S5P COBRA Sulphur Dioxide [L2__SO2CBR_] Readme

document number	S5P-BIRA-PRF-SO2CBR	
issue	1.0.0	
date	2022-09-14	
product version	V01.00.00	
status	Released	
Prepared by	N. Theys (BIRA-IASB)	SO2CBR product lead
Reviewed by		
Approved by		

CHANGE LOG

Reason for change	Issue	Revision	Date
First draft	1	0	14/09/2022

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 product and is applicable fto Offline (OFFL) products.

Product Identifier: **L2__SO2CBR_**

Example filename:

S5P_PAL__L2__SO2CBR_20220117T022855_20220117T041025_22086_02_010000_20220530T010253.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. For best data quality selection users are recommended to follow more stringent criteria, as listed in section 3.1.

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 for e.g. 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: 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
01.00.00	OFFL: orbit	Current version	- Initial version

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. The quality is of the data is deemed as poor if one of the following condition is not fulfilled:

1. snow_ice_flag < 0.5
2. sulfurdioxide_total_air_mass_factor_polluted > 0.15
3. sulfurdioxide_total_vertical_column > -0.0015 mol. m-2 and is not _Fill_Value.
4. selected_fitting_window_flag =1
5. cloud_fraction_crb < 0.3
6. solar_zenith_angle < 70°

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 < 70°.

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.

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`.

Contamination by volcanic SO_2

In case of an eruption, the initial covariance matrix used for the retrieval of SO_2 slant columns can 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 is not perfect and there are situations where the product quality is degraded for part of the orbit, often leading to negative biases. This will be improved in the next algorithm versions.

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.

Surface albedo climatology

The current surface albedo climatology 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. It is currently difficult to assess the exact impact on the SO_2 vertical column and it can only be evaluated when a higher resolution albedo climatology becomes available.

Snow-ice scenes

The snow-ice scenes are filtered out using a `qa_value` above 0.5 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.

4.2 Solved Data Quality Issues

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

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 algorithms 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 details 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:

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

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 the future, several changes are foreseen for the COBRA scheme:

a. Apply COBRA to window 2 and window 3, to get rid of the DOAS analysis completely.

b. Improve the robustness of the algorithm in case of an eruption, by the means of a covariance matrix fallback.

c. Possibly apply COBRA to NRT data.

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 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 01.00.00, a file class ‘PAL_’ was introduced. This replaces the file class ‘OFFL’ when the L1B > 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’ is replaced by ‘PAL_’.

7 Product Availability

The latest product release, version 01.00.00, 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:** 1.0.0; **date:** 2022-09-14.

[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:** 1.0.0; **date:** 2022-09-14.

[RD05] S5p Product Algorithm Laboratory L2 Processor File Format Guidelines. **source:** S&T; **ref:** ST-ESA-S5P_PAL-L2FFG-001; **issue:** 1.2; **date:** 2021-02-24.

[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/cm^2)
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
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