

# Precursors\_cci+

## ESA Climate Change Initiative (CCI)



### D4.2 Product User Guide (PUG)

Title	D4.2 Product User Guide (PUG)
Reference	Precursors_cci+_D4.2_PUG_01_01
Issue	01
Revision	01
Status	Final
Date of issue	29/05/2024
Document type	Deliverable

*This work has received funding from the European Space Agency under Contract No 4000138243 and is the result of the cooperation between the partners of the Precursors\_cci+ consortium.*





	FUNCTION	NAME	DATE
LEAD AUTHOR(S)	Scientists	F. Boersma L. Clarisse T. Danckaert I. De Smedt M. George N. Theys M. Van Damme P. Valks	18/01/2024
REVIEWED BY	Science Leaders	Michel Van Roozendael  Folkert Boersma	24/01/2024  23/01/2024
	ESA Project Officer	Simon Pinnock	13/03/2024
DISTRIBUTED TO	ESA Precursors_cci+ Consortium		





### DOCUMENT CHANGE RECORD

Issue	Revision	Date	Modified items	Observations
0	0	15/05/2023	Initial template	Creation of document
1	0	24/01/2024	First version	Some sections still incomplete
1	1	15/05/2024	Various minor changes + changes related to RID	





## Table of Contents

1.	PURPOSE AND SCOPE	7
2.	OVERVIEW OF THE PRECURSORS_CCI+ PRODUCTS	7
2.1.	Overview	7
2.2.	Terms and conditions for use of Precursors_CCI+ data	8
3.	NH <sub>3</sub> L3 FROM IASI	9
3.1.	Input data and algorithm	9
3.2.	File parameters	9
3.3.	Data access and volume	10
3.4.	Data reading examples	11
3.5.	Usage recommendations and known limitations	12
3.6.	Preliminary evaluation	12
3.7.	Contacts and citation	12
4.	CO L3 PRODUCT FROM IASI AND MOPITT	14
4.1.	Input data and algorithm	14
4.2.	File parameters	15
4.3.	Data access and volume	15
4.4.	Data reading examples	16
4.5.	Usage recommendations and known limitations	16
4.6.	Preliminary evaluation	17
4.7.	Contacts and citation	17
5.	NO <sub>2</sub> L3 FROM GOME, SCIAMACHY, GOME2-A/B/C, OMI AND TROPOMI	18
5.1.	Input data and algorithm	18
5.2.	File parameters	19
5.3.	Data access and volume	21





5.4.	Data reading examples	21
5.5.	Usage recommendations and known limitations	22
5.6.	Preliminary evaluation	22
5.7.	Contacts and citation	23
6.	<b>HCHO L3 FROM TROPOMI, OMI, GOME2, SCIAMACHY AND GOME</b>	<b>25</b>
6.1.	Input data and algorithm	25
6.2.	File parameters	26
6.3.	Data access and volume	28
6.4.	Data reading examples	28
6.5.	Usage recommendations and known limitations	29
6.6.	Preliminary evaluation	30
6.7.	Contacts and citation	30
7.	<b>SO<sub>2</sub> L3 FROM TROPOMI, OMI, SCIAMACHY AND GOME</b>	<b>32</b>
7.1.	Input data and algorithm	32
7.2.	File parameters	33
7.3.	Data access and volume	34
7.4.	Data reading examples	35
7.5.	Usage recommendations and known limitations	35
7.6.	Preliminary evaluation	37
7.7.	Contacts and citation	37
8.	<b>CHOCHO L3 FROM TROPOMI, OMI AND GOME2</b>	<b>38</b>
8.1.	Input data and algorithm	38
8.2.	File parameters	39
8.3.	Data access and volume	40
8.4.	Data reading examples	41
8.5.	Usage recommendations and known limitations	41





8.6.	Preliminary evaluation	42
8.7.	Contacts and citation	43
9.	ESA REFERENCE DOCUMENTS	44



## 1. Purpose and scope

The Precursors\_CCI+ project (<https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>) aims to develop long-term climate data records of precursor species of aerosol and ozone. The species include the short-lived atmospheric trace gases: nitrogen dioxide (NO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>), carbon monoxide (CO), formaldehyde (HCHO), ammonia (NH<sub>3</sub>) and glyoxal (CHOCHO). The data records originate from measurements from the satellite sounders IASI A/B/C, MOPITT, TROPOMI, OMI, GOME, GOME2 A/B/C and SCIAMACHY.

This product user guide serves as the primary reference for users seeking information to successfully employ the products in their work. The next section provides an overview of the different products. The remaining sections then provide detailed information on the products related to the each of the different species, including a short description of the input data and algorithms used, the file parameters, data access, format and reading examples, usage recommendation, preliminary evaluation/validation and finally contacts and citation.

## 2. Overview of the Precursors\_CCI+ products

### 2.1. Overview

All satellite products, summarized in Table 2-1 constitute gridded data sets (L3), on daily or monthly basis and with varying spatial resolutions depending on the species. In addition to datasets from individual satellite sounders, also certain merged datasets are available (tagged as MRG in the product identifier). All data sets are reported in NetCDF-4 CF format following CCI and GCOS standards, and are freely available on the Precursors\_CCI+ web site <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>.

**Table 2-1: List of the Precursors\_cci+ datasets and associated characteristics.**

Species	Product identifier	Source	Time period	Spatial res.	Temp. res.	L2 or L3 version
NH <sub>3</sub>	NH3_L3_IASI_A/B/C	ULB-LATMOS	2007-2023	0.125°	Monthly	ANNI v4.0.0
	NH3_L3_IASI_MRG	ULB-LATMOS	2007-2023	0.125°	Monthly	ANNI v4.0.0
CO	CO_L3_MRG	LATMOS-ULB	2007-2023	1°	Monthly	IASI L3 multiplatform v1.2 MOPITT L3 v9T
NO <sub>2</sub>	NO2_L3_S5P_TROPOMI	KNMI	2018-2021	0.5°/0.2°	Daily/Monthly	v2.3.1
	NO2_L3_EOS-AURA_OMI	KNMI	2004-2023	1.0°/0.5°/0.2°	Daily/Monthly	QA4ECV v1.1
	NO2_L3_METOP_GOME2_A/B/C	DLR/EUMETSAT	2007-2023	0.2°	Monthly	AC-SAF prototype DLR
	NO2_L3_ENVISAT_SCIAMACHY	DLR/KNMI/QA4ECV	2002-2012	0.2°	Monthly	QA4ECV v1.1
	NO2_L3_ERS2_GOME	DLR/KNMI/QA4ECV	1995-2003	0.2°	Monthly	QA4ECV v1.1
HCHO	HCHO_L3_S5P_TROPOMI	BIRA-IASB/DLR/ESA	2018-2023	0.25°/0.125°	Daily/Monthly	OFFL v0.3_02/CCI-p v1.0





	HCHO_L3_S5P_OMI	BIRA-IASB/QA4ECV	2005-2022	0.25°/0.125°	Daily/Monthly	QA4ECV v1.2/CCI+p v1.0
	HCHO_L3_METOP_GOME2_A/B/C	DLR/BIRA-IASB/EUMETSAT	2007-2023	0.25°	Monthly	GDP4.8/4.9 (-> AC SAF prototype)
	HCHO_L3_ENVISAT_SCIAMACHY	BIRA-IASB/QA4ECV	2003-2012	0.25°	Monthly	QA4ECV v1.2/CCI+p v1.0.
	HCHO_L3_ERS2_GOME	BIRA-IASB/QA4ECV	1996-2002	0.25°	Monthly	QA4ECV v1.2/CCI+p v1.0.
SO <sub>2</sub>	SO2_L3_S5P_TROPOMI	BIRA-IASB/DLR/ESA	2018-2023	0.25°/0.125°	Daily/Monthly	OFFL v0.3_02/CCI+p v1.0
	SO2_L3_S5P_OMI	BIRA-IASB	2005-2022	0.25°/0.125°	Daily/Monthly	CCI+p v1.0
	SO2_L3_ENVISAT_SCIAMACHY	BIRA-IASB	2003-2012	0.25°	Monthly	Tbd
	SO2_L3_ERS2_GOME	BIRA-IASB	1996-2002	0.25°	Monthly	Tbd
	SO2_L3_MRG	BIRA-IASB	1996-2023	0.25°	Monthly	Tbd
CHOCHO	CHOCHO_L3_S5P_TROPOMI	BIRA-IASB/ESA	2018-2023	0.25°	Daily/Monthly	S5P_PAL v x / CCI+p v1.0
	CHOCHO_L3_AURA_OMI	BIRA-IASB	2005-2022	0.25°	Daily/Monthly	CCI+p v1.0
	CHOCHO_L3_METOP_GOME2_A/B/C	BIRA-IASB	2007-2023	0.25°	Monthly	CCI+p v1.0

## 2.2. Terms and conditions for use of Precursors\_CCI+ data

The Precursors\_CCI+ datasets may be used by any user for any purpose, with the following terms and conditions:

- 1) Users of the CCI data are required to acknowledge the ESA Climate Change Initiative and Precursors\_CCI+ together with the individual data providers if the data are used in a presentation or publication. Please also cite any relevant dataset DOIs (see contacts and citation information in the respective sections).
- 2) Users of the CCI data are encouraged to interact with the CCI programme on use of the products, and to provide a copy of all reports and publications using the dataset. An offer of co-authorship should be considered, if the CCI data constitute a major component of a scientific publication.
- 3) Intellectual property rights (IPR) in the CCI data lie with the researchers and organisations producing the data.
- 4) Liability: No warranty is given as to the quality or the accuracy of the CCI data or its suitability for any use. All implied conditions relating to the quality or suitability of the information, and all liabilities arising from the supply of the information (including any liability arising in negligence) are excluded to the fullest extent permitted by law.





## 3. NH<sub>3</sub> L3 from IASI

### 3.1. Input data and algorithm

The IASI instrument is a Fourier transform spectrometer that measures the thermal infrared emission of the Earth-atmosphere system between 645 and 2760 cm<sup>-1</sup> with a spectral resolution of 0.5 cm<sup>-1</sup>. IASI provides global coverage of the Earth twice a day (at 9:30 and 21:30 mean local solar time) thanks to the combination of a wide swath width of 2 times 1100 km and a set of four simultaneous footprints of 12 km diameter at nadir.

The IASI NH<sub>3</sub> L2 product has been generated at ULB using the ANNI (Artificial Neural Network for IASI) v4 code. ANNI implements a two-step retrieval consisting of the calculation of a NH<sub>3</sub> hyperspectral range index, followed by the conversion of the index to a total column product using a neural network. This last step also includes calculation of diagnostic variables (averaging kernels, uncertainties). The input data for the retrieval algorithm consists of L1C spectra, parameters related to the geolocation (time, latitude, longitude) and observation (zenith angle), meteorological data (cloud and surface temperature, pressure, temperature, and humidity profiles). Clouds and surface temperature are derived from the L1C spectra using dedicated neural networks. Other meteorological data originates from ECMWF ERA5 reanalysis. All the data was postfiltered using the recommended L2 postfilter and keeping data with a cloud fraction below 10%.

The monthly L3 products are directly derived from the L2 by means of oversampled averaging on a 0.125° x 0.125°, where the exact IASI footprint on ground is taken into account. The L3 product consists of NH<sub>3</sub> total columns and associated random and systematic uncertainties for the morning and evening overpass separately. L3 are available for each instrument separately.

The merged product is formally identical to the NH<sub>3</sub> L3 product, but for each point in time 2007-2023 uses one IASI instrument only, thereby guaranteeing a consistent L3 product that uses approximately the same number of measurements for each month throughout. This L3 NH<sub>3</sub> CDR is generated based on observations from IASI/Metop-A (2008-2013), IASI/Metop-B, (2014-2019), IASI/Metop-C (2020-...) data. An alternative L3 merged product would consist of averaging multiple IASI datasets, but that would result in an uneven number of measurements (e.g. L3 that is based on less measurements before 2013 then after), and for this reason this alternative was not implemented.

The reader is invited to refer to the NH<sub>3</sub> ATBD available at <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> and to Clarisse et al. (2023) for a full description of the L2 and L3 algorithm.

### 3.2. File parameters

Tables 3.1 describe the variables contained in the L3 NH<sub>3</sub> netcdf files. The fill value is NaN for all.

**Table 3.1: Variables in the L3 NH<sub>3</sub> NetCDF files. N<sub>lat</sub> (1440) and N<sub>lon</sub> (2880) are the number of grid cells corresponding to the latitude and longitude respectively.**

Parameter	Unit	Dimension	Description
latitude	degrees	N <sub>lat</sub>	latitude



longitude	degrees	$N_{lon}$	longitude
nh3_total_column_ampm	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	L3 (AM+PM) NH <sub>3</sub> total column
nh3_total_column_am	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	L3 (AM) NH <sub>3</sub> total column
nh3_total_column_pm	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	L3 (PM) NH <sub>3</sub> total column
number_of_observations_ampm	1	$N_{lon} \times N_{lat}$	number of observations used for calculating the L3 (AM+PM)
number_of_observations_am	1	$N_{lon} \times N_{lat}$	number of observations used for calculating the L3 (AM)
number_of_observations_pm	1	$N_{lon} \times N_{lat}$	number of observations used for calculating the L3 (PM)
nh3_total_column_random_uncertainty_ampm	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	random total uncertainty on the NH <sub>3</sub> total column (AM+PM)
nh3_total_column_random_uncertainty_am	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	random total uncertainty on the NH <sub>3</sub> total column (AM)
nh3_total_column_random_uncertainty_pm	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	random total uncertainty on the NH <sub>3</sub> total column (PM)
nh3_total_column_systematic_uncertainty_ampm	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	systematic total uncertainty on the NH <sub>3</sub> total column (AM+PM)
nh3_total_column_systematic_uncertainty_am	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	systematic total uncertainty on the NH <sub>3</sub> total column (AM)
nh3_total_column_systematic_uncertainty_pm	$\text{mol.m}^{-2}$	$N_{lon} \times N_{lat}$	systematic total uncertainty on the NH <sub>3</sub> total column (PM)

### 3.3. Data access and volume

The L3 product can be accessed from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> following the filename convention:

IASI\_METOPX\_L3\_NH3\_YYYYMM\_ULB-LATMOS\_V4.0.0R.nc

where X=a, b or c and YYYYMM is the date of retrieval. The content of the netCDF files is listed in Tables 3.1.

Data volume is as follows:

- One file one sensor: 140 MB
- One year one sensor: 1.7 GB
- IASI-A complete: 22 GB
- IASI-B complete: 18 GB
- IASI-C complete: 6.8 GB
- Merged product: 25 GB

Note that the L2 source product (not described here) can be downloaded from AERIS (<https://iasi.aeris-data.fr/>) (one file per day).



## 3.4. Data reading examples

Examples on how to read the netcdf files are provided below, for Matlab or Python.

### Matlab code

```
% Filename of the NetCDF file
fn = 'IASI_METOPB_L3_NH3_202304_ULB-LATMOS_V4.0.0R.nc';

% Retrieve information about the file
fileinfo = ncinfo(fn);

% Read in all variables from the file
for i = 1:length(fileinfo.Variables)
    NH3.(fileinfo.Variables(i).Name) = ncread(fn, fileinfo.Variables(i).Name);
end

% Read in conversion factor from molesConversion to molecules/cm^2
conversion = fileinfo.Variables(3).Attributes(4).Value;

% Create a figure
figure;
imagesc(NH3.longitude, NH3.latitude, NH3.nh3_total_column_am' * conversion, [0 5e16]);
set(gca, 'ydir', 'normal');
colorbar;

% Load and display coastlines
load coastlines;
geoshow(coastlat, coastlon, 'DisplayType', 'line', 'Color', 'k');

% Set the title of the figure
title(fn, 'Interpreter', 'none');
```

### Python code

```
import os
import netCDF4 as nc
import matplotlib.pyplot as plt
import cartopy.crs as ccrs

# Full file path of the NetCDF file
file_path = 'IASI_METOPB_L3_NH3_202304_ULB-LATMOS_V4.0.0R.nc'

# Extract file name and directory path
file_name = os.path.basename(file_path)
directory_path = os.path.dirname(file_path)

# Open the NetCDF file
dataset = nc.Dataset(file_path)

# Create a dictionary to hold the data
NH3 = {}

# Read in all variables from the file
for variable in dataset.variables:
    NH3[variable] = dataset.variables[variable][:]

# Read NH3 total column data and conversion factor
conversion = dataset.variables['nh3_total_column_am'].getncattr('multiplication_factor_to_convert_to_molecules_per_cm2')

# Get longitude and latitude data
longitude = NH3['longitude'][:]
latitude = NH3['latitude'][:]

# Create a figure with specific size based on the grid's aspect ratio
fig = plt.figure(figsize=(14, 10 * (latitude.max() - latitude.min()) / (longitude.max() - longitude.min())))
ax = fig.add_subplot(1, 1, 1, projection=ccrs.PlateCarree())

# Display the data
image = plt.imshow(NH3['nh3_total_column_am'] * conversion, extent=[longitude.min(), longitude.max(), latitude.min(), latitude.max()], origin='lower', cmap='viridis')
plt.colorbar(image, label='NH3 Total Column (molec/cm²)')
plt.clim(0, 2e+16)

# Plot coastlines
ax.coastlines()

# Set the title of the figure using just the file name
plt.title(file_name)

# Show the plot
plt.show()
```



### 3.5. Usage recommendations and known limitations

- Users interested in long-term trends should use the merged L3 dataset. Users interested in short time periods are advised to use the L3 from the three instruments, taking caution that IASI-A started to drift from 2018 onwards, with gradually earlier overpass times as shown in Figure 3.1.
- Higher uncertainties are expected during the evening overpass over land (PM) at high latitude (>60° N or <60° S) and in absence of thermal contrast, which occurs mainly in the colder months (surface temperature below 10°C).
- Over land, it is recommended to use primarily the morning (AM) data, as the PM data is more noisy, and has known biases (some monthly L3 are substantially negative). Over ocean, both AM, PM and AM+PM L3 are recommended for use.
- When averaging the L3 data to larger time periods, to reduce the influence of months with very few measurements, it is recommended to disregard cells with fewer than 0.25 observations (using the appropriate number\_of\_observations variables).
- Mean random ( $\sigma_r$ ) and systematic ( $\sigma_s$ ) uncertainties are available for each grid cell. If a total uncertainty is desired, this can be obtained from the square root of the sum of squares of the different terms:

$$\sigma^2 = \sigma_r^2 + \sigma_s^2$$

- When averaging L3 data over larger temporal or spatial scales, an averaged uncertainty can be obtained from combine the different random ( $\sigma_{ri}$ ) and systematic ( $\sigma_{si}$ ) uncertainties as

$$\sigma^2 = \sum_{i=1}^n \left(\frac{\sigma_{ri}}{n}\right)^2 + \left(\sum_{i=1}^n \frac{\sigma_{si}}{n}\right)^2$$

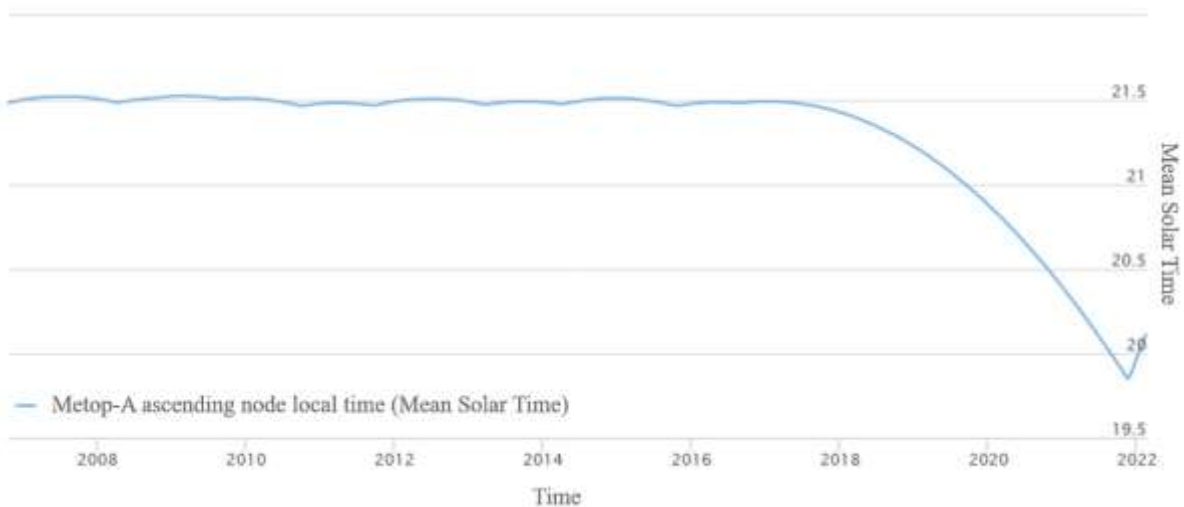


Figure 3.1 Mean Solar Time of Metop-A as a function of time, showing the drift after 2018. Source: <https://climserv.ipsl.polytechnique.fr/ixion/>

### 3.6. Preliminary evaluation

To be completed in the next version of the document.

### 3.7. Contacts and citation

Contacts:

- Lieven Clarisse: [Lieven.Clarisse@ulb.be](mailto:Lieven.Clarisse@ulb.be)
- Martin Van Damme: [Martin.Van.Damme@ulb.be](mailto:Martin.Van.Damme@ulb.be)





- Pierre-François Coheur: [Pierre.Coheur@ulb.be](mailto:Pierre.Coheur@ulb.be)

#### Dataset citation:

- Clarisse, L., Van Damme M., Coheur P. (2024), IASI L3 NH<sub>3</sub> ANNI v4 Climate Data Record, ESA Precursors CCI+, <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>

#### Paper citation:

- Clarisse, L., Franco, B., Van Damme, M., Di Gioacchino, T., Hadji-Lazaro, J., Whitburn, S., Noppen, L., Hurtmans, D., Clerbaux, C., and Coheur, P.: The IASI NH<sub>3</sub> version 4 product: averaging kernels and improved consistency, *Atmos. Meas. Tech.*, 16, 5009–5028, <https://doi.org/10.5194/amt-16-5009-2023>, 2023.

## 4. CO L3 product from IASI and MOPITT

### 4.1. Input data and algorithm

Our objective is to generate a homogenized and merged CO L3 CDR dataset. Satellite data from two instruments will be used as input: IASI and MOPITT. The IASI CO L2 (DOI:10.15770/EUM\_SAF\_AC\_0047) data consist in a reprocessed dataset, performed by EUMETSAT in 2021. MOPITT L2 data consist in the latest version available: v9T.

At the time of writing this document, CO L3 IASI-MOPITT merged product has not been produced yet. During the first cycle of the project, we produced an intermediate product, that we call “IASI multi-platform product”, that will not be distributed to any users. To generate this intermediate product, we post-processed IASI CO L2 data (in two steps: outliers filtering+reformatting and then cloud filtering) and averaged Metop-A, B and C data in order to get a new IASI CO L3 product: multi-platform monthly grids (L3). This new intermediate dataset is currently being validated by BIRA. This new IASI multi-platform dataset and MOPITT V9T data will then be combined, using corrective factors, in order to build a homogenized and merged CO L3 CDR dataset. It will consist in  $1^{\circ} \times 1^{\circ}$  monthly grids. The different steps are summarised in Figure 4.1.

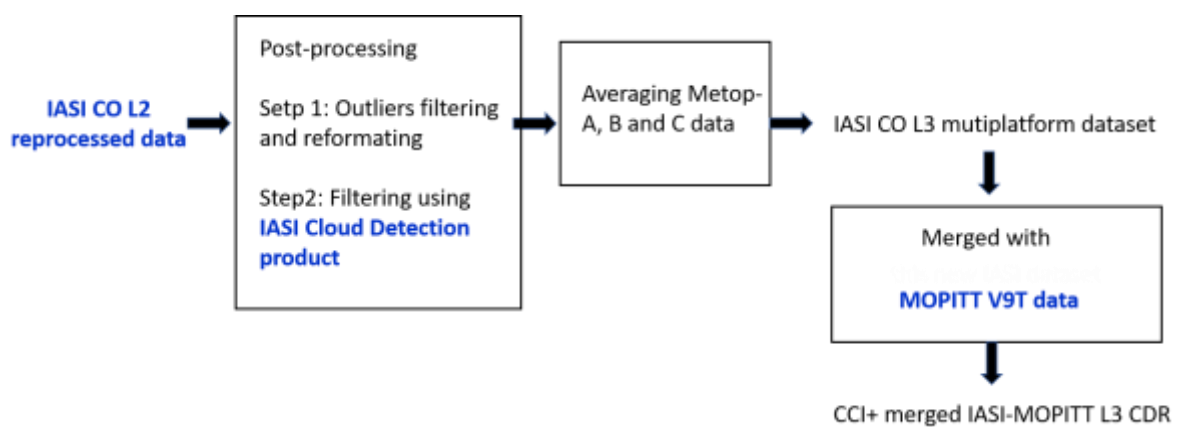


Figure 4.1: Graphic representation of the processing steps of the generation of the CCI+ IASI-MOPITT CO L3 merged product.

For this first version of PUG document, we will refer to the IASI CO L3 multiplatform product, as the file content should be identical or very similar. This will be completed in the second/final version of the document.

Information relative to IASI data can be found in the ATBD: [https://acsaf.org/docs/atbd/Algorithm\\_Theoretical\\_Basis\\_Document\\_IASI\\_CO\\_Mar\\_2023.pdf](https://acsaf.org/docs/atbd/Algorithm_Theoretical_Basis_Document_IASI_CO_Mar_2023.pdf). The related validation report will be available soon on the AC SAF webpage (to be updated in the next version of the document).

Information relative to MOPITT data can be found in the User Guide:

[https://www2.acom.ucar.edu/sites/default/files/documents/v9\\_users\\_guide\\_20220203.pdf](https://www2.acom.ucar.edu/sites/default/files/documents/v9_users_guide_20220203.pdf)

The related validation paper is: <https://amt.copernicus.org/preprints/amt-2021-370/amt-2021-370.pdf>

## 4.2. File parameters

Table 4.1 describes the variables to be contained in the L3 CO IASI-MOPITT netcdf files. The fill value is NaN for all.

Table 4.1: Variables to be in the L3 CO IASI-MOPITT NetCDF files. The dimensions N<sub>lat</sub> (180) and N<sub>lon</sub> (360) are the number of grid cells corresponding to the latitude and longitude respectively.

Parameter	Unit	Dimension	Description
latitude	degrees	N <sub>lat</sub>	latitude
longitude	degrees	N <sub>lon</sub>	longitude
co_total_column_day	mol.m <sup>-2</sup>	N <sub>lon</sub> × N <sub>lat</sub>	weighted average of the total carbon monoxide columns for daytime observations*
co_total_column_night	mol.m <sup>-2</sup>	N <sub>lon</sub> × N <sub>lat</sub>	weighted average of the total carbon monoxide columns for nighttime observations*
co_total_column_error_day	mol.m <sup>-2</sup>	N <sub>lon</sub> × N <sub>lat</sub>	error in the weighted average of the total carbon monoxide columns for daytime observations*
co_total_column_error_night	mol.m <sup>-2</sup>	N <sub>lon</sub> × N <sub>lat</sub>	error in the weighted average of the total carbon monoxide columns for nighttime observations*
number_of_observations_day	none	N <sub>lon</sub> × N <sub>lat</sub>	number of observations used for calculating the L3 (day)*
number_of_observations_night	none	N <sub>lon</sub> × N <sub>lat</sub>	number of observations used for calculating the L3 (night)*

At the time of writing this document (May 2024), the formula to calculate the merged IASI-MOPITT total columns is not finalized yet. The dataset is not produced yet. All the parameters in the table are subject to be modified.

## 4.3. Data access and volume

The CO L3 product will be accessed from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> following the filename convention:

CCI\_CO\_L3\_IASI\_MOPITT\_merged\_YYYYMM\_LATMOS-ULB\_V1.nc

where YYYYMM is the date of retrieval. The content of the netCDF files is listed in Tables 4.1.

Data volume is estimated as follows:

- One monthly file: 1,6 MB
- One year (12 files): 19 MB
- 2008-2022 period (15 years x 12 files = 180 files): 274 MB

## 4.4. Data reading examples

Here is an example of a python code to read one netcdf file and plot one CO monthly grid.

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

import numpy as np
from scipy.constants import Avogadro
import matplotlib.pyplot as plt
import cartopy.crs as ccrs
import cartopy.feature as cfeature
from netCDF4 import Dataset

# read CO grid for daytime from netcdf
nc = Dataset(YOUR_PATH/YYYY/CCI_CO_L3_IASI_MOPITT_merged_YYYYMM_LATMOS-ULB_V1.nc')
CO_grid_day = nc['/co_total_column_day'][:]
nc.close()

# convert grid (in mol.m-2) in molec./cm-2
CO_grid_day = CO_grid_day*Avogadro*1e-4

# transpose grid
CO_grid_day_t = CO_grid_day.transpose()

# create new figure
fig = plt.figure(figsize=(10, 5))

# specify projection
ax = plt.axes(projection=ccrs.PlateCarree())

# plot CO grid with pcolormesh
pcm = ax.pcolormesh(np.linspace(-180, 180, 361), np.linspace(-90, 90, 181),
                   CO_grid_day_t, cmap='jet', transform=ccrs.PlateCarree(), vmin=0, vmax=4E18)

#add coastlines
ax.add_feature(cfeature.BORDERS, linewidth=0.5)
ax.coastlines(linewidth=0.5)

# add colorbar
plt.colorbar(pcm, ax=ax, orientation='vertical', pad=0.1)

# add latitude/longitude gridflines
ax.gridlines(draw_labels=True, linewidth=0.5, color='gray', alpha=0.5, linestyle='--')

# define limits
ax.set_extent([-180, 180, -90, 90], crs=ccrs.PlateCarree())

# add grey background
ax.set_facecolor('lightgrey')

plt.show()
```

## 4.5. Usage recommendations and known limitations

To be completed in the next version of the document as the product is not produced yet.





## 4.6. Preliminary evaluation

To be completed in the next version of the document as the product is not produced yet

## 4.7. Contacts and citation

### Contact:

- Maya George: [maya.george@latmos.ipsl.fr](mailto:maya.george@latmos.ipsl.fr)
- Cathy Clerbaux: [cathy.clerbaux@latmos.ipsl.fr](mailto:cathy.clerbaux@latmos.ipsl.fr)

### Citation:

TBD



## 5. NO<sub>2</sub> L3 from GOME, SCIAMACHY, GOME2-A/B/C, OMI and TROPOMI

### 5.1. Input data and algorithm

NO<sub>2</sub> can be retrieved from a series of nadir UV-VIS sensors. We focus on the following sensors: GOME, SCIAMACHY, GOME2-A/B/C, OMI, TROPOMI.

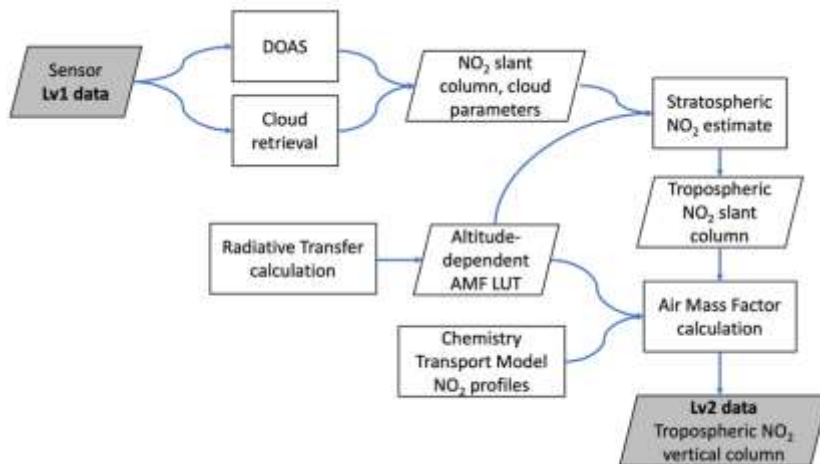
The NO<sub>2</sub> retrieval procedure consists of a procedure with three steps, and is performed for each level-1 spectrum:

- (1) Retrieval of the total NO<sub>2</sub> slant column density ( $N_s$ ) from the level-1 radiance and irradiance spectra measured by the sensor with the Differential Optical Absorption Spectroscopy (DOAS) method,
- (2) Separation of the  $N_s$  into a stratospheric ( $N_s^{strat}$ ) and tropospheric part ( $N_s^{trop} = N_s - N_s^{strat}$ ) based on information from a system incorporating slant columns and modelled or climatological information on tropospheric contributions,
- (3) Conversion of the tropospheric slant column density into a vertical column density ( $N_v^{trop}$ ) via a tropospheric air mass factor ( $M^{trop}$ ).

The central retrieval equation can thus be written as:

$$N_v^{trop} = \frac{N_s - N_s^{strat}}{M^{trop}} \tag{5-1}$$

In Eq. 5-1 the factor  $M^{trop}$  quantifies how the tropospheric slant column has been influenced by scattering, absorption, and reflection in the troposphere or by the Earth's surface.  $M^{trop}$  depends on the satellite viewing geometry, the presence of clouds or aerosols, surface properties, and the vertical distribution of NO<sub>2</sub> in the troposphere. The retrieval procedure therefore requires accurate, pixel-specific information on these geophysical parameters, and the NO<sub>2</sub> level-2 algorithms obtain this from the satellite measurements themselves, or from well-tested ancillary datasets.



**Figure 5-1: General overview of the processing chain for the retrieval of the level-2 tropospheric NO<sub>2</sub> columns.**

The L2 versions used to produce the L3 CDR are summarized in Table 2-1. The GOME-2-A, B, C data are directly taken from the AC SAF prototype product, developed in parallel of this project. For the other sensors, the slant columns are directly taken from the operational product (TROPOMI; van Geffen et al. (2022)), or from the QA4ECV products (OMI, SCIAMACHY, GOME; Boersma et al., 2018). The air mass factors, and stratospheric correction are calculated with consistent and updated settings. For further details on the L2 algorithm settings selected for this project, the reader is invited to refer to the NO<sub>2</sub> ATBD available at <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>.

The L3 products are directly derived from the L2 products. The L3 product consists of NO<sub>2</sub> tropospheric columns and associated uncertainties, detailed results, and input data. L3 are available for each instrument separately.

Two merged CDRs will also be provided, separately for morning and afternoon observations.

- Morning sensors: GOME/ERS-2, SCIAMACHY/ENVISAT, GOME2/METOP-A, B, C
- Afternoon sensors: OMI/AURA, TROPOMI/S5P

Since the spatial resolution and signal-to-noise ratio of the sensors are different, and significantly better for the afternoon sensors, different spatial resolutions are used. This is motivated by different user needs and applications, and various spatial resolutions can be generated. Table 5-1 provides the spatial and temporal resolution (preliminary) chosen for the morning and afternoon CDRs.

**Table 5-1: spatial resolution of the L3 NO<sub>2</sub> products for morning and afternoon sensors, at daily and monthly time resolution.**

Spatial resolution	Daily L3	Monthly L3
Morning CDR		0.5°/0.2°
Afternoon CDR	0.5°/0.2°	1.0°/0.5°/0.2°

The spatial resolution for daily L3 files needs to be coarser than for the monthly data, so that spatial averaging and gridding does not introduce major issues of spatial misallocation. KNMI (superobservations) and DLR (HARP) use similar tools for L3 gridding and averaging, but there are differences with respect to L3 uncertainty propagation and temporal weighting of daily (gridded) satellite data that make up the monthly mean (Rijsdijk et al. [2024]; E3UB-doc, *in prep.* [2024]).

## 5.2. File parameters

Table 5-2 describes the variables contained in the L3 NO<sub>2</sub> netcdf files. The fill value is NaN for all.

**Table 5-2: Variables in the L3 NO<sub>2</sub> NetCDF files.**

Parameter	Unit	Dimension	Description
latitude	degrees	lat	latitude



longitude	degrees	lon	longitude
latitude_bounds	degrees	lat, independent_2	grid latitude bounds
longitude_bounds	degrees	lon, independent_2	grid longitude bounds
time	days	time=1	effective day-of-month (number between 1-31)
effective_fractional_day	1	time, lat, lon	effective fractional day (number between 0-1)
layer	1	layer	layer dimension index
tm5_sigma_a	hPa	layer	pressure = tm5_sigma_a + surface_pressure * tm5_sigma_b
tm5_sigma_b	1		
surface_pressure	hPa	time, lat, lon	surface pressure
surface_albedo	1	time, lat, lon	surface LER (440 nm)
land_water_mask	1	time, lat, lon	0: water, 1: land, 2: land water transition
qa_L3	1	time, lat, lon	gridded data quality assurance value (0: not valid, 1: valid)
tropospheric_NO2_column_number_density	molec.c m <sup>-2</sup>	time, lat, lon	<b>NO<sub>2</sub> tropospheric vertical column number density</b>
tropospheric_NO2_column_number_density_temporal_std	molec.c m <sup>-2</sup>	time, lat, lon	Uncertainty on the NO <sub>2</sub> tropospheric vertical column number density associated with temporal standard deviation of L2 input data ( $\sigma_1$ ) within the cell
tropospheric_NO2_column_number_density_uncertainty	molec.c m <sup>-2</sup>	time, lat, lon	Uncertainty on the NO <sub>2</sub> tropospheric vertical column number density associated with time-averaged propagated uncertainty of L2 input data ( $\sigma_2$ )
tropospheric_NO2_column_number_density_uncertainty_kernel	molec.c m <sup>-2</sup>	time, lat, lon	Uncertainty on the NO <sub>2</sub> tropospheric vertical column number density associated with time-averaged propagated uncertainty of L2 input data, without the profile uncertainty contribution ( $\sigma_3$ )
tropospheric_NO2_column_number_density_uncertainty_random	molec.c m <sup>-2</sup>	time, lat, lon	Random uncertainty on the NO <sub>2</sub> tropospheric vertical column number density (HARP)
tropospheric_NO2_column_number_density_uncertainty_systematic	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the NO <sub>2</sub> tropospheric vertical column number density (HARP)
tropospheric_NO2_column_number_density_count	1	time, lat, lon	Effective number of observations per cell / fractional coverage $\alpha$ .
NO2_slant_column_number_density	molec.c m <sup>-2</sup>	time, lat, lon	NO <sub>2</sub> slant column number density
NO2_slant_column_number_density_uncertainty	molec.c m <sup>-2</sup>	time, lat, lon	NO <sub>2</sub> slant column number density uncertainty



NO2_slant_column_number_density_troposphere	molec.c m <sup>-2</sup>	time, lat, lon	Tropospheric NO2 slant column number density
NO2_stratospheric_column_number_density	molec.c m <sup>-2</sup>	time, lat, lon	Stratospheric NO2 vertical column density number density
tropospheric_NO2_column_number_density_amf	1	time, lat, lon	tropospheric air mass factor
cloud_fraction	1	time, lat, lon	effective cloud fraction
cloud_pressure	hPa	time, lat, lon	Effective cloud pressure (at optical centroid)
NO2_averaging_kernel	1	time, lat, lon, layer	Column averaging kernel

### 5.3. Data access and volume

The L3 product can be accessed from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> following the filename convention:

CCI+p-L3-NO2\_TC-instrument\_satellite\_L2version-KNMI-YYYYMM(DD)-L3version.nc

Example of file name for one TROPOMI monthly L3 file:

CCI+p-L3-NO2\_TC-TROPOMI\_S5P\_PAL\_v020301-KNMI-202310-fv0200.nc

The content of the netCDF files is listed in Table 5-3.

Data volume is as follows:

**Table 5-3: Size of one L3 file for morning and afternoon sensors, at daily and monthly time resolution.**

Spatial resolution	Daily L3	Monthly L3
Morning CDR		100Mb
Afternoon CDR	100Mb	2Gb

### 5.4. Data reading examples

```
import numpy as np
from netCDF4 import Dataset
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
from matplotlib import cm
from mpl_toolkits.axes_grid1 import make_axes_locatable

# Full file path of the NetCDF file
filename='CCI+p-L3-NO2_TC-TROPOMI_S5P_PAL_v020301-KNMI-202310-fv0200.nc'

# Map settings
proj_name='cyl'
Mspace=40
Pspace=30
parallels = np.arange(-90,90,Pspace)
meridians = np.arange(-180,180,Mspace)
LW=0.4
size_label=10

# Colorbar settings
ncolor=100
colormap = 'turbo'
scFactor=1e15
minScale = 0
maxScale = 10

# Variable to map and title
var2plot = "tropospheric_NO2_column_number_density"
title txt="Oct. 2023"
```



```
# Read the NetCDF file
with Dataset(filename, "r") as fid:
    NO2Map = fid[var2plot][:]
    NO2Map = NO2Map/scFactor
    varunit = fid[var2plot].units
    scaletitle='TROPOMI {} [{} {}]' .format(fid[var2plot].long_name,format(scFactor,'1.0e'),varunit)
    qa_Map = fid["qa_L3"][:]
    NO2Map[qa_Map<1]=np.nan

# Create the map
plt.figure(1,figsize=(12,12))
plt.title(scaletitle, loc='left',fontname="Times New Roman",fontweight="bold")
plt.annotate("ESA CCI+ ozone and aerosol precursors", xy=(0.01,0.015), xycoords='axes fraction',size=10,fontname="Times New Roman",fontweight="bold")
plt.title(title_txt, fontdict=None, loc='right',fontname="Times New Roman",fontweight="bold")

map = Basemap(projection=proj_name, llcrnrlon=-180, llcrnrlat=-90, urcrnrlon=180,
urcrnrlat=90,resolution='1',lat_0=0,lon_0=0)
map.drawcoastlines(linewidth=LW,color='grey')
map.drawcountries(linewidth=LW,color='whitesmoke')
map.fillcontinents(color='lightgrey',zorder=0)
map.drawmeridians(meridians,color='black',linewidth=LW/2)
map.drawparallels(parallels,color='black',linewidth=LW/2)

im = map.imshow(HCHO2Map[0,:,:], cmap=cm.get_cmap(colormap,ncolor),interpolation='none')
plt.colorbar(im, orientation="horizontal")
plt.clim(minScale, maxScale)

plt.show()
```

## 5.5. Usage recommendations and known limitations

- Users interested in long-term (10+ years) trends in NO<sub>2</sub> should use the merged L3 dataset. Users interested in short time periods are advised to use the L3 from each instrument.
- The main product is the **“tropospheric\_NO2\_column\_number\_density”**.
- A quality flag is provided for the L3 file, with the aim to filter cells containing not enough observations. The criterion for sufficient coverage is that the fractional coverage of a grid cell should exceed 0.1, which is equivalent to more than 3 days (10% of 30 or 31 days) with a fully covered grid cell to calculate a monthly mean. In coding, if tropospheric\_NO2\_column\_number\_density\_count ≥ 0.1, then qa\_L3=1. If tropospheric\_NO2\_column\_number\_density\_count < 0.1, then qa\_L3=0. Users are advised to use only qa\_L3==1.
- Various uncertainty estimates for NO<sub>2</sub> monthly mean L3 data are available for each grid cell. The first uncertainty estimate (σ<sub>1</sub>) represents the temporal standard deviation of the gridded L2 data within the cell and is a measure of the temporal variability of NO<sub>2</sub>. The second estimate (σ<sub>2</sub>) represents time-averaged propagated uncertainty of L2 input data (σ<sub>2</sub>) in the cell, and explicitly accounts for random as and systematic contributions to the total uncertainty.
- Additionally, estimates of average ‘random’ and ‘systematic’ uncertainties are provided for each grid cell.
- When averaging L3 data over larger temporal or spatial scales, an averaged uncertainty can be obtained from combining the different random (σ<sub>ri</sub>) and systematic (σ<sub>si</sub>) uncertainties as
  - $$\sigma^2 = \sum_{i=1}^n \left(\frac{\sigma_{ri}}{n}\right)^2 + \left(\sum_{i=1}^n \frac{\sigma_{si}}{n}\right)^2$$
- Mean averaging kernels and a priori vertical profiles are provided for each grid cell. This is only an approximation of these quantities with a vertical dimension. L2 data should preferably be used for an accurate exploitation of the averaging kernels.

## 5.6. Preliminary evaluation

Figure 5-1 below shows the monthly mean tropospheric NO<sub>2</sub> column at a resolution of 0.2° for January 2019. A complete L3 dataset for monthly mean tropospheric NO<sub>2</sub> columns has been generated for the entire period covered by TROPOMI v2.2 data, i.e. from May 2018 – December 2021. The evaluation will be completed in the next iteration of this document.



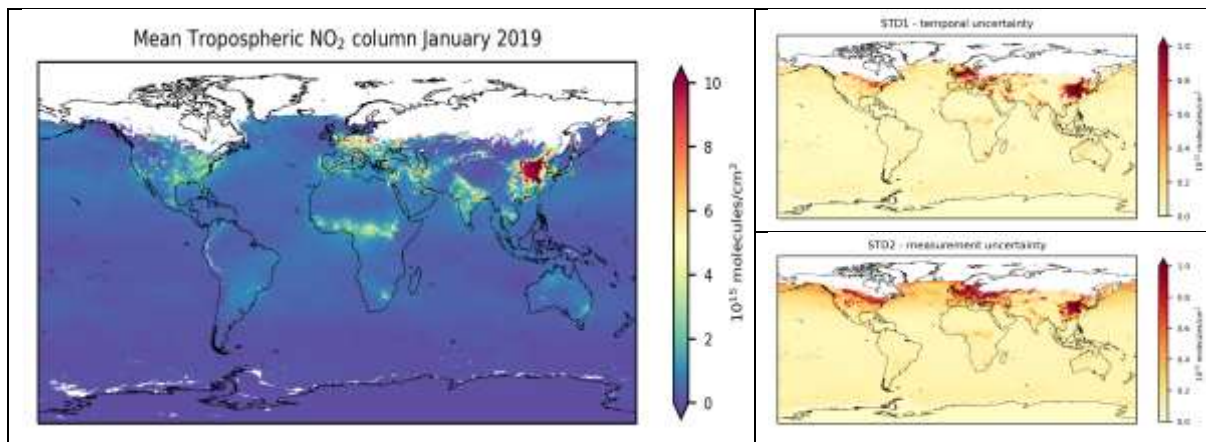


Figure 5-1: example of monthly mean TROPOMI tropospheric NO<sub>2</sub> columns (left) calculated with the superobservations approach at 0.2°×0.2°. The right panels show the uncertainty estimates associated with the monthly mean tropospheric columns. The upper right showing the (weighted) temporal standard deviation, a measure for the variability of NO<sub>2</sub> in time. The lower right panel shows the best estimate of how L2 uncertainties propagate in time to give a theoretical L3 uncertainty estimate.

## 5.7. Contacts and citation

### Contact:

- Isolde Glissenaar: [isolde.glissenaar@knmi.nl](mailto:isolde.glissenaar@knmi.nl)
- Isidora Anglou : [isidora.anglou@knmi.nl](mailto:isidora.anglou@knmi.nl)
- Folkert Boersma: [folkert.boersma@knmi.nl](mailto:folkert.boersma@knmi.nl)
- Henk Eskes: [henk.eskes@knmi.nl](mailto:henk.eskes@knmi.nl)

### Dataset citation:

- Glissenaar, I. A., Anglou, I., Boersma, K. F., & Eskes, H. (2024). ESA CCI+ TROPOMI L3 monthly mean NO<sub>2</sub> columns [Data set]. Royal Netherlands Meteorological Institute (KNMI). <https://doi.org/10.21944/CCI-NO2-TROPOMI-L3>

### Dataset availability:

- <https://www.temis.nl/airpollution/no2col/cci-no2-tropomi.php>

### Paper citations:

#### TROPOMI NO<sub>2</sub>

van Geffen, J., Eskes, H., Compornolle, 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<sub>2</sub> retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, *Atmos. Meas. Tech.*, 15, 2037–2060, <https://doi.org/10.5194/amt-15-2037-2022>, 2022.

#### OMI, SCIAMACHY, and GOME NO<sub>2</sub>

Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendaal, M., Wagner, T., Maasackers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C., and Compornolle, S. C.: Improving algorithms and uncertainty estimates for satellite NO<sub>2</sub> retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, *Atmos. Meas. Tech.*, 11, 6651–6678, <https://doi.org/10.5194/amt-11-6651-2018>, 2018.



Title: D4.2 Product User Guide (PUG)  
Issue 01 - Revision 01 - Status: Final  
Date of issue: 29/05/2024  
Ref.: Precursors\_cci+\_D4.2\_PUG\_01\_01

---





## 6. HCHO L3 from TROPOMI, OMI, GOME2, SCIAMACHY and GOME

### 6.1. Input data and algorithm

HCHO can be retrieved from a series of nadir UV-VIS sensors. We focus more particularly on the ESA and ESA Third Party Mission (ESA-TPM) sensors: GOME, SCIAMACHY, GOME-2, OMI, TROPOMI.

Figure 6-1 shows the flow diagram of the L2 HCHO DOAS retrieval algorithm, common to all UV-VIS sensors. The HCHO slant column ( $N_s$ ) fit is performed in the UV. Subsequently, to convert the slant column to a vertical column ( $N_v$ ), an air mass factor that accounts for the average light path through the atmosphere is calculated. For this purpose, several auxiliary data are read from external sources: cloud cover data, topographic information, surface albedo, and the a priori shape of the vertical HCHO profile in the atmosphere. Two types of air mass factors are calculated, including a correction for the cloud altitude, or not ( $M$  or  $M_{clear}$ ). Additional background normalization is required in the case of weak absorbers such as formaldehyde to compensate for unknown offsets ( $N_{s,0}$ ). Finally, a background vertical column derived from a CTM is added ( $N_{v,0}$ ). The HCHO vertical column ( $N_v$  or  $N_{v,clear}$ ) is obtained using the following equation:

$$N_v = \frac{N_s - \overline{N_{s,0}}}{M} + \overline{N_{v,0}} \quad 6-1$$

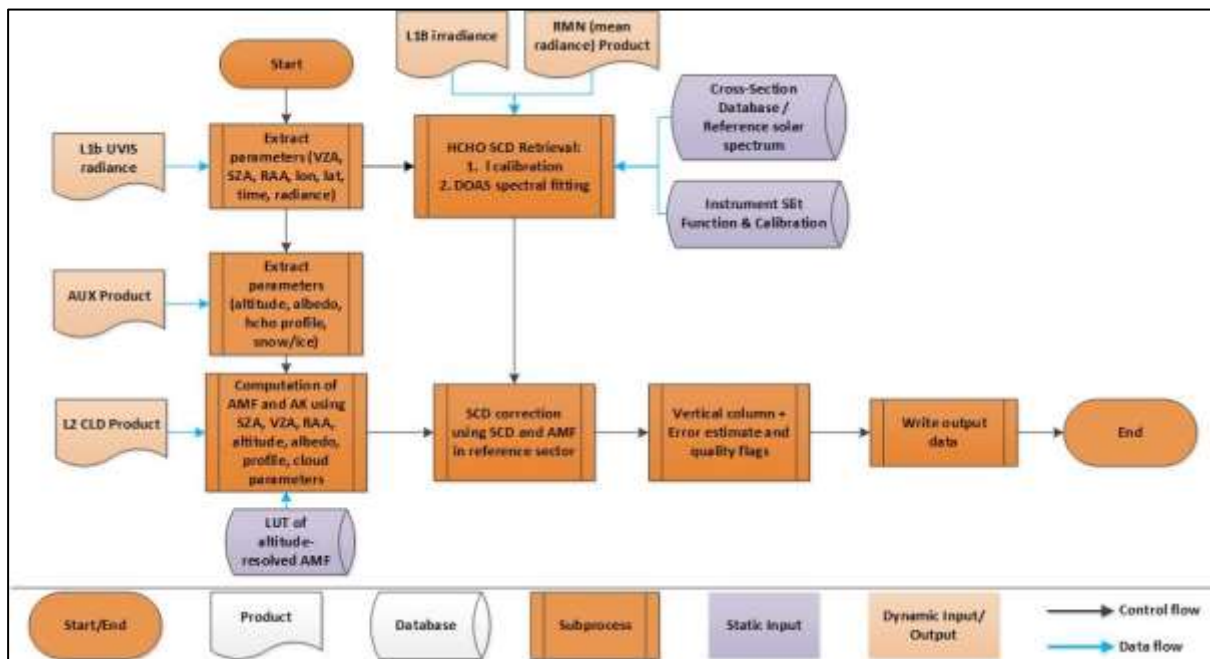


Figure 6-1: Flow Diagram of the DOAS retrieval algorithm for HCHO.

The L2 versions used to produce the L3 CDR are summarized in Table 2-1. The GOME-2-A, B, C data are directly taken from the AC SAF prototype product, developed in parallel of this

project. For the other sensors, the slant columns are directly taken from the operational product (TROPOMI), or from the QA4ECV products (OMI, SCIAMACHY, GOME). The air mass factors and background correction are re-calculated with consistent and updated settings. For further details on the L2 algorithm settings selected for this project, the reader is invited to refer to the HCHO ATBD available at <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>.

The L3 products are directly derived from the L2 products. The L3 product consists of HCHO tropospheric columns and associated random and systematic uncertainties, detailed results, and input data. L3 are available for each instrument separately.

Two merged CDRs will also be provided, separately for morning and afternoon observations.

- Morning sensors: GOME/ERS-2, SCIAMACHY/ENVISAT, GOME2/METOP-A, B, C
- Afternoon sensors: OMI/AURA, TROPOMI/S5P

Since the spatial resolution and signal-to-noise ratio of the sensors are different, and significantly better for the afternoon sensors, daily and monthly L3 files are provided for the afternoon sensors, while only monthly L3 files are provided for the morning sensors.

## 6.2. File parameters

Table 6- describe the variables contained in the L3 HCHO netcdf files. The fill value is NaN for all.

**Table 6-1: Variables in the L3 HCHO NetCDF files.**

Parameter	Unit	Dimension	Description
latitude	degrees	lat	latitude
longitude	degrees	lon	longitude
latitude_bounds	degrees	lat, independent_2	grid latitude bounds
longitude_bounds	degrees	lon, independent_2	grid longitude bounds
time	days	time=1	number of days since 1995-01-01
delta_time	seconds	time, lat, lon	seconds since time
layer	1	layer	layer dimension index
ctm_sigma_a	hPa	layer	pressure = ctm_sigma_a + surface_pressure * ctm_sigma_b
ctm_sigma_b	1		
surface_pressure	hPa	time, lat, lon	surface pressure
surface_albedo	1	time, lat, lon	surface LER (340 nm)
land_water_mask	1	time, lat, lon	0: water, 1: land, 2: land water transition
qa_L3	1	time, lat, lon	gridded data quality assurance value (0: not valid, 1: valid)



tropospheric_HCHO_column_number_density_clear	1	time, lat, lon	Background-corrected HCHO tropospheric vertical column number density, no cloud correction
tropospheric_HCHO_column_number_density	1	time, lat, lon	Background-corrected HCHO tropospheric vertical column number density
tropospheric_HCHO_column_number_density_uncertainty_random	molec.c m <sup>-2</sup>	time, lat, lon	Random uncertainty on the background-corrected HCHO tropospheric vertical column number density
tropospheric_HCHO_column_number_density_uncertainty_systematic	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the background-corrected HCHO tropospheric vertical column number density
tropospheric_HCHO_column_number_density_uncertainty_systematic_kernel	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the background-corrected HCHO tropospheric vertical column number density, without the profile uncertainty contribution
tropospheric_HCHO_column_number_density_clear_std	molec.c m <sup>-2</sup>	time, lat, lon	Standard deviation of the background-corrected HCHO tropospheric vertical column number density without cloud correction, within the cell
tropospheric_HCHO_column_number_density_std	molec.c m <sup>-2</sup>	time, lat, lon	Standard deviation of the background-corrected HCHO tropospheric vertical column number density within the cell
tropospheric_HCHO_column_number_density_count	1	time, lat, lon	Effective number of observations per cell
HCHO_slant_column_number_density_rms	1	time, lat, lon	DOAS rms
HCHO_slant_column_number_density	molec.c m <sup>-2</sup>	time, lat, lon	HCHO slant column number density
HCHO_slant_column_number_density_uncertainty	molec.c m <sup>-2</sup>	time, lat, lon	HCHO slant column number density uncertainty
HCHO_slant_column_number_density_corrected	molec.c m <sup>-2</sup>	time, lat, lon	Background-corrected HCHO slant column number density uncertainty
tropospheric_HCHO_column_number_density_background_correction	molec.c m <sup>-2</sup>	time, lat, lon	a priori HCHO vertical column density number density in the Pacific reference sector
tropospheric_HCHO_column_number_density_amf	1	time, lat, lon	tropospheric air mass factor
tropospheric_HCHO_column_number_density_amf_clear	1	time, lat, lon	tropospheric air mass factor, no cloud correction
cloud_fraction	1	time, lat, lon	effective cloud fraction
cloud_pressure	hPa	time, lat, lon	cloud top pressure
tropospheric_HCHO_column_number_density_apriori	molec.c m <sup>-2</sup>	time, lat, lon	a priori HCHO vertical column number density (integrated profile)
HCHO_volume_mixing_ratio_dry_air_apriori	ppv	time, lat, lon, layer	a priori HCHO vertical profile
HCHO_averaging_kernel	1	time, lat, lon, layer	Column averaging kernel



### 6.3. Data access and volume

The L3 product can be accessed from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> following the filename convention:

CCI+p-L3-HCHO\_TC-instrument\_satellite\_L2version-BIRA-YYYYMM(DD)-L3version.nc

Example of file name for one TROPOMI monthly L3 file:

MPC-L3-HCHO\_TC-TROPOMI\_S5P\_OFFL\_v0302-BIRA-202310-fv0200.nc

The content of the netCDF files is listed in Table 6-.

Data volume is as follows:

**Table 6-2: Size of one L3 file for morning and afternoon sensors, at daily and monthly time resolution.**

Spatial resolution	Daily L3	Monthly L3
Morning CDR		1000Mb
Afternoon CDR	250Mb	1000Mb

### 6.4. Data reading examples

```
import numpy as np
from netCDF4 import Dataset
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
from matplotlib import cm
from mpl_toolkits.axes_grid1 import make_axes_locatable

# Full file path of the NetCDF file
filename='MPC-L3-HCHO_TC-TROPOMI_S5P_OFFL_v0302-BIRA-202310-fv0200.nc'

# Map settings
proj_name='cyl'
Mspace=40
Pspace=30
parallels = np.arange(-90,90,Pspace)
meridians = np.arange(-180,180,Mspace)
LW=0.4
size_label=10

# Colorbar settings
ncolor=100
colormap = 'turbo'
scFactor=1e15
minScale = 0
maxScale = 20

# Variable to map and title
var2plot = "tropospheric_HCHO_column_number_density_clear"
title_txt="Oct. 2023"

# Read the NetCDF file
with Dataset(filename, "r") as fid:
    HCHO2Map = fid[var2plot][:]
    HCHO2Map = HCHO2Map/scFactor
    varunit = fid[var2plot].units
    scaletitle='TROPOMI {} [{} {}]' .format(fid[var2plot].long_name,format(scFactor,'1.0e'),varunit)
    qa_Map = fid["qa_L3"][:]
    HCHO2Map[qa_Map<1]=np.nan

# Create the map
plt.figure(1,figsize=(12,12))
plt.title(scaletitle, loc='left',fontname="Times New Roman",fontweight="bold")
plt.annotate("ESA CCI+ ozone and aerosol precursors", xy=(0.01,0.015), xycoords='axes fraction',size=10,fontname="Times New Roman",fontweight="bold")
plt.title(title_txt, fontdict=None, loc='right',fontname="Times New Roman",fontweight="bold")

map = Basemap(projection=proj_name, llcrnrlon=-180, llcrnrlat=-90, urcrnrlon=180, urcrnrlat=90,resolution='l',lat_0=0,lon_0=0)
map.drawcoastlines(linewidth=LW,color='grey')
map.drawcountries(linewidth=LW,color='whitesmoke')
```



```
map.fillcontinents(color='lightgrey',zorder=0)
map.drawmeridians(meridians,color='black',linewidth=LW/2)
map.drawparallels(parallels,color='black',linewidth=LW/2)

im = map.imshow(HCHO2Map[0,:,:], cmap=cm.get_cmap(colormap,ncolor),interpolation='none')
plt.colorbar(im,orientation="horizontal")
plt.clim(minScale,maxScale)

plt.show()
```

## 6.5. Usage recommendations and known limitations

- Users interested in long-term trends should use the merged L3 dataset. Users interested in short time periods are advised to use the L3 from each instrument.
- The main product is the “tropospheric\_HCHO\_column\_number\_density\_clear”.
- A quality flag is provided for the L3 file, with the aim to filter cells containing not enough observations. It is advised to use only qa\_L3==1. The qa\_L3 flag is based on the effective number of observations within the cell (tropospheric\_HCHO\_column\_number\_density\_count), after applying the L2 quality filtering. If the tropospheric\_HCHO\_column\_number\_density\_count is lower than the 5<sup>th</sup> percentile of the cell counts for the day/month, the qa\_L3 is set to 0.
- Mean random ( $\sigma_r$ ) and systematic ( $\sigma_s$ ) uncertainties are available for each grid cell. If a total uncertainty is desired, this can be obtained from the square root of the sum of squares of the different terms:

$$\sigma^2 = \sigma_r^2 + \sigma_s^2$$

- When averaging L3 data over larger temporal or spatial scales, an averaged uncertainty can be obtained from combine the different random ( $\sigma_{ri}$ ) and systematic ( $\sigma_{si}$ ) uncertainties as

$$\sigma^2 = \sum_{i=1}^n \left( \frac{\sigma_{ri}}{n} \right)^2 + \left( \sum_{i=1}^n \frac{\sigma_{si}}{n} \right)^2$$

- Mean averaging kernels and a priori vertical profiles are provided for each grid cell. This is only an approximation of these quantities with a vertical dimension. L2 data should preferably be used for an accurate exploitation of the averaging kernels.
- The example python script below uses the L3 AKs and ap Profiles to recalculate the AMFs, and compare the results with the provided L3 AMFs.

```
import numpy as np
from netCDF4 import Dataset
import matplotlib.pyplot as plt

# Initialisation
dryair = 28.9644;          #g/mole
ra = 6.0221367e23;      #molec/mole
g = 9.80665;            #m/s

# Full file path of the NetCDF file
filename='MPC-L3-HCHO_TC-TROPOMI_S5P_OFFL_v0302-BIRA_1D-20231031-fv0200.nc'

# Read the NetCDF file
with Dataset(filename, "r") as fid:
    sigma_a_Layer= fid["ctm_sigma_a"][:].flatten()
    sigma_b_Layer= fid["ctm_sigma_b"][:].flatten()
    nlayers = np.shape(sigma_a_Layer)[0]
    profile_Map = fid["HCHO_volume_mixing_ratio_dry_air_apriori"][0,...].flatten()
    profile_Map = np.reshape(profile_Map,[-1,nlayers])
    ak_Map = fid["HCHO_averaging_kernel"][0,...].flatten()
    ak_Map = np.reshape(ak_Map,[-1,nlayers])
    xa_Map = fid["tropospheric_HCHO_column_number_density_apriori"][0,...].flatten()
    sp_Map = fid["surface_pressure"][0,...].flatten()
    amf_Map = fid["tropospheric_HCHO_column_number_density_amf"][0,...].flatten()
    qa_Map = fid["qa_L3"][0,...].flatten()

total_size = np.shape(sp_Map)[0]

# Calculate the pressure at layer centers
pressure_Map = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))      # pressure at layer center [hPa]
for l in range(0,nlayers):
    pressure_Map[:,l] = (sigma_a_Layer[l] + sp_Map[:,l]*sigma_b_Layer[l])      # [hPa]

pi = np.full((total_size,nlayers+1),np.nan,dtype = np.dtype('f4'))      # pressure at layer borders [hPa]
pi[:,0]=sp_Map[:,l]      # surface pressure [hPa]
for l in range(0,nlayers-1):
    pi[:,l+1] = np.exp(np.log(pressure_Map[:,l])-(np.log(pressure_Map[:,l])-np.log(pressure_Map[:,l+1]))/2)
pi[:,nlayers]=np.exp( np.log(pi[:,nlayers-1])-2*( np.log(pi[:,nlayers-1])-np.log(pressure_Map[:,nlayers-1]) ) )

# Re-calculate the AMF using the a priori profile and the averaging kernel
dp = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))      # delta pressure per layer [hPa]
m_check = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))      # recalculate AMF
vcdal = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))      # a priori partial columns per layer [molec./cm2]
for l in range(0,nlayers):
    dp[:,l] = pi[:,l] - pi[:,l+1]
    vcdal[:,l] = 10*ra*np.multiply(profile_Map[:,l],dp[:,l])/(g*dryair)
```





```
## Check AMF
m_check[:,1] = ra*10*amf_Map[:]*np.multiply(ak_Map[:,1],np.multiply(profile_Map[:,1],dp[:,1]))/(g*dryair)
VCDA = np.nansum(vcdal,axis=1)
amf_check = np.nansum(m_check,axis=1)/VCDA

kok = (qa_Map>0)

# Plot the re-calculated AMF versus the L3 AMF
fig, ax =plt.subplots(1,1,figsize=(12,6))
plt.plot(amf_Map[kok],amf_check[kok], '.',c='black',markersize=0.2,fillstyle='full',linestyle='')
plt.grid(True)
plt.ylabel("Recalculated AMF using AK and ap Profile", fontsize=12, fontweight='bold', color='k')
plt.xlabel("L3 AMF", fontsize=12, fontweight='bold', color='k')
plt.title(filename, fontsize=10, fontweight='bold', color='k')
plt.xlim([0,3.5])
plt.ylim([0,3.5])
plt.show()
```

## 6.6. Preliminary evaluation

To be completed in the next version of the document.

## 6.7. Contacts and citation

To be completed in the next version of the document.

Contact:

- Isabelle De Smedt: [isabelle.desmedt@aeronomie.be](mailto:isabelle.desmedt@aeronomie.be)
- Jonas Vlietinck: [jonas.vlietienck@aeronomie.be](mailto:jonas.vlietienck@aeronomie.be)

Dataset citation:

- TBD

Paper citations:

TROPOMI HCHO (Operationnal version)

De Smedt, I., Pinardi, G., Vigouroux, C., Compernelle, 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., 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 and validation against MAX-DOAS network column measurements, *Atmos. Chem. Phys.*, 21, 12561–12593, <https://doi.org/10.5194/acp-21-12561-2021>, 2021.

OMI HCHO (QA4ECV version)

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernelle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedernana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veeffkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, *Atmos. Meas. Tech.*, 11, 2395–2426, <https://doi.org/10.5194/amt-11-2395-2018>, 2018.



Title: D4.2 Product User Guide (PUG)  
Issue 01 - Revision 01 - Status: Final  
Date of issue: 29/05/2024  
Ref.: Precursors\_cci+\_D4.2\_PUG\_01\_01

---





## 7. SO<sub>2</sub> L3 from TROPOMI, OMI, SCIAMACHY and GOME

### 7.1. Input data and algorithm

SO<sub>2</sub> can be retrieved from a series of nadir UV-VIS sensors. We focus more particularly on the following sensors: GOME, SCIAMACHY, OMI, TROPOMI.

Figure 6-1 shows the flow diagram of the L2 SO<sub>2</sub> COBRA vertical column retrieval algorithm, common to all UV-VIS sensors. First, the measured top-of-atmosphere radiances in the UV are analyzed and yield the slant columns (SCD). The SO<sub>2</sub> fit is performed using COBRA in multiple fitting windows, characterized by different SO<sub>2</sub> absorption strengths. Depending on the retrieved columns amounts in the different fitting windows, a selection of the final SCD is performed. Subsequently, to convert the slant column to a vertical column ( $VCD=SCD/AMF$ ), an air mass factor (AMF) that accounts for the average light path through the atmosphere is calculated. For this purpose, several auxiliary data are read from external sources: cloud cover data, topographic information, surface albedo, and the a priori shape of the vertical SO<sub>2</sub> profile in the atmosphere. Two types of air mass factors are calculated, one for an SO<sub>2</sub> profile from a chemical transport model and another one for box profiles with prescribed SO<sub>2</sub> plume heights, representative of volcanic scenarios. The AMF module also includes calculation of errors and column averaging kernels.

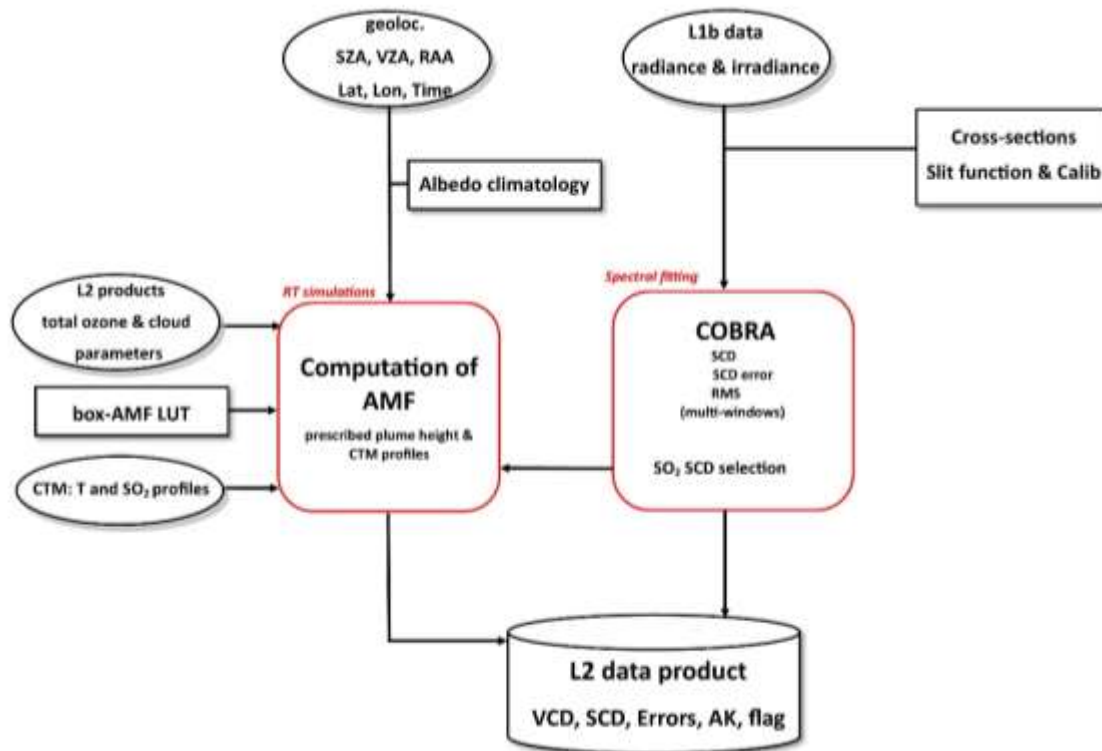


Figure 7-1: Flow diagram of the SO<sub>2</sub> column retrieval algorithm.



For further details on the L2 algorithm settings selected for this project, the reader is referred to the SO<sub>2</sub> ATBD available at <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>.

*NOTE: for SO<sub>2</sub>, the generation of L3 products has not started yet and all the following is subject to change. The text is a simple adaptation of section 6 on HCHO CDR.*

The L3 products are directly derived from the L2 products. The L3 product consists of SO<sub>2</sub> total columns for CTM and box profiles, and associated random and systematic uncertainties, detailed results, and input data. L3 are available for each instrument separately and a merged CDR will also be provided. The spatial and temporal resolution for the single-mission and merged CDR is tbd.

## 7.2. File parameters

Table 6- describe the variables contained in the L3 SO<sub>2</sub> netcdf files. The fill value is NaN for all.

**Table 7-1: Variables in the L3 SO<sub>2</sub> NetCDF files (preliminary).**

Parameter	Unit	Dimension	Description
latitude	degrees	lat	latitude
longitude	degrees	lon	longitude
latitude_bounds	degrees	lat, independent_2	grid latitude bounds
longitude_bounds	degrees	lon, independent_2	grid longitude bounds
time	days	time=1	number of days since 1995-01-01
delta_time	seconds	time, lat, lon	seconds since time
layer	1	layer	layer dimension index
ctm_sigma_a	hPa	layer	pressure = ctm_sigma_a + surface_pressure * ctm_sigma_b
ctm_sigma_b	1		
surface_pressure	hPa	time, lat, lon	surface pressure
surface_albedo	1	time, lat, lon	surface LER (340 nm)
land_water_mask	1	time, lat, lon	0: water, 1: land, 2: land water transition
qa_L3	1	time, lat, lon	gridded data quality assurance value (0: not valid, 1: valid)
total_SO2_column_number_density_clear	1	time, lat, lon	SO <sub>2</sub> total vertical column number density, no cloud correction
total_SO2_column_number_density	1	time, lat, lon	SO <sub>2</sub> total vertical column number density
total_SO2_column_number_density_uncertainty_random	molec.c m <sup>-2</sup>	time, lat, lon	Random uncertainty on the SO <sub>2</sub> total vertical column number density



total_SO2_column_number_density_uncertainty_systematic	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the SO2 total vertical column number density
total_SO2_column_number_density_uncertainty_systematic_kernel	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the SO2 total vertical column number density, without the profile uncertainty contribution
total_SO2_column_number_density_clear_std	molec.c m <sup>-2</sup>	time, lat, lon	Standard deviation of the SO2 total vertical column number density without cloud correction, within the cell
total_SO2_column_number_density_std	molec.c m <sup>-2</sup>	time, lat, lon	Standard deviation of the SO2 total vertical column number density within the cell
total_SO2_column_number_density_count	1	time, lat, lon	Effective number of observations per cell
SO2_slant_column_number_density_rms	1	time, lat, lon	COBRA rms
SO2_slant_column_number_density	molec.c m <sup>-2</sup>	time, lat, lon	SO2 slant column number density
SO2_slant_column_number_density_uncertainty	molec.c m <sup>-2</sup>	time, lat, lon	SO2 slant column number density uncertainty
total_SO2_column_number_density_amf	1	time, lat, lon	tropospheric air mass factor
total_SO2_column_number_density_amf_clear	1	time, lat, lon	tropospheric air mass factor, no cloud correction
cloud_fraction	1	time, lat, lon	effective cloud fraction
cloud_pressure	hPa	time, lat, lon	cloud top pressure
total_SO2_column_number_density_apriori	molec.c m <sup>-2</sup>	time, lat, lon	a priori SO2 vertical column number density (integrated profile)
SO2_volume_mixing_ratio_dry_air_apriori	ppv	time, lat, lon, layer	a priori SO2 vertical profile
SO2_averaging_kernel	1	time, lat, lon, layer	Column averaging kernel
total_SO2_column_number_density_box	1	time, lat, lon, 3	SO2 total vertical column number density, for box profiles at 1, 7, 15 km

### 7.3. Data access and volume

The L3 product can be accessed from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> following the filename convention:

CCI+p-L3-SO2\_TC-instrument\_satellite\_L2version-BIRA-YYYYMM(DD)-L3version.nc

Example of file name for one TROPOMI monthly L3 file:

MPC-L3-SO2\_TC-TROPOMI\_S5P\_OFFL\_v0302-BIRA-202310-fv0200.nc

The content of the netCDF files is listed in Table 6-.

Data volume is as follows: TBD



## 7.4. Data reading examples

```
import numpy as np
from netCDF4 import Dataset
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
from matplotlib import cm
from mpl_toolkits.axes_grid1 import make_axes_locatable

# Full file path of the NetCDF file
filename= 'MPC-L3-SO2_TC-TROPOMI_S5P_OFFL_v0302-BIRA-202310-fv0200.nc'

# Map settings
proj_name='cyl'
Mspace=40
Pspace=30
parallels = np.arange(-90,90,Pspace)
meridians = np.arange(-180,180,Mspace)
LW=0.4
size_label=10

# Colorbar settings
ncolor=100
colormap = 'turbo'
scFactor=1e15
minScale = 0
maxScale = 20

# Variable to map and title
var2plot = "total_SO2_column_number_density "
title_txt="Oct. 2023"

# Read the NetCDF file
with Dataset(filename, "r") as fid:
    SO22Map = fid[var2plot][:]
    SO22Map = SO22Map/scFactor
    varunit = fid[var2plot].units
    scaletitle='TROPOMI {} [{} {}]' .format(fid[var2plot].long_name, format(scFactor, '1.0e'), varunit)
    qa_Map = fid["qa_L3"][:]
    SO22Map[qa_Map<1]=np.nan

# Create the map
plt.figure(1,figsize=(12,12))
plt.title(scaletitle, loc='left',fontname="Times New Roman",fontweight="bold")
plt.annotate("ESA CCI+ ozone and aerosol precursors", xy=(0.01,0.015), xycoords='axes fraction',size=10,fontname="Times New Roman",fontweight="bold")
plt.title(title_txt, fontdict=None, loc='right',fontname="Times New Roman",fontweight="bold")

map = Basemap(projection=proj_name, llcrnrlon=-180, llcrnrlat=-90, urcrnrlon=180, urcrnrlat=90,resolution='l',lat_0=0,lon_0=0)
map.drawcoastlines(linewidth=LW,color='grey')
map.drawcountries(linewidth=LW,color='whitesmoke')
map.fillcontinents(color='lightgrey',zorder=0)
map.drawmeridians(meridians,color='black',linewidth=LW/2)
map.drawparallels(parallels,color='black',linewidth=LW/2)

im = map.imshow(SO22Map[0,:,:], cmap=cm.get_cmap(colormap, ncolor),interpolation='none')
plt.colorbar(im, orientation="horizontal")
plt.clim(minScale, maxScale)

plt.show()
```

## 7.5. Usage recommendations and known limitations

- Users interested in long-term trends should use the merged L3 dataset. Users interested in short time periods are advised to use the L3 from each instrument.
- The main product is the “total\_SO2\_column\_number\_density”. For volcanic applications, the product to be used is the “total\_SO2\_column\_number\_density\_box”.
- A quality flag for the main product “total\_SO2\_column\_number\_density” is provided for the L3 file, with the aim to filter cells containing not enough observations. It is advised to use only qa\_L3==1. The qa\_L3 flag is based on the effective number of observations within the cell (total\_SO2\_column\_number\_density\_count), after applying the L2 quality filtering. If the total\_SO2\_column\_number\_density\_count is lower than the 5<sup>th</sup> percentile of the cell counts for the day/month, the qa\_L3 is set to 0.
- Mean random ( $\sigma_r$ ) and systematic ( $\sigma_s$ ) uncertainties are available for each grid cell. If a total uncertainty is desired, this can be obtained from the square root of the sum of squares of the different terms:

$$\sigma^2 = \sigma_r^2 + \sigma_s^2$$



- When averaging L3 data over larger temporal or spatial scales, an averaged uncertainty can be obtained from combine the different random ( $\sigma_{ri}$ ) and systematic ( $\sigma_{si}$ ) uncertainties as

$$\sigma^2 = \sum_{i=1}^n \left(\frac{\sigma_{ri}}{n}\right)^2 + \left(\sum_{i=1}^n \frac{\sigma_{si}}{n}\right)^2$$

- Mean averaging kernels and a priori vertical profiles are provided for each grid cell. This applies only to the main product **“total\_SO2\_column\_number\_density”**. This is only an approximation of these quantities with a vertical dimension. L2 data should preferably be used for an accurate exploitation of the averaging kernels.
- The example python script below uses the L3 AKs and ap Profiles to recalculate the AMFs, and compare the results with the provided L3 AMFs.

```
import numpy as np
from netCDF4 import Dataset
import matplotlib.pyplot as plt

# Initialisation
dryair = 28.9644;           #g/mole
ra = 6.0221367e23;        #molec/mole
g = 9.80665;              #%/s

# Full file path of the NetCDF file
filename= 'MPC-L3-SO2_TC-TROPOMI_S5P_OFFL_v0302-BIRA-202310-fv0200.nc'

# Read the NetCDF file
with Dataset(filename, "r") as fid:
    sigma_a_Layer= fid["ctm_sigma_a"][:].flatten()
    sigma_b_Layer= fid["ctm_sigma_b"][:].flatten()
    nlayers = np.shape(sigma_a_Layer)[0]
    profile_Map = fid["SO2_volume_mixing_ratio_dry_air_apriori"][0,...].flatten()
    profile_Map = np.reshape(profile_Map,[-1,nlayers])
    ak_Map = fid["SO2_averaging_kernel"][0,...].flatten()
    ak_Map = np.reshape(ak_Map,[-1,nlayers])
    xa_Map = fid["total_SO2_column_number_density_apriori"][0,...].flatten()
    sp_Map = fid["surface_pressure"][0,...].flatten()
    amf_Map = fid["total_SO2_column_number_density_amf"][0,...].flatten()
    qa_Map = fid["qa_L3"][0,...].flatten()

total_size = np.shape(sp_Map)[0]

# Calculate the pressure at layer centers
pressure_Map = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))      # pressure at layer center [hPa]
for l in range(0,nlayers):
    pressure_Map[:,l] = (sigma_a_Layer[l] + sp_Map[:,l]*sigma_b_Layer[l])        # [hPa]

pi = np.full((total_size,nlayers+1),np.nan,dtype = np.dtype('f4'))            # pressure at layer borders [hPa]
pi[:,0]=sp_Map[:,l]                                                             # surface pressure [hPa]
for l in range(0,nlayers-1):
    pi[:,l+1] = np.exp(np.log(pressure_Map[:,l])-(np.log(pressure_Map[:,l])-np.log(pressure_Map[:,l+1]))/2)
pi[:,nlayers]=np.exp( np.log(pi[:,nlayers-1])-2*( np.log(pi[:,nlayers-1])-np.log(pressure_Map[:,nlayers-1]) ) )

# Re-calculate the AMF using the a priori profile and the averaging kernel
dp = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))              # delta pressure per layer [hPa]
m_check = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))          # recalculate AMF
vcda = np.full((total_size,nlayers),np.nan,dtype = np.dtype('f4'))            # a priori partial columns per layer [molec./cm2]
for l in range(0,nlayers):
    dp[:,l] = pi[:,l] - pi[:,l+1]
    vcda[:,l] = 10*ra*np.multiply(profile_Map[:,l],dp[:,l])/(g*dryair)
    # Check AMF
    m_check[:,l] = ra*10*amf_Map[:,l]*np.multiply(ak_Map[:,l],np.multiply(profile_Map[:,l],dp[:,l]))/(g*dryair)
VCDA = np.nansum(vcda,axis=1)
amf_check = np.nansum(m_check,axis=1)/VCDA

kok = (qa_Map>0)

# Plot the re-calculated AMF versus the L3 AMF
fig, ax =plt.subplots(1,1,figsize=(12,6))
plt.plot(amf_Map[kok],amf_check[kok], '.',c='black',markersize=0.2,fillstyle='full',linestyle='')
plt.grid(True)
plt.ylabel("Recalculated AMF using AK and ap Profile", fontsize=12, fontweight='bold', color='k')
plt.xlabel("L3 AMF", fontsize=12, fontweight='bold', color='k')
plt.title(filename, fontsize=10, fontweight='bold', color='k')
plt.xlim([0,3.5])
plt.ylim([0,3.5])
plt.show()
```





## 7.6. Preliminary evaluation

To be completed in the next version of the document.

## 7.7. Contacts and citation

### Contact:

- Nicolas Theys: [theys@aeronomie.be](mailto:theys@aeronomie.be)

### Dataset citation:

- TBD

### Paper citation:

- 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.

## 8. CHOCHO L3 from TROPOMI, OMI and GOME2

### 8.1. Input data and algorithm

CHOCHO can be retrieved from a series of nadir UV-VIS sensors. We focus on the following sensors: GOME2, OMI and TROPOMI.

Figure 8-1 shows the flow diagram of the L2 DOAS retrieval algorithm for CHOCHO, which is common to all UV-VIS sensors. First, we fit a CHOCHO slant column  $N_s$ , based on VIS radiance measurements in the 435-460nm window. Next, we compute an air mass factor  $M$  which accounts for the average light path through the atmosphere for the wavelengths in the fitting window, which we use to convert the slant column into a vertical column  $N_v$ . To compute the air mass factor, we rely on auxiliary data from a number of external sources: cloud cover data, topographic information, surface albedo, and the a priori shape of the CHOCHO vertical profile in the atmosphere. The effect of clouds is not modelled in the L2 air mass factor computation. Instead, pixels with significant cloud cover (cloud fraction > 20%) are excluded from the L3 product. Finally, a background normalization  $\Delta_{bg}$  is performed to correct for detector row or viewing angle biases in the retrieved slant columns. We then obtain the CHOCHO vertical column from the equation

$$N_v = \frac{N_s - \Delta_{bg}}{M} \quad 8-1$$

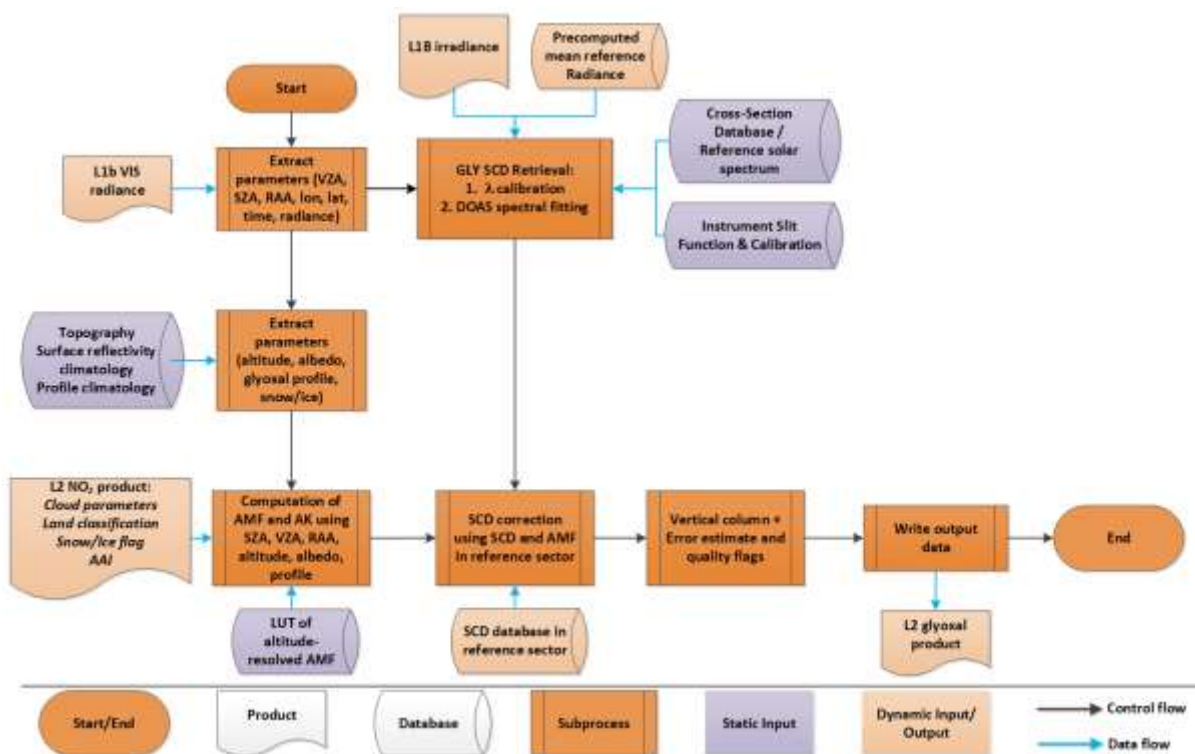


Figure 8-1 Flow diagram of the CHOCHO DOAS retrieval algorithm.

For TROPOMI, the L3 CDR are produced using the slant column fitting results from the S5P-PAL L2 data. For the GOME-2 and OMI sensors, level 1 data have fully reprocessed. Air mass factors and background correction values have been recalculated using consistent settings for all sensors. More details on the L2 algorithm settings can be found in the CHOCHO ATBD available at <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>.

The L3 products are directly derived from the L2 products. L2 values are binned into regular longitude/latitude grids, where each L3 grid cell is assigned the area-weighted mean of all overlapping L2 pixels. The L3 product consists of CHOCHO tropospheric columns and associated random and systematic uncertainties, detailed results, and input data. L3 data are available for each instrument separately. In addition, two merged CDR's will be provided:

- a morning overpass CDR containing data from GOME2/METOP-A,B,C instruments,
- an afternoon overpass CDR containing merged data from OMI/AURA and TROPOMI/S5P sensors.

Since the afternoon sensors have a significantly higher signal-to-noise ratio and spatial resolution, we provide daily and monthly grids for the afternoon sensors, while only monthly grids are available for the GOME-2 sensors.

## 8.2. File parameters

Table 8-1 describes the content of the L3 CHOCHO netCDF files. The fill values are NaN for all real-valued fields, and -1 for integer-valued fields / flags.

**Table 8-1: Variable content of L3 CHOCHO NetCDF files.**

Parameter	Unit	Dimension	Description
latitude	degrees	lat	latitude
longitude	degrees	lon	longitude
latitude_bounds	degrees	lat, independent_2	grid latitude bounds
longitude_bounds	degrees	lon, independent_2	grid longitude bounds
time	days	time=1	number of days since 1995-01-01
delta_time	seconds	time, lat, lon	seconds since time
ctm_sigma_a	hPa	layer	pressure = ctm_sigma_a + surface_pressure * ctm_sigma_b
ctm_sigma_b	1		
surface_pressure	hPa	time, lat, lon	surface pressure
surface_albedo	1	time, lat, lon	surface DLER (448 nm)
land_water_mask	1	time, lat, lon	0: water, 1: land, 2: land water transition
qa_L3	1	time, lat, lon	gridded data quality assurance value (0: not valid, 1: valid)
tropospheric_C2H2O2_column_number_density	1	time, lat, lon	Background-corrected glyoxal tropospheric vertical column number density





tropospheric_C2H2O2_column_number_density_uncertainty_random	molec.c m <sup>-2</sup>	time, lat, lon	Random uncertainty on the background-corrected glyoxal tropospheric vertical column number density
tropospheric_C2H2O2_column_number_density_uncertainty_systematic	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the background-corrected glyoxal tropospheric vertical column number density
tropospheric_C2H2O2_column_number_density_uncertainty_systematic_kernel	molec.c m <sup>-2</sup>	time, lat, lon	Systematic uncertainty on the background-corrected glyoxal tropospheric vertical column number density, without the profile uncertainty contribution
tropospheric_C2H2O2_column_number_density_std	molec.c m <sup>-2</sup>	time, lat, lon	Standard deviation of the background-corrected glyoxal tropospheric vertical column number density within the cell
tropospheric_C2H2O2_column_number_density_count	1	time, lat, lon	Effective number of observations per cell
C2H2O2_slant_column_number_density_rms	1	time, lat, lon	DOAS rms
C2H2O2_slant_column_number_density	molec.c m <sup>-2</sup>	time, lat, lon	glyoxal slant column number density
C2H2O2_slant_column_number_density_uncertainty	molec.c m <sup>-2</sup>	time, lat, lon	glyoxal slant column number density uncertainty
C2H2O2_slant_column_number_density_corrected	molec.c m <sup>-2</sup>	time, lat, lon	Background-corrected glyoxal slant column number density uncertainty
tropospheric_C2H2O2_column_number_density_background_correction	molec.c m <sup>-2</sup>	time, lat, lon	a priori glyoxal vertical column density number density in the Pacific reference sector
tropospheric_C2H2O2_column_number_density_amf	1	time, lat, lon	tropospheric air mass factor
cloud_fraction	1	time, lat, lon	effective cloud fraction
tropospheric_C2H2O2_column_number_density_apriori	molec.c m <sup>-2</sup>	time, lat, lon	a priori glyoxal vertical column number density (integrated profile)
C2H2O2_volume_mixing_ratio_dry_air_apriori	ppv	time, lat, lon, layer	a priori glyoxal vertical profile
C2H2O2_averaging_kernel	1	time, lat, lon, layer	Column averaging kernel

### 8.3. Data access and volume

The L3 product is available in netCDF format, and can be accessed from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/> following the filename convention:

CCI+p-L3-CHOCHO\_TC-instrument\_satellite\_L2version-BIRA-YYYYMM(DD)-L3version.nc

Example of file name for one TROPOMI monthly L3 file:

CCI+p-L3-CHOCHO\_TC-TROPOMI\_S5P\_PAL\_v03-BIRA-202310-fv0200.nc

The content of the netCDF files is listed in Table 8-1.





## 8.4. Data reading examples

```

import numpy as np
from netCDF4 import Dataset
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
from matplotlib import cm
from mpl_toolkits.axes_grid1 import make_axes_locatable

# Full file path of the NetCDF file
filename='CCI+p-L3-CHOCHO_TC-TROPOMI_S5P_PAL_v03-BIRA-202310-fv0200.nc'

# Map settings
proj_name='cyl'
Mspace=40
Pspace=30
parallels = np.arange(-90,90,Pspace)
meridians = np.arange(-180,180,Mspace)
LW=0.4
size_label=10

# Colorbar settings
ncolor=100
colormap = 'turbo'
scFactor=1e15
minScale = 0
maxScale = 20

# Variable to map and title
var2plot = "tropospheric_C2H2O2_column_number_density"
title_txt="Oct. 2023"

# Read the NetCDF file
with Dataset(filename, "r") as fid:
    CHOCHO2Map = fid[var2plot][:]
    CHOCHO2Map = CHOCHO2Map/scFactor
    varunit = fid[var2plot].units
    scaletitle='TROPOMI {} [{} {}]'.format(fid[var2plot].long_name,format(scFactor,'1.0e'),varunit)
    qa_Map = fid["qa_L3"][:]
    CHOCHO2Map[qa_Map<1]=np.nan

# Create the map
plt.figure(1,figsize=(12,12))
plt.title(scaletitle, loc='left',fontname="Times New Roman",fontweight="bold")
plt.annotate("ESA CCI+ ozone and aerosol precursors", xy=(0.01,0.015), xycoords='axes fraction',size=10,fontname="Times New Roman",fontweight="bold")
plt.title(title_txt, fontdict=None, loc='right',fontname="Times New Roman",fontweight="bold")

map = Basemap(projection=proj_name, llcrnrlon=-180, llcrnrlat=-90, urcrnrlon=180, urcrnrlat=90,resolution='1',lat_0=0,lon_0=0)
map.drawcoastlines(linewidth=LW,color='grey')
map.drawcountries(linewidth=LW,color='whitesmoke')
map.fillcontinents(color='lightgrey',zorder=0)
map.drawmeridians(meridians,color='black',linewidth=LW/2)
map.drawparallels(parallels,color='black',linewidth=LW/2)

im = map.imshow(CHOCHO2Map[0,:,:], cmap=cm.get_cmap(colormap, ncolor),interpolation='none')
plt.colorbar(im, orientation="horizontal")
plt.clim(minScale, maxScale)

plt.show()

```

## 8.5. Usage recommendations and known limitations

- Users interested in long-term trends should use the merged L3 dataset. Users interested in short time periods are advised to use the L3 from each instrument.
- The main product is the “**tropospheric\_CHOCHO\_column\_number\_density**”.
- A quality flag is provided for the L3 file, with the aim to filter cells which do not contain enough observations. It is advised to use only qa\_L3==1. The qa\_L3 flag is based on the effective number of observations within the cell (tropospheric\_CHOCHO\_column\_number\_density\_count), after applying the L2 quality filtering. If the tropospheric\_CHOCHO\_column\_number\_density\_count is lower than the 5<sup>th</sup> percentile of the cell counts for the day/month, the qa\_L3 is set to 0.
- Mean random ( $\sigma_r$ ) and systematic ( $\sigma_s$ ) uncertainties are available for each grid cell. If a total uncertainty is desired, this can be obtained from the square root of the sum of squares of the different terms:

$$\sigma^2 = \sigma_r^2 + \sigma_s^2$$

- When averaging L3 data over larger temporal or spatial scales, the uncertainty on the average can be obtained by combining the different random ( $\sigma_{ri}$ ) and systematic ( $\sigma_{si}$ ) uncertainties as

$$\sigma^2 = \sum_{i=1}^n \left(\frac{\sigma_{ri}}{n}\right)^2 + \left(\sum_{i=1}^n \frac{\sigma_{si}}{n}\right)^2$$



- Mean random and systematic uncertainties are available for each grid cell. If a total uncertainty is desired, this can be obtained from the square root of the sum of squares of the different terms.
- Mean averaging kernels and a priori vertical profiles are provided for each grid cell. This is only an approximation of these quantities with a vertical dimension. L2 data should preferably be used for an accurate exploitation of the averaging kernels.
- The example python script below uses the L3 AKs and ap Profiles to recalculate the AMFs, and compare the results with the provided L3 AMFs.

```
import numpy as np
from netCDF4 import Dataset
import matplotlib.pyplot as plt

# Initialisation
dryair = 28.9644;          #g/mole
ra = 6.0221367e23;       #molec/mole
g = 9.80665;             #m/s

# Full file path of the NetCDF file
filename='CCI+p-L3-CHOCHO TC-TROPOMI S5P PAL v03-BIRA-20231031-fv0200.nc'

# Read the NetCDF file
with Dataset(filename, "r") as fid:
    sigma_a_Layer= fid["ctm_sigma_a"][:].flatten()
    sigma_b_Layer= fid["ctm_sigma_b"][:].flatten()
    nlayers = np.shape(sigma_a_Layer)[0]
    profile_Map = fid["C2H2O2_volume_mixing_ratio_dry_air_apriori"][0,...].flatten()
    profile_Map = np.reshape(profile_Map, [-1, nlayers])
    ak_Map = fid["C2H2O2_averaging_kernel"][0,...].flatten()
    ak_Map = np.reshape(ak_Map, [-1, nlayers])
    xa_Map = fid["tropospheric_C2H2O2_column_number_density_apriori"][0,...].flatten()
    sp_Map = fid["surface_pressure"][0,...].flatten()
    amf_Map = fid["tropospheric_C2H2O2_column_number_density_amf"][0,...].flatten()
    qa_Map = fid["qa_L3"][0,...].flatten()

total_size = np.shape(sp_Map)[0]

# Calculate the pressure at layer centers
pressure_Map = np.full((total_size, nlayers), np.nan, dtype = np.dtype('f4')) # pressure at layer center [hPa]
for l in range(0, nlayers):
    pressure_Map[:, l] = (sigma_a_Layer[l] + sp_Map[:]*sigma_b_Layer[l]) # [hPa]

pi = np.full((total_size, nlayers+1), np.nan, dtype = np.dtype('f4')) # pressure at layer borders [hPa]
pi[:, 0]=sp_Map[:]; # surface pressure [hPa]
for l in range(0, nlayers-1):
    pi[:, l+1] = np.exp(np.log(pressure_Map[:, l]) - (np.log(pressure_Map[:, l]) - np.log(pressure_Map[:, l+1]))/2)
pi[:, nlayers]=np.exp( np.log(pi[:, nlayers-1]) - 2*( np.log(pi[:, nlayers-1]) - np.log(pressure_Map[:, nlayers-1]) ) )

# Re-calculate the AMF using the a priori profile and the averaging kernel
dp = np.full((total_size, nlayers), np.nan, dtype = np.dtype('f4')) # delta pressure per layer [hPa]
m_check = np.full((total_size, nlayers), np.nan, dtype = np.dtype('f4')) # recalculate AMF
vcdal = np.full((total_size, nlayers), np.nan, dtype = np.dtype('f4')) # a priori partial columns per layer [molec./cm2]
for l in range(0, nlayers):
    dp[:, l] = pi[:, l] - pi[:, l+1]
    vcdal[:, l] = 10*ra*np.multiply(profile_Map[:, l], dp[:, l])/(g*dryair)
    # Check AMF
    m_check[:, l]
= ra*10*amf_Map[:]*np.multiply(np.multiply(ak_Map[:, l], np.multiply(profile_Map[:, l], dp[:, l]))/(g*dryair)
VCDa = np.nansum(vcdal, axis=1)
amf_check = np.nansum(m_check, axis=1)/VCDa

kok = (qa_Map>0)

# Plot the re-calculated AMF versus the L3 AMF
fig, ax =plt.subplots(1,1,figsize=(12,6))
plt.plot(amf_Map[kok], amf_check[kok], '.', c='black', markersize=0.2, fillstyle='full', linestyle='')
plt.grid(True)
plt.ylabel("Recalculated AMF using AK and ap Profile", fontsize=12, fontweight='bold', color='k')
plt.xlabel("L3 AMF", fontsize=12, fontweight='bold', color='k')
plt.title(filename, fontsize=10, fontweight='bold', color='k')
plt.xlim([0, 3.5])
plt.ylim([0, 3.5])
plt.show()
```

## 8.6. Preliminary evaluation

To be completed in the next version of the document.





## 8.7. Contacts and citation

### Contact:

- Thomas Danckaert ([thomas.danckaert@aeronomie.be](mailto:thomas.danckaert@aeronomie.be))
- Jonas Vlietinck ([jonas.vlietinck@aeronomie.be](mailto:jonas.vlietinck@aeronomie.be))

### Dataset citation:

- TBD

### Paper citation:

- Lerot, C., Hendrick, F., Van Roozendael, M., Alvarado, L. M. A., Richter, A., De Smedt, I., Theys, N., Vlietinck, J., Yu, H., Van Gent, J., Stavrakou, T., Müller, J.-F., Valks, P., Loyola, D., Irie, H., Kumar, V., Wagner, T., Schreier, S. F., Sinha, V., Wang, T., Wang, P., and Retscher, C.: Glyoxal tropospheric column retrievals from TROPOMI – multi-satellite intercomparison and ground-based validation, *Atmos. Meas. Tech.*, 14, 7775–7807, <https://doi.org/10.5194/amt-14-7775-2021>, 2021.



## 9. ESA Reference documents

Available from <https://climate.esa.int/en/projects/precursors-for-aerosols-and-ozone/>

- Algorithm Theoretical Basis Document: Precursors\_cci+\_D2.1\_ATBD\_01\_00 D2.1
- End to End ECV Uncertainty Budget: Precursors\_cci+\_D2.2\_E3UB\_01\_00
- Product Validation and Intercomparison Report:
- Climate Assessment Report: