S5P COBRA Sulphur Dioxide [L2__SO2CBR_] Readme

document number	S5P-BIRA-PRF-SO2CBR	
issue	2.2.0	
date	2024-04-16	
product version	V02.00.01	
status	Released	
Prepared by	N. Theys (BIRA-IASB) J. Vlietinck (BIRA-IASB)	SO2CBR product lead
Reviewed by		
Approved by		

CHANGE LOG

Reason for change	Issue	Revision	Date
First draft	1	0	14/09/2022
Updates of tables 1 and 2	1	1	18/01/2023
Document update for SO2CBR v2.0.0	2	0	20/12/2023
Update of the qa_value definition	2	1	12/04/2024
Document update for product version V02.00.01	2	2	16/04/2024

1 Summary

This is the Product Readme File (PRF) for the Copernicus Sentinel 5 Precursor Tropospheric Monitoring Instrument (S5P/TROPOMI) COBRA sulfur dioxide Level 2 data v2 product and is applicable to Offline (OFFL) products.

Product Identifier: **L2__SO2CBR_**

Example filename:

S5P_PAL__L2__SO2CBR_20200626T041008_20200626T055138_14000_03_020000_20231214T202248.nc

The Readme file describes the current processing baseline, product and quality limitations, and product availability status.

The data file contains the `sulfurdioxide_total_vertical_column` which gives the total atmospheric column between the surface and the top of troposphere, and `sulfurdioxide_total_vertical_column{1,7,15} km` which are total SO_2 columns assuming 1km thick box profiles at ground level, centered at 7km and at 15km a.s.l. The respective random error originating from the spectral fit is provided in the `sulfurdioxide_total_vertical_column_precision` and systematic error in the field `sulfurdioxide_total_vertical_column_trueness`. Similarly, random and systematic error estimates are also provided for the SO_2 columns for the assumed box profiles. As a user guideline for the data quality a `qa_value` is given with the data and is applicable only to `sulfurdioxide_total_vertical_column`. In order to avoid misinterpretation of the data quality, it is recommended at the current stage to only use those pixels with a `qa_value` above 0.5.

Important note: investigations based on SO2CBR v2 reprocessed data revealed that the qa_value in the data files is calculated in a sub-optimal way. It is recommended to recalculate the qa_value variable following the calculation rule given in section 3.1, and then apply the pixels selection of qa_value larger than 0.5.

Note that the SO_2 data product may be used in different ways, and different fields in the file are relevant depending on the application. The averaging kernels are provided and should be used e.g., for comparisons with models or profile measurements.

Independent validation using satellite and ground-based SO_2 measurements concludes that the COBRA OFFL SO_2 data is compliant with the requirements as defined in [RD01], see Table 1.

Parameter	Data product	Vertical Resolution	Bias	Random
SO2	SO2 enhanced	Total column	30%	0.3 DU
SO2	Total SO2	Total column	50%	3 DU

Table 1: Data quality target for the Sentinel-5 Precursor TROPOMI L2 SO_2 product. Note that a distinction is made between volcanic SO_2 conditions (referred as ‘Enhanced’ in Table 1) and SO_2 pollution scenarios in the boundary layer (‘Total’).

2 Processing baseline description

The history of the SO2CBR processor versions is detailed in **Table 2**.

Processor Version	In operation from	In operation until	Relevant improvements
01.00.01	19 July 2022 onwards	Initial version	Initial version
02.00.00	7 May 2018 onwards	Current version	Better treatment of volcanic SO_2 plumes and mitigation of the impact of large SO_2 columns on the covariance matrix calculation. Improved covariance matrix calculation and retrieval error estimates.
02.00.01	7 May 2018 onwards	Current version	This version includes small bug fixes for processor version 02.00.00. Only a limited number of orbits needed to be reprocessed with 02.00.01

Table 2: History of SO_2 processor versions

3 Product Quality

3.1 Recommendations for data usage

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 low 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, contamination by volcanic SO_2, and important variables out of range (e.g., the Air Mass Factor (AMF)).

The `qa_value` is a continuous variable, ranging from 0 (error) to 1 (good quality). Currently pixels are considered of good quality `qa_value` above 0.5. As mentioned in section 1, the qa_value variable in the files is not calculated in an optimal way. It is recommended to recalculate the qa_value based on the qa_value calculation rule described in Table 3. A python script to recalculate the qa_value is given in Annex.

Description	warning	error	qa_value multipl. fact.	Comment
Solar zenith angle out of range error.		sza_range_error	0	solar_zenith_angle>85
Solar zenith angle out of range warning	high_sza_warning		1-(0.9226-cos(SZA))	only for SZA>65
Slant column fit returned error, no values can be computed.		slant_column_density_error	0	fill value in fitted_slant_columns_win1
Airmass factor could not be computed.		airmass_factor_error	0	fill value in sulfurdioxide_total_air_mass_factor_polluted
Vertical column density could not be computed.		vertical_column_density_error	0	fill value in sulfurdioxide_total_vertical_column
Cloud fraction above threshold. cloud_fraction_intensity_weighted (CRF_SO2)>0.5	cloud_warning		(1-CRF_SO2)	only for cloud_fraction_intensity_weighted >0.5
Pixel contains snow/ice.	snow_ice_warning		0.49	snow_ice_flag=1
Slant column from COBRA not applied	back_corr_warning		0.5 0.75	sulfurdioxide_cobra_flag=0 sulfurdioxide_cobra_flag=1
Possible contamination by volcanic SO2	SO2_volcanic_warning		0.6 (wdw2) / 0.2 (wdw3)	selected_fitting_window_flag =2 or =3
VCD negative		VCD negative error	0	sulfurdioxide_total_vertical_column < - 0,0045 mol.m-2
AMF out of range (AMF<0.15)	AMF_range_warning		0.49	sulfurdioxide_total_air_mass_factor_polluted <0.15

Table 3: SO2CBR qa_value calculation rule. The qa_value is obtained by multiplying the initial qa_value (set to 1) with the multiplication factors corresponding to the different warnings and errors.

Please note that `qa_value` applies only to `sulfurdioxide_total_vertical_column` (anthropogenic SO_2). The L2 SO_2 product also includes volcanic SO_2 products for prescribed SO_2 plume heights at 1, 7, 15 km (`sulfurdioxide_total_vertical_column_{1,7,15}km`), relevant in case of volcanic emissions. In that case, the only filtering criteria needed is SZA < 65°.

For further details, data users are encouraged to read the description associated with this data product, in section 5.

3.2 Validation results

Initial validation results performed at BIRA-IASB ([RD01]) concludes that the TROPOMI SO_2 COBRA columns are of general good quality.

Compared to the TROPOMI SO_2 products from the operational and scientific PCA algorithms, the results from SO2CBR are improved over clean areas and at high latitudes. For large emission hotspots, all products are consistent. Compared to SO2CBR v1, the SO2CBR v2 improves the results over high latitudes and artefacts related to covariance matrix corruption by elevated SO_2 loadings (from volcanic eruptions).

The agreement of SO2CBR VCD data with MAX-DOAS (Xianghe and Mohali stations) and Pandora (Mexico City and Wakkerstroom) instruments is generally very good. No discernible biases can be identified from the comparison. More validation will be performed in the future to obtain more insight in the accuracy of the SO2CBR columns.

4 Data Quality Remarks

4.1 Known Data Quality Issues

Currently, the following data quality issues are known, not covered by the quality flags, and should be kept in mind when looking at the SO_2 product itself and also at validation results.

Bands 3-4 and 6 spatial miss-alignment

The band 3-4 (450 pixels per scanline) footprints are not fully aligned with the band 6 (448 pixels per scanline) footprints. In the worst case, the misalignment can be in the order of half a ground pixel. The OCRA cloud algorithm retrieves the cloud fraction at bands 3 and 4 and interpolates it linearly, according to the covered area, to band 6. This is an *a priori* to the ROCINN algorithm which works in band 6. Over heterogeneous scenes the mis-registration might have a large impact on the data quality. The cloud height and optical thickness retrieved in band 6 are interpolated back to the band 3 footprints. Due to missing overlap with the band 6 footprints, the first pixel in band 3 (no overlap) does not contain cloud data and the second pixel in band 3 (only partial overlap), contains cloud products with reduced quality. This is also reflected in the cloud data `qa_value`.

A-priori profiles from TM5 model

The current version of the TM5 Chemistry Transport Model (CTM) does not include SO_2 emissions over the large hotspot region of Norilsk, Siberia. Consequently, the SO_2 columns are likely underestimated over Norilsk for low albedo conditions.

Snow-ice scenes

The snow-ice scenes are filtered out using a `qa_value` above 0.6 but the current algorithm is processing the data anyway. A proper treatment of snow-ice scenes is not part of the current algorithm version, and climatological values for the surface albedo are used for the AMF calculation. Therefore, the resulting VCDs are largely overestimated, and the data should not be used. A next algorithm version will include a better treatment of snow-ice scenes in the AMF calculation.

Offsets

Local offsets (negative or positive) not completely corrected, may arise in certain regions. In general, negative offsets can be observed over bright scenes (e.g., salt lakes) while positive offsets are seen over dark scenes.

qa_value

The qa_value variable is not optimal. It is recommended to recalculate the qa_value based on the calculation rule described in section 3.1.

4.2 Solved Data Quality Issues

Upon processor version updates, corrected data quality issues will be listed here.

Contamination by volcanic SO_2

For SO2CBR v1, in case of an eruption, the initial covariance matrix used for the retrieval of SO_2 slant columns could be strongly affected by the spectra containing SO_2 absorption. In principle, this is mitigated by excluding the corresponding spectra from the covariance calculations for the next iterations. However, this procedure was found far than perfect and there are situations where the product quality was degraded for part of the orbit, often leading to negative biases. In the SO2CBR v2, this has been largely improved by using DOAS retrieved slant column densities (from the operational product) to prefilter the spectra for the initial covariance matrix calculation. Most of the related artefacts are solved by this procedure but still there are situations (extended end dispersed plumes) where the number of spectra for the covariance matrix calculation is less than usual or where a small negative bias is observed in the retrieved SCD (due to remaining SO_2 in the set of clean spectra). Note that the user of the product is interested by comparing the SO2 results from COBRA and DOAS, the slant columns from the (DOAS) operational retrievals are stored in the data files. The VCDs can be calculated easily by scaling the SCDs with the air mass factors (also in the files).

Surface albedo climatology

For SO2CBR v1, the surface albedo climatology used has a spatial resolution of 0.5° x 0.5°, and a time resolution of 1 month. This resolution is known to be too coarse compared to the much higher spatial resolution of S5p TROPOMI pixels. This has an impact on the accuracy of the SO_2 vertical column (mostly for the polluted scenario) through the AMF calculation. For SO2CBR v2, we make use of the TROPOMI surface DLER V2.1 product, available on the S5P PAL portal ([ER02]).

4.3 Data Features

This section describes some characteristics of the data that might seem anomalous, however they are physically correct and not related to any problem.

Pixel geolocation around North Pole (feature)

The solar irradiance is measured on a daily basis over the North Pole at a reference azimuth angle to remove seasonal effects on the measurements. To this end, a yaw manoeuvre is executed when the instrument is still in radiance mode, causing possible distortion on the scanlines observed during this manoeuvre (i.e., crossing scanlines, "bow-tie" ground pixel shape instead of rectangular). This occurs at most during the last 26 seconds of radiance measurements in few orbits (7-9 per week). Though this may seem anomalous, it is physically correct, and not related to any problem on the data geolocation.

4.4 Mission Operations Change

A change in the Copernicus Sentinel 5P operations scenario, increasing the spatial resolution from 7.0 km to 5.5 km along track for all measurements, became operational starting from 6 August 2019, orbit 9388.

5 Algorithm Change Record

In this section, we aim at giving a short description of the algorithm and planned evolution.

The SO_2 Covariance-Based Retrieval Algorithm (COBRA) is an improved version of the existing TROPOMI operational SO_2 L2 algorithm, and there are several algorithm components in common. As a starting point, it is useful to remind the main algorithm parts of the L2 SO_2 operational algorithm (for a detailed description of the L2__SO2____ algorithm, please refer to the ATBD ([RD02]):

1. Slant column density retrievals: after a wavelength calibration step, the measured radiances are analyzed with the Differential Optical Absorption Spectroscopy (DOAS) technique in three fitting windows: 312-326 nm (wdw1, default), 325-335 nm (wdw2), 360-390 nm (wdw3).
2. Background correction: to account for possible SCD offsets, the retrieved SCDs are corrected using a dedicated background correction processor. It is acting separately for each row and fitting window and is updated using measured (presumably SO_2 free) SCDs of previous days. Note that the background correction processor also calculates (dynamically) radiance reference spectra used for the DOAS analysis.
3. Air mass factor calculations: based on SCDs results, a selection of one of the three fitting windows is made as final SCD. To convert the SCD into vertical column, air mass factors are calculated for 4 different a priori profiles: 1 profile for polluted scenario (from the TM5 chemical transport model), 3 box profiles for volcanic cases peaking at 1, 7, 15 km height. The AMFs are calculated using a look-up-table of height-resolved air mass factors and accounts for dependence on observation geometry, surface reflectance and clouds. The AMF module is the final step of the SO_2 product generation and also include the calculation of the so-called column averaging kernels as well as the product error estimations.

The current SO_2 COBRA algorithm is changing the slant column density retrieval in the default fitting window only. The SCD, background correction and AMF results in window 2 and window3 are the same as in the operational algorithm. More precisely, COBRA is improving the SCDs both in terms of noise and biases and removes the need for a post-processing background correction. The algorithm is described in detail in [RD03]. In short, the COBRA scheme starts from the wavelength calibrated radiances and retrieves a single fitted parameter: the SO_2 slant column:

$$\widehat{SCD} = (k^T S^{-1} k)^{-1} k^T S^{-1} (y - \bar{y}) \quad (1)$$

In this expression, k is the row-dependent SO_2 cross section vector over the fitting range, y is the measurement vector (log intensity radio from radiance and irradiance measurements) of the pixel to be analyzed. S and $\bar{y}$ are the covariance and mean vector of a set of clean (SO_2-free) spectra. The idea of COBRA is to select a set of measured SO_2-free spectra, representative of the background variability of the spectra and use the inverse of the covariance matrix as a weight to optimally retrieve the target species (SO_2 in this case). Although the method is simple in principle, the algorithm is taking care of calculating separately the covariance matrix (and $\bar{y}$) for each orbit, each row and for 6 scanline segments. A precaution is also implemented to remove the spectra from the clean set that are contaminated by SO_2, in the form of an iteration process. Finally, from the fit, it is also possible to calculate a retrieval error (precision estimate) by:

$$\widehat{SCD}_{err} = \chi . \sqrt{(k^T S^{-1} k)^{-1}} \quad (2)$$

with the normalized root-mean-square χ:

$$\chi^2 = \frac{1}{N-1}.\left(\Delta y^T S^{-1} \Delta y\right) \qquad (3)$$

where $\Delta y = y - \bar{y} - k.\widehat{SCD}$ (residual) and N being the number of wavelengths in the fitting interval. This variable is typically close to 1, except for conditions of bad fit quality (χ>>1).

Compared to the operational algorithm, there is one important deviation that is the fitting window used. By default, COBRA uses 310.5-326 nm, instead of 312-326 nm for window 1. This further improves the retrieval noise. To cope with this change, the look-up-table of height-resolved AMFs for window 1 has been regenerated, for a representative wavelength of 311.7 nm (instead of 313 nm as before). All the rest of the algorithm is the same as in the operational algorithm, except for the estimation of the SCD systematic error estimation in window. Because COBRA improves the SCD results compared to DOAS, the formulation of the SCD systematic uncertainty has been slightly adapted to 0.04DU+10% of the SCD.

In COBRA version 2, several changes have been applied to the retrieval scheme:

a. A better retrieval error estimation based on the normalized root mean square (Eq. 3).
b. Improved selection of clean spectra, reducing artefacts at high latitudes.
c. Improvement of the robustness of the algorithm in case of an eruption, by prefilter the spectra affected by strong SO_2 columns. This is performed at the first iteration using the DOAS results from the operational product. If not enough clean spectra are found (less than 100) to estimate a covariance matrix, the algorithm falls back on the operational (DOAS-based) product results. It should be noted that the DOAS slant columns are stored in the file, for comparison purpose.
d. An improved selection of fitting windows is used, to switch to the fitting windows 2 and 3more easily. This is important mostly for SO_2 plumes in the lower troposphere.
e. The surface reflectance data base has been updated to the latest S5P DLER data set. For this, the minLER monthly climatology (0.125°×0.125°) is used at 328nm and 380nm.

6 Data Format

The product is stored as NetCDF4 file. The NetCDF4 file contains both the data and the metadata for the product.

For OFFL data the product is stored as a single file per satellite orbit. Processing of near-real time (NRTI) data is currently not foreseen.

Details of the data format are provided in the Product File Specification (PFS) document ([RD04]).

6.1 Data format changes

The file format follows a netCDF4 structure with content organized according to the S5P-PAL guidelines ([RD05]).

In processor version 02.00.00, a file class ‘PAL_’ was introduced. This replaces the file class ‘OFFL’ or ‘RPRO’ when the L1B to L2 processing takes place on the S5P-PAL system. This is reflected in the resulting output files through the ‘file_class’ global attribute and through the output file name, where ‘OFFL’ or ‘RPRO’ is replaced by ‘PAL_’.

7 Product Availability

The latest product release, version 02.00.00, is currently implemented in the so-called pre-operational environment of the S5P-PAL system ([ER01]). The data will become available through the data portal of the PAL system ([ER02]). Details will be provided in a future update of this document.

8 References

[RD01] S5P/TROPOMI SO2CBR Validation Report, **source**: BIRA; **ref**: S5P-BIRA-L2-VR-SO2CBR; **issue:** 2.0.0; **date:** 2023-12-20.

[RD02] Sentinel-5 precursor/TROPOMI Level 2 Algorithm Theoretical Basis Document Sulphur Dioxide SO2, **source**: BIRA-IASB; **ref:** S5P-BIRA-L2-400E-ATBD; **url:** https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-ATBD-SO2-TROPOMI

[RD03] S5p L2 COBRA paper: Theys, N., Fioletov, V., Li, C., De Smedt, I., Lerot, C., McLinden, C., Krotkov, N., Griffin, D., Clarisse, L., Hedelt, P., Loyola, D., Wagner, T., Kumar, V., Innes, A., Ribas, R., Hendrick, F., Vlietinck, J., Brenot, H., and Van Roozendael, M.: A sulfur dioxide Covariance-Based Retrieval Algorithm (COBRA): application to TROPOMI reveals new emission sources, Atmos. Chem. Phys., 21, 16727–16744, https://doi.org/10.5194/acp-21-16727-2021, 2021.

[RD04] S5P/TROPOMI SO2CBR Product Format Specification, **source**: BIRA; **ref**: S5P-L2 BIRA-PFS-SO2CBR; **issue:** 2.0; **date:** 2023-12-14.

[RD05] S5p Product Algorithm Laboratory L2 Processor File Format Guidelines. **source:** S&T; **ref:** ST-ESA-S5P_PAL-L2FFG-001; **issue:** 1.4; **date:** 2023-03-27.

[ER01] https://www.s5p-pal.com/

[ER02] https://data-portal.s5p-pal.com/

Abbreviations and acronyms

AMF	Air Mass Factor
ATBD	Algorithm Theoretical Basis Document
BIRA-IASB	Royal Belgian Institute for Space Aeronomy
COBRA	Covariance-Based Retrieval Algorithm
CTM	Chemical Transport Model
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit (1 DU: 2.69 x 10^{16} molec/cm²)
ESA	European Space Agency
L2	Level-2
MAX-DOAS	Multi Axis Differential Optical Absorption Spectroscopy
NRT	Near-real time
OFFL	Offline
PAL	Product Algorithm Laboratory
PFS	Product Format Specificatiopn
PRF	Product Readme File
SCD	Slant Column Density
S5P	Sentinel-5 Precursor
SZA	Solar Zenith Angle
SO2	Sulfur dioxide
TROPOMI	Tropospheric Monitoring Instrument
VCD	Vertical Column Density

Annex: python script to recalculate the qa_value variable

```
import numpy as np

import netCDF4 as nc


l2="S5P_PAL__L2__SO2CBR_20240401T010111_20240401T024241_33506_03_020001_20240411
T035411.nc"

with nc.Dataset(l2,'r+') as ncl2:

    sza=ncl2["/PRODUCT/SUPPORT_DATA/GEOLOCATIONS/solar_zenith_angle"][:]

    snow=ncl2["/PRODUCT/SUPPORT_DATA/INPUT_DATA/snow_ice_flag"][:]

    amf=ncl2["/PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/
            sulfurdioxide_total_air_mass_factor_polluted"][:]

   crf=ncl2["/PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/
        cloud_fraction_intensity_weighted"][:]

    fitwin=ncl2['/PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/selected_fitting_window_flag'][:]

    vcd=ncl2['/PRODUCT/sulfurdioxide_total_vertical_column'][:]

    scd=ncl2['/PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/
          sulfurdioxide_slant_column_corrected'][:]

    cobraflag=ncl2["/PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/sulfurdioxide_cobra_flag"][:]

    #create an initial qa value matrix filled with ones

   qa=np.ones(sza.shape)

    #when value of pixel is invalid (NAN), then qa to zero

    qa[np.isnan(sza) | np.isnan(crf) | np.isnan(amf)  | np.isnan(vcd) ]=0

    #when sza>85 of pixel is invalid (NAN), then qa to zero

    qa[sza>85]=0

    #when vcd<-0.0045 of pixel is invalid (NAN), then qa to zero

    qa[vcd<-0.0045]=0

    qa[(sza<=85) & (sza>65)]=qa[(sza<=85) & (sza>65)]*(0.0774+np.cos(np.radians(sza[(sza<=85)
       &  (sza>65)])))

    #the following lines reduces the qa value step by step

    qa[snow==1]= qa[snow==1]*0.49

    qa[amf<0.15]=qa[amf<0.15]*0.49

    qa[fitwin==2]=qa[fitwin==2]*0.6

    qa[fitwin==3]=qa[fitwin==3]*0.2

    qa[crf>0.5]=qa[crf>0.5]*(1-crf[crf>0.5])

    qa[cobraflag==1]=qa[cobraflag==1]*0.75

    qa[cobraflag==0]=qa[cobraflag==0]*0.5

    ncl2["/PRODUCT/qa_value"][:] = (100*qa).astype(np.uint8)/100
```