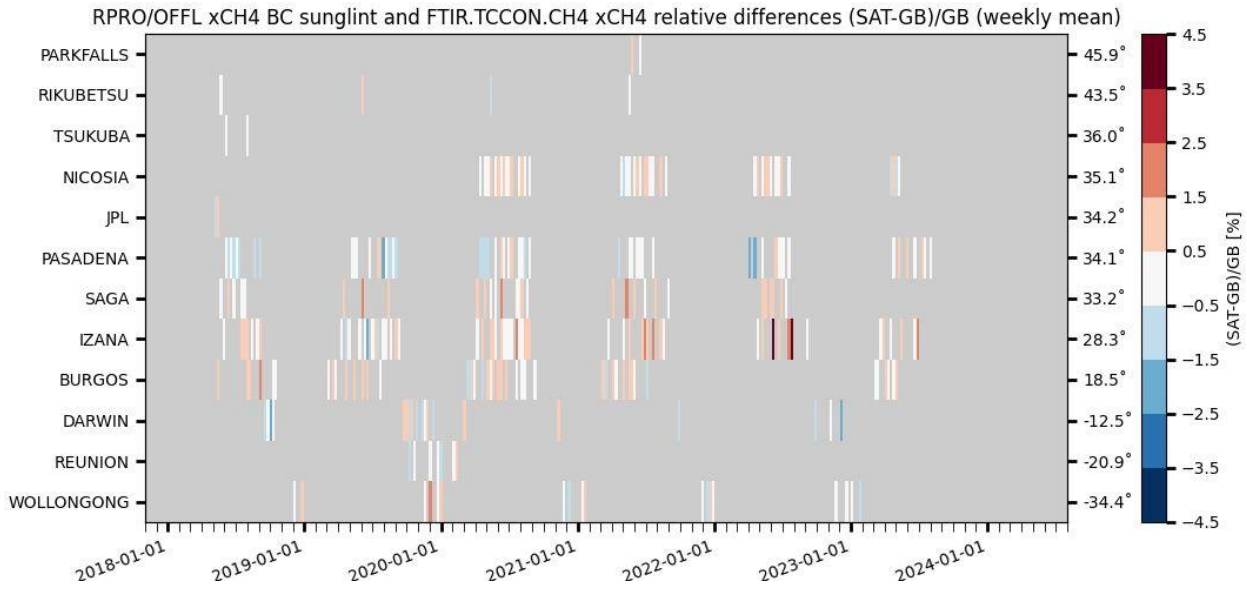


Figure 98: Mosaic plots of relative biases between S5P scientific L2_CH4 product and ground-based CH_4 TCCON column data for the bias corrected methane products.

The systematic difference between the S5P scientific L2_CH4 bias-corrected product and the TCCON XCH_4 data is on average 0.21 %, well within the mission requirements. The 1σ spread of the relative difference between the S5P and the TCCON data around the mean value is 0.67 %, also below the mission requirements (precision <1%). The individual values for the different stations are indicated in **Figure 93**.

For NDACC, six sites have co-locations with sun-glint pixels, of which only four have sufficient co-locations to build a robust statistic. These sites are a mixture of island sites or sites close to a coastline. Although the number of co-locations is low, the bias and precision are close to the values of the standard S5P bias-corrected XCH_4 product.

	SRON-NDACC xCH4 bc sun-glint, 100km 3h				
station	**#**	**rel. std**	**correlation**	**rel diff bias(%)**	**std rel diff (%)**
IZANA	**164**	**0.7**	**0.66**	**-1.18**	**0.99**
MAUNA.LOA.HI	**31**	**1.5**	**0.24**	**1.52**	**1.54**
LA.REUNION.MAIDO	**108**	**1.1**	**0.78**	**-0.93**	**0.59**
WOLLONGONG	**63**	**1.3**	**0.72**	**-0.49**	**0.87**
		1.46	**0.56**	**0.36**	**1.33**

Table 15 – Overview of statistical quality indicators for the co-located S5P sun-glint pixels and NDACC. All quality indicators are of the same order of magnitude as the values for the standard S5P bias corrected product.

13 Validation Results: L2_CLOUD

13.1 L2_CLOUD products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_CLOUD product identified in **Table 1**: the (radiometric) Cloud Fraction (CF), the Cloud Top Height (CTH)/Cloud Height (CH), and the Cloud Optical Thickness (COT)/ Cloud Albedo (CA). There are actually two sub-products within the L2_CLOUD files: OCRA/ROCINN_CAL (providing CF, CTH, COT) and OCRA/ROCINN_CRB (providing CF, CH, CA). Shorthand notation used here for both sub-products is CLOUD CAL and CLOUD CRB. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. The results presented here cover only processor version 2.4.1 and higher, which includes the data of the full-mission reprocessing (designated "RPRO" here). For validation results on other previous processor versions, the reader is referred to ROCVR version 17.1.0, available at https://mpc-vdaf.tropomi.eu/ Subsection 13.3 focuses on the validation of the L2_CLOUD OFFL product while Subsection 13.4 demonstrates evidence that NRTI and OFFL data do not differ significantly and that their respective validations yield similar conclusions.

13.1.1 Differences between validation results before and after version 2.4.1 reprocessing

Up to and including ROCVR-17, the validation of the S5P cloud products was performed on a mix of processor versions (1.1.7 up to 2.4). Starting from ROCVR-18, following the full mission reprocessing, the validation of the S5P cloud products is performed on purely 2.4 data. We performed a comparative analysis in February 2023, using the same methodology as in ROCVR-17 for both datasets. Users that want to change to the data set including the 2.4 reprocessing should be aware of the following:

- With the inclusion of the reprocessed 2.4 data, **artificial jumps** in cloud fraction and cloud height, as reported up to ROCVR-17, are resolved.
- Considering the difference in **scaled radiometric cloud fraction** of **OCRA/ROCINN CRB** versus **FRESCO-S** over Cloudnet sites:
    - The difference dispersion improves at high-latitude sites when using the reprocessed 2.4 data.
    - Compared to processor version 1 data, the difference dispersion of processor 2.4 data increases at other sites.
- Considering the difference in **cloud height** of **OCRA/ROCINN CRB** versus **FRESCO-S** over Cloudnet sites
    - The difference dispersion improves when using the reprocessed 2.4 data.
    - The mean difference is now about -500 m (CRB lower than FRESCO). This is larger than with processor version 1 data, but smaller than with ROCINN CRB version 2.2 data.
- Considering the difference in cloud height of **OCRA/ROCINN CAL** versus **FRESCO-S** over Cloudnet sites
    - The difference dispersion improves when using the reprocessed 2.4 data.
    - Over land sites, the ROCINN CAL cloud height is above FRESCO-S in summer, and below in winter.
    - Over coastal or island sites, ROCINN CAL cloud top height is above FRESCO-S.
- Comparing ROCINN_CAL with Cloudnet cloud top height, and FRESCO with Cloudnet cloud mid height, we find a similar bias and dispersion for the 2.4 data as for the data with version 2.2 or higher.
- Comparing ROCINN_CRB with Cloudnet cloud mid height, we find an improved, less negative, bias compared to processor version 2.2 or higher, and a similar dispersion.

13.2 Validation approach

13.2.1 Overview of changes

Changes since ROCVR#22.

- Added one Cloudnet location: Cabauw.
- In the latitudinal zonal mean plots of cloud sRCF and cloud height, FRESCO data with snow-ice is filtered out.

Changes since ROCVR#21

- Added 3 Cloudnet locations: Savilahti, Eriswil, Soverato.

Changes since ROCVR#19.

- Added comparisons of OCRA/ROCINN_CAL CLOUD mean height (CMH) vs CLOUDNET cloud mean height.
- Replaced ROCINN_CAL Cloud top height (CTH) vs FRESCO cloud height (CH) results with ROCINN_CAL CMH vs FRESCO CH, to harmonize with the ROCINN_CAL CMH vs CLOUDNET CMH. (Large but anticipated impact on results).

Changes since ROCVR#18.

- Harmonized settings of ‘S5P CLOUD vs CLOUDNET’ with those of ‘S5P CLOUD vs FRESCO’ regarding qa_value and cloud fraction filtering (small impact on results).
- Removed version-specific information (several figures, date lines, …) which became obsolete after the mission reprocessing.
- Added a section with discussion on differences between validation results before and after the 2.4 reprocessing.

13.2.2 Ground-based networks

CLOUDNET lidar/radar data

S5P TROPOMI L2_CLOUD cloud data have been routinely compared at 27 ground-based stations (

Table 16) to reference lidar/radar data from the cloud target classification product of the CLOUDNET and ARM ground-based networks [ER_Cloudnet]. Cloud base height, cloud top height and a vertical cloud classification profile (resolution <100 m) are provided each 30 s, typically.

Comparison settings

For the comparisons between S5P and CLOUDNET data, the approach is slightly adapted compared to ROCVR-17, to harmonize with the settings used for the S5P CLOUD vs S5P FRESCO comparisons (see later). The impact on the resultant quality indicators is small, and the collocation numbers are slightly higher. Here follows a description of the comparison settings.

S5P L2_CLOUD pixels are selected if

- the `qa_value > 0.25`,
- if the product is CLOUD CAL: `CF > 0.05`,
- if the product is CLOUD CRB: the scaled radiometric cloud fraction `sRCF=(CF*CA/0.8) > 0.1`,
- the pixel encompasses the CLOUDNET station,

with CF the radiometric cloud fraction, CA the cloud albedo and sRCF the scaled radiometric cloud fraction. It is not possible to use a filter on scaled radiometric cloud fraction for CLOUD CAL, since this product does not provide a cloud albedo.
The CLOUDNET data is selected if

- the station is cloud covered (according to CLOUDNET) at least 50% of the 1200 s temporal interval centered at the TROPOMI overpass time,
- and the standard deviation of CLOUDNET cloud height is smaller than 0.5 km within this temporal interval.

Note that there is no filtering of multilayer clouds. The cloud mean height (CMH) or cloud top height is calculated from CLOUDNET cloud types 1-7.

We present here also comparisons of the S5P TROPOMI FRESCO, which employs an alternative cloud retrieval algorithm, with CLOUDNET, using the same comparison settings as for CLOUD CRB. Summarizing, the following quantities are compared:

- L2_CLOUD ROCINN_CAL CTH with CLOUDNET CTH,
- L2_CLOUD ROCINN_CAL CMH with CLOUDNET CMH. ROCINN_CAL CMH is derived here as (CBH+CTH)/2, both of which are available in the L2_CLOUD data files,
- L2_CLOUD ROCINN_CRB CH with CLOUDNET CMH,
- FRESCO-S CH with CLOUDNET CMH.

Table 16 – List of ground-based stations providing the cloud classification data product, and used in this study: 24 CLOUDNET stations and 4 ARM (Atmospheric Radiation Measurement) stations. Data is collected from ACTRIS using the EVDC harvester.

Station	Location	Network	Latitude (N)	Longitude (E)
Ny-Ålesund	Svalbard	CLOUDNET	78.932	11.921
Summit	Greenland	NOAA/ARM	72.60	-38.42
Andoya	Norway	ARM	69.14	15.68
Kenttarova	Finland	CLOUDNET	67.987	24.23
Savilahti	Finland	CLOUDNET	62.892	27.634
Hyytiala	Finland	CLOUDNET	61.84	24.29
Norunda	Sweden	CLOUDNET	60.09	17.48
Mace Head	Ireland	CLOUDNET	53.325	-9.9
Lindenberg	Germany	CLOUDNET	52.211	14.13
Cabauw	Netherlands	CLOUDNET	51.97	4.926
Leipzig	Germany	CLOUDNET	51.35	12.43
Chilbolton	United Kingdom	CLOUDNET	51.145	-1.437
Jülich	Germany	CLOUDNET	50.909	6.413
Palaiseau	France	CLOUDNET	48.713	2.208
Munich	Germany	CLOUDNET	48.15	11.57
Schneefernerhaus	Germany	CLOUDNET	47.42	10.98
Eriswil	Switzerland	CLOUDNET	47.071	7.873
Galati	Romania	CLOUDNET	45.435	28.037
Bucharest	Romania	CLOUDNET	44.35	26.03
Potenza	Italy	CLOUDNET	40.6	15.72
Graciosa	Azores	ARM	39.092	-28.026
Soverato	Italy	CLOUDNET	38.69	16.545
Granada	Spain	CLOUDNET	37.16	-3.605
Mindelo	Cape Verde	CLOUDNET	16.879	-24.995
Iquique	Chile	CLOUDNET	-20.54	-70.18
Maido	La Reunion	CLOUDNET	-21.08	55.38
Villa Yacanto	Argentina	ARM	-32.13	-64.73
Punta-Arenas	Chile	CLOUDNET	-53.135	-70.883

13.2.3 Satellites

13.2.4 Alternative S5P cloud algorithms

S5P FRESCO

The support product S5P TROPOMI FRESCO cloud height is also compared to CLOUDNET observations, and directly with CLOUD CRB and CLOUD CAL at the CLOUDNET stations. This helps to judge if discrepancies between S5P CLOUD CRB and CLOUDNET are specific to the adopted cloud retrieval algorithm or are of more general nature. The S5P FRESCO support product is not publicly disseminated separately, but is used as input for e.g., the S5P NO2 retrieval. Earlier versions of the algorithm are described in e.g., [Koelemeijer 2001]. Like CLOUD CRB, FRESCO-S models a cloud as a Lambertian reflector. Information on cloud pressure and effective cloud fraction is derived from the reflectance in and around the O_2 A band. As opposed to CLOUD CRB, where cloud albedo is retrieved, in FRESCO-S, the cloud albedo is assumed to be 0.8 or the TOA reflectance at 758 nm if the reflectance is larger than 0.8. We note that at small cloud fractions, the surface albedo is adapted to prevent negative cloud fractions.

Comparison settings

Given the different assumption for cloud albedo in the CLOUD CRB and FRESCO retrieval models, CLOUD CRB CF and FRESCO CF are not directly comparable. Instead, we compare here the cloud fractions rescaled to cloud albedo=0.8: sRCF = CF*CA/0.8, the scaled radiometric cloud fraction, CF the radiometric cloud fraction and CA the cloud albedo.

S5P CLOUD pixels and S5P FRESCO pixels covering CLOUDNET stations were extracted. Common overpasses were considered. For CLOUD, qa_value ≥ 0.25 was applied the minimal setting proposed in the Readme file. For FRESCO, qa_value > 0.5 was applied. Additionally, (new since ROCVR#18) FRESCO pixels containing snow/ice are removed (i.e., only keeping pixels with ‘snow-free land’ or ‘ocean’). Additionally, for the cloud height comparisons, sRCF > 0.1 was by default applied for both CLOUD CRB and FRESCO. In the section 13.3.5.3 ‘dependence on influence quantities’, we investigate the impact of sRCF on the cloud height difference, and here we also check the setting sRCF > 0.

13.2.5 Field campaigns and modelling support

None for this report.

13.3 Validation of L2_CLOUD OFFL

13.3.1 Recommendations for data usage

Data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms.

The qa_value summarizes the quality of the product by taking into consideration several aspects like the spectral channel quality flags from L1B data, geometry limitations (e.g. not reliable retrievals for SZA>75°), inhomogeneous scene warnings, spatial mis-registration, high residual of the fitting process etc. As a first recommendation, the PRF proposes to only use those TROPOMI pixels associated with a qa_value >= 0.5. If cloud retrievals over snow and ice should be included, only data with qa_value >= 0.25 should be used. This allows including some snow/ice scenes while still filtering out retrievals with very low Root-Mean-Square (RMS). For application in trace gas retrievals, the lower value of qa_value >= 0.25 could be used in order to increase the data yield. It should be noted that this recommendation is only preliminary, may be subject to change and is not based on any extensive analysis. The most appropriate usage of the CLOUD product depends on the use-case and should be based on the qa_value in conjunction with the retrieval diagnostics and warning and error flags provided in the CLOUD product and summarized in Table 5-1 of the CLOUD ATBD. These inputs together represent a much more detailed and more appropriate information than the qa_value alone. For the validation performed in this work, we use the more relaxed constraint of qa_value >= 0.25.

Some of the known data quality issues are not covered by the quality flags and have been considered when interpreting the validation results reported hereafter (see also the Product Readme File (PRF)). Those issues are:

1. insensitivity to very thin clouds,
2. treatment of multi-layer clouds,
3. unknown straylight impact in the NIR,

13.3.2 Status of validation

This section presents a summary of the key validation results obtained by the MPC VDAF and by S5PVT AO projects. Up-to-date validation results and consolidated validation reports are available through the MPC VDAF Portal at http://mpc-vdaf.tropomi.eu.

The validation vs. CLOUDNET ground-based data uses S5P L2_CLOUD RPRO+OFFL 02.04.01-02.06.01 data. This covers the period from 2018-04-30 until August 2024. The version updates to 02.06.01 includes several changes, including new cloud-free maps and a new co-registration scheme (see the CLOUD PRF). CLOUDNET data from 28 stations were considered in this analysis (1 station added since ROCVR#21). Note that the station Summit was excluded, as it did not provide sufficiently valid co-locations, which can probably be explained by snow-ice filtering of the cloud products.

13.3.3 Quality screening numbers

Error! Reference source not found. presents the fraction of pixels with qa_value ≥ 25 % for S5P CLOUD CAL. A decrease is visible for the sites Granada, Mindelo and Iquique since v2.6.1. This can be related to a higher occurrence of flag value 'cloud_inhomogeneity_warning' for these sites. A similar figure is obtained for S5P CLOUD CRB (not shown).

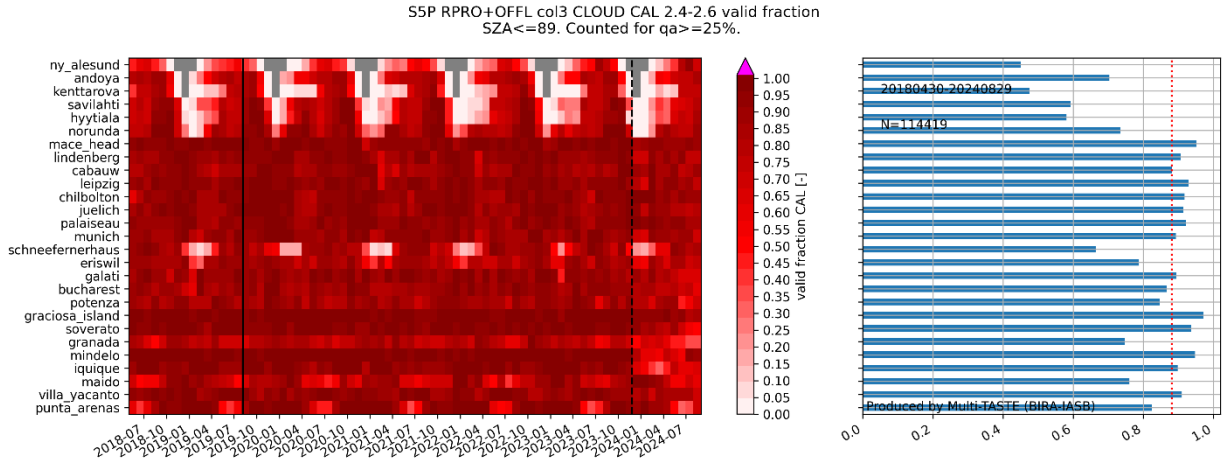


Figure 99: Fraction of valid pixels (qa_value ≥ 25 %) for S5P CLOUD CAL.

13.3.4 Radiometric cloud fraction (L2_CLOUD CAL & L2_CLOUD CRB)

13.3.4.1 Bias

Comparison with alternative algorithm S5P FRESCO

In Compernolle et al., (2021), the latitudinal variation of zonal means of CLOUD CAL and CRB v1 cloud fraction is also compared with MODIS geometric cloud fraction. A similar variation is found between the products, but as expected, the geometric cloud fraction is higher than the radiometric cloud fraction of the S5P CLOUD products.

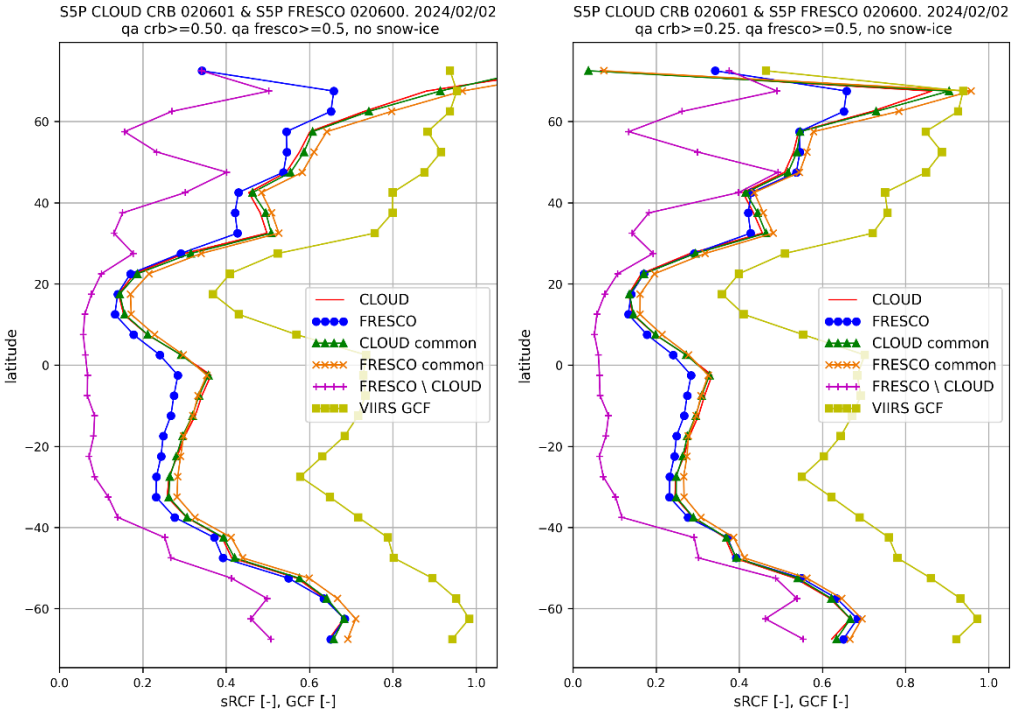


Figure 100. Latitudinal variation of zonal means of scaled radiometric cloud fraction of S5P OCRA/ROCINN CRB and S5P FRESCO-S, and geometrical VIIRS cloud fraction, for day 2024/02/02. Two quality filter criteria are used for CRB: the standard qa_value>0.5, and a looser `qa_value` > 0.25.

Figure 100 presents the latitudinal variation of zonal means of scaled radiometric cloud fraction of S5P CLOUD OCRA/ROCINN_CRB and S5P FRESCO at one day, tested with the two proposed quality filter criteria for CRB: the standard qa_value>0.5, and a looser `qa_value` > 0.25. The behaviour of both settings is similar, indicating that this less strict quality filtering can indeed be used. The difference between CRB sRCF and FRESCO sRCF is small, except near the poles. VIIRS geometrical cloud fraction (GCF) is also indicated, displaying a similar variation but a higher value.

Figure 101 presents boxplots of the difference (top left) and normed relative difference (top right) between rescaled cloud fractions of CLOUD CRB and FRESCO at the CLOUDNET stations, as well as the evolution of the monthly median difference over the CLOUDNET stations (bottom, left) and the overall median difference (bottom right). The loose criterion qa_value≥0.25 was applied, but this has only a small impact on the resulting figures compared to the qa_value≥0.5 criterion. Overall, the median differences are negative (CRB lower than FRESCO) but always within the 20% bias requirement. Larger median differences are seen for Iquique, Mindelo, Mace Head, Soverato, Graciosa Island and Ny-Ålesund, which are all coastal or island sites. Since CLOUD v2.6.1, an increase in CRB cloud fraction seems visible for sites Graciosa Island, Soverato and Granada.

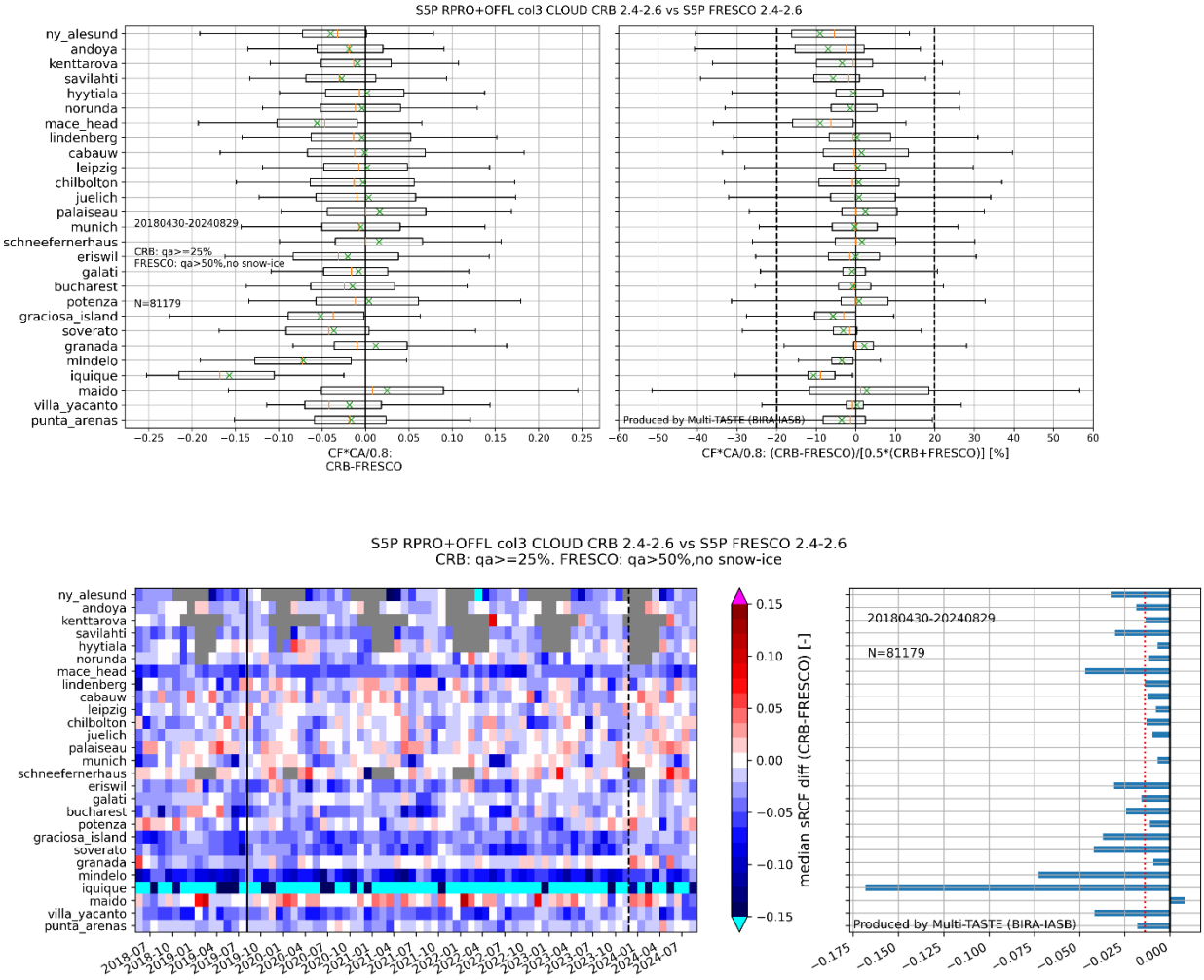


Figure 101. Top left. Box plots of S5P CLOUD CRB OFFL cloud fraction minus S5P FRESCO OFFL cloud fraction, after both have been rescaled to a cloud albedo of 0.8. Left: same but for the normed relative difference. The 20% bias requirement is indicated in dashed lines. Bottom left: monthly median of scaled cloud fraction difference, per month and per station. Indicated is the pixel size switch (2019/08/06) and the update to CLOUD v2.6.1 (2023/11/26). Bottom right: Median of scaled cloud fraction differences per station. The red dashed line gives the median of these median differences.

13.3.4.2 Dispersion

Figure 102 (left) presents the evolution of the difference dispersion (½ IP68 (CRB minus FRESCO)) over the CLOUDNET stations, per month. **Figure 102** (right) presents the overall dispersion per station. At all stations, the ½ IP68 slightly exceeds the dispersion requirement of 0.05, while there is a high dispersion at the island site Maido and Cabauw.

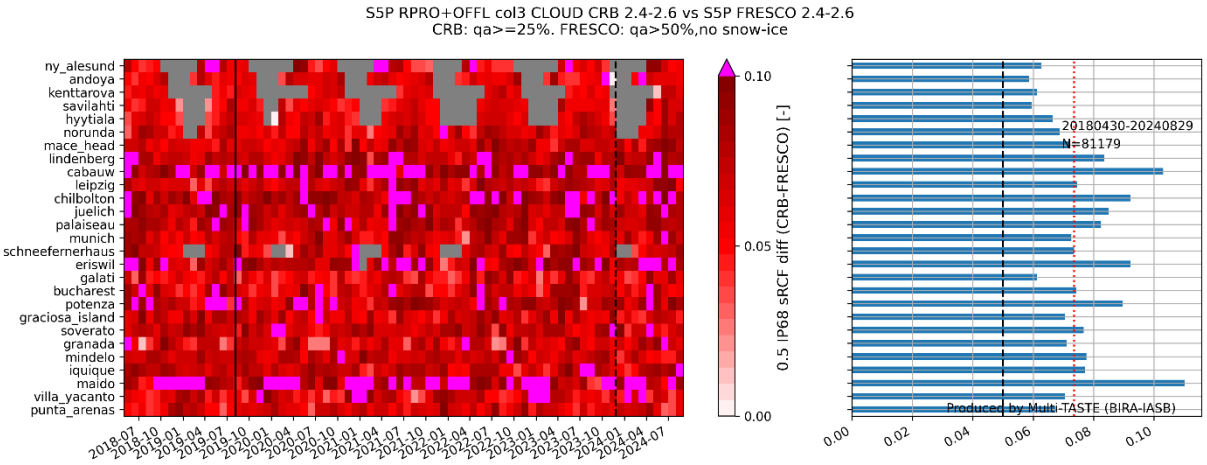


Figure 102. Left. Monthly ½ IP68 of scaled cloud fraction difference of CLOUD OFFL vs FRESCO OFFL, per month and per station. Indicated is the pixel size switch (2019/08/06) and the update to CLOUD v2.6.1 (2023/11/26). Right: ½ IP68 of scaled cloud fraction difference (CLOUD CRB-FRESCO), per station.

13.3.4.3 Dependence on influence quantities

The S5P L2_CLOUD cloud fraction gets unphysically high values at very large SZAs (above 85 degrees) due to very weak illumination. The other cloud parameters might also be affected for high SZAs due to limitation in the RTM treatment of spherical atmosphere. The high surface albedo above snow and/or ice covered surfaces is a challenge for cloud retrievals. Note that a very large SZA implies a measurement above the polar region, and therefore snow-ice covered surfaces are likely.

13.3.4.4 Drifts, cycles and shorter term variability

None reported.

13.3.4.5 Geographical patterns

None reported.

13.3.5 Cloud top height (L2_CLOUD CAL) and cloud height (L2_CLOUD CRB)

13.3.5.1 Bias

Comparison with alternative S5P cloud height retrieval FRESCO

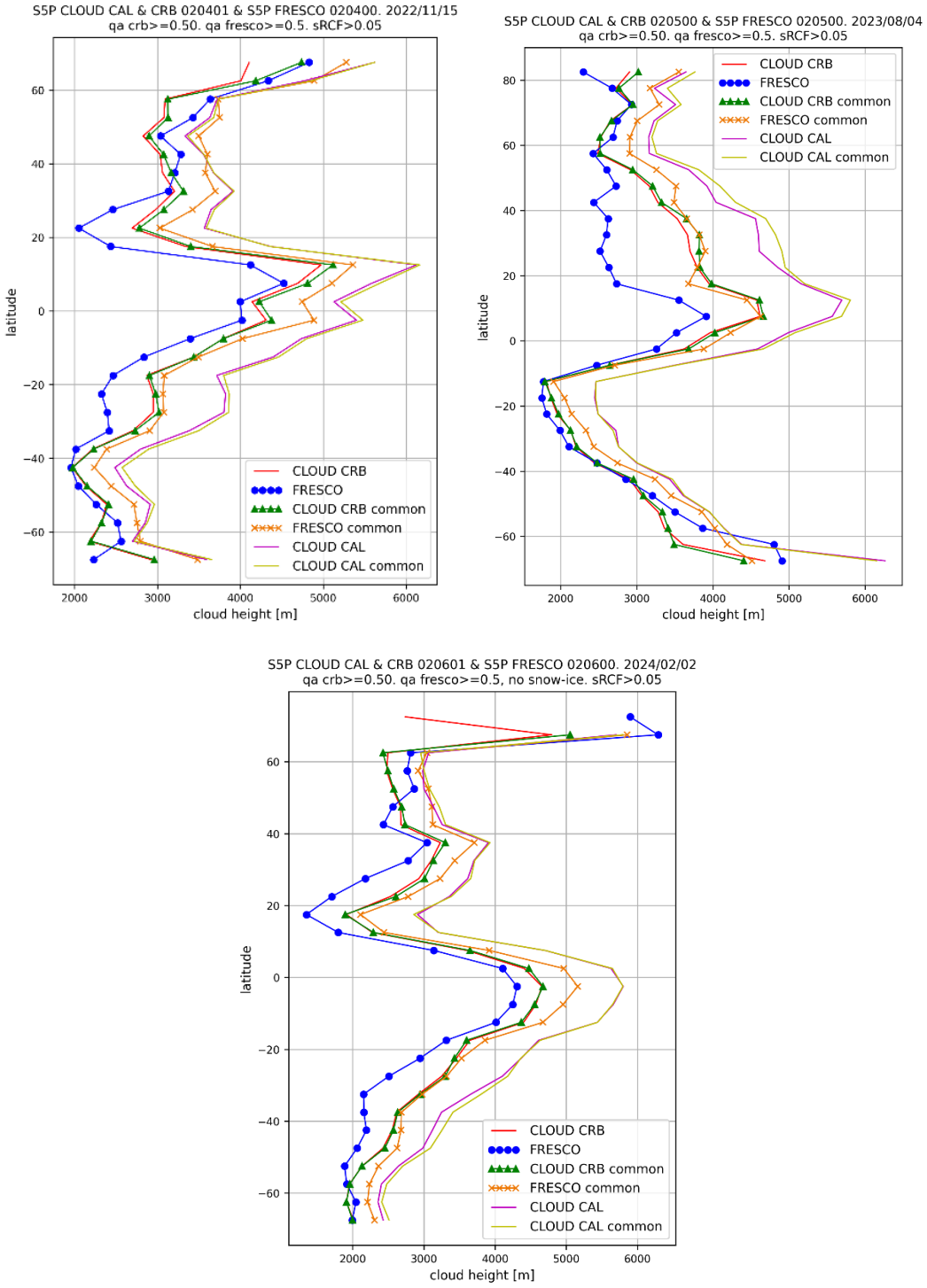


Figure 103: Latitudinal variation of zonal means of cloud height of S5P CLOUD OCRA/ROCINN_CRB, S5P FRESCO and ROCINN_CAL at two days. Selection criteria: qa_value > 0.5, no invalid values, sRCF > 0.05. 'common' means the subset of valid pixels common to CLOUD and FRESCO.

Figure 103 presents the latitudinal variation of zonal means of cloud height of S5P CLOUD OCRA/ROCINN_CRB and S5P FRESCO at a single day. If no common subset of pixels is considered, FRESCO is on average lower than ROCINN_CRB, but if the common subset of pixels is taken, it is mostly higher. In other words, with the FRESCO qa > =0.5, sRCF > 0.05 criteria more pixels are retained which have typically a low cloud height. ROCINN_CAL cloud height is higher than both FRESCO and CRB.

S5P CLOUD CRB CH was compared with S5P FRESCO CH, over the CLOUDNET stations (**Figure 104**).

At most stations, a median difference of about -0.5 km (CRB lower than FRESCO) or about -30% (median normed relative difference). Clear exceptions (CRB closer to FRESCO) occur at the stations Iquique, Mindelo, Graciosa Island and Mace Head, which are island or coastal sites. This is likely related to the larger difference at low cloud fraction between both cloud products: see section 13.3.5.3 where the Iquique case is further investigated.

The monthly median difference is plotted in **Figure 104**, bottom left. Smaller negative values of CRB-FRESCO might be discerned this summer compared to previous summers, which might be related to the ROCINN version change.

We also compare S5P CLOUD CAL CMH with S5P FRESCO CH. For most sites, the overall median difference is close to -0.5 km, but a seasonal cycle is visible, with the difference being maximal in winter months and minimal in summer. Exceptions are Iquique, Mindelo, Graciosa Island and Mace Head, where the difference is less negative or even positive. An increase in CLOUD CAL CMH vs FRESCO CH might be present for the site Maido since v2.6.1. Moreover, CLOUD CAL CMH minus FRESCO CH clearly reaches more positive values this summer than previous summers, and less negative values this winter, which might be related to the version change.

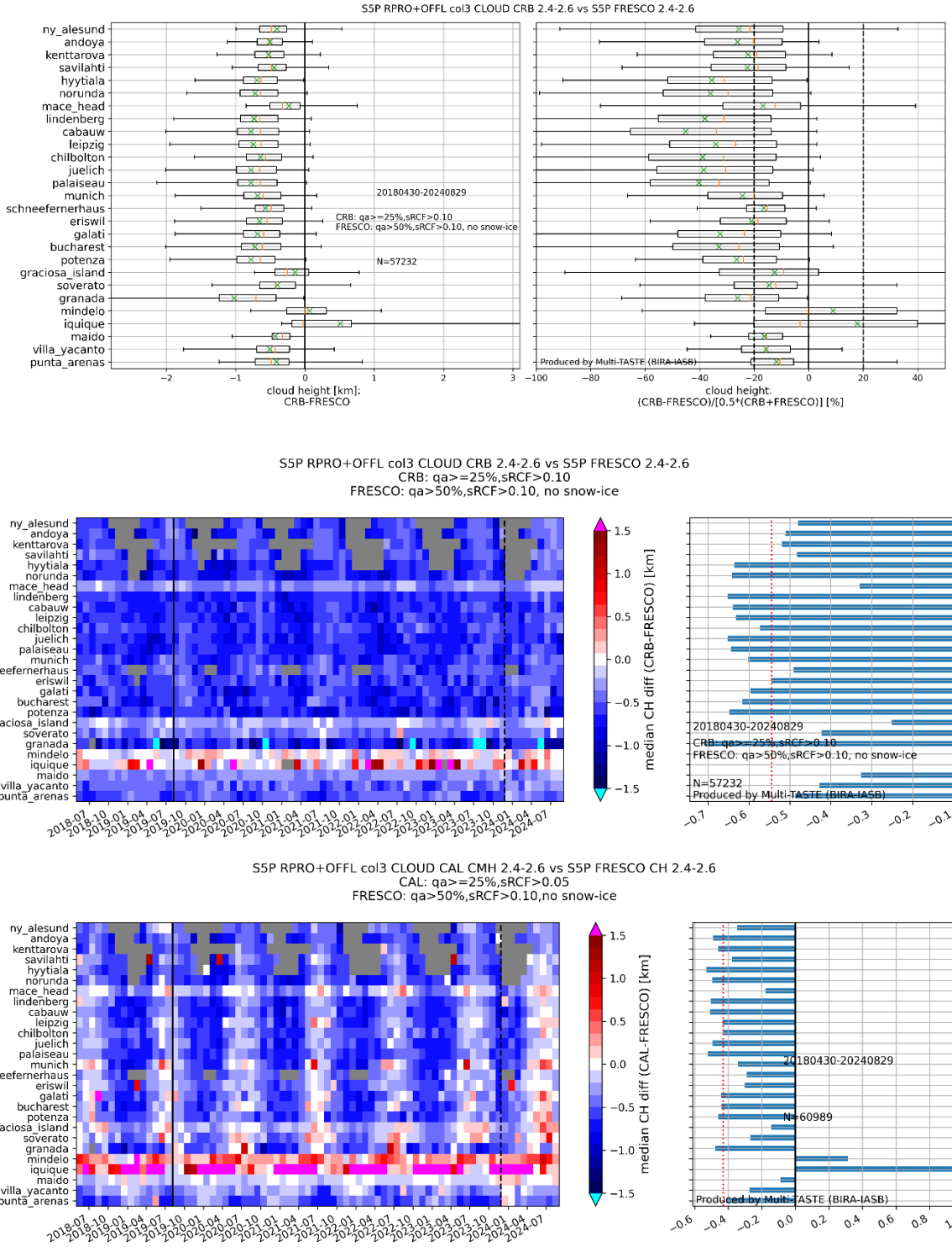


Figure 104: Top. Boxplots of S5P CLOUD CRB CH minus S5P FRESCO CH (left) and of the normed difference (right). Middle. Monthly median difference per station of CRB CH-FRESCO CH (left), and overall median difference (right). Bottom. Monthly median difference per station of CAL CMH-FRESCO CH (left), and overall median difference (right). Indicated is the pixel size switch (2019/08/06) and the update to CLOUD v2.6.1 (2023/11/26).

Comparison with CLOUDNET cloud top height and cloud height

Here, we check how the effective cloud heights of CLOUD CAL, CLOUD CRB and FRESCO are related to the cloud heights as obtained from CLOUDNET.

L2_CLOUD CAL cloud top height is generally below the CLOUDNET cloud top height. This can be seen in **Figure 105**, which presents boxplots per station of absolute scale and relative difference (top panels) and of the monthly median difference (bottom panel).

The bias depends on the cloud height. In distribution plots of CLOUDNET cloud top heights two modes are typically visible (see e.g., Compernolle, 2021, Fig. 10), where the higher mode contains more ice cloud and multilayer clouds. We consider a CLOUDNET CTH of 4 km as the threshold between low (<4 km) and high (>4 km) clouds. To obtain an overall value for bias and relative bias, we calculate the median over all per-station medians of the difference and of the relative difference, for all clouds, low and high clouds. (See **Table 17**). For high clouds, CAL CTH is 3 km below Cloudnet's CTH, while it is 0.3 km below for low clouds.

Table 17. Statistics, calculated over the stations of S5P CLOUD OFFL CAL CTH-CLOUDNET CTH, S5P CLOUD OFFL CAL CMH-CLOUDNET CMH, S5P CLOUD OFFL CRB CH-CLOUDNET CMH and S5P FRESCO OFFL CH-CLOUDNET CMH. Measurements up to 2024/05 are included. The following statistics are provided: the median over all stations of (i) median difference, (ii) median relative difference, (iii) 0.5 IP68 (difference) and (iv) Pearson R. The statistics are calculated for 'all clouds', 'high clouds' (Cloudnet CTH>4 km) and 'low clouds' (Cloudnet CTH<4km).

comparison	med(med(SAT-GND)) [km]	med(med((SAT-GND))/GND) [%]	med(0.5IP68(SAT-GND)) [km]	med(R)
CAL CTH vs Cloudnet CTH	-1.3	-27	1.9	0.75
CAL CTH vs Cloudnet CTH, high	-2.8	-37	1.9	0.57
CAL CTH vs Cloudnet CTH, low	-0.3	-14	0.6	0.42
CAL CMH vs Cloudnet CMH	-0.4	-15	1.1	0.75
CAL CMH vs Cloudnet CMH, high	-0.6	-12	1.7	0.54
CAL CMH vs Cloudnet CMH, low	-0.3	-16	0.5	0.47
CRB CH vs Cloudnet CMH	-0.5	-22	1.2	0.72
CRB CH vs Cloudnet CMH, high	-1.0	-19	1.9	0.47
CRB CH vs Cloudnet CMH, low	-0.3	-19	0.5	0.36
FRESCO CH vs Cloudnet CMH	0.0	0	1.3	0.68
FRESCO CH vs Cloudnet CMH, high	-0.3	-6	1.9	0.49
FRESCO CH vs Cloudnet CMH, low	0.2	13	0.6	0.40

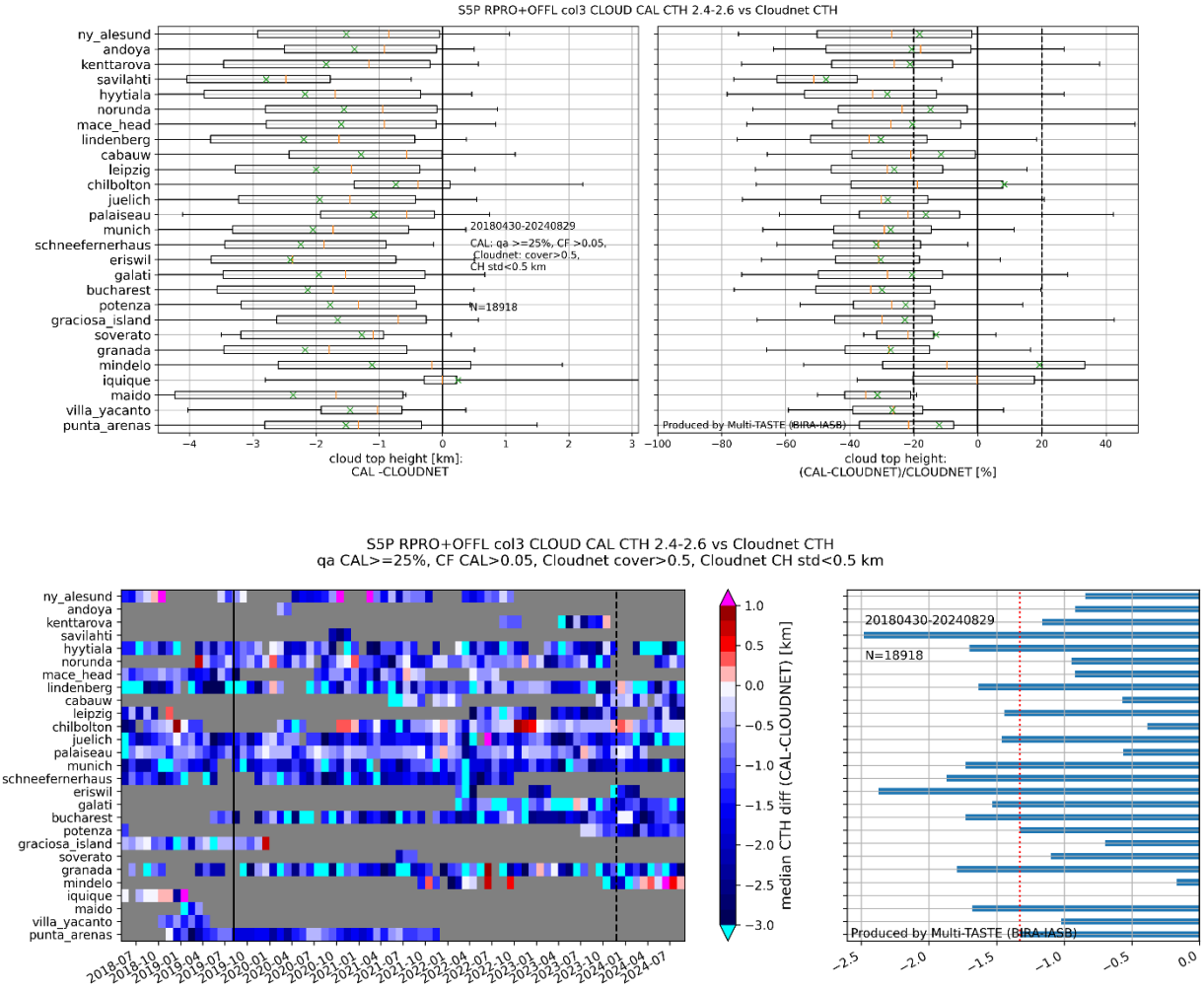


Figure 105: Top panel: Boxplots of S5P L2_CLOUD CAL CTH minus CLOUDNET CTH, per station. Boxplot conventions: box bounds are at first and third quartile. Orange line is median. Whiskers are at 5 and 95 percentiles. Green cross is mean. Right panel: Similar as upper panel but now for the relative difference. Sensing date range is indicated on the plot. Bottom left panel: monthly median of CLOUD CAL CTH-CLOUDNET CTH, per station. Pixel size switch (2019/08/06) and the update to CLOUD v2.6.1 (2023/11/26) are indicated on the plot. Right panel: median difference per station. The red dashed line gives the median over the median differences.

L2_CLOUD CAL cloud mean height (CMH) is generally below the CLOUDNET cloud mean height. This can be seen in **Figure 106**, which presents boxplots per station of absolute scale and relative difference (top panels) and of the monthly median difference (bottom panel). There seems to be a higher positive bias this winter compared to previous winters, which might be related to the change to v2.6.1.

To obtain an overall value for the bias and the relative bias, we calculate the median over all per-station medians of the difference (CAL CMH-CLOUDNET CMH) and of the relative difference. We calculate this for all clouds, low and high clouds. CAL CMH is 0.6 km below Cloudnet CMH for high clouds, and 0.3 km below Cloudnet CMH for low clouds.

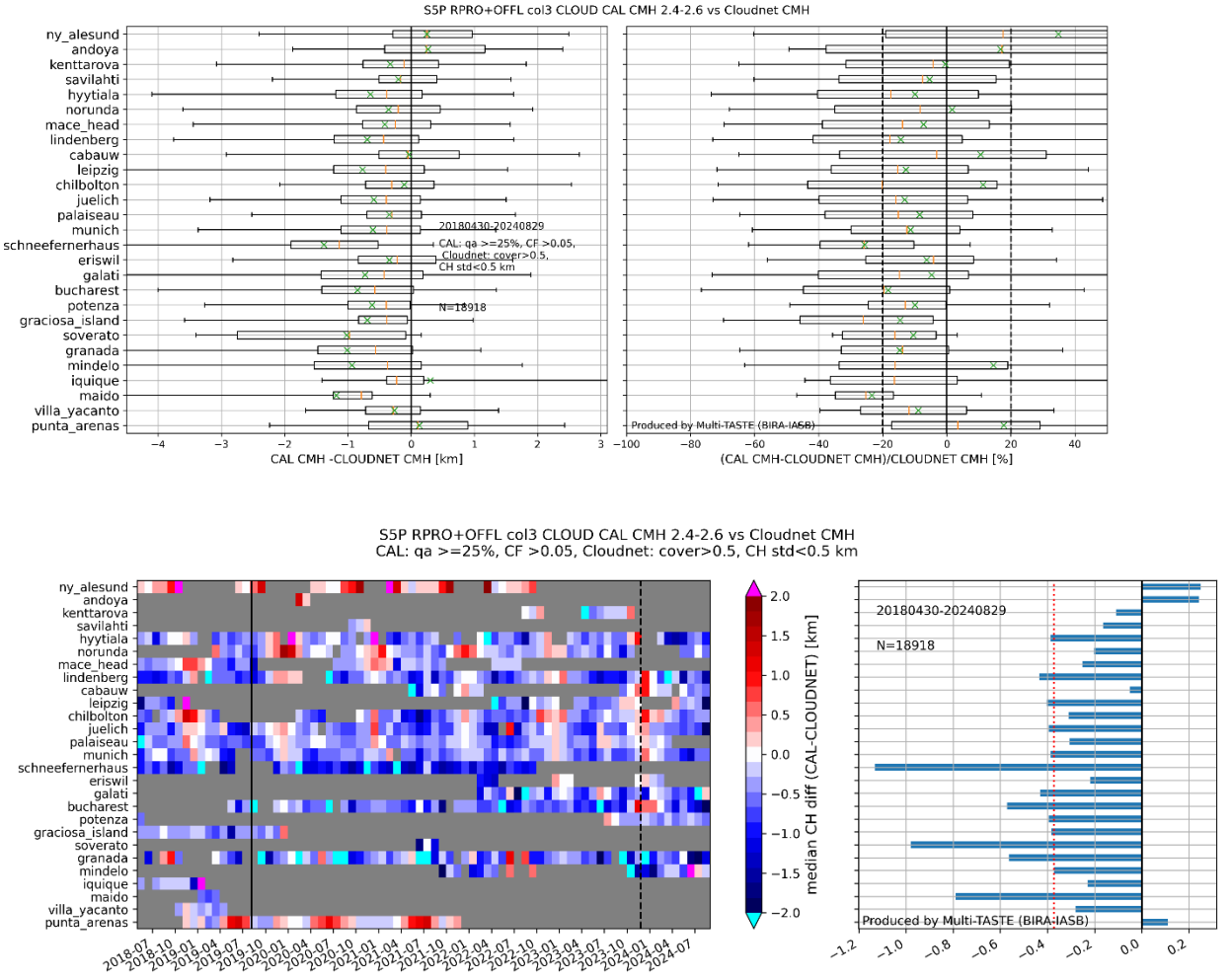


Figure 106: Top panel: Boxplots of S5P L2_CLOUD CAL CMH minus CLOUDNET CMH, per station. Boxplot conventions: box bounds are at first and third quartile. Orange line is median. Whiskers are at 5 and 95 percentiles. Green cross is mean. Right panel: Similar as upper panel but now for the relative difference. Sensing date range is indicated on the plot. Bottom left panel: monthly median of CLOUD CAL CMH-CLOUDNET CMH, per station. Pixel size switch (2019/08/06) and the update to CLOUD v2.6.1 (2023/11/26) are indicated on the plot. Right panel: median difference per station. The red dashed line gives the median over the median differences.

L2_CLOUD CRB cloud height is generally below the CLOUDNET cloud mean height. This can be seen in Figure 107, which presents boxplots per station of absolute scale and relative difference (top panels) and of the monthly median difference (bottom panel).

To obtain an overall value for the bias and the relative bias, we calculate the median over all per-station medians of the difference (CRB CH-CLOUDNET CMH) and of the relative difference. We calculate this for all clouds, low and high clouds. CRB CH is 1.0 km below Cloudnet CMH for high clouds, and 0.3 km below Cloudnet CMH for low clouds.

Figure 108 displays boxplots of the difference and relative difference between S5P FRESCO CH and CLOUDNET CMH, and the monthly medians per station. The overall bias (median over the station median differences) is very close to zero, but there is considerable interstation scatter and a seasonal cycle at several sites (e.g., Juelich, Palaiseau), with FRESCO being higher in winter and lower in summer.

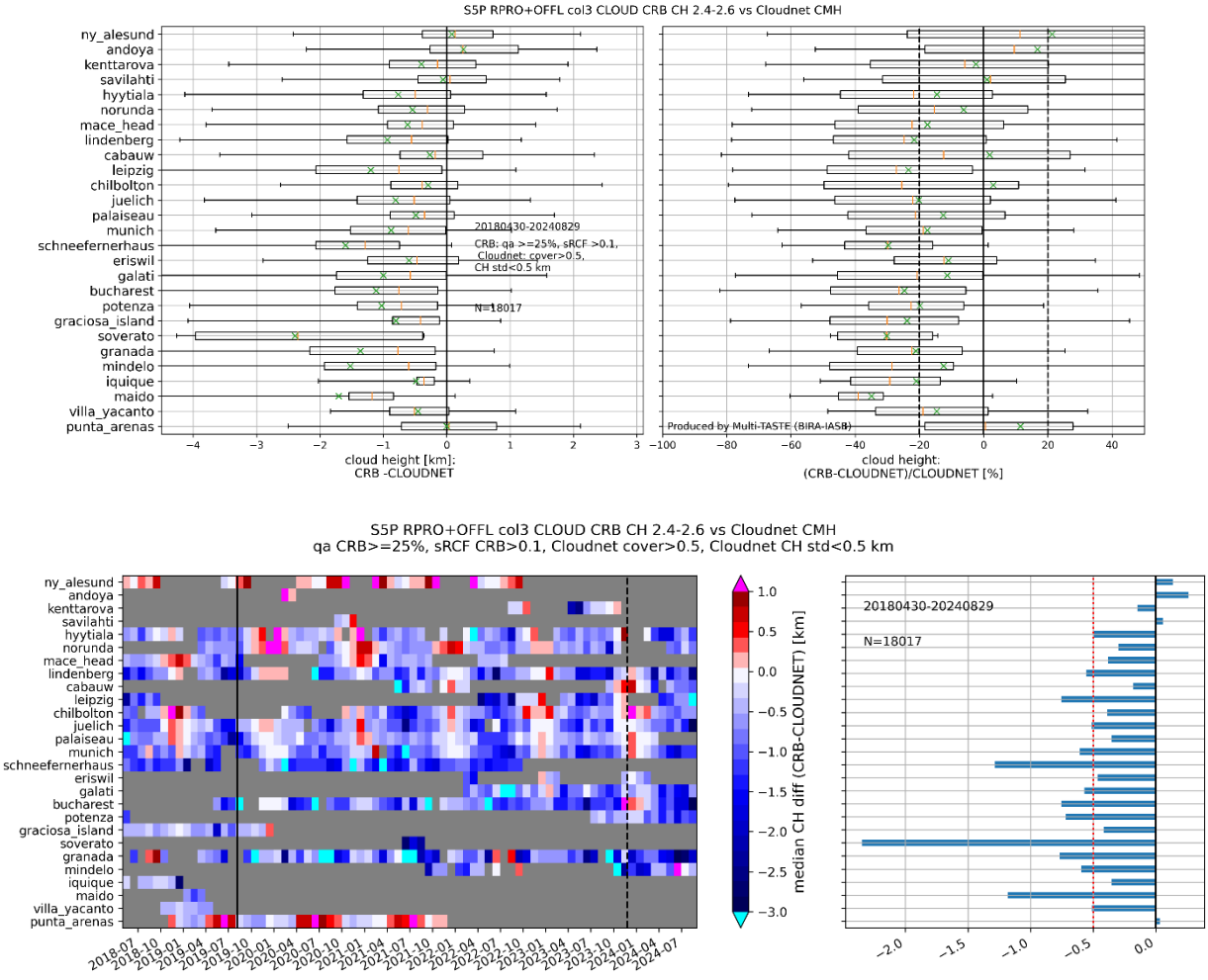


Figure 107: Upper panel: Boxplots of S5P L2_CLOUD CRB RPRO+OFFL CH minus CLOUDNET CH (upper left) and of the relative difference (upper right), per station. The same conventions as for **Table 17** apply. Sensing time range is indicated on the figure. Bottom left panel: monthly median of CLOUD CRB CH-CLOUDNET CMH, per station. Pixel size switch (2019/08/06) and update to CLOUD v2.6.1 (2023/11/26) are indicated on the plot. Bottom right panel: median difference per station. The dashed red line indicates the median over the median differences.

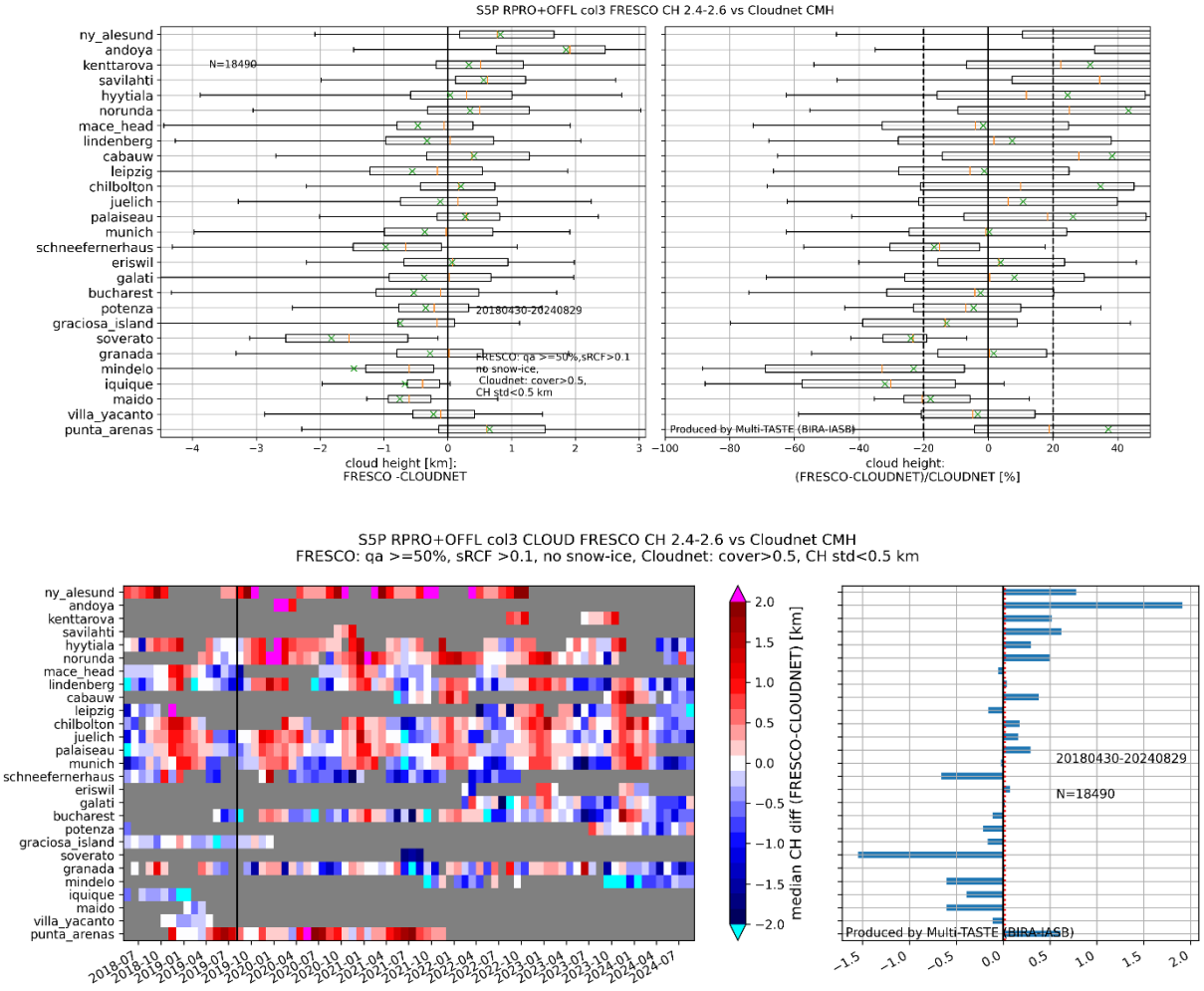


Figure 108: Upper panel: Boxplots of S5P L2_CLOUD FRESCO RPRO+OFFL v1.3-v1.4 CH minus CLOUDNET CH (upper left) and of the relative difference (upper right), per station. The same conventions as for **Table 17** apply. Sensing time range is indicated on the figure. Bottom left panel: monthly median of FRESCO CH-CLOUDNET CH, per station. Pixel size switch (2019/08/06) and update to CLOUD v2.6.1 (2023/11/26) are indicated on the plot. Bottom right panel: Median difference per station. The red dashed line indicates the median of the median differences.

13.3.5.2 Dispersion

Comparison with alternative S5P cloud height retrievals

The comparison of S5P CLOUD CRB CH vs S5P FRESCO CH reveals a low ½ IP68 close to the dispersion requirement of 0.5 km at most stations (Figure 109, right). Exceptions are Iquique and Granada where the dispersion is higher. The median dispersion over all stations is slightly below the dispersion requirement.

Figure 109 shows the evolution of monthly ½ IP68 with time. The dispersion of CAL CMH vs FRESCO CH is typically higher, mostly slightly exceeding the dispersion requirement. This could be related to the different model assumptions (layer model vs Lambertian model). Iquique is a strong outlier with a dispersion reaching 2.5 km.

The deviating behaviour at Iquique is likely related to the low-lying FRESCO cloud heights at low cloud fraction (see section 13.3.5.3). Since CLOUD v2.6.1, a dispersion increase vs FRESCO is visible for both CRB as CAL at site Maido.

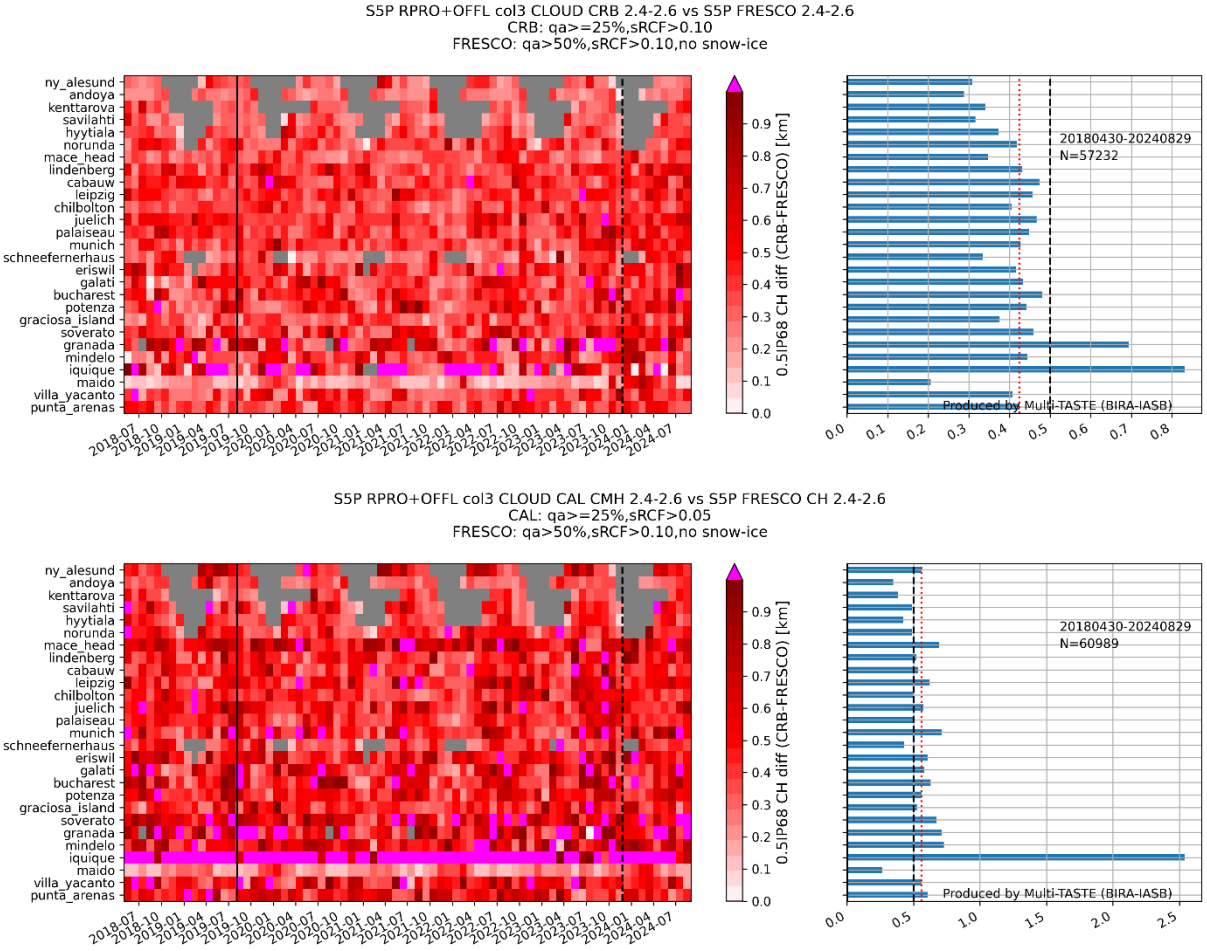


Figure 109: Top. Left: Monthly ½ IP68 of the cloud height difference of CLOUD CRB OFFL vs FRESCO OFFL, per month and per station. Indicated are the pixel size switch (2019/08/06) and the upgrade to CLOUD v2.6.1 (2023/11/26). Right: ½ IP68 of S5P CLOUD CRB CH minus S5P FRESCO CH. Sensing date range is indicated on the figure. Bottom. Similar but for CLOUD CAL CMH vs FRESCO CH.

Comparison with CLOUDNET cloud top height and cloud height

From the width of the boxplots in **Figure 105** to **Figure 108**, it can be inferred that the dispersion of S5P CLOUD CAL CTH minus CLOUDNET CTH, S5P CLOUD CAL CMH minus CLOUDNET CMH, S5P CLOUD CRB CH minus CLOUDNET CH and S5P FRESCO CH minus CLOUDNET CMH exceeds the upper limit for error dispersion (500 m). However, also CLOUDNET CTH random error, and comparison error, contribute to the difference dispersion, and these contributions have not been quantified yet.

Figure 110 displays the monthly dispersion per station of S5P CLOUD CAL CTH minus CLOUDNET CTH, S5P CLOUD CAL CMH minus CLOUDNET CMH, S5P CLOUD CRB CH minus CLOUDNET CMH, and FRESCO CH minus CLOUDNET CMH.

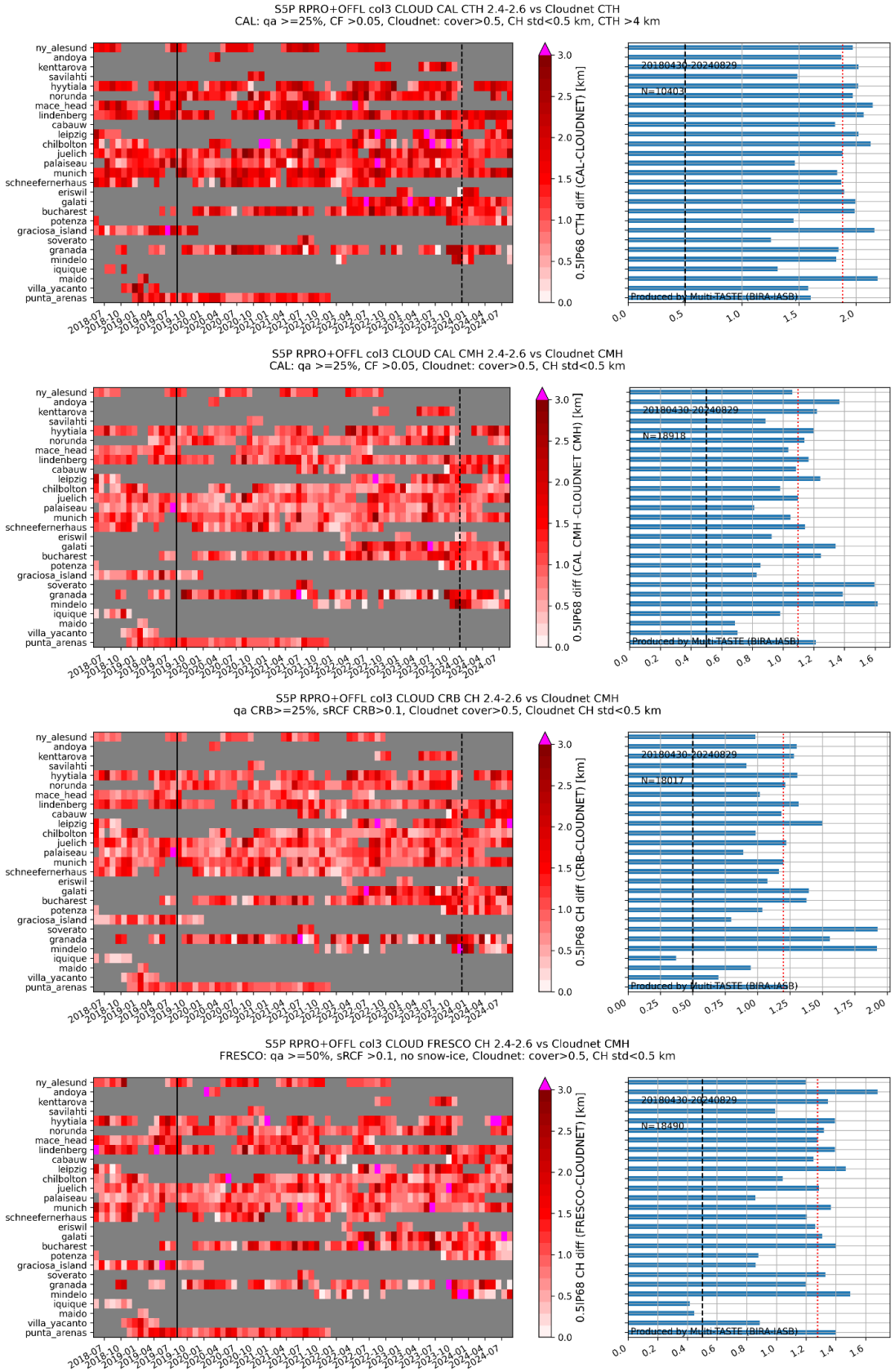


Figure 110. Monthly dispersion (½IP68) per station of S5P CLOUD CAL CTH minus CLOUDNET CTH (top), S5P CLOUD CAL CMH minus CLOUDNET CMH, S5P CLOUD CRB CH minus CLOUDNET CH, and S5P FRESCO CH minus CLOUDNET CH (bottom). Indicated are the pixel size switch (2019/08/06) and the update to CLOUD v2.6.1 (2023/11/26).

Overall values of dispersion are obtained by taking the median over all per-station ½ IP68 values (see **Table 17**). Distinguishing between high (CLOUDNET CTH >4km) and low (CLOUDNET CTH<4 km) clouds

- The dispersion of CLOUD CAL CTH vs CLOUDNET CTH is 1.9 km for high clouds and 0.6 km for low clouds. The overall value is 1.9 km.
- The dispersion of CLOUD CAL CMH vs CLOUDNET CMH is lower compared to the previous case: 1.7 km for high clouds and 0.5 km for low clouds. Note also that the value for 'all clouds' has decreased significantly to 1.1 km.
- For CLOUD CRB CH, the difference dispersion with CLOUDNET CH is 1.9 km for high clouds. It is 0.5 km for low clouds.
- For FRESCO CH the dispersion is 1.9 km for high clouds. For low clouds, the dispersion is ~0.5 km.
- The Pearson-R correlation with CLOUDNET was calculated for the three cloud products all clouds, and for low or high clouds separately (see **Table 17**). As can be expected, correlations are lower for low or high clouds separately, given the lower dynamical range. Highest correlations are obtained for CLOUD CAL vs CLOUDNET.

13.3.5.3 Dependence on influence quantities

The bias and dispersion results of CRB cloud height vs FRESCO cloud height above are provided after first filtering both products on sRCF>0.1. Figure 111 shows that the agreement between CRB and FRESCO worsens for lower CRB sRCF. For low sRCF points, there is more scatter, and there is a systematic bias visible of low sRCF points with FRESCO cloud height being lower than CRB cloud height. Therefore, it is likely that the filter sRCF>0.1 has an important impact on the results. Note that for the site Iquique, this filter is not sufficient to remove all low-lying FRESCO points, which probably explains the deviating nature in bias compared to the other comparisons.

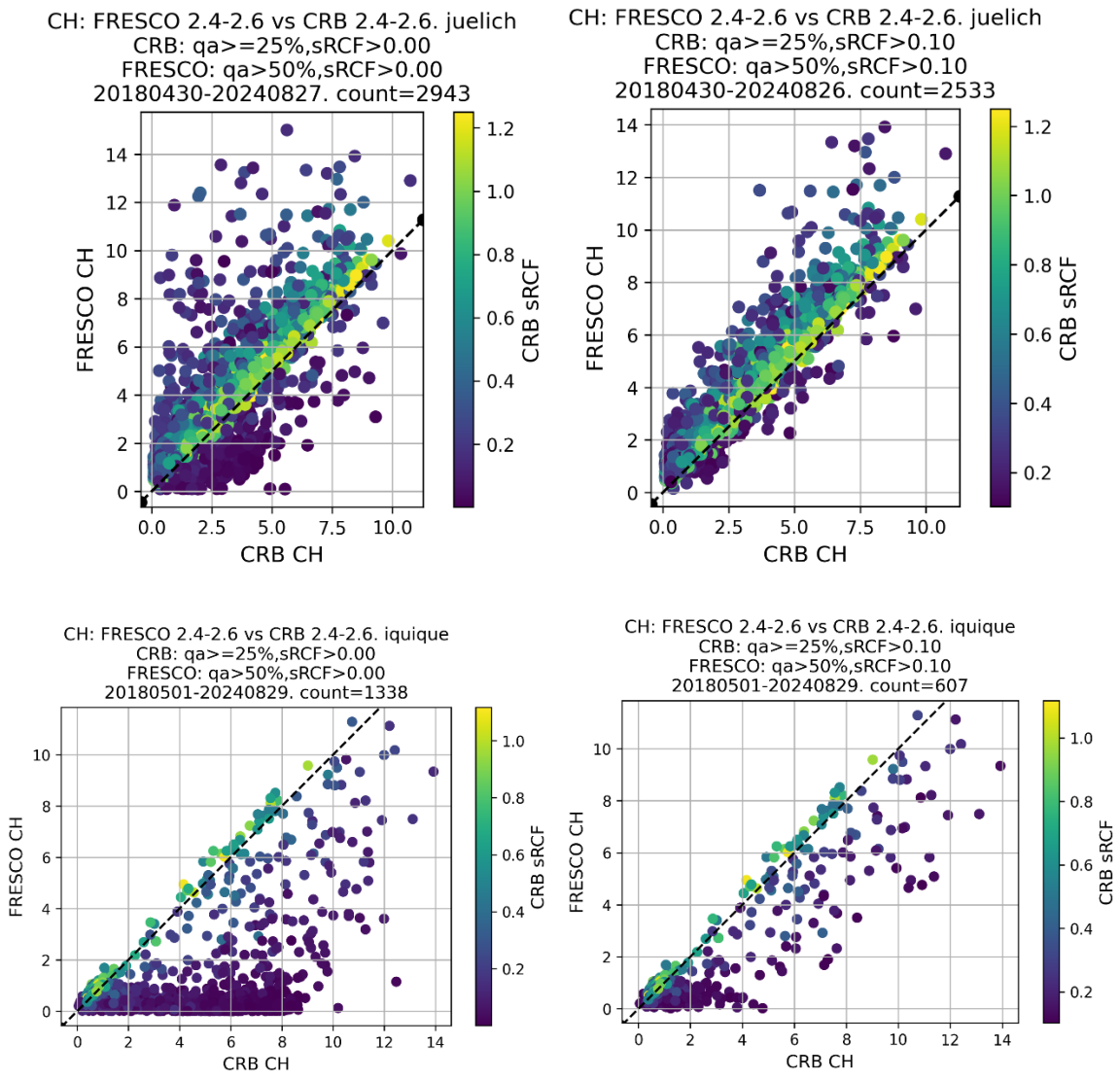


Figure 111. Scatter plot between FRESCO and CLOUD CRB cloud height, with points colored according to CRB sRCF, for the sites Juelich and Iquique, and for two screening settings: sRCF > 0. For both CRB and FRESCO, and sRCF > 0.1 (which is the default to calculate bias and dispersion).

13.3.5.4 Short term variability

Nothing to report.

13.3.5.5 Geographical patterns

Nothing to report.

13.4 Comparison of L2_CLOUD NRTI and OFFL products

This section investigates if L2_CLOUD NRTI and OFFL are significantly different. Comparisons of CLOUD CAL vs CLOUDNET and CLOUD CRB vs CLOUDNET, available at the VDAF-AVS, were intercompared for OFFL and NRTI for several sites, for the full NRTI period and specifically for CLOUD v2.6.1. Differences in statistics are small, in the order of 0.1-0.2 km (inspection at 2024/09/02; data not shown).

14 Validation Results: L2_AER_AI

14.1 L2_AER_AI products and requirements

This section reports on the validation of the following geophysical variables of the S5P TROPOMI L2_AER_AI UV aerosol absorbing index products identified in **Table 1**. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. The NRTI and OFFL processors producing very similar data products, only validation of the L2_AER_AI NRTI product is reported hereafter. Subsection 14.4 demonstrates evidence that NRTI and OFFL data do not differ significantly and that their respective validations yield similar conclusions.

14.2 Validation approach

The UV aerosol index (UVAI) is not a geophysical quantity that can be directly compared to independent measurements from ground or to model results. The way to validate this index is to compare it to coincident satellite measurements from different sensors. For the validation of S5P TROPOMI UVAI, measurements from EOS-Aura OMI and Suomi-NPP OMPS are well suited for that purpose.

In addition to the validation using satellite observations, the S5P TROPOMI UVAI data products can also be checked for internal consistency. For example, the following tests can be performed:

a) the dependence of the UVAI on the observation geometry (in particular on the SZA and the VZA of the measurement) can be investigated;
b) the UVAI values for clear sky and low aerosol amount should be close to zero;
c) the geographical patterns of the UVAI can be compared to those of other measurements, e.g., trace gas distributions of large biomass burning plumes or volcanic plumes.

It should be noted that the S5P TROPOMI AER_AI data product, the UVAI is calculated for three wavelength pairs, 388 / 354 nm, 380 / 340 nm, and 367 / 335 nm. The first pair allows for a direct comparison to the UVAI from OMI (which is also calculated for 388 / 354 nm). The third pair is designed to be compatible with the AI to be calculated for the Sentinel-5 mission.

14.2.1 Ground-based networks

As stated above, satellite UVAI data cannot be directly compared to ground-based measurements.

14.2.2 Satellites

S5P TROPOMI UV aerosol index data are compared to the aerosol indices obtained from EOS-Aura OMI and Suomi-NPP OMPS. Both OMI and OMPS have similar afternoon overpass times as compared to TROPOMI. With OMI the same wavelength pair (388 / 354 nm) can be compared.

14.2.3 Field campaigns and modelling support

As stated above, no direct comparison of the UVAI to non-satellite measurements is possible.

14.3 Validation of L2_AER_AI NRTI

14.3.1 Recommendations for data usage followed

In order to avoid misinterpretation of the data quality and to avoid the effects of sun glint, it is recommended to only use those TROPOMI pixels associated with a `qa_value` above 0.8. The variables `aerosol_index_340_380_precision` and `aerosol_index_354_388_precision` can also be used to diagnose the quality of the UVAI. These are new data product fields and are under evaluation.

For further details, data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product, all available on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms [ER_CoperATBD].

14.3.2 Status of validation

This section presents updated validation results obtained as a part of the S5P Mission Performance Centre (MPC) and by S5P Validation Team (S5PVT) AO projects.

The validation of S5P TROPOMI L2_AER_AI data presented here is based on comparisons with similar aerosol indices from the EOS-Aura OMI and Suomi-NPP OMPS satellite missions. Both OMI and OMPS have similar afternoon overpass times as compared to TROPOMI and with OMI the same wavelength pair (354/388 nm) can be compared. One example for desert dust is shown in **Figure 112**. The structures are quite similar with nearly the same values for the desert plume, with OMI appearing to exhibit slightly higher values and as expected, OMI captures fewer details of the plume structure. The largest differences in OMI and TROPOMI are observed for the values near 1.0; the transition on the swath edge is also smoother for the TROPOMI data.

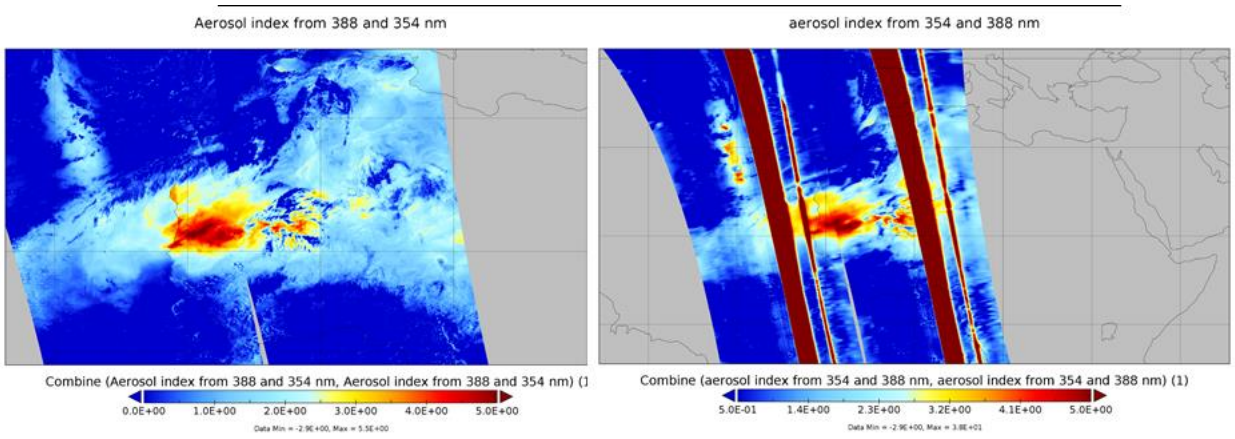


Figure 112: Comparison of S5P TROPOMI RPRO UVAI (orbits 24007 & 24008, left) and OMI OMAERO UV Aerosol Index (orbits 95097 & 95098, right) for Saharan dust on 1 June 2022. In general very good agreement is found (the stripes in north-south direction in the OMI data are caused by the OMI row anomaly and should be ignored).

Comparison results between S5P TROPOMI and OMPS UVAI are shown in **Figure 113** and **Figure 114** below (courtesy of Omar Torres and Changwoo Ahn, NASA-GSFC). Good agreement is found, especially between the operational TROPOMI product and the OMPS LER product. The spread of the S5P TROPOMI values is similar as the OMPS values (assuming LER clouds). Most of the larger spread can be explained by the larger number of the (smaller) TROPOMI ground pixels within the considered area (see also **Figure 114**). From this comparison, it is concluded that the S5P TROPOMI UVAI is also within the requirement

for random errors of 0.1 UVAI units. It should be noted that the standard deviation of the OMPS Mie product is systematically smaller due to the more realistic assumptions about clouds and surface reflectance.

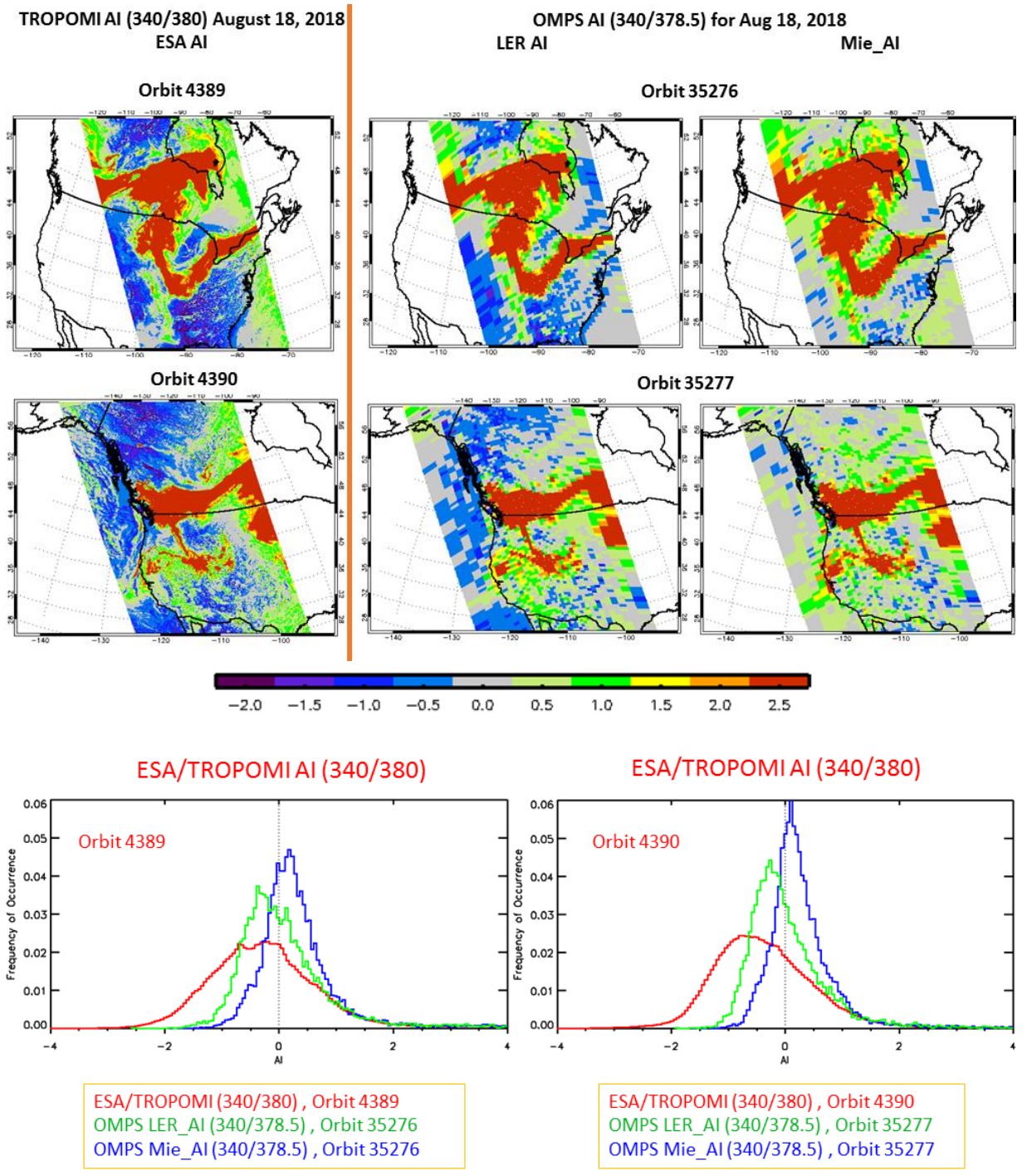


Figure 113: Comparison of UVAI from TROPOMI and OMPs for an observation of a biomass burning plume (18 Aug. 2018). For OMPS UVAI are calculated assuming either LER or Mie clouds. The UVAI for Mie clouds yields results that are more consistent. TROPOMI and OMPS LER results show very good agreement. However, the frequency distribution is broader for TROPOMI observations (courtesy of Omar Torres and Changwoo Ahn, NASA-GSFC).

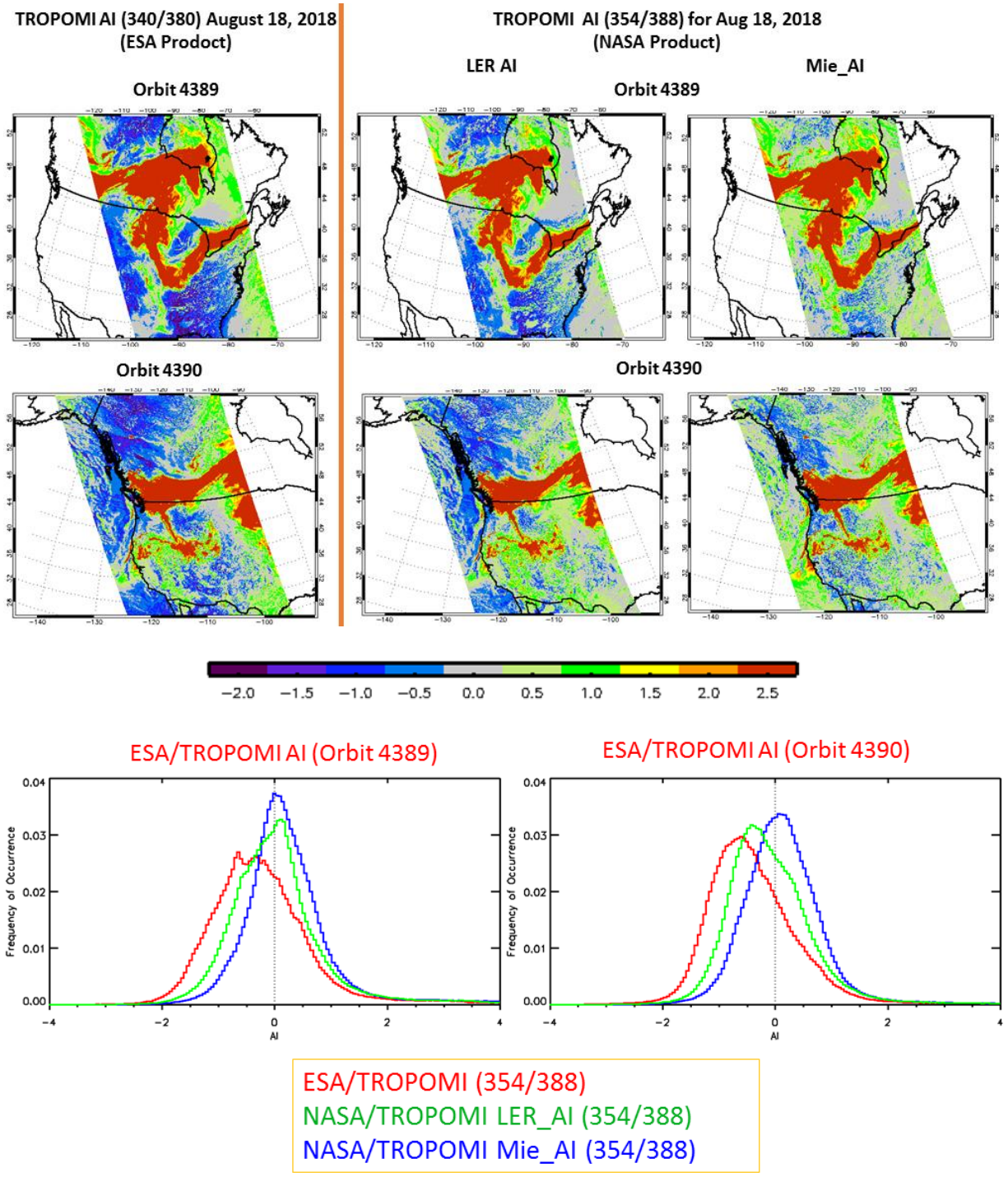


Figure 114: Comparison of UVAI from the operational TROPOMI UVAI product and TROPOMI UVAI analysed using the OMPS algorithm for an observation of a biomass burning plume (18 Aug. 2018). Like in the previous comparison, good agreement is found, but still the frequency distribution for the operational product is slightly broader than for the results using the OMPS algorithms (courtesy of Omar Torres and Changwoo Ahn, NASA-GSFC).

From the performed validation studies, it is concluded that the L2_AER_AI UVAI from S5P TROPOMI is of very good quality and fulfils the requirements. The spread (see also **Figure 116**) of the UVAI should be further investigated. Investigations are underway to possibly improve this spread by using a more realistic cloud model (Mie) and surface reflectance.

14.3.3 Bias

The very small difference (about 0.5 UVAI units) between S5P TROPOMI and other instruments measuring aerosol index (OMI and OMPS) is within the requirements.

14.3.4 Dispersion

The S5P TROPOMI UVAI is very probably within the requirement for random errors of 0.1 UVAI unit.

14.3.5 Dependence on influence quantities

There is a slight cross-track dependence of -0.25 (West – East side of TROPOMI swath), which is related to the use of the LER model in the retrieval. It should be noted that this cross-track dependence decreases with increasing UVAI values.

14.3.6 Short term variability

The global mean aerosol index is evaluated to give an overall indication of the stability of the data product. The global mean is calculated for all pixels on day with full global coverage and it is not expected to vary from day-to-day. A time series of the global mean is given for the TROPOMI UVAI for both wavelength pairs and for the NRTI and OFFL data streams. The period of 20 July 2018 through February 2024 is shown in **Figure 115**. Similar values are found for both wavelength pairs. A very slight degradation of -0.1 UVAI units is found over the whole period. But this change is still within requirements.

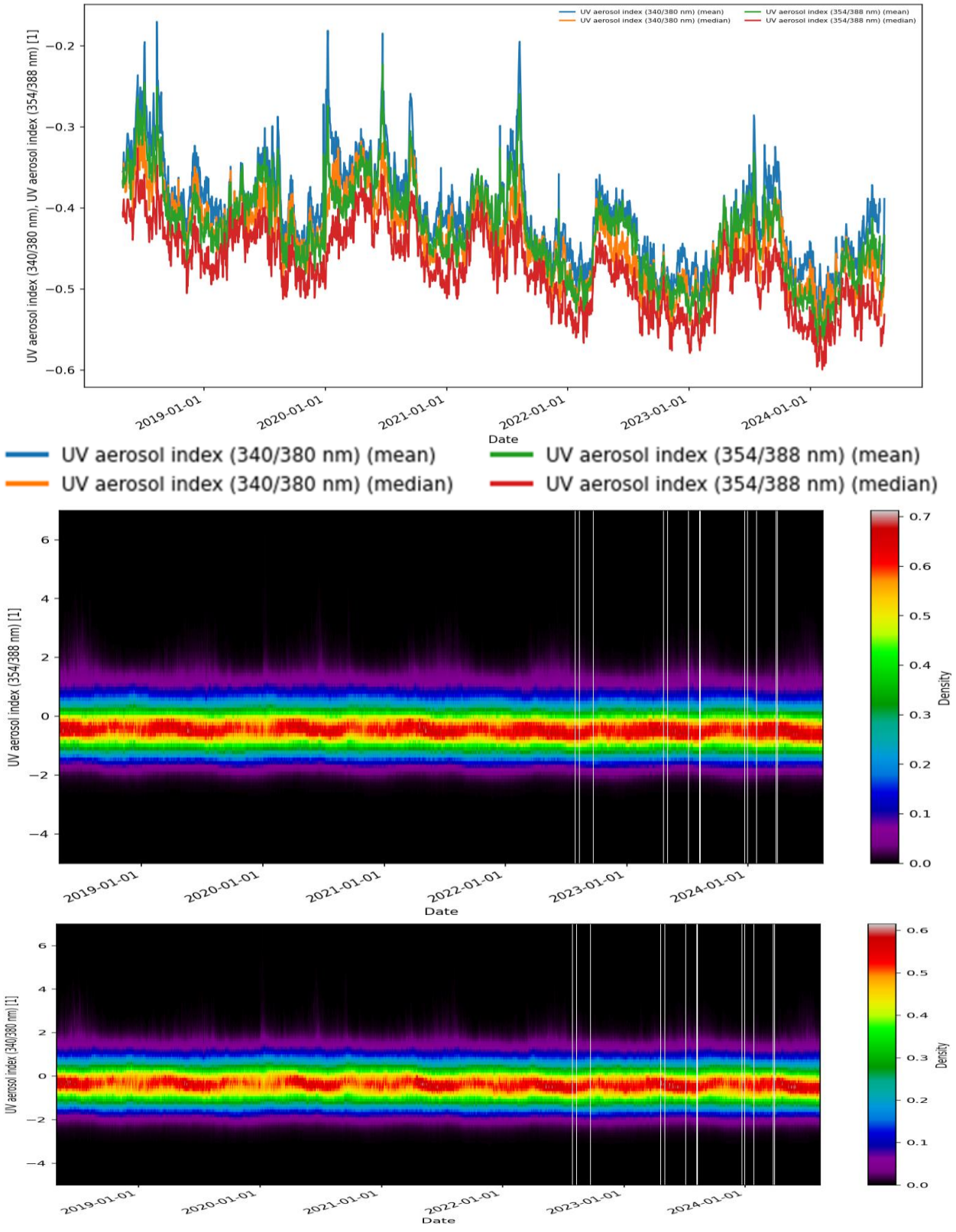


Figure 115: Comparison of the global daily mean and median for both L2_AER_AI UVAI wavelength pairs (340/380 and 354/388 nm), for the reprocessed data set from 20 July 2018 through August 2024. Middle and bottom: Corresponding histograms for both wavelength pairs.

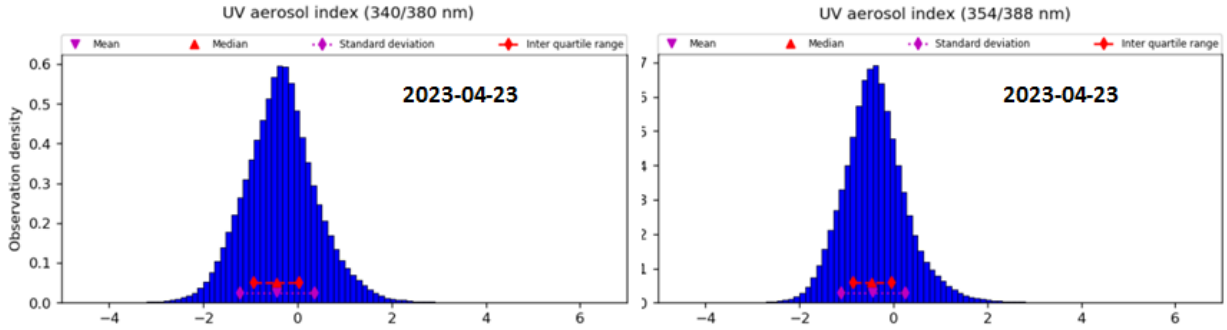


Figure 116: Frequency distribution of the UVAI (left: 340/380nm, right: 354/388 nm) for April 23, 2023.

14.3.7 Geographical patterns

There are no obvious geographical features. For pixels (partially) covered by clouds with a small horizontal extent and a non-homogeneous vertical structure, these clouds are non-Lambertian and result in positive values similar to that of absorbing aerosol. It should also be noted that for many fully clouded scenes, aerosols might be located below the clouds and are therefore invisible for the satellite instrument.

14.3.8 Other features

The slight (increasing) negative bias of the S5P TROPOMI results should be monitored. The temporal variation of the spread should also be investigated.

14.4 Equivalence of L2_AER_AI NRTI and OFFL products

Figure 117 below shows a comparison for a selected orbit on October 3, 2018. For this orbit, the L2_AER_AI UV aerosol absorbing index for both wavelength pairs are very similar for the OFFL and NRTI products. Based on this comparison and the comparison of the global means shown before, the close similarity in behaviour of both the NRTI and OFFL data streams indicates that the validation results for the NRTI data product are also valid for the OFFL data product.

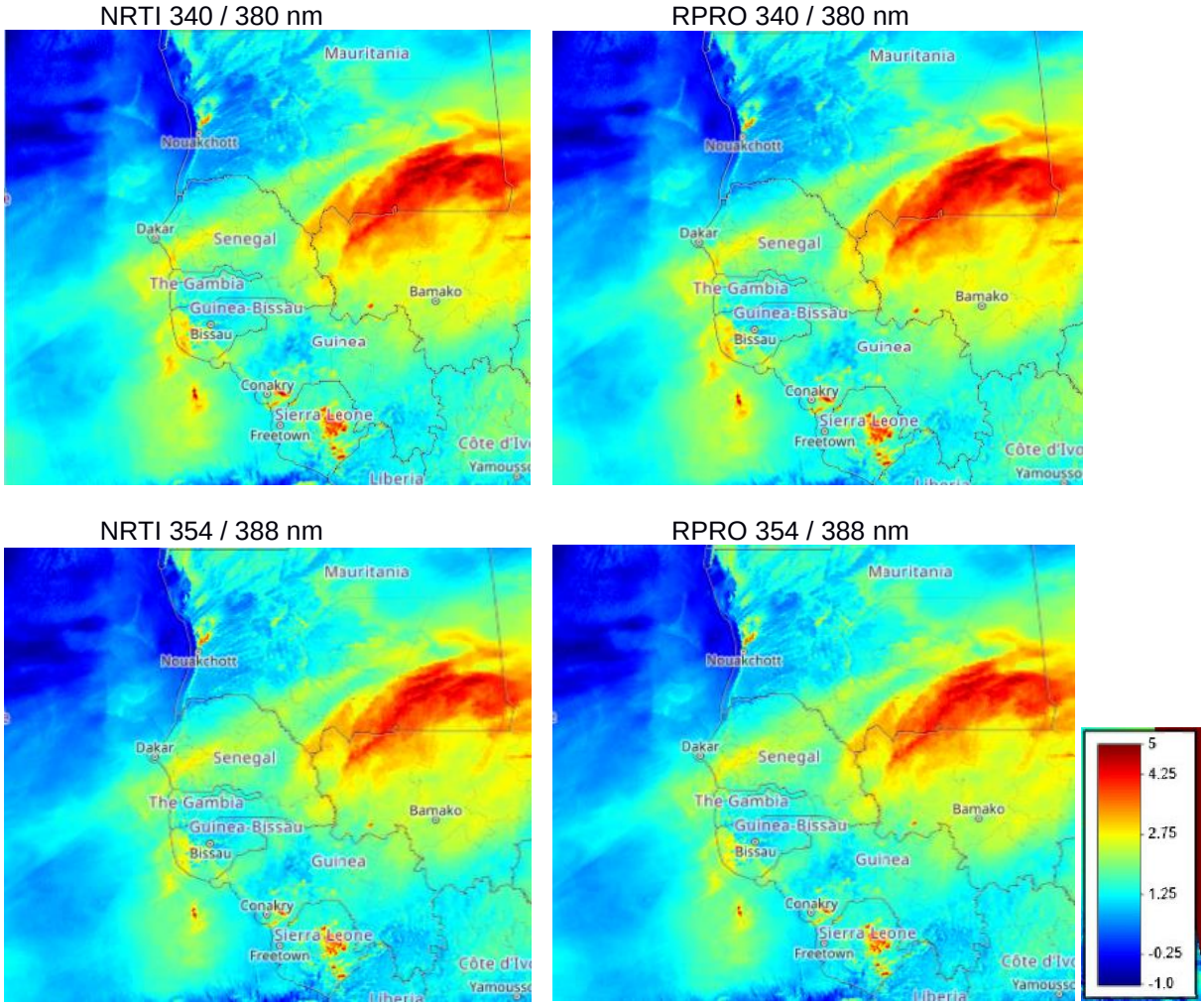


Figure 117: Comparison of the S5P TROPOMI near real time and reprocessed UVAI for a selected area on October 4, 2022 for the two wavelength pairs (top: 340 / 380 nm, bottom: 354 / 388 nm).

15 Validation Results: L2_AER_LH

15.1 L2_AER_LH products and requirements

This section reports on the validation of the S5P TROPOMI L2_AER_LH aerosol layer height (ALH) product as identified in **Table 1**. Validation results are discussed with respect to the product quality targets outlined in **Table 3**. Most validation results are shown for the L2_AER_LH OFFL product (v02.06.00) which was implemented in November 2023. In section 6.3.3, also validation results for the version 1 ALH results are presented, which are representative for the ALH data before November 2023. In general, very good agreement between the offline and near-real-time product is found (see e.g. previous versions of the validation reports).

15.2 Validation approach

The ALH presented here is computed only for known aerosol layer, which, lacking an AOT product, is done by selecting high UV AI values (larger than 0) and a cloud fraction filter. This means that mainly desert dust, smoke and volcanic plumes will be processed. Therefore, the validation focused on selected desert dust cases, fires plumes and volcanic eruptions.

15.2.1 Ground-based networks

Validation of the TROPOMI ALH with ground-based networks is desirable, since satellite-to-satellite comparisons have their own specific limitations. However, currently, the number of ground-based observations co-located with TROPOMI during suitable aerosol events is very limited to non-existent. A small number of observations using selected ground-based lidars in southern and eastern Europe (see below) were compared to TROPOMI ALH, but more comparisons from lidar and ceilometer networks may help provide more statistically representative results

The European Aerosol Research Lidar Network, EARLINET (https://www.earlinet.org), was founded in 2000 as a research project for establishing a quantitative, comprehensive, and statistically significant database for the horizontal, vertical, and temporal distribution of aerosols on a continental scale (Pappalardo et al., 2014). Since then EARLINET has continued to provide the most extensive collection of ground-based data for the aerosol vertical distribution over Europe. EARLINET observations are performed on a regular schedule for daytime and nighttime measurements. In addition to these systematic measurements for the consolidation of a European aerosol climatology, further observations are devoted to monitoring special events over the continent, such as Saharan dust outbreaks, forest fires, photochemical smog, and volcanic eruptions. The EARLINET database represents the largest collection of ground-based data of the vertical aerosol distribution on a continental scale. The main information stored in the files of the EARLINET database is the vertical distribution backscatter and aerosol extinction coefficients. The basic issue in this validation approach is the difficulty in identifying good spatiotemporal collocations between EARLINET lidar stations observations and TROPOMI/S5P overpasses.

15.2.2 Satellites

S5P TROPOMI aerosol layer height data were compared to the stereoscopic plume height product from MISR and to the weighted extinction height provided by CALIOP. The stereoscopic plume height product from MISR is an offline product that can be computed for selected fire plumes, using a freely available code (MINX). It makes use of the nine available viewing directions of MISR, which senses a scene from different directions during an overpass. This provides stereoscopic height information for a scene with enough contrast. The MINX code has to be processed manually, and also the fire plumes have to be hand-picked and selected digitally by hand. In this document plumes from 115 fires in 2018, prepared and provided by D. Griffin from the Environment and Climate Change Canada institute, are compared with TROPOMI ALH. Furthermore, the weighted extinction height from CALIOP on Calipso are compared to TROPOMI ALH for collocated pixels. All pixels were selected where Calipso was closer to S5P than 100 km and the sensing time of CALIOP and TROPOMI was less than three hours apart. The resulting number of pixels (about 2.5 million pixels in from May 2018 – March 2019) were screened for clouds and selected for aerosols. This resulted in about 1 million pixels over the oceans and 0.5 million pixels over land. The results of the comparisons are presented below.

A few more satellite products are available for comparison with the TROPOMI ALH. GOME-2 provides the Absorbing Aerosol Height (AAH), which is layer height product that is computed for selected pixels with high UV-AI, representing thick absorbing aerosol plumes. The AAH is comparable to the ALH since it also uses the depth of oxygen absorption lines in the O2-A band to derive the height of scattering layer. However, it differs from the ALH in that it only uses one or a few absorption lines and the continuum, while the TROPOMI ALH fits about 3,500 lines in the O2-A band, which should make it more accurate than the AAH. A similar product as the GOME AAH is available from EPIC on DSCVR. This product can be expected to have similar accuracy as the GOME AAH, but since DSCVR is parked in Lagrangian point L1 between the Sun and the Earth, it can deliver aerosol layer height at a one-hour time resolution. This would make it possible to monitor the evolution of aerosol layer heights and cover the time differences between overpasses of e.g. Calipso and MIRS, and TROPOMI.

15.2.3 Field campaigns and modelling support

So far, no field campaigns have been planned to validate the ALH.

15.3 Validation of L2_AER_LH

15.3.1 Recommendations for data usage followed

The ALH is very sensitive to cloud contamination. However, aerosols and clouds can be difficult to distinguish, and ALH is computed for all FRESCO effective cloud fractions smaller than 0.05. Since the ALH is sensitive to elevated scattering layers, and cloud layers are generally optically (much) thicker than aerosol layers, not discriminating between clouds and aerosol will strongly bias the ALH towards cloud layer heights. Cloud masks are available from FRESCO and VIIRS, and are strongly recommended to filter for residual clouds. A sun-glint mask is also available to screen sun-glint regions, which are not filtered beforehand. These and other sources of uncertainties are indicated with the `qa_value`. Use of pixels with a `qa_value` below 0.5 is not recommended.

The variables `aerosol_mid_pressure_precision` and `aerosol_mid_height_precision` can also be further used to diagnose the quality of the ALH.

For further details, data users are encouraged to read the Product Readme File (PRF), Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD) associated with this data product on https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms [ER_CoperATBD].

15.3.2 Status of validation

This section presents validation results obtained as a part of Validation Team (S5PVT) AO projects and development tests during the update of the forward model.

The validation of S5P TROPOMI L2_AER_LH data presented here is based on comparisons with MISR and CALIOP. In **Table 18**, the details of five selected cases are presented, which were compared to the CALIOP weighted extinction height. A sixth case of very high altitude smoke from intense biomass burning in Australia in early 2020 (see section 6.3.9) shows a notable difference with CALIOP measurements, showing a limitation of the S5P L2_ALH product.

Table 18 – Case studies for desert dust cases with Calipso co-locations

Date	Type of Case	TROPOMI orbit	Calipso orbit start time [Day/Night]
2018-06-01	Desert Dust	3280	14:28:34 [D]
2017-07-31		4343	09:55:28 [D]
2020-06-17		13878	15:07:45 [D]
2020-09-07	Smoke	15045	20:49:39 [D]
2020-02-02	pollution	12058	18:19:17 [D]

In the three images below, the desert dust cases are illustrated:

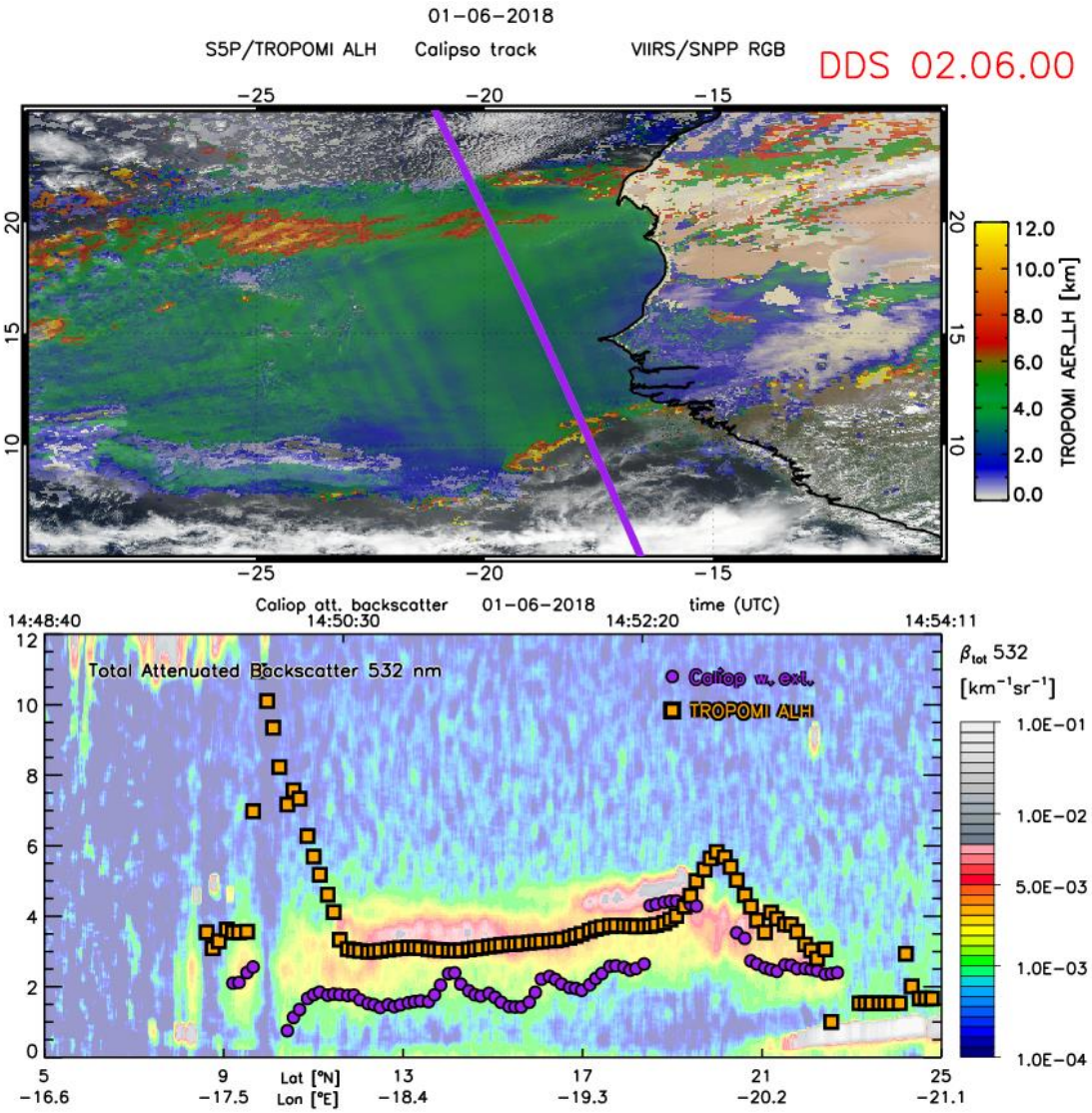


Figure 118: ALH (v02.06.00) for a desert dust case on 1 June 2018 retrieved by TROPOMI (top panel) and the CALIOP total attenuated backscatter image (bottom panel). The purple line and dots show the CALIPSO track and CALIOP weighted extinction averaged along the track, respectively. In the bottom panel the ALH retrieved by TROPOMI along the CALIOP orbit track is shown by the orange squares.

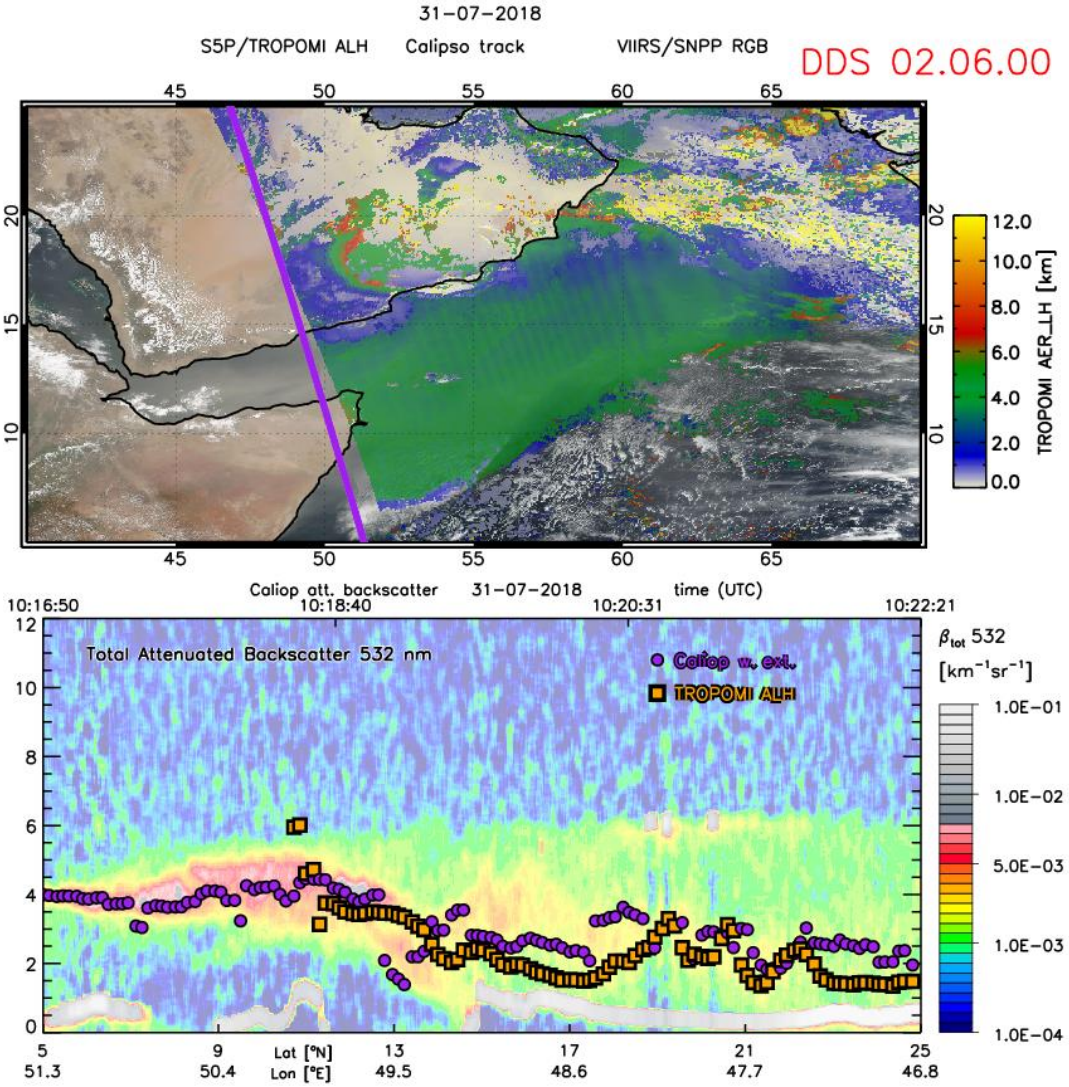


Figure 119: Same as above, but for 31 July 2018 near the Arabian Peninsula.

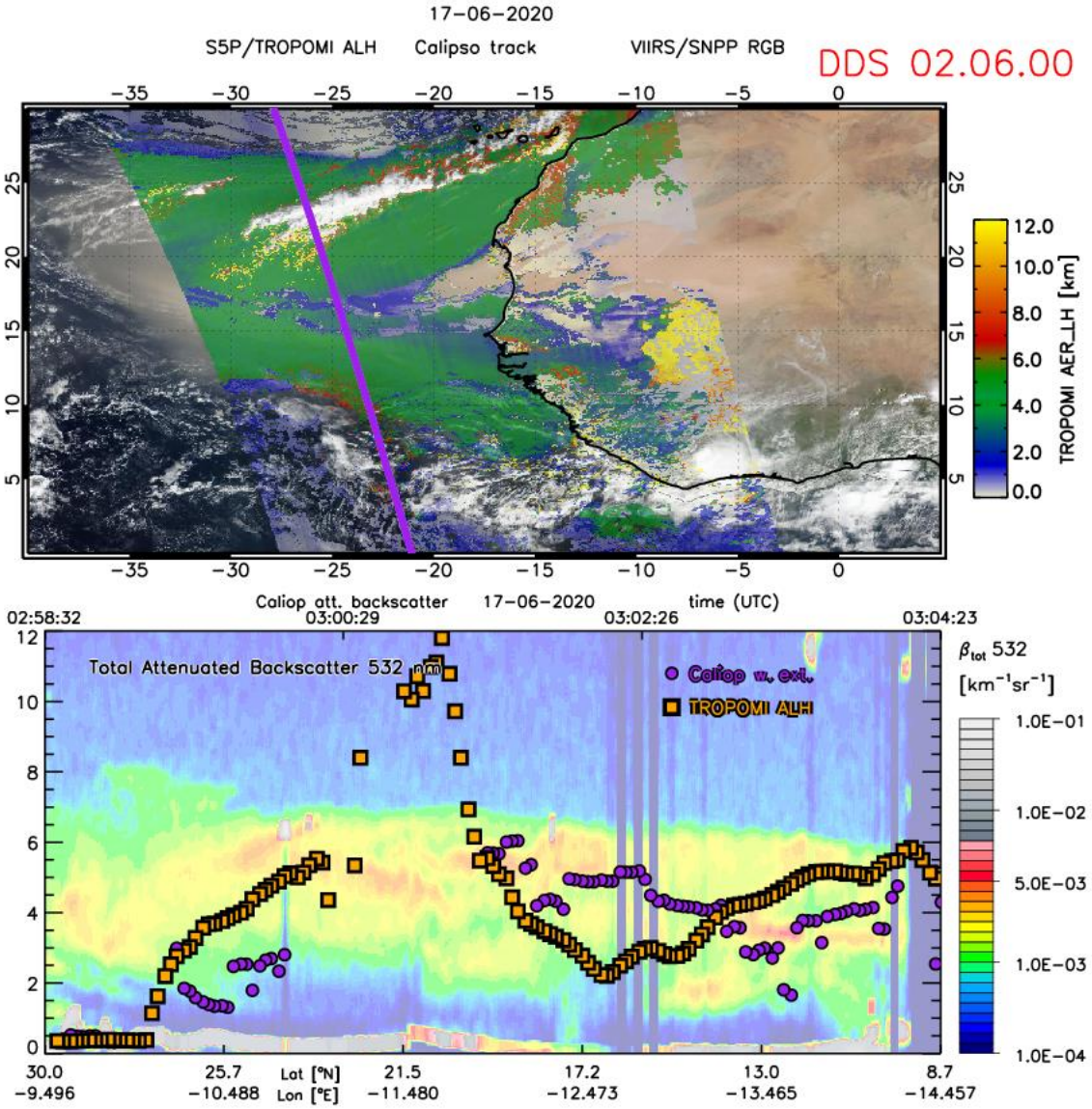


Figure 120: Same as above, but for 17 June 2020.

First, the images show that the maximum attenuated backscatter measured by CALIOP, at 532 nm, is not a good indicator of the plume centre height. The maximum total attenuated backscatter is often at the plume top. The weighted extinction height is computed from the level-2 aerosol extinction profiles. Here, aerosol extinction is computed in cloud-free areas using a feature mask, distinguishing (among others) aerosol and cloud layers. Each well-defined aerosol layer and aerosol-free layer is split in 100 m height segments to allow for averaging over complex layer structures along the CALIOP path. The average extinction height is then computed by (Nanda, *et al.,* 2018):

$$Z_{ext} = \frac{\sum_{i=1}^{n} b_{ext,i} \cdot Z_i}{\sum_{i=1}^{n} b_{ext,i}}$$

where Z_i is the height from sea level in the ith lidar vertical level (in km), and $b_{ext,i}$ is the aerosol extinction coefficient (in km^{-1}) at the same level.

The weighted extinction height is an indication of the maximum of the extinction, and is often related to the centre of a plume if the attenuation of the beam is small. However, for strongly attenuated beams, the weighted extinction height is biased to the top of the plume. **Error! Reference source not found.** show that the weighted extinction height correlates well with the TROPOMI ALH for dust. The retrieval over land remains challenging, especially over bright surfaces. However, the surface albedo fitting (since v02.06.00) has clearly improved the retrievals over land considerably, with also many retrievals where earlier version of the algorithm failed. TROPOMI generally shows comparable plumes heights with CALIOP. However, clouds strongly bias the TROPOMI ALH towards the cloud altitude. Therefore, additional cloud screening, which is available in the product in the form of flags, is essential for the user to retrieve proper aerosol layer heights.

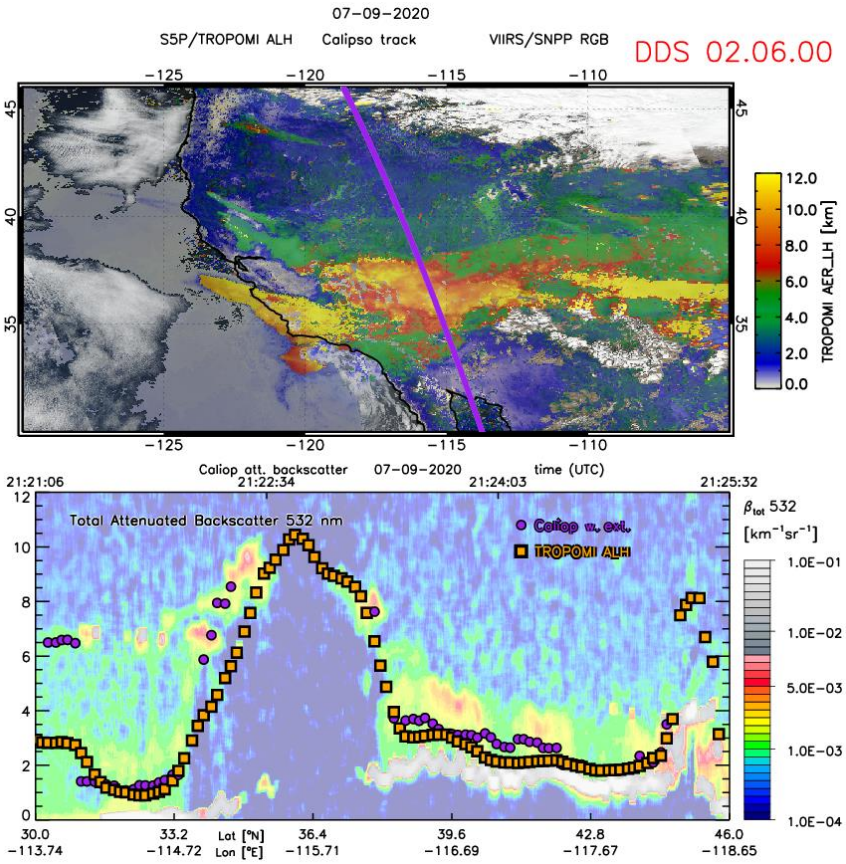


Figure 121: ALH for the Californian wildfires case on 7 September 2018 retrieved by TROPOMI (top panel) and the CALIOP total attenuated backscatter image (bottom panel). The purple line and dots show the CALIPSO track and CALIOP weighted extinction averaged along the track, respectively. In the bottom panel the ALH retrieved by TROPOMI along the CALIOP orbit track is shown by the orange squares.

In **Error! Reference source not found.**, the TROPOMI ALH is shown and compared to CALIOP weighted extinctions for a rather challenging multi-layer smoke case over California. The smoke plume is very dense and the TROPOMI ALH retrieval closely follows the CALIOP weighted extinction retrieval over the entire domain shown, which is remarkable, given the large vertical variation of the smoke, between less than a kilometre to more than ten kilometres altitude. Furthermore, the majority of the retrievals are over land, and the plume is very well covered with nearly all pixels converging to a plausible ALH. The smoke in this case covered valleys in the north to the domain shown and raised to high altitude, which is very well depicted by the ALH, showing many of the vertical features of the smoke distribution.

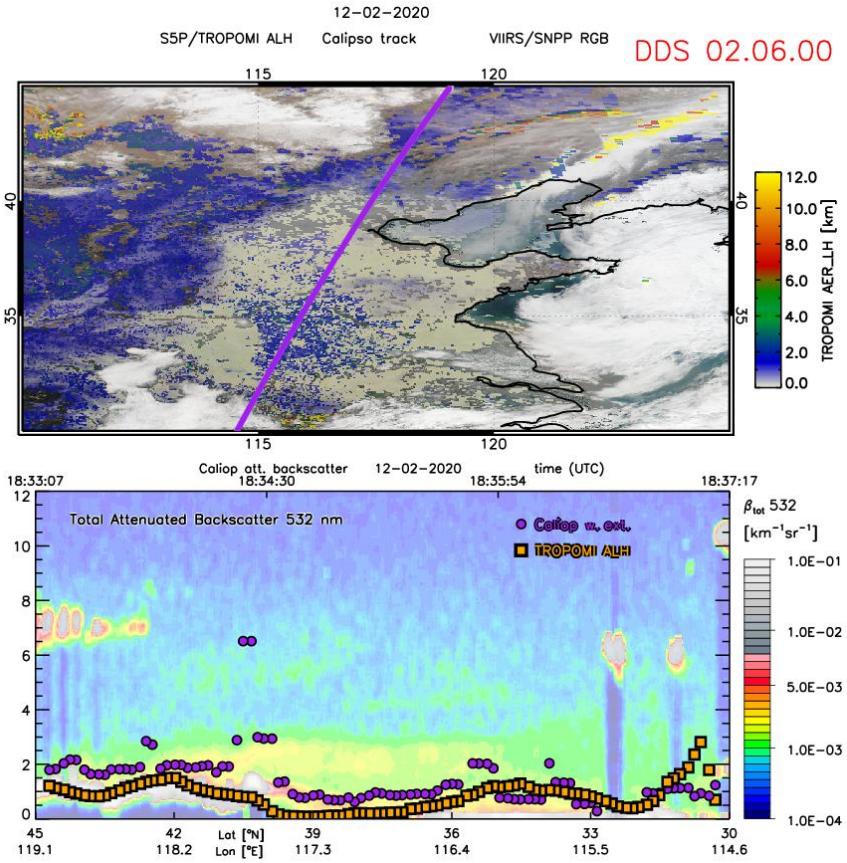


Figure 122: ALH for pollution over China on 12 February 2020 retrieved by TROPOMI (top panel) and the CALIOP total attenuated backscatter image (bottom panel). The purple line and dots show the CALIPSO track and CALIOP weighted extinction averaged along the track, respectively. In the bottom panel the ALH retrieved by TROPOMI along the CALIOP orbit track is shown by the orange squares.

The TROPOMI ALH is shown and compared to CALIOP weighted extinctions for a case over East Asia with severe pollution from the Beijing and Shanghai industrial regions. This case is interesting, because the TROPOMI ALH retrievals are very close to the ground, which is very likely for these aerosols, and the small, industrial aerosols are mostly scattering aerosols, in contrast to desert dust and smoke, which have a strong absorbing component as well. Since version 2, all cloud-free scenes are processed for ALH and this example shows that scattering aerosol close to the ground are apparently also well captured by the ALH.

The five cases represent rather thick desert dust, smoke and pollution plumes, all of which were selected for their strong signal. These are not representative for the global average ALH, but for verification purposes, we use these cases to investigate the change in ALH for different versions. For the current version (v02.06.00), a comparison with Caliop weighted extinction is presented in Figures 114 and 115.

A linear regression with Caliop Weighted extinction shows a good correlation with TROPOMI ALH. Interestingly, since the introduction of the surface albedo fit, the majority of the land retrievals for the selected cases are within the requirements, shown by the grey band in Figure 114. Also the ocean retrievals show a good correlation, but a larger spread. This may be due to sun glint, which is something that needs to be investigated.
A histogram of the difference between Caliop weighted extinction and TROPOMI ALH is shown in Fig. 115 for the same cases. It shows again that the ocean retrievals match very well between Caliop and

TROPOMI, but with a larger spread than over land. Over land, however, a bias of the TROPOMI ALH retrieval can be observed of about 0.7 km towards the surface, compared to Caliop weighted extinction.

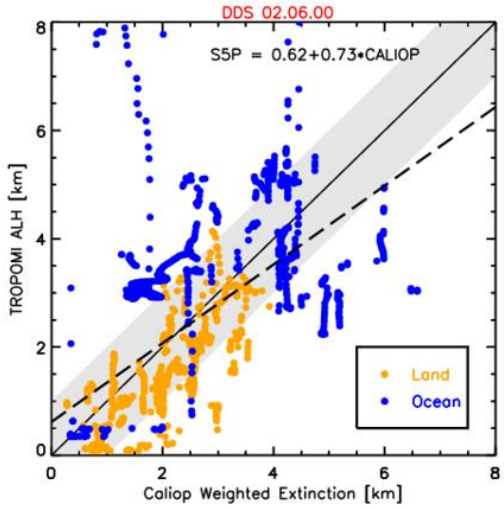


Figure 123: Comparison of ALH from TROPOMI and CALIOP weighted extinction height for the cases presented**Error! Reference source not found.**. Pixels over land are plotted in orange and pixels over water in blue.

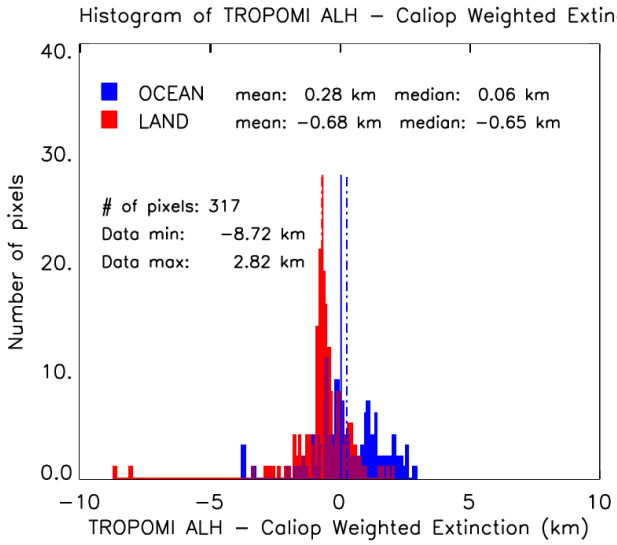


Figure 124: Histogram of the difference between the ALH from TROPOMI and CALIOP weighted extinction for the cases presented**Error! Reference source not found.**. Pixels over land are plotted in red and pixels over water in blue.

15.3.3 Version 1 data

Below the validation results for version 1 data are retained for comparison with the current data version. The comparison between CALIOP and TROPOMI was extended to all collocated pixels within 100 km and within 3 hours of each other, yielding about 1 million pixels over the ocean and 0.5 million over land, see **Figure 125** (left). The figure shows that the TROPOMI ALH is systematically lower than CALIOP weighted extinction heights. The retrieved ALH from TROPOMI differs from CALIOP weighted extinction height by 1.0 km on average, with a standard deviation of 1.97 km. More than 50% of the TROPOMI ALH retrievals over the ocean have an absolute difference with CALIOP weighed extinction height less than 1.0 km. Retrievals over land have a larger difference, with -2.41 km on average and a median of -1.75 km. The results are very skewed over land, with very large negative values dictating the average — this is indicated by the very large standard deviation of 3.56 km. 50% of the selected colocations over land have an absolute

difference with CALIOP weighted extinction height less than approximately 1.0 km. On the right, a similar histogram is shown, but now for only those pixels that have a minimal cost function, or χ^2, smaller than 1E5. The χ^2 represents the goodness-of-fit of the modelled sun-normalised radiances to the observations in the O2-A band, and therefore is a measure of the representativeness of the model (of a simple one aerosol layer atmosphere with known surface reflectance) to reality. Smaller χ^2 indicate a better fit. The retrievals over land generally have much higher χ^2, and therefore are less reliable. The right panel in **Figure 125** show the results for pixels with a χ^2 than can be expected to be a reasonably good fit. The differences between TROPOMI ALH and CALIOP weighted extinction height then reduce to -0.62 km over ocean and -1.2 km over land.

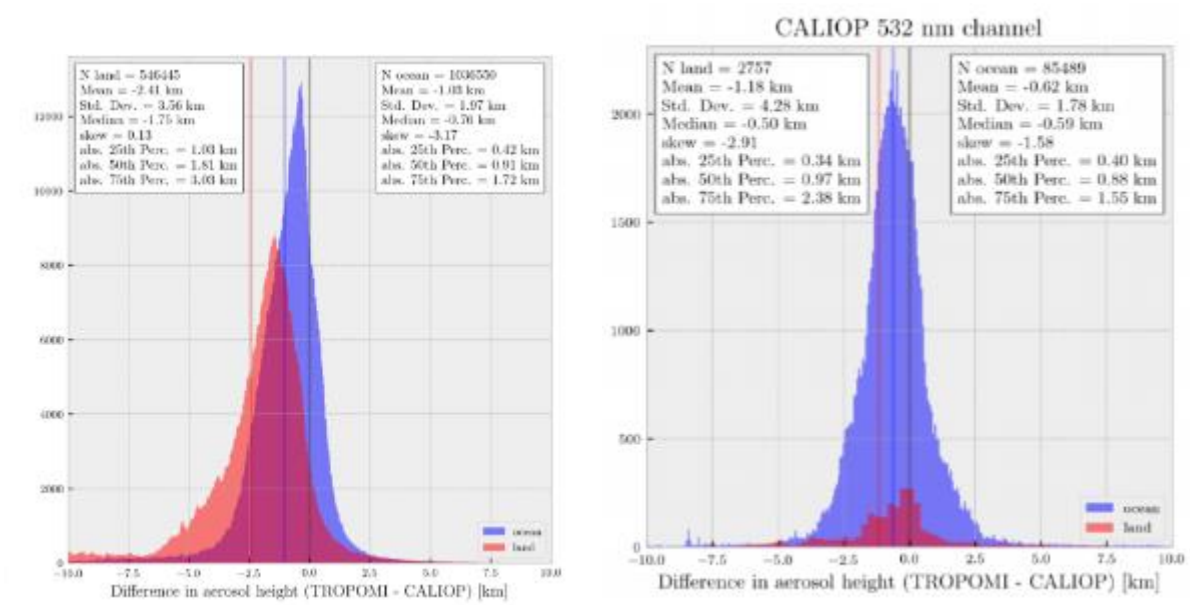


Figure 125: Histogram of differences between CALIOP weighted extinction height and TROPOMI ALH from collocated data between 1 May 2018 and 28 February 2019 (left). The right panel shows the same histogram, but for pixels that were screened for a minimal cost function (chi-squared) smaller than 1E5.

Additional validation of the TROPOMI ALH was provided by Environment and Climate Change Canada. TROPOMI ALH was compared to MISR stereoscopic plume height and CALIOP “layer_base_altitude” and “layer_top_altitude” products for 115 fire plumes in 2018 over northern America (Griffin et al, 2019). The results are summarized in **Figure 126** and **Figure 127**. The maximum plume heights above ground level for the 2018 fires in North America are, on average, 2 km (ranging between 0.4 and 5.5 km) and 1.6 km (ranging between 0.01 and 8.4 km) for MISR and TROPOMI, respectively. The mean plume heights (above ground level) within one fire plume are on average 1.4 km (ranging between 0.3 and 3.2 km for MISR) and 0.8 km (ranging between 0.01 and 2.8 km for TROPOMI). Overall, TROPOMI’s maximum and mean plume height is on average 0.59±1.3 km and 0.55±0.74 km lower than the plume height derived from MISR, respectively.

The difference between the plume height observed by TROPOMI and CALIOP depends significantly on the thickness of the plume (as derived from CALIOP). Thicker plumes seem to be better captured by TROPOMI and the thicker the plume the smaller the difference between the CALIOP and TROPOMI plume height. TROPOMI was biased low in comparison to CALIOP for thin smoke plumes (thickness of less than 1.5 km) and TROPOMI ALH is on average 2.1 km lower. Much better agreement and a higher correlation between the two satellite datasets is found for thicker plumes. The mean difference reduces with the thickness of the plumes, the mean difference between the TROPOMI and CALIOP mid aerosol layer is just 50 m for very thick plumes (>3 km). The geometrically thick plumes are typically optically thicker

plumes, too. The reason for the reduced bias with increasing layer thickness is probably the sensitivity of the TROPOMI AER_LH algorithm to the scattering layer in the scene, which is more and more dominated by the surface if the aerosol layer is optically thinner. Currently, a simple Lambertian Equivalent Reflection (LER) database from GOME-2 is used in the ALH retrieval to fit the observations to the simulated reflectances. An improvement is expected when a (directional) LER database from TROPOMI becomes available.

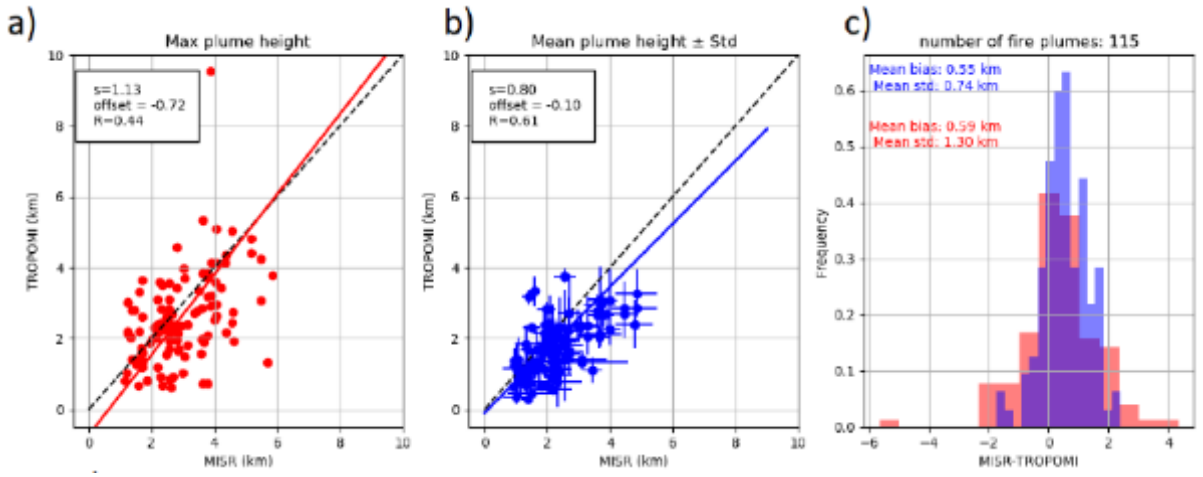


Figure 126: Comparison of TROPOMI ALH (processor version before v02.05.00) and MISR plume height for 115 fires over Nothern America in 2018. See Griffin et al, 2019 for details.

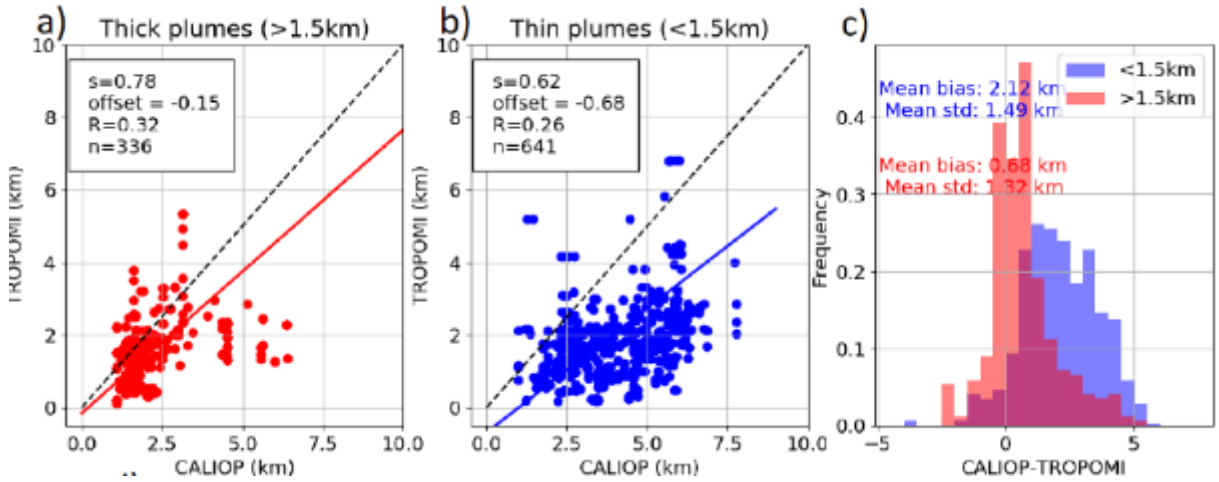


Figure 127: Comparison of TROPOMI ALH (processor version before v02.05.00) and CALIOP average aerosol layer height (top minus bottom of aerosol layer as defined by the feature mask) for collocated pixels near fires over Northern America in 2018. See Griffin et al, 2019, for details.

The validation of S5P TROPOMI L2_AER_LH data presented below is based on comparisons with EARLINET ground-based lidar stations. Data from fifteen stations were considered in this analysis and were compared to S5P L2_AER_LH (RPRO +OFFL) v02.06.00 data for the period from May 2018 until August 2024. Regarding the use of the reprocessed TROPOMI AER_LH data product, the Product Algorithm Laboratory (PAL) was implemented. The geographical distribution of the selected EARLINET stations depicted in **Figure 128** indicates the domain of applicability of the validation results. All participating stations (red circles) operate high-performance multi-wavelength lidar systems. The location of the stations across the Mediterranean basin is an ideal test environment for TROPOMI ALH features due to their proximity to the Sahara Desert and Europe, with frequently observed events of mineral dust and smoke particles. The TROPOMI aerosol layer height product can be examined under a complete set of different atmospheric conditions. The ALH product was tested over both land and sea. Over land, the TROPOMI ALH product has decreased detection capabilities than over the sea surfaces since, over bright surfaces, the retrieval algorithm becomes increasingly sensitive to errors in the surface albedo features. This validation was performed by the team from the Aristotle University of Thessaloniki (AUTH). For the routine validation of the S5P/TROPOMI aerosol layer height retrievals, the automated validation server in LAP-AUTH deployed within the QA4EO project (Work Package 2191; Michailidis et al. 2021b) collects S5P ALH data and correlative measurements to identify suitable co-locations, compares to the co-located data, and produces S5P data quality indicators. The approach followed is mostly based on the previous expertise and methodology that have been developed and applied in EARLINET for the GOME2 cal/val activities (Michailidis et al., 2021a). Detailed information about the validation methodology and current status of the validation results can be found in Michailidis et al. (2023).

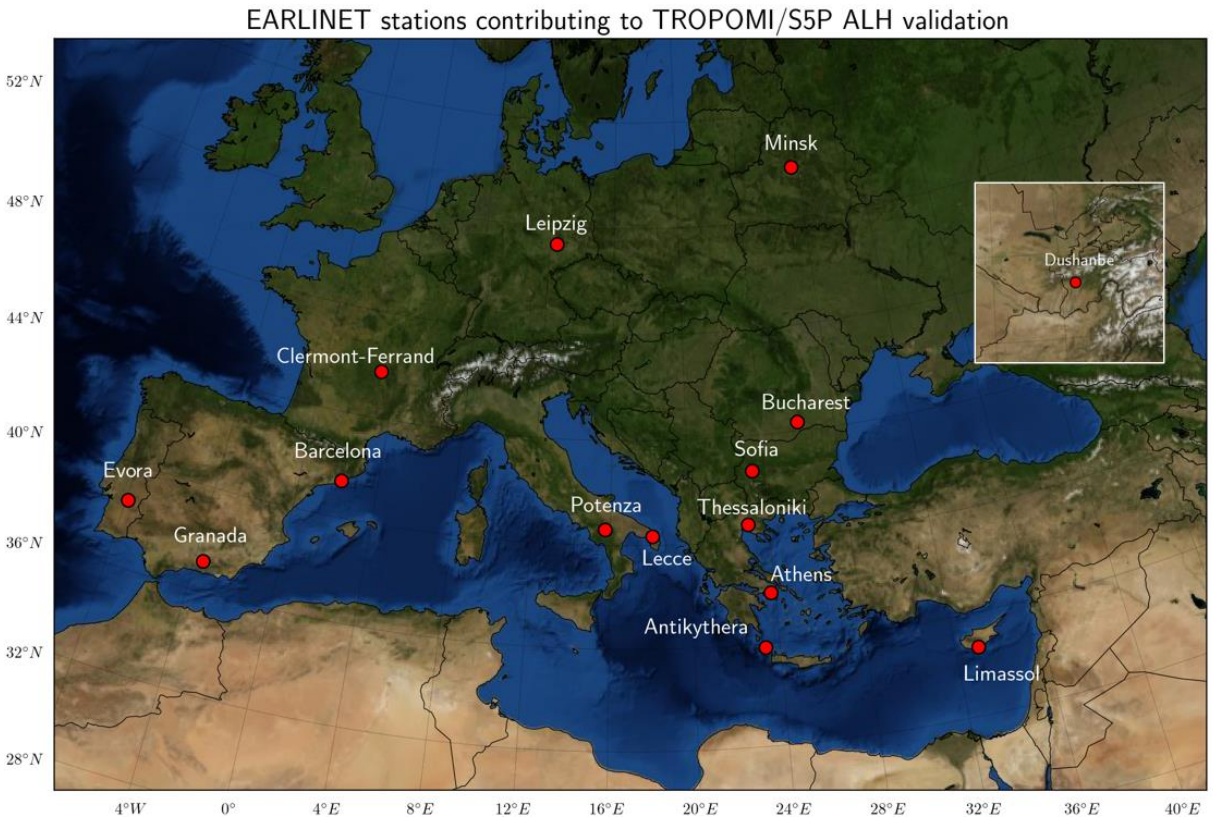


Figure 128: Geographical distribution of EARLINET ground-based stations for which co-locations with S5P L2 AER_LH data were used (period May 2018 – August 2023).

Table 19: Locations of EARLINET lidar stations and their geographical coordinates.

Station	Code	Country	Longitude, latitude, elevation
Antikythera	AKY	Greece	23.31°E, 35.86°N, 193m
Athens	ATZ	Greece	23.78°E, 37.96°N, 212m
Barcelona	BRC	Spain	2.12°E, 41.3930°N, 115 m
Bucharest	INO	Romania	26.03°E, 44.34°N, 93 m
Clermont-Ferrand	PUY	France	3.11°E, 45.7610°N, 420 m
Dushanbe	DUS	Tajikistan	68.85° E, 38.55°N, 864 m
Évora	EVO	Portugal	7.91°W, 38.56°N, 293m
Granada	GRA	Spain	3.60° W, 37.16°N, 680m
Lecce	SAL	Italy	18.10°E, 40.33°N, 30m
Leipzig	LEI	Germany	12.43°E, 51.35°N, 125 m
Limassol[1,2]	LIM	Cyprus	33.04°E, 34.67°N, 10m
Minsk	MAS	Belarus	27.60°E, 53.91°N, 200 m
Potenza	POT	Italy	15.72°E, 40.60°N, 760m
Sofia	SOF	Bulgaria	23.38°E, 42.65°N, 550 m
Thessaloniki	THE	Greece	22.95°E, 40.60°N, 50 m

[1]Cyprus University of Technology (CUT) [before Oct 2020]
[2]Leibniz Institute for Tropospheric Research and ERATOSTHENES Centre of Excellence [after Oct 2020]

15.3.4 Validation approach

The AUTH team used the aerosol layer height retrieved from ground-based lidar systems within EARLINET to validate the TROPOMI ALH product. TROPOMI observations, co-located with the ground-based EARLINET measurements, are found by selecting all filtered and averaged TROPOMI pixels within a radius of 150 km around each station and with a maximal time difference of 4h. Pixels with an associated quality assurance value of less than 0.5 were excluded. This filter does not remove all pixels considered unusable. Some pixels with unphysically low or high ALH information are still present. Some of the known data quality issues are not covered by the quality flag criterion of 0.5 and additional flags should be applied to the data during the validation analysis reported hereafter. Those issues include the insensitivity to very high altitudes aerosol layers and the treatment of remaining high-altitude clouds. The strong possibility of remaining clouds in the TROPOMI field-of-view is one of the reasons why an optimal spatial collocation with the lidar measurement is not achieved for every target pixel. Moreover, a common source of uncertainty when dealing with lidar data is the system's overlap height (Siomos et al., 2018), which determines the altitude above which a profile contains trustworthy values. In the validation analysis we consider the effect of the overlap in the estimation of the ALH from a lidar measurement. Routine validation is done using the automated ALH validation system operated at LAP-AUTH. The TROPOMI pixel selection scheme and flags applied in the presented validation study, were made following the recommendation by the Product User Manual (PUM) and Algorithm Theoretical Basis Document (ATBD). Detailed information about the validation methodology and current status of the validation results can be found in Michailidis et al. (2023).

The results of the comparison between the TROPOMI L2_AER_LH (RPRO+OFFL; v. 02.06.00) and EARLINET-derived aerosol heights for all the identified collocated cases between May 2018 and August 2024, are shown in **Figure 129** in the scatterplot of TROPOMI ALH against EARLINET ALH_{bsc} for all common cases (over land: N = 117, over ocean: N = 97) used for the intercomparison. By defining a weighted height from EARLINET aerosol backscatter profile products (ALH_{bsc}), the quantitative validation at pixels over the selected EARLINET stations illustrates that TROPOMI ALH over ocean is consistent with ALH_{bsc}, with a high correlation coefficient R=0.66 and mean bias 1.46 km. Over land, less good agreement is found (R = 0.56 and mean bias −0.49 km). The yellow solid line is the linear fit line between the datasets. The colour scale indicates the averaged TROPOMI aerosol index values. Overall, the TROPOMI ALH retrievals are systematically lower than the corresponding lidar heights in both clusters. This can also be seen in **Figure 129** (bottom), which presents histogram plots of their absolute differences, which appear smaller for ocean pixels (left Figure in red) and increase for the for land pixels (right Figure in green).

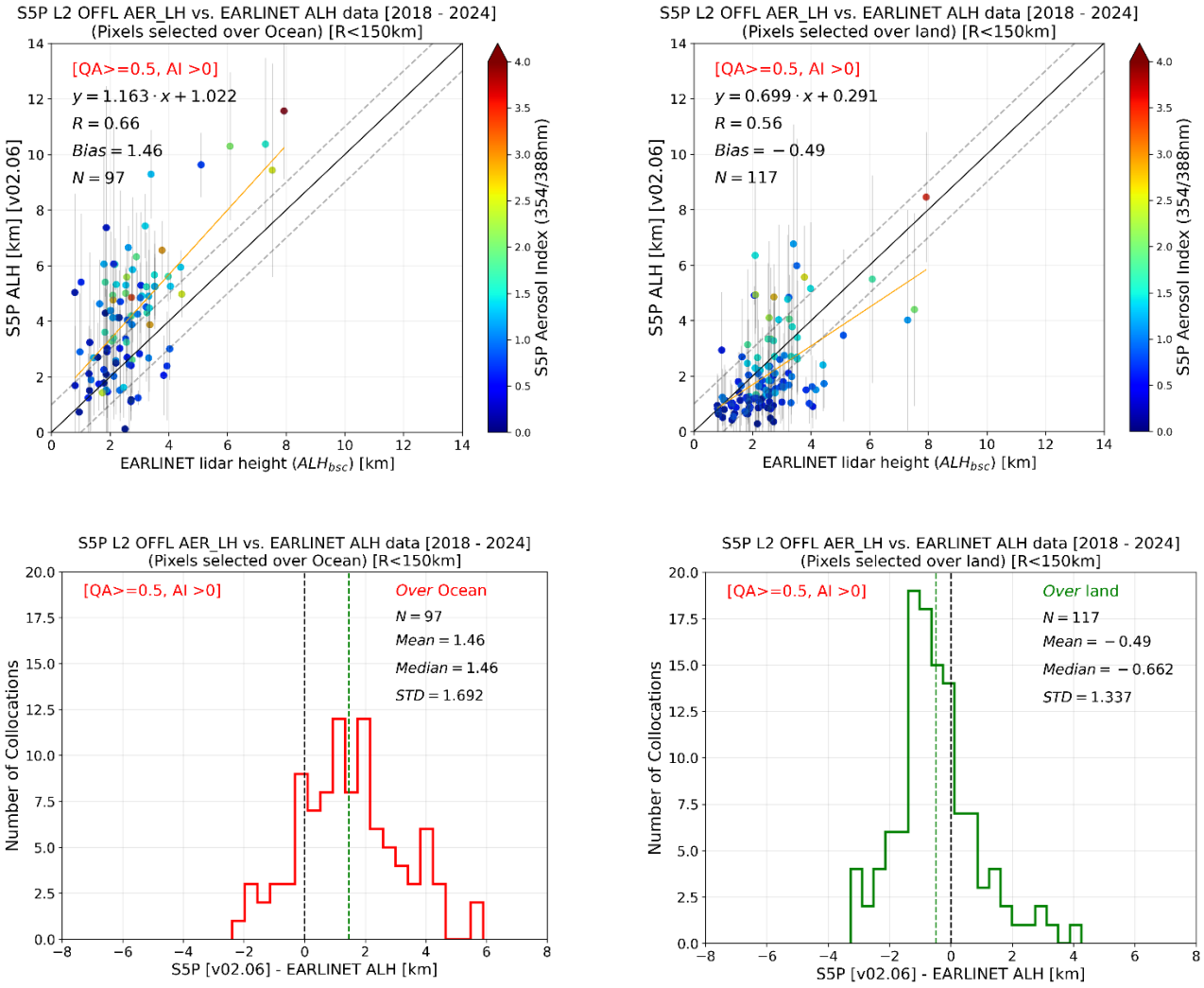


Figure 129: *Top*: TROPOMI ALH (v02.06.00) against EARLINET data for pixels over ocean (left) and land (right) for the period from May 2018 to August 2024. Colour-scale indicates TROPOMI retrieved UVAI (388/354nm) values, vertical error bars indicate the standard deviation of the spatial average of the TROPOMI ALH pixels. EARLINET weighted extinction height is consistently lower than TROPOMI ALH. *Bottom:* Histogram of absolute difference between TROPOMI and EARLINET ALHs, shown in red over the ocean pixels (left) and green for land pixels (right).

Validation statistics for collocated pairs are shown in **Table 21** for TROPOMI and EARLINET for correlation coefficient R, slope and intercept of a linear regression and mean and median absolute biases.

Table 20: Statistics on the comparison of the common subset of L2_AER_LH and co-located EARLINET stations.

	Number of cases	Slope	Y-intercept	Pearson R	Bias±std (km)
Over ocean	97	1.16	1.02	0.66	1.46 ±1.69
Over land	117	0.69	0.29	0.56	-0.49±1.33

To test the performance of the new S5P AER_LH processor version (v02.06) against previous versions (v02.04, 02.05), an analysis was carried out for the 2018 – 2023 period, using the archived collocated TROPOMI-EARLINET datasets for a detailed comparison. The reprocessed TROPOMI AER_LH V02.06 data product provided by the Product Algorithm Laboratory (PAL) was used on this scope. Selected cases (dust and smoke aerosols) were analyzed demonstrating the promising performance of the proposed algorithm under various aerosol loading scenarios over different surface surfaces (ocean/land).

i) **S5P AER_LH** product **performance over Ocean:**

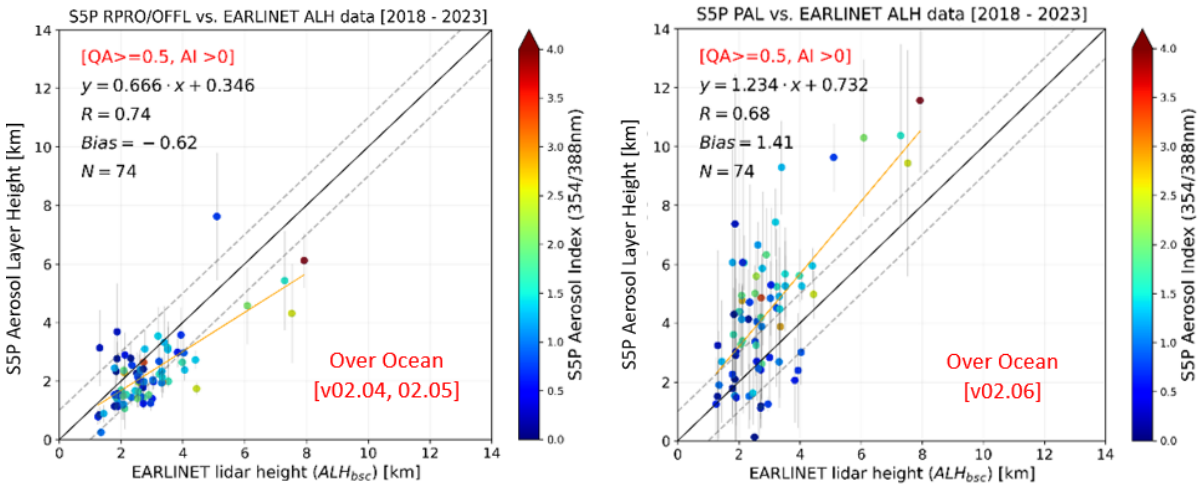


Figure 130: TROPOMI ALH plotted against EARLINET data over land, using S5P AER_LH v02.04 & v02.05 algorithms (left), and S5P AER_LH v02.06 algorithm (right). Color-scale indicates TROPOMI-retrieved UVAI values, error bars correspond to the spatial variability of the averaged TROPOMI ALH pixels. Data filters: QA≥0.5, AI>0, within 150 km from ground-based stations.

ii) S5P AER_LH product **performance over Land:**

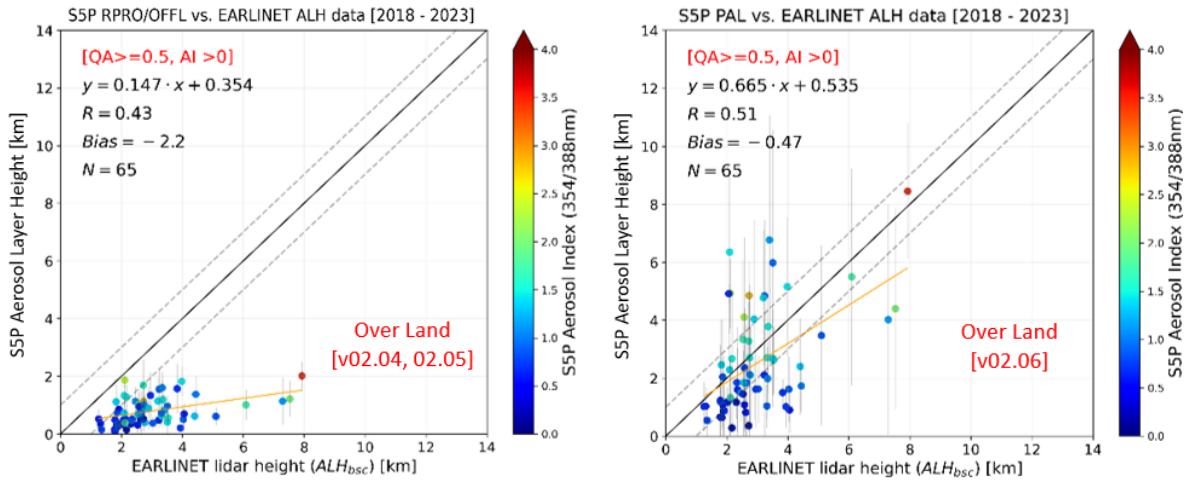


Figure 131: Same as previous, including only Land surface pixels.

The results of the comparison between the different TROPOMI L2 AER_LH algorithm versions against EARLINET-derived aerosol heights for collocated points between 2018 and 2023 are presented in Fig. 1 and Fig. 2 separately over ocean and land surface, respectively. The colour scale indicates the averaged TROPOMI aerosol index values. Regarding the latest 02.06 version, the quantitative validation at pixels over the selected EARLINET stations illustrates that TROPOMI ALH over ocean is consistent with ALHbsc, with a correlation coefficient R=0.68 and mean bias 1.41 km (Prev. version: R: 0.74 /MB: -0.62km). Over land, an agreement of R = 0.51 and mean bias -0.47 km is found (Prev. version: R: 0.43 /MB: -2.2km). Specific flags have been applied [QA≥0.5, UVAI>0] within a distance radius of 150km around the EARLINET stations. The statistical metrics are also shown in the inbox text and also are given in Table 1 for ocean and land satellite retrievals. It is clearly seen that with the implementation of the new version (v02.06) the number of processed pixels has increased significantly, both on land and at sea. It is also important to note that the distribution shows a significant increase in height levels. The mean layer heights are on average higher for TROPOMI over ocean and lower over land surface compared against lidar. The obvious overestimation of the aerosol layer height retrieved by the current algorithm from TROPOMI over ocean is probably related to the cloud-affected scenes, leading to biased or non-convergent retrievals. Thresholds for cloud-filtering are not rather strict where optically thin clouds can bias the retrieval of aerosols. Additionally, according to the findings, the previously observed large bias over land has been reduced since the surface albedo fitting was included in the retrieval procedure in the 02.06.00 processor version.

Table 21: Statistics of the TROPOMI-EARLINET ALH data set comparison, for different algorithm processor versions, over ocean and land surface respectively.

AER_LH Version	Number of points	Mean Bias±std [km]	Pearson (R)
Over Ocean			
v02.04 / 02.05	74	-0.62±0.92	0.74
v02.06*	74	1.41±1.77	0.68
Over Land			
v02.04 / 02.05	65	-2.2±1.23	0.43
v02.06*	65	-0.47±1.59	0.51

**The reprocessed TROPOMI AER_LH V02.06 data product provided by the Product Algorithm Laboratory (PAL).*

Overall, the above quantitative statistical results suggest that the data quality of S5P ALH was improved by the updated algorithm version. Further examination is required to support these observations, investigating the effect of the presence of clouds on the retrieval as well as assessing whether the new version detects the upper limit of the highest detected layer when compared to the lidar signals. Specific sensitivity tests need to be performed to investigate the biases and standard deviation in retrieved aerosol height for typical cases of cloud contamination - optically thin and thick cirrus and low-altitude broken cumulus clouds - using different flagging masks.

15.3.5 Case studies

Assessment of updated AER_LH algorithm version (v02.06.00): Extreme biomass burning event during Greek summer wildfires in August 2023.

In late August 2023, a series of wildfires occurred in northern Greece, causing extensive damage to thousands of square kilometers in the Evros region and releasing a significant amount of smoke into the troposphere. These emissions led to severe air pollution conditions and significantly reduced visibility in the affected area for several days. The TROPOMI sensor closely monitored these fires and tracked the movement of the smoke as it dispersed towards the Thessaloniki Lidar station. To demonstrate the effectiveness of the new algorithm processing version (v02.06.00) of the TROPOMI ALH product, a typical day with elevated aerosol levels was selected. The corresponding AER_LH dataset has been processed via the processor version 2.6.0 (labeled as TEST product) and provided to the AUTH team for the validation study. The specific incident corresponds to simultaneous observations from the Thessaloniki lidar station during the transport of smoke originating from the Dadia forest fire, close to Alexandroupoli city, in northern Greece. The performance evaluation of TROPOMI ALH was performed at various spatial distances around the lidar station and on different types of surfaces (sea and land). The spatial distribution of S5P ALH retrievals for versions 02.05.00 and 02.06.00 is illustrated in **Figure 132**, covering a wide area centered on the Thessaloniki site. Three colored circles represent radii of 50, 100 and 150 km around the station.

This case is also emphasizing the issue related to the limitation of satellite measurements over land areas where the effect of the surface reflectance is dominant. The contrast observed between land and sea regarding the retrieval of the ALH product and the surface albedo values is obvious, as can be seen from the color scale in the ALH retrievals. Mean TROPOMI ALH averaged over circles centered at the Thessaloniki site with increasing radius, as a function of that radius (expressed in km). According to the KNMI team responsible for the operational use of the ALH product, the surface albedo was added in the Optimal Estimation Fit state vector for improved ALH over land Updated quality assurance (QA) in the new version.

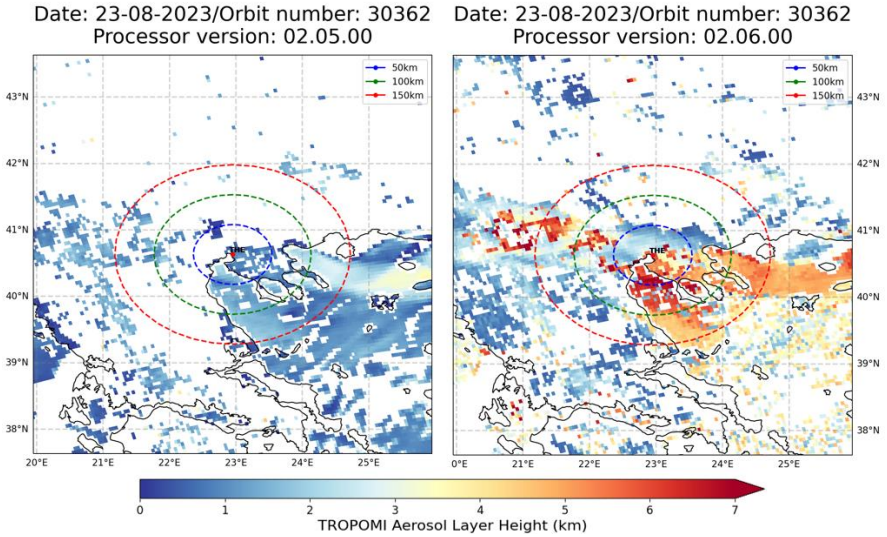


Figure 132: TROPOMI ALH spatial distribution on 23 August 2023, based on (left) V02.05.00 and (right) V02.06.00 algorithm processing version. The Thessaloniki EARLINET station is depicted with a red star. The color-coded concentric circles denote radii distances of 50, 100 and 150km from Thessaloniki lidar station.

Figure 133 displays frequency histograms for the two versions of the processing algorithms (v02.05 in green, v02.06 in magenta) individually over sea and land surface. It is evident that the adoption of the new version has notably augmented the number of recovered pixels, both on land and at sea. Notably, the distribution of pixels retrieved with the new version over sea exhibits a marked increase in height levels. We employ our validation procedure using satellite retrievals distinctly over land and water surfaces to further illustrate the recognized TROPOMI ALH issues over land.

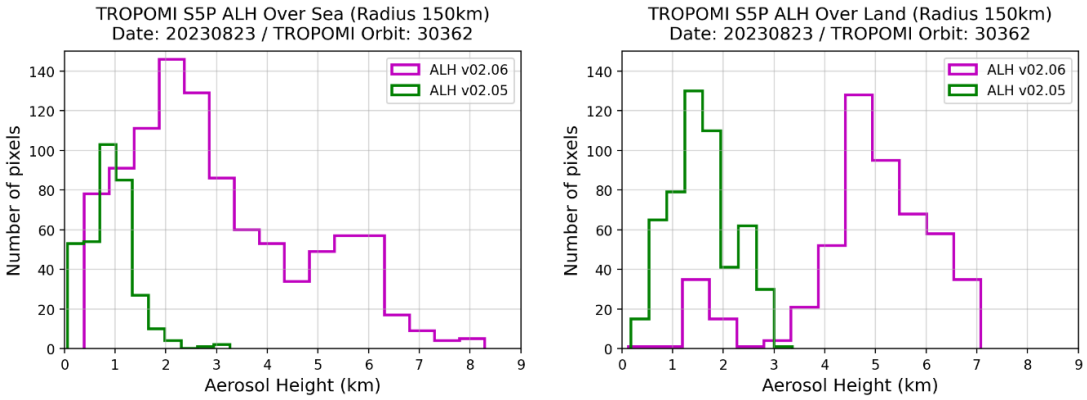


Figure 133: Frequency histograms of TROPOMI ALH retrievals for the version v02.05.00 (green) and v02.06.00 (magenta).

Table 22 shows the results of the statistical analysis for different spatial radius limits around the station, as well as for a different type of surface compared to the previous version. Regarding the v02.06.00, over sea the average value of TROPOMI ranges between 4.8 – 6 km, while over land surface the average value is close to 3 km. To have as representative a comparison between TROPOMI-Lidar observations as possible, a radius distance of 50 km around the Thessaloniki station was used. Recall that so far in earlier validation reports, a radius of 150km has been used due to the reduced number of converged pixels over land.

Table 22: Statistical metrics for the TROPOMI ALH averaged over circles with increasing radius around the Thessaloniki site.

TROPOMI/S5P ALH Over Sea surface					
Radius ≤ 50km		**Radius ≤ 100km**		**Radius ≤ 150km**	
V.02.06.00	**V.02.05.00**	**V.02.06.00**	**V.02.05.00**	**V.02.06.00**	**V.02.05.00**
6.04±1.46 (31)	1.52±0.42 (31)	5.45±1.45 (183)	1.61±0.47 (185)	4.81±1.35 (514)	1.58±0.62 (533)
TROPOMI/S5P ALH Over Land surface					
3.09±1.62 (307)	0.81±0.47 (100)	3.08±1.73 (671)	0.85±0.49 (229)	3.03±1.73 (986)	0.9±0.47 (339)

Figure 135 illustrates the mean aerosol backscatter profiles at 1064 nm derived from THELISYS for the time between 11:00-12:00 UTC. The analysis reveals a clear single layer between 4 and 6km with maximum backscatter value ~6.8 $Mm^{-1} \cdot sr^{-1}$ and centre of layer mass at 4.9km. For the selected case, a notable discrepancy between lidar and TROPOMI layer identification was found, the reasons for which warrant further investigation. Lidar measurements observed two stratified plumes, one around 0.5-2.3 km and one at 4-6 km. The mean layer heights are on average higher in altitude for TROPOMI over ocean and lower over land surface compared against lidar measurements, respectively. The statistical metrics are given in **Table 22** for ocean and land satellite retrievals separately in a distance radius of 150km around the Thessaloniki lidar station. An elevated single well-developed thick layer was found, and smoke particles appeared in the free troposphere between 4 and 6km above sea level (a.s.l). The averaged backscatter profile at 1064 nm is shown for the time close to S5P overpass. Two distinct layers were detected, the first one layer between 0.8-2.32km and the second at 4.32-5.5km. As with the previous examples, here we also find a discrepancy between both lidar aerosol heights and the satellite estimation of the aerosol height. Clearly the AER_LH is placed much lower than both calculated altitudes of the lidar profile. Applying the v02.06.00 the TROPOMI observations report an average layer at 3.09±1.62 km over land and 6.04±1.46 km over sea, while for the version 02.05.00 the layer is marked at 0.81±0.47 km and 1.52±0.42 km respectively. The calculated ALHbsc from the lidar profile places it at 4.2km. The center of mass of each of the two detected layer is illustrated with orange color.

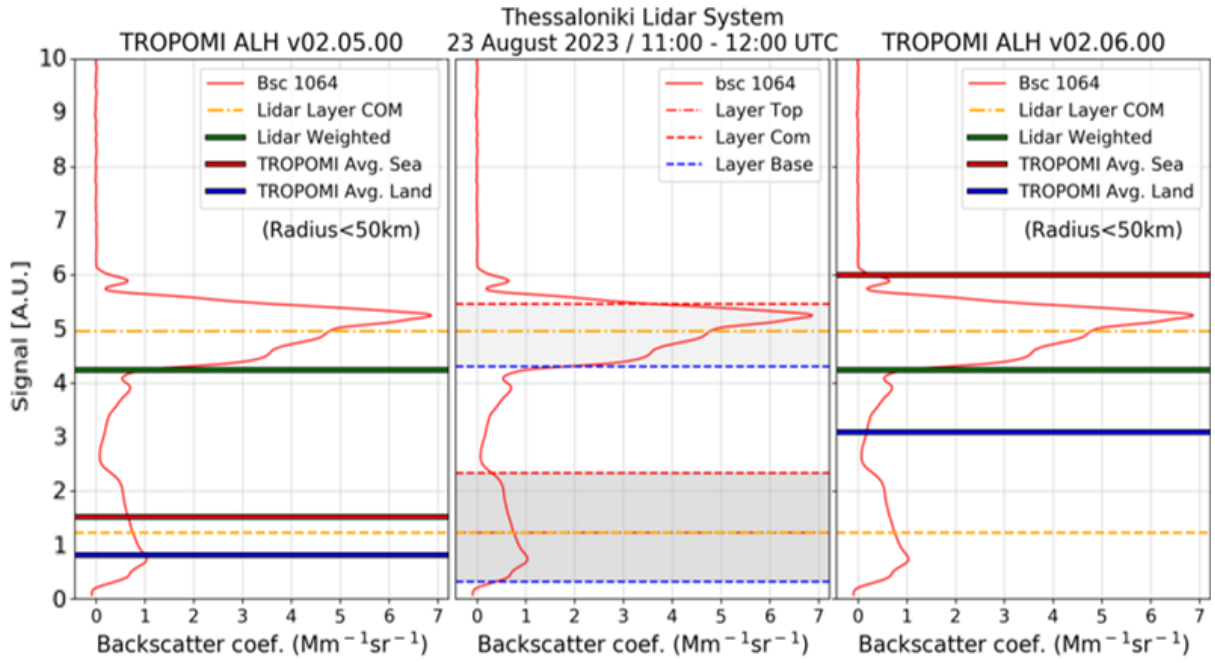


Figure 134. Averaged Backscatter lidar profile obtained at 1064nm close to Sentinel-5P overpass (11:00-12:00). The corresponding ALH values are marked with different colors (within a radius of 50 km from the lidar station).

15.3.6 Bias

The systematic difference between S5P TROPOMI ALH and MISR aerosol plume height is about 600 m. This is mostly due to differences in the sensitivity of the instruments and the differences in the algorithms. A difference of about 500 m (lower for TROPOMI) is expected from simulations. TROPOMI ALH is sensitive to the centroid aerosol layer height. Furthermore, TROPOMI ALH is more accurate for thicker plumes, between CALIOP and TROPOMI layer height decrease to only 50 m. The TROPOMI ALH is well within the requirements of 100 hPa for the bias.

From the comparison with ground-based observations, a slightly larger difference is found. By defining a backscatter-weighted aerosol height from EARLINET aerosol backscatter profile products (ALHbsc), the quantitative validation at pixels over the selected EARLINET stations illustrates that TROPOMI ALH is consistent with ALHbsc, with a high correlation R=0.66 (R=0.56) and mean bias about 1.46±1.69 km (-0.49±1.33 km) over ocean and land pixels respectively.

As a final point, it appears that aerosol layer altitudes retrieved from TROPOMI are systematically lower than altitudes from the lidar retrievals. There is a bias, which is related to the use of the LER model in the retrieval. From the start of the operational data record (30 April 2018), a steadily negative bias of the S5P TROPOMI ALH mainly outside the requirements (bias > 1km). This finding needs further investigation to possibly improve this spread by using a more realistic surface reflectance.

15.3.7 Dispersion

The S5P TROPOMI ALH dispersion is large due to cloud contamination and surface effects. With rigorous cloud screening, 50 % of the pixels are already within 1 km of the CALIOP weighted extinction height. Accounting for the expected bias, this is within the requirements of 50 hPa. However, this preliminary conclusion needs further investigation and confirmation.

15.3.8 Dependence on influence quantities

The TROPOMI ALH is strongly dependent on subpixel clouds, and cloud filtering remains essential. The user is strongly encouraged to use all available cloud filters. Bright surfaces have a strong effect on the ALH. To address this, in November 2023 a new version of the ALH algorithm was introduced (2.6.0) that included surface fitting in the optimal estimation routine. This resulted in a much larger number of successful retrievals especially over land (see **Figure 135**). Other features in the figure are the natural seasonal variability and a few episodes of intense wildfires, especially around the beginning of the year 2020. This was caused by the extreme wildfires in New South Wales, producing large plumes of smoke at very high altitudes, which were visible for several weeks (see section 3.3.9).

Many factors can play a crucial role in the comparisons between TROPOMI and ground-based lidar observations that can neither be attributed solely to the S5P data, nor to pure area-averaging differences. The possibility of remaining clouds in the TROPOMI retrievals is one of the reasons why an optimal spatial collocation with the lidar measurement is not achieved for every target pixel. In addition, the strong underestimation of the aerosol layer height retrieved by the current algorithm from TROPOMI over land is probably related to the surface reflectivity climatology used in the forward model, leading to biased or non-convergent retrievals over land. Over bright surfaces, a systematic and strong negative bias (up to 80%) is observed compared to ground-based LIDAR observations. Bright surfaces have a strong effect on the ALH, and very high ALH (altitude up to 12 km) often occur over the Saharan desert. These should be filtered, but a filtering scheme is currently not available. Sun-glint produces high UV-AI values and are processed for ALH. These ALH values show up in an overview plot, but are easily filtered using the sun-glint filters. In addition, the ALH for aerosol-free sun-glint areas are close to zero (altitude) as expected. Another factor that can affect the satellite-ground based intercomparison of measurements or products is the topography. The complex topography, in terms of geographical characteristics, and the horizontal distance between the TROPOMI pixels and the ground-based lidar sites are, however, features that should also be examined when inter-comparing EARLINET and TROPOMI aerosol layer heights (Michailidis et al., 2023).

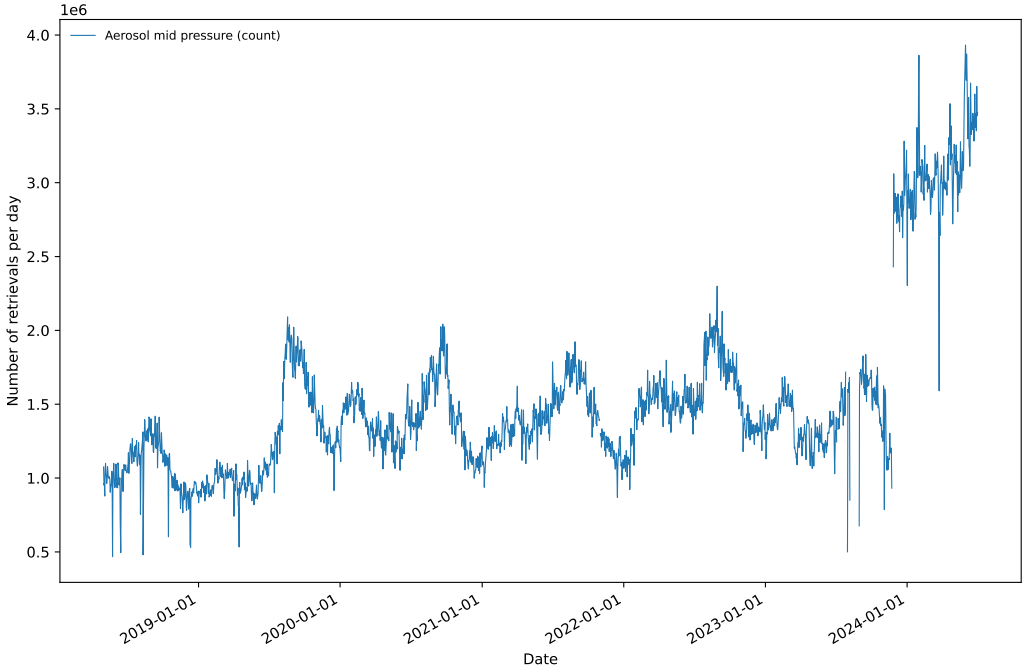


Figure 135: Number of successfully processed ALH pixels per day from June 2018 to August 2024. After the change to version 2.6 in November 2023, the number of successful retrievals increased because more retrievals succeed over land.

The comparison to ground based observations indicate that the S5P L2_AER_LH reports unphysically low values over land due to strong illumination. The high surface albedo above land surfaces is a challenge for ALH retrievals.

15.3.9 Short term variability

The high correlation coefficients between TROPOMI and ground-based observations demonstrate the ability of the TROPOMI observations to properly capture the temporal variability of tropospheric aerosol plumes. The TROPMI ALH data product is strongly event-driven and more detailed assessments on the short-term variability cannot be provided at the moment.

15.3.10 Geographical patterns

There are no obvious geographical features.

15.3.11 Other features

A limitation of the S5P ALH product has become apparent following the severe bushfires in New South Wales during the 2019-2020 fire season. Hundreds of severe wildfires have consumed an estimated 18.6 million hectares in the southeast of Australia. The smoke and gases from these fires were well visible in several S5P products, including UV-AI, ALH and HCHO and CO total column. In **Figure 136** a screen shot shows the S5P ALH on 11 January 2020 over the south Pacific as displayed on the S5P-TROPOMI-KNMI-Level 2 Product Maps webpage. It shows the extent of the fire ash plume from the fires, as well as the altitude as derived by the AER_LH product algorithm.

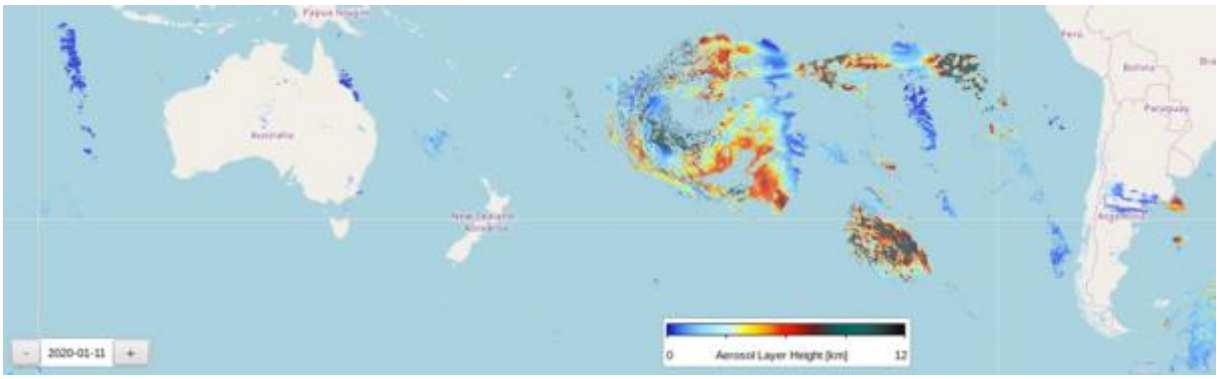


Figure 136: TROPOMI AER_LH product (processor version before v02.05.00) on 11 January 2020 over the south Pacific, showing the altitude as derived by the S5P AER_LH algorithm of the fires smoke from Australian bush fires.

The smoke provides an opportunity to compare the AER_LH with CALIOP measurements, since the extent of the smoke plume is so large that the CALIPSO satellite track intersects with the plume almost daily. An inspection of CALIOP quick-looks revealed much higher altitudes of the smoke derived by CALIOP than by TROPOMI.

A comparison of the CALIOP backscatter data and AER_LH data as before is presented below for 11 January 2020. In **Figure 137** the AER_LH product for 11 January 2020 is plotted again over a VIIRS RGB picture, showing the smoke over clouds and in clear sky (the ALH is retrieved only in clear sky pixels). The maximum altitude in the AER_LH data is about 13 km. However, the CALIOP data show much higher altitudes for the plume. In **Figure 138** the CALIOP total attenuated backscatter at 532 nm is shown for the yellow track shown in **Figure 137**. The plume can be seen around about 44^{o}S and 110^{o}E at an altitude between about 17 and 21 km, which is much higher than the S5P AER_LH. The AER_LH retrievals from TROPOMI are shown in the curtain plot as black and white dots as before. Clearly, the AER_LH is much lower than the altitude of the smoke plume.

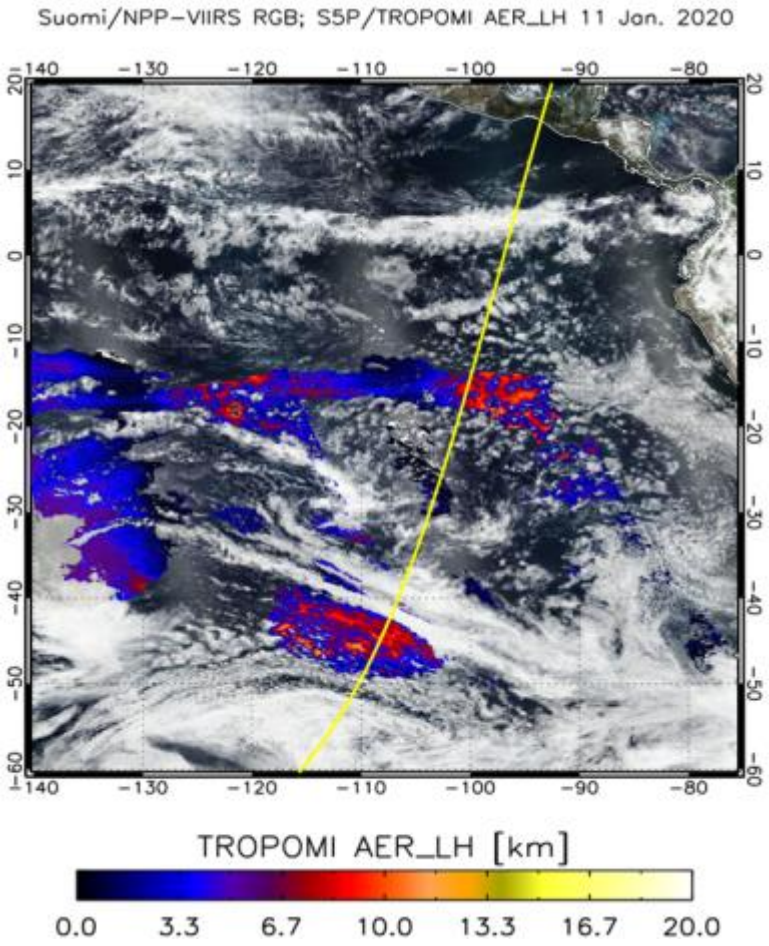


Figure 137: NPP/VIIRS RGB image with S5P/TROPOMI AER_LH on 11 January over the south Pacific with the CALIPSO track of that day overlaid in yellow.

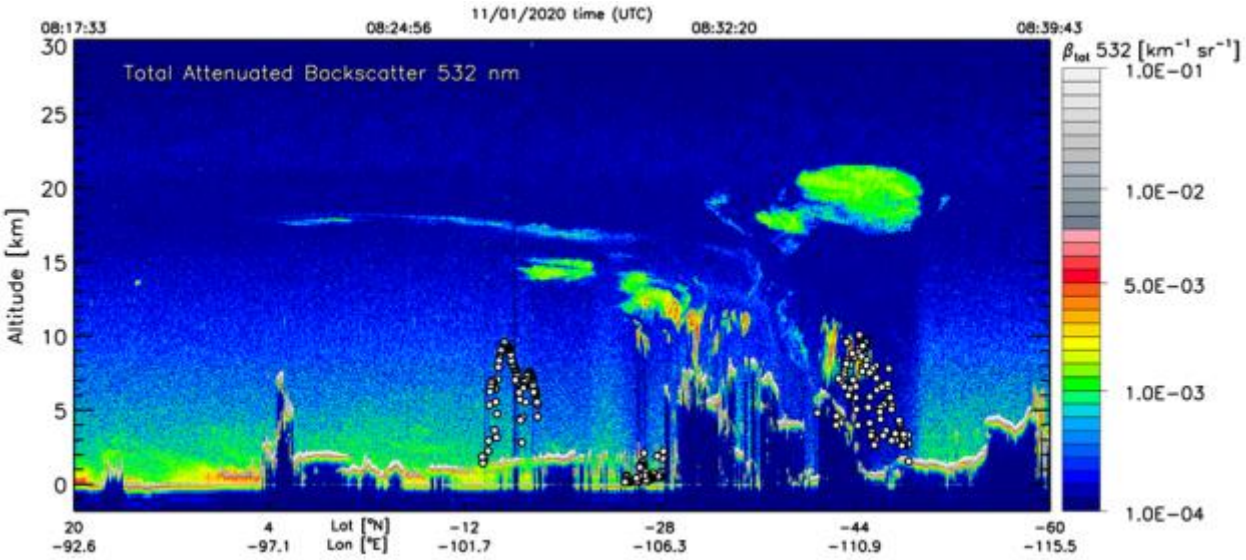


Figure 138: CALIOP aerosol backscatter rat 532 nm along the track shown in **Figure 137** (eastern of Australia on 11 January 11, 2020). The yellow dots represent the TROPOMI ALH (processor version before v02.05.00).

The exact reason for the much lower altitude retrieved by the AER_LH algorithm is not clear, but it is obvious that altitudes above 20 km were not anticipated. The pressures at these altitudes are about 93 hPa (17 km) to 50 hPa (21 km) (Anderson *et al.*, 1982). The AER_LH neural network (NN) was trained to perform within pressures of 1000-75 hPa, so the sensitivity of the algorithm to aerosols at this altitude is low at best. In the weeks after 11 January 2020 the plume kept clearly visible in CALIOP data and rose to even higher altitudes, up to even 30 km. At that altitude, air pressures can be expected to be as low as 10 hPa. The AER_LH algorithm was not created to retrieve ALH at such low air pressures. A new NN may be trained to incorporate these extreme low air pressures. The need for such an extension will have to be investigated, as the occurrence of high altitude smoke like the case presented here seems rather rare. Furthermore, simulations will have to be performed first to test whether the AER_LH algorithm is at all sensitive to aerosols at such a high altitude, before this is to be included in the NN and operational algorithm.

Another issue that can play a role here is cloud contamination. As can be seen from **Figure 137** and **Figure 138**, the area is very cloudy and the algorithm is known to be very sensitive to (residual) cloud contamination, and this will also bias the ALH low.

16 References

The validation activities and requirements applying to the operational phase of the S5P mission are described in the *S5P Cal/Val Plan for the Operational Phase* [S5P-CSCOP], the *S5P Geophysical Validation Requirements Document* [S5PVT-Req], the *Copernicus Sentinels 4 and 5 Mission Requirements Traceability Document* [S4/5-MRTD], and the recommendations formulated by ESL-L2 developers in their *Algorithm Theoretical Basis Documents* available on the ESA Copernicus Sentinel Online website [ER_CoperATBD].

16.1 Reference documents

[S5PVT-Req] Requirements for the Geophysical Validation of Sentinel-5 Precursor Products
source: ESA; **ref:** S5P-RS-ESA-SY-164; **issue**; **date:** 2014-05-21

[S5P-CSCOP] ESA-EOPG-CSCOP-PL-0073, Sentinel-5 Precursor Calibration and Validation Plan for the Operational Phase
source: ESA; **ref:** ESA-EOPG-CSCOP-PL; **issue:** 1; **revision:** 1; **date:** 2017-11-06

[S4/5-MRTD] Copernicus Sentinels 4 and 5 Mission Requirements Traceability Document
source: ESA; **ref:** EOP-SM/2413/BV-bv; **issue:** 2; **revision:** 0; **date:** 2017-07-07

[CEOS-Nom] CEOS/ISO:19159 - Committee on Earth Observation Satellites (CEOS): general calibration and validation resources publicly available on http://calvalportal.ceos.org / ISO TS 19159-1:2014(en), Geographic information - Calibration and validation of remote sensing imagery sensors and data — Part 1: Optical sensors

[JCGM-GUM] GUM: Joint Committee for Guides in Metrology (JCGM/WG 1) 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in a measurement (GUM)

[JCGM-VIM] VIM/ISO:99 Joint Committee for Guides in Metrology (JCGM/WG 2) 200:2012 & ISO/IEC Guide 99-12:2007, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)

[S5P-NomL1] Terms, definitions and abbreviations for TROPOMI L01b data processor;
source: KNMI; **ref:** S5P-KNMI-L01B-0004-LI; **issue:** 3.0.0; **date:** 2013-11-08

[S5P-NomA] Terms and symbols in the TROPOMI Algorithm Team;
source: KNMI; **ref:** SN-TROPOMI-KNMI-049; **issue:** 0.1.2; **date:** 2013-03-11

16.2 Peer-reviewed articles

Basher, R. E., Review of the Dobson spectrophotometer and its accuracy, Rep. 13, WMO Global Ozone Res. and Monit. Proj., Geneva, Dec. [Available at http://www.esrl.noaa.gov/gmd/ozwv/dobson/papers/report13/report13.html], 1982.

Bauwens, M., S. Compernolle, et al.: Impact of coronavirus outbreak on NO_2 pollution assessed using TROPOMI and OMI observations, Geophys. Res. Letters, Vol. 47, Iss. 11, https://doi.org/10.1029/2020GL087978, 8 May 2020.

Bognar, K., Zhao, X., Strong, K., Boone, C., Bourassa, A., Degenstein, D., Drummond, J., Duff, A., Goutail, F., Griffin, D., Jeffery, P., Lutsch, E., Manney, G., McElroy, C., McLinden, C., Millán, L., Pazmino, A., Sioris, C., Walker, K., and Zou, J.: Updated validation of ACE and OSIRIS ozone and NO2 measurements in the Arctic using ground-based instruments at Eureka, Canada, Journal of Quantitative Spectroscopy and Radiative Transfer, 238, 106 571, https://doi.org/https://doi.org/10.1016/j.jqsrt.2019.07.014, 2019.

Buchwitz, M., Reuter, M., Schneising, O., Hewson, W., Detmers, R. G., Boesch, H., Hasekamp, O., Aben, I., Bovensmann, H., Burrows, J., Butz, A., Chevallier, F., Dils, B., Frankenberg, C., Heymann, J., Lichtenberg, G., De Mazière, M., Notholt, J., Parker, R., Warneke, T., Zehner, C., Griffith, D. W. T., Deutscher, N., Kuze, A., Suto, H., and Wunch, D.: Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO2 and XCH4 CRDP3 data set, Remote Sens. Environ., 203, 276–295, https://doi.org/10.1016/j.rse.2016.12.027, 2017.

Buchwitz, M., Aben, I., Armante, R., Boesch, H., Crevoisier, C., Di Noia, A., Hasekamp, O. P., Reuter, M., Schneising-Weigel, O., and Wu, L.: Algorithm Theoretical Basis Document (ATBD) – Main document for Greenhouse Gas (GHG: CO2 and CH4) data set CDR 3 (2003–2018), C3S project, 2019.

Compernolle, S., Argyrouli, A., Lutz, R., Sneep, M., Lambert, J.-C., Fjæraa, A. M., Hubert, D., Keppens, A., Loyola, D., O'Connor, E., Romahn, F., Stammes, P., Verhoelst, T., and Wang, P.: Validation of the Sentinel-5 Precursor TROPOMI cloud data with Cloudnet, Aura OMI O2–O2, MODIS, and Suomi-NPP VIIRS, Atmos. Meas. Tech., 14, 2451–2476, https://doi.org/10.5194/amt-14-2451-2021, 2021.

Compernolle, S., Verhoelst, T., Pinardi, G., Granville, J., Hubert, D., Keppens, A., Niemeijer, S., Rino, B., Bais, A., Beirle, S., Boersma, F., Burrows, J. P., De Smedt, I., Eskes, H., Goutail, F., Hendrick, F., Lorente, A., Pazmino, A., Piters, A., Peters, E., Pommereau, J.-P., Remmers, J., Richter, A., van Geffen, J., Van Roozendael, M., Wagner, T., and Lambert, J.-C.: Validation of Aura-OMI QA4ECV NO2 climate data records with ground-based DOAS networks: the role of measurement and comparison uncertainties, Atmos. Chem. Phys., 20, 8017–8045, https://doi.org/10.5194/acp-20-8017-2020, 2020.

D'Amico, G., Amodeo, A., Mattis, I., Freudenthaler, V., and Pappalardo, G.: EARLINET Single Calculus Chain – technical – Part 1: Pre-processing of raw lidar data, Atmos. Meas. Tech., 9, 491–507, https://doi.org/10.5194/amt-9-491-2016, 2016.

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., Hedelt, P., Hendrick, F., Irie, H., Kumar, V., Lambert, J.-C., Langerock, B., Lerot, C., Liu, C., Loyola, D., Piters, A., Richter, A., Rivera Cárdenas, C. I., Romahn, F., Ryan, R. G., Sinha, V., Theys, N., Vlietinck, J., Wagner, T., Wang, T., Yu, H., and Van Roozendael, M.: Comparative assessment of TROPOMI and OMI formaldehyde observations against MAX-DOAS network column measurements, Atmos. Chem. Phys., 21, 12561–12593, https://doi.org/10.5194/acp-2021-378, 2021.

Errera, Q. and Fonteyn, D.: Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements, J. Geophys. Res., 106, 12253–12265, doi:10.1029/2001JD900010, 2001.

Eskes, H., van Geffen, J., Sneep, M., Veefkind, P., Niemeijer, S. and Zehner, C.: S5P Nitrogen Dioxide v02.03.01 intermediate reprocessing on the S5P-PAL system: Readme file, 1.0, 2021.

Garane, K., Lerot, C., Coldewey-Egbers, M., Verhoelst, T., Koukouli, M. E., Zyrichidou, I., Balis, D. S., Danckaert, T., Goutail, F., Granville, J., Hubert, D., Keppens, A., Lambert, J.-C., Loyola, D., Pommereau, J.-P., Van Roozendael, M., and Zehner, C.: Quality assessment of the Ozone_cci Climate Research Data Package (release 2017) – Part 1: Ground-based validation of total ozone column data products, Atmos. Meas. Tech., 11, 1385–1402, https://doi.org/10.5194/amt-11-1385-2018, 2018.

Garane, K., Koukouli, M.-E., Verhoelst, T., Fioletov, V., Lerot, C., Heue, K.-P., Bais, A., Balis, D., Bazureau, A., Dehn, A., Goutail, F., Granville, J., Griffin, D., Hubert, D., Keppens, A., Lambert, J.-C., Loyola, D., McLinden, C., Pazmino, A., Pommereau, J.-P., Redondas, A., Romahn, F., Valks, P., Van Roozendael, M., Xu, J., Zehner, C., Zerefos, C., and Zimmer, W.: TROPOMI/S5ptotal ozone column data: global ground-based validation and consistency with other satellite missions, Atmos. Meas. Tech., https://doi.org/10.5194/amt-12-5263-2019, 2019.

Hendrick, F., Barret, B., Van Roozendael, M., Boesch, H., Butz, A., De Mazière, M., Goutail, F., Hermans, C., Lambert, J.-C., Pfeilsticker, K., and Pommereau, J.-P.: Retrieval of nitrogen dioxide stratospheric profiles from ground-based zenith-sky UV-visible observations: validation of the technique through correlative comparisons, Atmos. Chem. Phys., 4, 2091–2106, doi:10.5194/acp-4-2091-2004, 2004.

Hendrick, F., Pommereau, J.-P., Goutail, F., Evans, R. D., Ionov, D., Pazmino, A., Kyrö, E., Held, G., Eriksen, P., Dorokhov, V., Gil, M., and Van Roozendael, M.: NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations, Atmos. Chem. Phys., 11, 5975–5995, https://doi.org/10.5194/acp-11-5975-2011, 2011.

Herman, J., A. Cede, E. Spinei, G. Mount, M. Tzortziou, and N. Abuhassan, NO_2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res., Vol. 114, D13307, doi:10.1029/2009JD011848, 2009.

Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J.-L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell III, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., van Gijsel, J. A. E., Van Malderen, R., von der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, Atmos. Meas. Tech., 9, 2497-2534, https://doi.org/10.5194/amt-9-2497-2016, 2016.

Hubert, Daan, Klaus-Peter Heue, Jean-Christopher Lambert, Tijl Verhoelst, Marc Allaart, Steven Compernolle, Patrick D. Cullis, Angelika Dehn, Christian Félix, Bryan J. Johnson, Arno Keppens, Debra E. Kollonige, Christophe Lerot, Diego Loyola, Maznorizan Mohamad, Maria Paulete Pereira Martins, Ankie Piters, Henry Selkirk, Anne M. Thompson, Pepijn Veefkind, Holger Vömel, Jacquelyn C. Witte, and Claus Zehner, TROPOMI tropospheric ozone column data : Geophysical assessment and comparison to ozonesondes, GOME-2B and OMI, Atmos. Meas. Tech., https://doi.org/10.5194/amt-14-7405-2021, 2021.

Inness, A., Flemming, J., Heue, K.-P., Lerot, C., Loyola, D., Ribas, R., Valks, P., van Roozendael, M., Xu, J., and Zimmer, W.: Monitoring and assimilation tests with TROPOMI data in the CAMS system: near-real-time total column ozone, Atmos. Chem. Phys., 19, 3939–3962, https://doi.org/10.5194/acp-19-3939-2019, 2019.

Keckhut, P.; McDermid, S.; Swart, D.; McGee, T.; Godin-Beekmann, S.; Adriani, A.; Barnes, J.; Baray, J.-L.; Bencherif, H.; Claude, H.; di Sarra, A.; Fiocco, G.; Hansen, G.; Hauchecorne, A.; Leblanc, T.; Lee, C.; Pal, S.; Megie, G.; Nakane, H.; Neuber, R.; Steinbrecht, W. & Thayer, J. Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change *J. Environ. Monit.,* **2004***, 6*, 721-733.

Keppens, A., J.-C. Lambert, J. Granville, D. Hubert, T. Verhoelst, S. Compernolle, B. Latter, B. Kerridge, R. Siddans, A. Boynard, J. Hadji-Lazaro, C. Clerbaux, C. Wespes, D.R. Hurtmans, P.-F. Coheur, J. C. A. van Peet, R. J. van der A, K. Garane, M. E. Koukouli, D. S. Balis, A. Delcloo, R. Kivi, R. Stübi, S. Godin-Beekmann, M. Van Roozendael, and C. Zehner: Quality assessment of the Ozone_cci Climate Research Data Package (release 2017): 2. Ground-based validation of nadir ozone profile data products, Atmos. Meas. Tech., 11, 3769-3800, https://doi.org/10.5194/amt-11-3769-2018, 2018.

Keppens, A., Compernolle, S., Verhoelst, T., Hubert, D., and Lambert, J.-C.: Harmonization and comparison of vertically resolved atmospheric state observations: Methods, effects, and uncertainty budget, *Atmos. Meas. Tech.*, 12, 4379–4391, https://doi.org/10.5194/amt-12-4379-2019, 2019.

Keppens, A., Compernolle, S., Hubert, D., Verhoelst, T., Granville, J., and Lambert, Removing Prior Information from Remotely Sensed Atmospheric Profiles by Wiener Deconvolution Based on the Complete Data Fusion Framework, *Remote Sensing*, 14(9), 2197; https://doi.org/10.3390/rs14092197 - 04 May 2022.

Keppens, A., Di Pede, S., Hubert, D., Lambert, J.-C., Veefkind, P., Sneep, M., De Haan, J., ter Linden, M., Leblanc, T., Compernolle, S., Verhoelst, T., Granville, J., Nath, O., Fjæraa, A. M., Boyd, I., Niemeijer, S., Van Malderen, R., Smit, H. G. J., Duflot, V., Godin-Beekmann, S., Johnson, B. J., Steinbrecht, W., Tarasick, D. W., Kollonige, D. E., Stauffer, R. M., Thompson, A. M., Dehn, A., and Zehner, C.: 5 years of Sentinel-5P TROPOMI operational ozone profiling and geophysical validation using ozonesonde and lidar ground-based networks, Atmos. Meas. Tech., 17, 3969–3993, https://doi.org/10.5194/amt-17-3969-2024, 2024.

Kerr, J. B., C. T. McElroy, and R. A. Olafson, Measurements of ozone with the Brewer ozone spectrophotometer, in Proceedings of the Quadrennial Ozone Symposium, Boulder, Colorado, edited by J. London, pp. 74–79, Natl. Cent. for Atmos. Res., Boulder, Colorado, 1981.

Kerr, J. B., I. A. Asbridge, and W. F. J. Evans, Intercomparison of total ozone measured by the Brewer and Dobson spectrophotometers at Toronto, J. Geophys. Res., 93, 11,129–11,140, doi:10.1029/JD093iD09p11129, 1988.

Lambert, J.-C., M. Van Roozendael, J. Granville, P. Gerard, P. Peeters, P.C. Simon, H. Claude and J. Staehelin, Comparison of the GOME ozone and NO_2 total amounts at mid-latitude with ground-based zenith-sky measurements, in Atmospheric Ozone - 18th Quad. Ozone Symp., L'Aquila, Italy, 1996, R. Bojkov and G. Visconti (Eds.), Vol. I, pp. 301-304, 1997.

Lambert, J.-C., M. Van Roozendael, M. De Mazière, P.C. Simon, J.-P. Pommereau, F. Goutail, A. Sarkissian, and J.F. Gleason, Investigation of pole-to-pole performances of spaceborne atmospheric chemistry sensors with the NDSC, Journal of the Atmospheric Sciences, Vol. 56, pp. 176-193, https://doi.org/10.1175/1520-0469(1999)056<0176:IOPTPP>2.0.CO;2 , 1999.

Leblanc, T., R. J. Sica, J. A. E. van Gijsel, S. Godin-Beekmann, A. Haefele, T. Trickl, G. Payen, and G. Liberti (2016), Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 2: Ozone DIAL uncertainty budget, Atmos. Meas. Tech., 9(8), 4051-4078, 2016.

Leblanc, T.; Brewer, M. A.; Wang, P. S.; Granados-Mu noz, M. J.; Strawbridge, K. B.; Travis, M.; Firanski, B.; Sullivan, J. T.; McGee, T. J.; Sumnicht, G. K.; Twigg, L. W.; Berkoff, T. A.; Carrion, W.; Gronoff, G.; Aknan, A.; Chen, G.; Alvarez, R. J.; Langford, A. O.; Senff, C. J.; Kirgis, G.; Johnson, M. S.; Kuang, S. & Newchurch, M. J. Validation of the TOLNet lidars: the Southern California Ozone Observation Project (SCOOP) Atmos. Meas. Tech., 11, 6137-6162, 2018.

Loew, A., W. Bell, L. Brocca, C. E. Bulgin, J. Burdanowitz, X. Calbet, R. V. Donner, D. Ghent, A. Gruber, T. Kaminski,J. Kinzel, C. Klepp, J.-C. Lambert, G. Schaepman-Strub, M. Schröder, and T. Verhoelst, Validation practices for satellite based earth observation data across communities, Rev. Geophys., DOI: 10.1002/2017RG000562, 2017.

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., 13, 3561–3580, 2020, https://doi.org/10.5194/amt-13-3561-2020, 2020.

Michailidis, K., Koukouli, M.-E., Siomos, N., Balis, D., Tuinder, O., Tilstra, L. G., Mona, L., Pappalardo, G., and Bortoli, D.: First validation of GOME-2/MetOp absorbing aerosol height using EARLINET lidar observations, Atmos. Chem. Phys., 21, 3193–3213, https://doi.org/10.5194/acp-21-3193-2021, 2021a.

Michailidis, K., MariLiza Koukouli, & Dimitris Balis. Validation of ESA EO Aerosol Height products with EARLINET Lidar observations (v1.0). Zenodo. https://doi.org/10.5281/zenodo.5563420, 2021b.

Michailidis, K., Koukouli, M.-E., Balis, D., Veefkind, J. P., de Graaf, M., Mona, L., Papagianopoulos, N., Pappalardo, G., Tsikoudi, I., Amiridis, V., Marinou, E., Gialitaki, A., Mamouri, R.-E., Nisantzi, A., Bortoli, D., João Costa, M., Salgueiro, V., Papayannis, A., Mylonaki, M., Alados-Arboledas, L., Romano, S., Perrone, M. R., and Baars, H.: Validation of the TROPOMI/S5P aerosol layer height using EARLINET lidars, Atmos. Chem. Phys., 23, 1919–1940, https://doi.org/10.5194/acp-23-1919-2023, 2023.

Mona, L., Amodeo, A., Pandolfi, M., and Pappalardo, G.: Saharan dust intrusions in the Mediterranean area: three years of Raman lidar measurements, J. Geophys. Res.-Atmos., 111, d16203, doi:10.1029/2005JD006569, 2006.

Nanda, S., Veefkind, J. P., de Graaf, M., Sneep, M., Stammes, P., de Haan, J. F., Sanders, A. F. J., Apituley, A., Tuinder, O., and Levelt, P. F.: A weighted least squares approach to retrieve aerosol layer height over bright surfaces applied to GOME-2 measurements of the oxygen A band for forest fire cases over Europe, Atmos. Meas. Tech., 11, 3263-3280, https://doi.org/10.5194/amt-11-3263-2018, 2018.

Nanda, S., de Graaf, M., Veefkind, J. P., Sneep, M., ter Linden, M., Sun, J., and Levelt, P. F.: A first comparison of TROPOMI aerosol layer height (ALH) to CALIOP data, Atmos. Meas. Tech., 13, 3043–3059, https://doi.org/10.5194/amt-13-3043-2020, 2020.

Pinardi, G., M. Van Roozendael, F. Hendrick, N. Theys, N. Abuhassan, A. Bais, A. Blechschmidt, F. Boersma, A. Cede, J. Chong, U. Friess, J. Granville, J. R. Herman, R. Holla, J. Hovila, H. Irie, Y. Kanaya, N. Kouremeti, J.-C. Lambert, J. Ma, E. Peters, A. Piters, O. Postylyakov, A. Richter, J. Remmers, M. Tiefengraber, P. Valks, T. Vlemmix, T. Wagner, and F. Wittrock: Validation of tropospheric NO2 columns measurements from GOME-2 and OMI using MAX-DOAS and direct sun network observations, Atmos. Meas. Tech., 13, 6141–6174, https://doi.org/10.5194/amt-13-6141-2020, 2020.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico,G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: to-wards an advanced sustainable European aerosol lidar network,Atmos. Meas. Tech., 7, 2389–2409, https://doi.org/10.5194/amt-7-2389-2014, 2014.

Pommereau, J. and Goutail, F.: O3 and NO2 ground-based measurements by visible spectrometry during Arctic winter and spring, Geophys. Res. Lett., 15, 891–894, https://doi.org/10.1029/GL015i008p00891, 1988.

Richter, A., M. Weber, J. P. Burrows, J.-C. Lambert, and A. van Gijsel, Validation Strategy for Satellite Observations of Tropospheric Reactive Gases, Annals of Geophysics, Vol. 56, 10.4401/AG-6335, 2013.

Rodgers, C. D. Inverse Methods for Atmospheric Sounding. *World Scientific,* 2000.

Rodgers, C. D. & Connor, B. J. Intercomparison of remote sounding instruments. *J. Geophys. Res.,* 2003*, 108*, 4116.

Roscoe, H., Johnston, P., Van Roozendael, M., Richter, A., Sarkissian, A., Roscoe, J., Preston, K., Lambert, J.-C., Hermans, C., De Cuyper, W., Dzienus, S., Winterrath, T., Burrows, J., Goutail, F., Pommereau, J.-P. D'Almeida, E., Hottier, J., Coureul, C., Ramon, D., Pundt, I., Bartlett, L., McElroy, C., Kerr, J., Elokhov, A., Giovanelli, G., Ravegnani, F., Premuda, M., Kostadinov, I., Erle, F.,Wagner, T., Pfeilsticker, K., Kenntner, M., Marquard, L., Gil, M., Puentedura, O., Yela, M., Arlander, W., Kåstad Høiskar, B., Tellefsen, C., Karlsen Tørnkvist, K., Heese, B., Jones, R., Aliwell, S., and Freshwater, R.: Slant column measurements of O3 and NO2 during the NDSC intercomparison of zenith-sky UV-visible spectrometers in June 1996, Journal of Atmospheric Chemistry, 32, 281–314, 1999.

Schneider, M., Hase, F., and Blumenstock, T.: Ground-based remote sensing of HDO/H2O ratio profiles: introduction and validation of an innovative retrieval approach, Atmos. Chem. Phys., 6, 4705-4722, https://doi.org/10.5194/acp-6-4705-2006, 2006.

Schneising, O., Buchwitz, M., Reuter, M., Bovensmann, H., Burrows, J. P., Borsdorff, T., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Hermans, C., Iraci, L. T., Kivi, R., Landgraf, J., Morino, I., Notholt, J., Petri, C., Pollard, D. F., Roche, S., Shiomi, K., Strong, K., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor, Atmos. Meas. Tech., 12, 6771–6802, https://doi.org/10.5194/amt-12-6771-2019, 2019.

Sha, M. K., Langerock, B., Blavier, J.-F. L., Blumenstock, T., Borsdorff, T., Buschmann, M., Dehn, A., De Mazière, M., Deutscher, N. M., Feist, D. G., García, O. E., Griffith, D. W. T., Grutter, M., Hannigan, J. W., Hase, F., Heikkinen, P., Hermans, C., Iraci, L. T., Jeseck, P., Jones, N., Kivi, R., Kumps, N., Landgraf, J., Lorente, A., Mahieu, E., Makarova, M. V., Mellqvist, J., Metzger, J.-M., Morino, I., Nagahama, T., Notholt, J., Ohyama, H., Ortega, I., Palm, M., Petri, C., Pollard, D. F., Rettinger, M., Robinson, J., Roche, S., Roehl, C. M., Röhling, A. N., Rousogenous, C., Schneider, M., Shiomi, K., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., Uchino, O., Velazco, V. A., Vigouroux, C., Vrekoussis, M., Wang, P., Warneke, T., Wizenberg, T., Wunch, D., Yamanouchi, S., Yang, Y., and Zhou, M.: Validation of methane and carbon monoxide from Sentinel-5 Precursor using TCCON and NDACC-IRWG stations, Atmos. Meas. Tech., 14, 6249–6304, https://doi.org/10.5194/amt-14-6249-2021, 2021.

Siomos, N., Balis, D. S., Voudouri, K. A., Giannakaki, E., Filioglou, M., Amiridis, V., Papayannis, A., and Fragkos, K.: Are EARLINET and AERONET climatologies consistent? The case of Thessaloniki, Greece, Atmos. Chem. Phys., 18, 11885–11903, https://doi.org/10.5194/acp-18-11885-2018, 2018.

Smit, H.; Straeter, W.; Johnson, B.; Oltmans, S.; Davies, J.; Tarasick, D.; Hoegger, B.; Stübi, R.; Schmidlin, F.; Northam, T.; Thompson, A.; Witte, J.-C.; Boyd, I. & Posny, F. Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Juelich Ozone Sonde Intercomparison Experiment (JOSIE) *J. Geophys. Res.,* **2007***, 112*, 1-18.

Tilstra, L. G., de Graaf, M., Wang, P., and Stammes, P.: In-orbit Earth reflectance validation of TROPOMI on board the Sentinel-5 Precursor satellite, Atmos. Meas. Tech., 13, 4479–4497, 2020, https://doi.org/10.5194/amt-13-4479-2020, 2020.

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., and Veefkind, J. P.: Sentinel-5P TROPOMI NO$_2$ retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15, 2037-2060, https://amt.copernicus.org/articles/15/2037/2022/, 2022.

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.

Vandaele, A. C., Fayt, C., Hendrick, F., Hermans, C., Humbled, F., Van Roozendael, M., Gil, M., Navarro, M., Puentedura, O., Yela, M., Braathen, G., Stebel, K., Tørnkvist, K., Johnston, P., Kreher, K., Goutail, F., Mieville, A., Pommereau, J.-P., Khaikine, S., Richter, A., Oetjen, H., Wittrock, F., Bugarski, S., Frieß, U., Pfeilsticker, K., Sinreich, R., Wagner, T., Corlett, G., and Leigh, R.: An intercomparison campaign of ground-based UV-visible measurements of NO2, BrO, and OClO slant columns: Methods of analysis and results for NO2, J. Geophys. Res. Atmos., 110, https://doi.org/10.1029/2004JD005423, 2005.

Verhoelst, T., Granville, J., Hendrick, F., Köhler, U., Lerot, C., Pommereau, J.-P., Redondas, A., Van Roozendael, M., and Lambert, J.-C.: Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons, Atmos. Meas. Tech., 8, 5039-5062, https://doi.org/10.5194/amt-8-5039-2015, 2015.

Verhoelst, T., S. Compernolle, G. Pinardi, J.-C. Lambert, H.J. Eskes, K.-U. Eichmann, A.M. Fjæraa,J. Granville, S. Niemeijer, A. Cede, M. Tiefengraber, F. Hendrick, A. Pazmiño, A. Bais, A. Bazureau, K.F. Boersma, K. Bognar, A. Dehn, S. Donner, A. Elokhov, M. Gebetsberger, F. Goutail, M. Grutter de laMora, A. Gruzdev, M. Gratsea, G.H. Hansen, H. Irie, N. Jepsen, Y. Kanaya, D. Karagkiozidis, R. Kivi,K. Kreher, P.F. Levelt, C. Liu, M. Müller, M. Navarro Comas, A.J.M. Piters, J.-P. Pommereau, T. Portafaix, O. Puentedura, R. Querel, J. Remmers, A. Richter, J. Rimmer, C. Rivera Cárdenas, L. Saavedra de Miguel, V. P. Sinyakov, K. Strong, M. Van Roozendael, J.P. Veefkind, T. Wagner, F. Wittrock,M. Yela González, and C. Zehner, Ground-based validation of the Copernicus Sentinel-5p TROPOMI NO_2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., https://doi.org/10.5194/amt-14-481-2021, 2021.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J., Jones, N., Kivi, R., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI/S5P formaldehyde validation using an extensive network of ground-based FTIR stations, Atmos. Meas. Tech., 13, 3751–3767, https://doi.org/10.5194/amt-2020-30, 2020.

von Clarmann, T. Validation of remotely sensed profiles of atmospheric state variables: strategies and terminology. Atmos. Chem. Phys., 2006, 6, 4311-4320, https://doi.org/ 10.5194/acp-6-4311-2006.

Yela, M., Gil-Ojeda, M., Navarro-Comas, M., Gonzalez-Bartolomé, D., Puentedura, O., Funke, B., Iglesias, J., Rodríguez, S., García, O., Ochoa, H., and Deferrari, G.: Hemispheric asymmetry in stratospheric NO2 trends, Atmospheric Chemistry and Physics, 17, 13 373–389, https://www.atmos-chem-phys.net/17/13373/2017/ , 2017.

16.3 Electronic references

[ER_TROPOMI]	TROPOMI website	http://www.tropomi.eu
[ER_VDAF]	TROPOMI Validation Website / Validation Data Analysis Facility	http://mpc-vdaf.tropomi.eu
[ER_VDAF-AVS]	Validation Data Analysis Facility Automated Validation Server	http://mpc-vdaf-server.tropomi.eu
[ER_CoperATBD]	Copernicus Sentinel-5p products and algorithms documents webpage	https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms
[ER_MPS]	TROPOMI Portal for Instrument and Calibration	http://mps.tropomi.eu
[ER_L2QC]	TROPOMI Portal for Level-2 Data Quality Control	http://mpc-l2.tropomi.eu
[ER_S5PVT]	S5P Validation Team AO projects	https://earth.esa.int/web/guest/pi-community/apply-for-data/ao-s
[ER_CoperEC]	Copernicus Programme website	http://www.copernicus.eu
[ER_CoperESA]	ESA Copernicus website	http://www.esa.int/copernicus
[ER_CAMS]	Copernicus Atmosphere Monitoring Service (CAMS) website	http://atmosphere.copernicus.eu
[ER_C3S]	Copernicus Climate Change Service (C3S) website	http://climate.copernicus.eu
[ER_CEOS-Nom]	CEOS Cal/Val Terms and Definitions	http://calvalportal.ceos.org/documents/10136/551648/IASB-BIRA+Metrology+Terms+and+Definitions
[ER_GUM]	Guide to the expression of uncertainty in a measurement (GUM)	http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
[ER_VIM]	International Vocabulary of Metrology (VIM)	http://www.bipm.org/en/publications/guides/vim.html
[ER_CODA]	CODA Atmospheric Toolbox	https://atmospherictoolbox.org/coda

ESA FRM Projects Websites

[ER_FRM4DOAS]	Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations project website	http://frm4doas.aeronomie.be
[ER_FRM4GHG]	Fiducial Reference Measurements for Ground-Based Infrared Greenhouse Gas Observations project website	http://frm4ghg.aeronomie.be
[ER_Pandonia]	Fiducial Reference Measurements for Ground-Based Direct-Sun Air-Quality Observations project	http://pandonia.net

Monitoring Networks Websites and Data Centres

[ER_ACTRIS]	European Research Infrastructure for the observation of Aerosol, Clouds, and Trace gases website	http://www.actris.eu
[ER_Cloudnet]	Cloudnet remote sensing network website	https://cloudnet.fmi.fi
[ER_COCCON]	Collaborative Carbon Column Observing Network (COCCON) website	https://www.imk-asf.kit.edu/english/COCCON.php
[ER_EARLINET]	European Aerosol Research Lidar Network (EARLINET) website	http://www.earlinet.org
[ER_EUBREWNET]	COST Action for a coherent network of European Brewer Spectrophotometer monitoring stations (EUBREWNET) website	http://www.eubrewnet.org
[ER_EUMETNET]	European Meteorological Services Network (EUMETNET) website	http://eumetnet.eu
[ER_EVDC]	ESA Validation Data Centre (EVDC) website	http://evdc.esa.int
[ER_NDACC]	Network for the Detection of Atmospheric Composition Change (NDACC) website	http://ndacc.org
[ER_NOVAC]	Network for Observation of Volcanic and Atmospheric Change (NOVAC) website	http://novac-community.org/
[ER_SHADOZ]	Southern Hemisphere ADditional OZonesonde programme website	https://tropo.gsfc.nasa.gov/shadoz
[ER_TCCON]	Total Carbon Column Observing Network (TCCON) website	https://tccon-wiki.caltech.edu
[ER_TOLnet]	Tropospheric Ozone Lidar Network (TOLnet) website	http://www-air.larc.nasa.gov/missions/TOLNet
[ER_WOUDC]	World Ozone and Ultraviolet Data Centre (WOUDC) website	http://woudc.org

17 Acknowledgements

This Section acknowledges the authors of this report in charge of the ATM-MPC S5P Routine Operations validation service (**Table 23**), the operators of S5P validation facilities, the providers of Fiducial Reference Measurements and other validation data, and the support provided by the Agencies.

17.1 ATM-MPC S5P Routine Operations Validation Service

Table 23 – Responsibilities for the ATM-MPC S5P routine operations validation service: Product Validation Coordinators responsible for validation and reporting per data product (third column), and Product Validation Contributors participating in the validation and reporting per data product (fourth column).

S5P Product ID	Geophysical Quantity	Product Coordinator for Routine Operations Validation Activities	Product Contributors to Routine Operations Validation Activities
L1B	Radiance and irradiance	A. Ludewig (KNMI)	M. Coldewey-Egbers (DLR)
L2_O3	O_3 total column	T. Verhoelst (BIRA-IASB)	K. Garane (AUTH) K.-P. Heue (DLR) J. van Gent (BIRA-IASB)
L2_O3_PR	O_3 vertical profile	A. Keppens (BIRA-IASB)	P. Veefkind (KNMI) O. Tuinder (KNMI)
L2_O3_TCL	O_3 tropospheric column	D. Hubert (BIRA-IASB)	K.-P. Heue (DLR)
L2_NO2	NO_2 stratospheric column	K.-U. Eichmann (IUPB)	T. Verhoelst (BIRA-IASB)
	NO_2 tropospheric column		S. Compernolle (BIRA-IASB) H. Eskes (KNMI) G. Pinardi (BIRA-IASB) P. Valks (DLR) J. van Geffen (KNMI) T. Verhoelst (BIRA-IASB)
	NO_2 total column		
L2_SO2	SO_2 total column	T. Wagner (MPI-C)	P. Hedelt (DLR) N. Theys (BIRA-IASB)
L2_SO2_LH	SO_2 layer height	T. Wagner (MPI-C)	A. Pseftogkas (AUTH) M.-E. Koukouli (AUTH)
L2_HCHO	HCHO total column	K.-U. Eichmann (IUPB)	S. Compernolle (BIRA-IASB) I. De Smedt (BIRA-IASB) C. Erciyes (DLR) C. Vigouroux (BIRA-IASB)
L2_CO	CO total column	B. Langerock (BIRA-IASB)	M.K. Sha (BIRA-IASB)
L2_CH4	CH_4 total column	M.K. Sha (BIRA-IASB)	B. Langerock (BIRA-IASB) M.C. Martinez Velarte (SRON)
L2_CLOUD	Cloud Fraction	S. Compernolle (BIRA-IASB)	R. Lutz (DLR) P. Wang (KNMI)
	Cloud Height		A. Argyrouli (DLR)
	Cloud Optical Thickness		P. Wang (KNMI)
L2_AER_AI	Aerosol Absorbing Index	T. Wagner (MPI-C)	D.C. Stein Zweers (KNMI)
L2_AER_LH	Aerosol Layer Height		M. de Graaf (KNMI) K. Michailidis (AUTH)

17.2 S5P validation facilities

The ATM-MPC Validation Data Analysis Facility (VDAF) hosted at BIRA-IASB by S. Compernolle, J.-C. Lambert and B. Langerock, runs the S5P TROPOMI Automated Validation Server (VDAF-AVS) [ER_VDAF-AVS] developed and operated jointly by s[&]t and at BIRA-IASB. The VDAF-AVS server is based on the HARP toolset developed and maintained by S. Niemeijer at s[&]t. The VDAF also hosts the dedicated S5P TROPOMI validation website [ER_VDAF].

Part of the validation work for trace gases data relies on the Multi-TASTE versatile validation system, developed and operated at BIRA-IASB by S. Compernolle, J. Granville, D. Hubert, A. Keppens, J.-C. Lambert, and T. Verhoelst. Multi-TASTE has been supported by the Belgian Federal Science Policy Office (BELSPO), with *ad hoc* support provided by the EC, ESA and EUMETSAT for specific satellite validation and metrology applications.

Part of the total ozone and aerosol validation work makes use of the validation facility operated at AUTH, and developed by D. Balis, K. Garane, ML. Koukouli and K. Michailidis with support from ESA and EUMETSAT.

The ESA Atmospheric Validation Data Centre (EVDC) [ER_EVDC], hosted at the Norwegian Institute for Air Research (NILU) under the supervision of A.M. Fjæraa, is acknowledged for facilitating access to the validation data from ground-based monitoring networks and field campaigns.

17.3 Validation data

The ground-based data used in this study was obtained as part of the Brewer and Dobson ozone column monitoring networks ([ER_WOUDC], [ER_EUBREWNET]), the Network for the Detection of Atmospheric Composition Change (NDACC) [ER_NDACC], Southern Hemisphere Additional Ozonesonde programme (SHADOZ) [ER_SHADOZ], and the Total Carbon Column Observation Network (TCCON) [ER_TCCON], all contributors to WMO's Global Atmosphere Watch (GAW). Data archived in the associated data centres and lists of associated data originators are publicly available.

Instrument PIs, the scientific teams and the staff at the stations are thanked warmly for special processing efforts and faster data delivery dedicated to TROPOMI validation:

- Rapid delivery O_3 column data from the LATMOS_RT facility at IPSL/UVSQ, WMO's Ozone Mapping Centre in Thessaloniki and WOUDC in Toronto was gathered in the framework of the S5PVT AO project VALTOZ (ID #28568, PI D. Balis, AUTH, Co-Is ML. Koukouli, E. Zyrichidou, J.-C. Lambert, T. Verhoelst, J. Granville, A. Pazmiño, F. Goutail, J.-P. Pommereau, A. Bazureau, and C. Zerefos).
- Rapid delivery O_3 profile data from the SHADOZ network was organised in the framework of the S5PVT AO project CHEOPS-5p (ID #28587, PIs A. Keppens and J.-C. Lambert, BIRA-IASB, Co-Is D. Balis, D. Hubert, W. Steinbrecht, T. Stavrakou, A. Delcloo, S. Godin-Beekmann, T. Leblanc, R. Stübi, A.M. Thompson, T. Verhoelst, G. Ancellet, and V. Duflot). Rapid delivery ozonesonde profile data were also provided by KNMI (A. Piters, M. Allaart) and NOAA (B.J. Johnson).

- Rapid delivery NO_2 data from NDACC MAX-DOAS and ZSL-DOAS stations was gathered in the framework of the S5PVT AO projects CESAR (ID #28596, PI A. Apituley, KNMI) and NIDFORVAL (ID #28607, PI G. Pinardi, BIRA-IASB). The LATMOS SAOZ_RT team (A. Pazmino, A. Bazureau, F. Goutail, and J.-P. Pommereau) at IPSL/UVSQ/UPMC/CNRS is thanked for the near-real-time processing and delivery of ZSL-DOAS SAOZ data. ESA's FRM programme and LuftBlick/U. Innsbruck (A. Cede, M. Gebetsberger and M. Tiefengraber) are acknowledged for the rapid delivery of total NO_2 data from the Pandonia Global Network (PGN).
- Rapid delivery HCHO data from NDACC FTIR and MAX-DOAS stations was gathered in the framework of the S5PVT AO projects CESAR (ID #28596, PI A. Apituley, KNMI) and NIDFORVAL (ID #28607, Co-PIs G. Pinardi and C. Vigouroux, BIRA-IASB). This work could not be possible without the work of the FTIR and DOAS spectra and/or data providers: Carlos Augusto Bauer Aquino (IFRO); Cornelis Becker (SAHO); Thomas Blumenstock and Amelie Röhling (KIT, IMK-ASF); Martine De Mazière, Christian Hermans, François Hendrick, Michel Van Roozendael and Minqiang Zhou (BIRA); Omaira García (AEMET); Michel Grutter, Claudia Rivera, Alejandro Bezanilla, César Guarin and Wolfgang Stremme (UNAM); James Hannigan and Ivan Ortega (NCAR); Pascal Jeseck and Yao Té (LERMA-IPSL); Nicholas Jones and Clare Paton-Walsh (Univ. Wollongong), Rigel Kivi (FMI), Erik Lutsch and Kim Strong (Univ. Toronto); Maria Makarova and Anatoly Poberovskii (Univ. St. Petersburg); Emmanuel Mahieu and Christian Servais (Univ. Liège); Jean-Marc Metzger (Univ. Reunion Island); Isamu Morino and Hideaki Nakajima (NIES); Isao Murata (Univ. Tohoku); Tomoo Nagahama (ISEE); Justus Notholt, Mathias Palm and Holger Winkler (Univ. Bremen); Markus Rettinger and Ralf Sussman (KIT, IMK-IFU); John Robinson and Dan Smale (NIWA); Pucai Wang, Youwen Sun and Cheng Liu (CAS); Ankie Piters (KNMI); Thomas Wagner, Sebastian Donner and Julia Remmers (MPIC).
- Rapid delivery of CO and CH4 data from COCCON FTIR stations was gathered in the framework of the ESA projects COCCON-PROCEEDS and Sentinel-5p MPC-VDAF.
- Rapid delivery CO and CH_4 data from TCCON FTIR stations was gathered in the framework of the S5PVT AO project TCCON4S5P (ID #28603, PI M. Kumar Sha, BIRA-IASB).
- Rapid delivery of NDACC data is partly supported by the CAMS-27 data procurement service contracted by ECMWF for the validation of the Copernicus Atmospheric Monitoring Service (CAMS).

CLOUDNET classification product was obtained via the European Research Infrastructure for the observation of Aerosol, Clouds, and Trace gases (ACTRIS) [ER_ACTRIS] and EVDC. Data was processed at the Department of Meteorology, University of Reading, UK, and at the Finnish Meteorological Institute. They acknowledge funding from the EU's Horizon 2020 programme under grant agreement No 654109 and the Cloudnet project (EU contract EVK2-2000-00611).

Automated Lidars and Ceilometers (ALC) data was obtained as part of the E-PROFILE observation programme run in the framework of the European Meteorological Services Network (EUMETNET) [ER_EUMETNET].

EUMETSAT AC-SAF and DLR are acknowledged for the provision of MetOp-A, MetOp-B and MetOp-C GOME-2 ozone and cloud data.

EARLINET is acknowledged for providing aerosol LIDAR profiles available at https://data.earlinet.org/. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654109 and previously from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°262254.KNMI is acknowledged for the provision of EOS-Aura OMI O_3, NO_2, HCHO and UVAI data. The OMI QA4ECV data records are an outcome of the EC FP7-SPACE-2013-1 project No 607405: Quality Assurance for Essential Climate Variables (QA4ECV).

NASA/GSFC is acknowledged for the provision of (i) Suomi-NPP OMPS radiance, O_3 and UVAI data, (ii) Suomi-NPP VIIRS cloud data obtained with a pre-production code run specifically for limited S5P team analysis, (iii) EOS-Aqua MODIS cloud fraction, cloud top height and cloud optical thickness data, and (iv) MISR and CALIOP aerosol layer height data.

17.4 Agency support

The ESA/Copernicus ATM-MPC S5P Routine Operations Validation Service is supported jointly by ESA, the Belgian Federal Science Policy Office (BELSPO) through BIRA-IASB, the Netherlands Space Office (NSO), and the German Aerospace Centre (DLR). S5PVT Announcement of Opportunity (AO) projects [ER_S5PVT] having contributed to this report are funded by several national agencies from Europe, Canada, China, Japan and the USA.

18 Terms, definitions and abbreviated terms

18.1 Terms and definitions

accuracy	closeness of agreement between a quantity value obtained by measurement and the true value of the measurand; note that it is not a quantity and it is not given a numerical quantity value [JCGM-VIM]
bias	(1) systematic error of indication of a measuring system [JCGM-VIM] (2) estimate of a systematic measurement error [JCGM-VIM]
error	(1) measured quantity value minus a reference quantity value [JCGM-VIM] (2) difference of quantity value obtained by measurement and true value of the measurand (CEOS/ISO)
influence quantity	quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result [JCGM-VIM]
Level-1b data	calibrated, geo-located Earth reflectance and radiance spectra in all spectral bands; solar irradiance data, annotation data and references to used calibration data
Level-2 data	geophysical measurand at the same resolution and geolocation as the Level 1 source data from which it is derived
Level-3 data	data or retrieved geophysical parameters (i.e. derived from Level 1 or 2 data products) mapped on uniform space-time grid scales, usually with some completeness and consistency. Such re-sampling may include averaging, compositing, kriging, use of Kalman filters…
measurand	quantity intended to be measured [JCGM-VIM]
measurement bias	estimate of a systematic measurement error [JCGM-VIM]
measurement error	measured quantity value minus a reference quantity value [JCGM-VIM]
measurement uncertainty	non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used [JCGM-VIM]
precision	closeness of agreement between quantity values obtained by replicate measurements of a quantity on the same or similar object under specified conditions [JCGM-VIM]
random error	component of measurement error that in replicate measurements varies in an unpredictable manner; note that random measurement error equals measurement error minus systematic measurement error [JCGM-VIM]
relative standard uncertainty	standard measurement uncertainty divided by the absolute value of the measured quantity value [JCGM-VIM]
stability	ability of a measuring system to maintain its metrological characteristics constant with time [JCGM-VIM]
systematic error	component of measurement error that in replicate measurements remains constant or varies in a predictable manner [JCGM-VIM]
uncertainty	non-negative parameter that characterizes the dispersion of the quantity values that are being attributed to a measurand, based on the information used [JCGM-VIM]
validation	(1) the process of assessing, by independent means, the quality of the data products derived from the system outputs (CEOS/ISO) (2) verification where the specified requirements are adequate for an intended use [JCGM-VIM]
verification	the provision of objective evidence that a given data product fulfils specified requirements; note that, when applicable, measurement uncertainty should be taken into consideration [JCGM-VIM]

18.2 Acronyms and abbreviations

A(A)I	Aerosol (Absorbing) Index
AC-SAF	Atmospheric Composition Satellite Application Facility
ACTRIS	European Research Infrastructure for the observation of Aerosol, Clouds, and Trace gases
AK	averaging kernel
ALC	Automated Lidars and Ceilometers network
ALH	Aerosol Layer Height
AMF	Air Mass Factor
AO	Announcement of Opportunity
ARM	Atmospheric Radiation Measurement program
ATBD	Algorithm Theoretical Basis Document
AVS	Automated Validation Server
AUTH	Aristotle University of Thessaloniki
BELSPO	Belgian Federal Science Policy Office
BIRA-IASB	Royal Belgian Institute for Space Aeronomy
C3S	Copernicus Climate Change Service
CAL	Clouds As Layers
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CAMS	Copernicus Atmosphere Monitoring Service
CCD	Convective Cloud Differential method
CCI	Climate Change Initiative
CESAR	Cabauw Experimental Research Site for Atmospheric Research
CF	Cloud Fraction (fractional cloud cover)
CHEOPS-5p	Validation of Copernicus HEight-resolved Ozone data Products from Sentinel-5p
CLOUDNET	Cloud properties monitoring Network
COCCON	Collaborative Carbon Column Observing Network
COT	Cloud Optical thickness
CRB	Clouds as Reflecting Boundaries
CRG	Climate Research Group
C(T)H	Cloud (Top) Height
DFS	Degree of Freedom of the Signal
DIAL	Differential Absorption Lidar
DLR	German Aerospace Centre / Deutsches Zentrum für Luft- und Raumfahrt
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit
EARLINET	European Aerosol Research Lidar Network
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
EOS	Earth Observing System
EPS	EUMETSAT Polar System

ESA	European Space Agency
ESL	Expert Support Laboratory
EU	European Union
EUMETNET	European Meteorological Services Network
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EVDC	ESA Atmospheric Validation Data Centre
FRM	Fiducial Reference Measurement
FTIR	Fourier Transform Infra-Red
FWHM	Full Width at Half Maximum
GE_LER	Geometry-dependent Effective Lambert-Equivalent Reflectivity
G3W	Global Greenhouse Gas Watch
GAW	Global Atmosphere Watch
GOME(-2)	Global Ozone Monitoring Experiment(-2)
GOSAT(-2)	Greenhouse gases Observing SATellite(-2)
GSFC	Goddard Space Flight Center
GUM	Guide to the Expression of Uncertainty in Measurement
IPSL/UVSQ	Institut Pierre-Simon Laplace / Université de Versailles Saint-Quentin-en-Yvelines
IUP-UB	Institute of Environmental Physics - University of Bremen
JCGM	Joint Committee for Guides in Metrology
KNMI	Koninklijk Netherlands Meteorologisch Instituut / Royal Dutch Meteorological Institute
LATMOS	Laboratoire Atmosphères, Milieux, Observations Spatiales
LER	Lambert-Equivalent Reflectivity
Lidar	LIght Detection And Ranging
MAX-DOAS	Multi Axis Differential Optical Absorption Spectroscopy
MetOp	polar orbiting Meteorological Operational satellite
MPC	Mission Performance Centre
MPI-C	Max Planck Institute for Chemistry
NASA	National Aeronautics and Space Administration
NDACC	Network for the Detection of Atmospheric Composition Change
NIDFORVAL	S5P NItrogen Dioxide and FORmaldehyde VALidation using NDACC and complementary FTIR and UV-Vis DOAS ground-based remote sensing data
NOVAC	Network for Observation of Volcanic and Atmospheric Change
NILU	Norsk Institutt for Luftforskning / Norwegian Institute for Air Research
NOAA	National Oceanic and Atmospheric Administration
NRT	Near Real Time
NSO	Netherlands Space Office
PANDORA	not an acronym; direct Sun UV-visible spectrometer
PGN	Pandora Global Network
OFFL	Off-line
OMI	Ozone Monitoring Instrument
OMPS	Ozone Mapper and Profiling Suite

PAL	Product Algorithm Laboratory
PDGS	Payload Data Ground Segment
PGN	Pandonia Global Network
PI	Principal Investigator
PICS	Pseudo-Invariant Calibration Site
PRF	Product Readme File
PUM	Product User Manual
QA4EO	Quality Assurance framework for Earth Observation
QC	Quality Control
QI	Quality Indicator
QWG	Quality Working Group
RAL	Rutherford Appleton Laboratory
S5P	Sentinel-5 Precursor
S5PVT	Sentinel-5 Precursor Validation Team
SAOZ	Système d'Analyse par Observation Zénithale
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY
SHADOZ	Southern Hemisphere ADditional OZonesonde programme
SRON	Netherlands Institute for Space Research
Suomi-NPP	Suomi National Polar-orbiting Partnership
TCCON	Total Carbon Column Observing Network
TCCON4S5P	Validation of S5P Methane and Carbon Monoxide with TCCON Data
TOLNet	Tropospheric Ozone Lidar Network
TROPOMI	Tropospheric Monitoring Instrument
UTLS	Upper Troposphere / Lower Stratosphere
UVAI	Ultraviolet aerosol absorbing index
VDAF	Validation Data Analysis Facility
VIIRS	Visible Infrared Imaging Radiometer Suite
VIM	International Vocabulary of Metrology
WMO	World Meteorological Organization
WOUDC	World Ozone and Ultraviolet Data Centre
ZSL-DOAS	Zenith-Scattered-Light Differential Optical Absorption Spectroscopy

END OF DOCUMENT