CENTER FOR ASTROPHYSICS HARVARD & SMITHSONIAN

VERSION: 2.1
RELEASE DATE: January 6, 2026

# TROPOSPHERIC EMISSIONS: MONITORING OF POLLUTION (TEMPO) PROJECT

## Trace Gas and Cloud Level 2 and 3 Data Products: User Guide

**January 6, 2026**

**DOI:** 10.5067/doc/tempo/L2andL3UserGuide/V2.1

# AUTHORS

Gonzalo González Abad (Center for Astrophysics | Harvard & Smithsonian)

Caroline Nowlan (Center for Astrophysics | Harvard & Smithsonian)

Huiqun (Helen) Wang (Center for Astrophysics | Harvard & Smithsonian)

Heesung Chong (Center for Astrophysics | Harvard & Smithsonian)

John Houck (Center for Astrophysics | Harvard & Smithsonian)

Xiong Liu (Center for Astrophysics | Harvard & Smithsonian)

Kelly Chance (Center for Astrophysics | Harvard & Smithsonian)

**CONTRIBUTIONS:**

TEMPO Science Team, Center for Astrophysics | Harvard & Smithsonian

TEMPO Ground Systems Team, Center for Astrophysics | Harvard & Smithsonian

# REVISION HISTORY

| Version | Item | Description | Release date |
|---|---|---|---|
| 1.0 | All | Initial version | May 20, 2024 |
| 1.1 | Table 3.3 | Correct main_data_quality_flag logic description | September 18, 2024 |
| 1.2 | Validation status; nomenclature; trace gas usage and known issues text. | Update validation status to provisional; update nomenclature to make consistent with ATBDs; add text on use of negative trace gas columns; add surface pressure issue to known issues | December 6, 2024 |
| 2.0 | Known issues; cloud and trace gas usage; file description; reference to NRT product; time description. | Update improvements, known issues and usage information in NO2, HCHO and CLDO4; update file descriptions; add reference to NRT products and documentation; add description of GPS time. | September 15, 2025 |
| 2.1 | NO2, HCHO and CLDO4 known issues; changes from previous version; NO2 and CLDO4 algorithm description; version 4 release date. | Minor updates to version 4 processing added for NO2 and HCHO; add additional known issues for NO2, HCHO and CLDO4; modify CLDO4 destriping description; add destriping to NO2 algorithm description; correct version 4 release date. | January 6, 2026 |

# PRODUCT and SCIENCE DATA PROCESSING CENTER PIPELINE VERSION

| Product Version Designation | Science Data Processing Center Pipeline Version | Release |
|---|---|---|
| V03 | 4.4.0 | May 20, 2024; first public release of TEMPO Level 2 and Level 3 products |
| V04<br>NRT V02 | 4.7.0 | September 17, 2025; TEMPO standard version 4 and NRT version 2 |
| V04<br>NRT V02 | 4.8.0 | January 6, 2026; TEMPO standard version 4 and NRT version 2 (mid-version minor updates) |

## Table of Contents

# 1 Introduction

## 1.1 The TEMPO mission

Tropospheric Emissions: Monitoring of Pollution (TEMPO) [Zoogman et al., 2017] is a geostationary satellite mission designed to measure air quality over North America during daylight hours at high spatial resolution ($2 \times 4.75$ km$^2$ at the center of field of regard) and with a temporal resolution of one hour or less. Launched on 7 April 2023, TEMPO obtained first light on 2 August 2023 and began nominal operations after passing its post-launch acceptance review in October 2023.

## 1.2 Science data processing overview

The TEMPO data products are generated by the Science Data Processing Center (SDPC) at the Smithsonian Astrophysical Observatory (SAO). All products, once generated at SAO, are pushed to NASA's Earth Observing System Data and Information System (EOSDIS) Cumulus cloud for public distribution by the NASA Atmospheric Science Data Center (ASDC). These include Level 1B (L1B) files (calibrated solar irradiance and geolocated radiance spectra), Level 2 (L2) files (cloud and trace gas products at the native ground pixel footprint for the East-West granules which make up part of a TEMPO scan) and Level 3 (L3) files (cloud and trace gas products on a regular grid created from all Level 2 granules constituting an entire East-West scan).

The SDPC generates TEMPO products from Level 0 (raw data) in the following order: (1) L1B; (2) L2 clouds; (3) L2 trace gases; and (4) L3 products. This is because L2 clouds require L1B as input, L2 trace gases require both L1B and L2 clouds as inputs, and L3 products require L2 products as inputs. Occasionally, L2 and L3 product files are missing when corresponding L1B files exist because either the spectra are not intended for trace gas retrievals or other conditions exist that prevent nominal L2 processing.

## 1.3 Data versioning and release history

Updates to the SDPC operational pipeline result in periodic new data releases. When this occurs, a new collection is created that includes all data products. The collection version number can be found in the data file name and has format VXX, where XX denotes the collection version. Since first light on 2 August 2023, the SAO science team has been working to improve the processing pipeline and science products with assistance from the TEMPO validation team. On 5 February 2024, a limited set of preliminary unvalidated TEMPO Version 1 (V01) products was released to the public for users to become familiar with the file format and content. On 26 February 2024, the SDPC pipeline was updated to produce Version 2 (V02) products for the public release of the V02 Level 1B (irradiance and radiance) products. On 14 May 2024, the SDPC was updated to produce Version 3 (V03) Level 1, 2, and 3 products. Version 3 products constituted the first public release of TEMPO Level 2 and 3 products. These were declared at the Provisional maturity level on 9 December 2024. The TEMPO validation document defines the Provisional level as "the performance has been demonstrated through a large, but still (seasonally or otherwise) limited number of independent measurements. The analysis is sufficient for limited qualitative determinations of product fitness-for-purpose, and the product is potentially ready for testing by operational users and may be suitable for scientific publication." [TEMPO Validation Team, 2023].

On 17 September 2025, the SDPC pipeline was updated to produce Version 4 (V04) Level 1, 2, and 3 products. Version 4 consists of several updates to calibration and retrieval algorithms which should result in more reliable Level 2 and 3 cloud, $NO_2$ and HCHO products. These changes include improved Level 1 radiometric calibration and stray light estimation, the update of the atmospheric a priori model from GEOS-CF version 1 to version 2, improved surface albedo inputs, updates to codes that read surface pressure and cloudy-sky scattering weights, and destriping of $O_2$-$O_2$ and $NO_2$ slant columns. Products will remain at the provisional validation level until a full validation of Version 4 data occurs. Users should consult the relevant data landing pages for current product maturity.

The Version 4 L1B RADT (twilight radiances) were released on 6 January 2026. At this time, the algorithm team also made minor updates to V04 L1B radiances and irradiances and trace gases ($NO_2$ destriping and HCHO radiance reference calculation). Product versions remained the same (V04 and NRT V02). These updates are applied in forward processing from 6 January 2026 and in V04 reprocessing from 2 August 2023 to 16 September 2025.

TEMPO also provides low-latency (< 3 hours) near real-time (NRT) data products. The NRT version 2 (V02) products are produced in parallel with the standard version 4 products. Further information on NRT products can be found in the TEMPO NRT user guide and ATBD supplement [Nowlan et al., 2025]. Usage recommendations for the NRT products are largely identical to those for the standard products provided in this user guide.

## 1.4 User guide description

This document provides a user guide for the Version 4 Level 2 and Level 3 trace gas products nitrogen dioxide (NO2) and formaldehyde (HCHO), and Version 4 Level 2 and Level 3 clouds (CLDO4). Table 1.1 lists the products described in this document. Note that the user guides for Level 1, total ozone, and ozone profile products are provided separately.

This guide contains a brief description of each algorithm, with further details available in the corresponding Algorithm Theoretical Basis Documents (ATBDs) available from the TEMPO project page at the NASA ASDC.

Users are recommended to take particular note of the "known issues" listed for each product in this user guide, as well as the recommendations for use and data filtering criteria. In addition, please note that the cloud fraction and cloud pressure from the CLDO4 product are used in the $NO_2$ and HCHO algorithm. As a result, the known issues listed for the cloud product will influence the $NO_2$ and HCHO products.

**Table 1.1.** TEMPO Level 2 and 3 products described in this document

| Product | Level | Description | Spatial Resolution and Coverage | Validation maturity level |
|---|---|---|---|---|
| NO2 | 2 | Nitrogen dioxide tropospheric and stratospheric vertical columns | Native TEMPO ground pixel, one granule | Provisional |
| NO2 | 3 | Nitrogen dioxide tropospheric and stratospheric vertical columns on regular grid | All Level 2 granules merged for a single scan, on 0.02° × 0.02° regular grid | Provisional |
| HCHO | 2 | Formaldehyde vertical columns | Native TEMPO ground pixel, one granule | Provisional |
| HCHO | 3 | Formaldehyde vertical columns on regular grid | All Level 2 granules merged for a single scan, on 0.02° × 0.02° regular grid | Provisional |
| CLDO4 | 2 | Effective cloud fraction (ECF) and cloud optical centroid pressure (OCP) | Native TEMPO ground pixel, one granule | Provisional |
| CLDO4 | 3 | ECF and OCP on regular grid | All Level 2 granules merged for a single scan, on a 0.02° × 0.02° regular grid | Provisional |

# 2 Data Organization and Format

## 2.1 Level 2 products

TEMPO Level 2 data products are organized in terms of the scan number (e.g., S004) and granule number (e.g., G01) for each day. Note that file names use UTC time. As some scans during daylight hours over North America may occur during or after UTC midnight, some files from late in the day may be labeled with the date of the next day.

Each scan is a sweep from the East to the West, and nominally takes one hour to complete, with shorter scans occurring over a more limited East-West region during morning and evening (~40 minutes) or during special high-time resolution operations. The data collected during each scan is divided into different granules to keep the data volume for each file manageable. Each granule contains data collected during a short time range (~6.7 minutes in nominal operations). Each granule file contains the full field of regard in the North-South direction, but only a portion of the East-West coverage of a full scan.

The granule number (e.g., G01, G02 etc.) does not always correspond to the same geographic location, but rather to the order in which the granule was collected for the scan. On some days, Level 2 trace gas and cloud products may begin with scan numbers higher than one (e.g., at S002) if twilight measurements (for city lights) were made.

## 2.2 Level 3 products

TEMPO Level 3 products are generated using an area-weighted gridding approach using all granules from an East-West scan. For each variable of interest, the algorithm maps the corresponding Level 2 variable onto a 0.02° × 0.02° regular grid over the domain of 168°W – 13°W and 14°N – 73°N. Gridding is performed for all data without filtering by quality flag, cloud cover, etc. As the Level 3 grid is at a somewhat finer spatial sampling than that of individual TEMPO ground pixels, recommendations for Level 2 data usage and filtering should also be applied for Level 3 data under most circumstances (these recommendations are discussed later in Sections 3.3 and 4.3).

## 2.3 Filenames

Table 2.1 shows the TEMPO Level 2 and 3 file naming convention.

The Level 2 NO2, HCHO and CLDO4 filenames have a format of

TEMPO_{GAS}_L2_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc

with YYYY (year), MM (month), DD (day), HH (hour), NN (minute), and SS (second) denoting the UTC time stamp at the beginning of the time period covered by the file, and XXX and YY denoting the scan and granule numbers, respectively. The segment GAS represents the product: CLDO4, NO2 or HCHO. (Note that curly brackets are added here for display purposes.) For example, the TEMPO Level 2 Version 4 $NO_2$ product for Scan 17 Granule 03 with a start UTC time of 00:15:04 on 10 May 2024 has a name of TEMPO_NO2_L2_V04_20240510T001504Z_S017G03.nc

Following a similar convention, the TEMPO Level 3 filenames have a format of

TEMPO_{GAS}_L3_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z_S{XXX}.nc

**Table 2.1** TEMPO Level 2 and 3 file naming convention.

| Product | Filename format |
|---|---|
| NO2 Level 2 | TEMPO_NO2_L2_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc |
| NO2 Level 3 | TEMPO_NO2_L3_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}.nc |
| HCHO Level 2 | TEMPO_HCHO_L2_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc |
| HCHO Level 3 | TEMPO_HCHO_L3_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}.nc |
| CLDO4 Level 2 | TEMPO_CLDO4_L2_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc |
| CLDO4 Level 3 | TEMPO_CLDO4_L3_V04_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}.nc |

## 2.4 Format

TEMPO products are in self-explanatory NetCDF4 format. The content of a NetCDF file can be reviewed by a command “ncdump -h {FILENAME}” on Linux systems or explored on a graphical interface using the software “Panoply,” distributed by NASA Goddard Institute for Space Studies (https://www.giss.nasa.gov/tools/panoply/). NetCDF files can be read using FORTRAN, C, Python, Matlab, IDL, R, etc.

Each file contains several groups (named “product”, “geolocation”, “qa_status”, and “support_data”) and global attributes. Each group contains a number of variables. Each variable is described by its attributes.

Dimensions used for the variables are defined in each NetCDF file. For Level 2 files, the *xtrack* = 2048 dimension is fixed along the North-South direction. The first and last few pixels in the North-South direction contain fill values as they are not suitable for retrieval. The dimension *mirror_step* denotes the East-West direction and can vary by granule (for nominal operations, there are typically 131 or 132 mirror steps per granule). The *corner* = 4 dimension is used for the 4 corners of each ground pixel. The dimension *swt_level* gives the number of vertical layers in the atmosphere for variables with an altitude dimension. In V04, *swt_level* = 72, corresponding to the number of vertical layers in the GEOS-CF v2 data which are used as ancillary input to TEMPO retrievals.

Level 3 files have dimensions of *longitude*, *latitude* and *time*. (In all cases, the *time* dimension is always equal to one for the Level 3 files.)

Key variables for Level 2 and 3 files in this release are provided below in the relevant sections. A full list of variables can be found in the appendices, or by displaying a file's contents using software that can read NetCDF files.

## 2.5 Observation time standard

TEMPO Level 2 files contain the time of observation of each mirror step in the variable *time* within the *geolocation* group. TEMPO observation timestamps in the file are in GPS time standard, and represented by seconds since the reference GPS epoch of 6 January 1980 00:00:00 UTC. GPS time is not corrected for leap seconds. As of September 2025, this results in the TEMPO observation timestamps being ahead of UTC time by 18 seconds.

# 3 Trace gas products ($NO_2$ and HCHO)

## 3.1 Algorithm overview

The TEMPO pipeline for trace gas retrievals provides a common framework for the retrieval of $NO_2$ and HCHO slant column densities (SCD) (also used for $O_2$-$O_2$ for clouds). SCDs are derived using least-squares minimization to directly fit a modeled radiance to the observed spectrum. $NO_2$ and HCHO also share the algorithm for the calculation of air mass factors (AMFs) used to derive the vertical column density (VCD). The specific features of each product's retrieval are described in Sections 3.1.1 and 3.1.2.

### 3.1.1 $NO_2$

The TEMPO $NO_2$ algorithm consists of four parts: (1) $NO_2$ SCD retrieval, (2) AMF calculation, (3) SCD destriping; and (4) stratospheric/tropospheric separation into stratospheric and tropospheric VCDs. The SCD and AMF calculations are based on SAO's OMI heritage spectral fitting and AMF algorithms [González Abad et al., 2015]. The stratospheric/tropospheric calculation is based on the methodology described by Geddes et al. [2018].

The $NO_2$ retrieval uses a solar irradiance spectrum as a source in a 405 - 465 nm fitting window, simultaneously fitting $NO_2$, $O_3$, $H_2O$ vapor, $O_2$-$O_2$ and liquid $H_2O$ absorption, as well as a Ring spectrum, under-sampling correction, wavelength shift with respect to the irradiance reference spectrum and 4th order closure polynomials.

### 3.1.2 HCHO

The TEMPO HCHO algorithm consists of three parts: (1) HCHO differential SCD (dSCD) retrieval using a radiance reference source term, (2) AMF calculation, and (3) a background correction to add back the HCHO column contained in the radiance reference source spectrum to the final VCD. TEMPO's HCHO retrieval is based on SAO's OMI heritage algorithms.

The HCHO retrieval uses a modeled radiance built around a source term calculated as the mean radiance reference derived for each cross track (North-South) position as the average of the radiances for a whole scan (all East-West positions). The HCHO spectral fit uses a 328.5 - 356.5 nm fitting window, simultaneously fitting HCHO, $O_3$, $NO_2$, BrO and $O_2$-$O_2$, as well as a Ring spectrum, under-sampling correction, wavelength shift with respect to the reference spectrum and 3rd order closure polynomials. A background correction is performed to account for HCHO signals included in the radiance reference spectra which are used as source terms.

### 3.1.3 Air mass factor calculation

The AMF calculations for $NO_2$ and HCHO use precomputed look-up tables of scattering weights at 47 vertical layers parameterized as functions of viewing geometry, $O_3$ profile, surface albedo and surface pressure. The treatment of cloudy pixels uses the cloud fraction and cloud pressure from the TEMPO CLDO4 product and the independent pixel approximation [Martin et al., 2002].

For surface reflectance, TEMPO uses geometry-dependent Lambertian equivalent reflectance (GLER). Currently, over land, the GLER is based on the MODIS bidirectional reflectance kernels (v6.1). Over the oceans, the GLER is based on the Cox-Munk function. Snow and ice can dramatically change surface albedo and significantly influence cloud and trace gas results. Snow and ice fraction is specified through the Interactive Multisensor Snow and Ice Mapping system (IMS). In the future, we anticipate implementation of GLER derived from TEMPO for TEMPO clouds and trace gases.

Ancillary data for meteorological variables (surface pressure, temperature-pressure profile, humidity, 2-m winds) and trace gas vertical profiles ($NO_2$, HCHO) are from the GEOS-CF version 2 forecast or analysis available at the time of SDPC processing. These are instantaneous hourly forecasts or analyses sampled at 0.25° longitude × 0.25° latitude (from the model run on a cubed sphere at c360 horizontal resolution). Typically, forecasts within 24 hours of the TEMPO observations are used in forward operational processing, but the more accurate analyses are used in Version 4 reprocessing (2 August 2023 – 16 September 2025). Hourly climatologies for each month provide a backup if the forecast or analysis is not available, though this rarely occurs. The NetCDF file attribute *apriori_source* indicates whether a forecast, analysis or climatology is used.

The scattering weights, trace gas profiles and temperature profiles provided in the trace gas product files are on the same 72 vertical layers as the GEOS-CF v2 model.

## 3.2 Key Variables

This section lists the key variables in the trace gas Level 2 and 3 files from the *geolocation* and *product* groups. Many other useful variables for filtering and interpreting data can be found in the *support_data* and *qa_statistics* groups.

For a comprehensive list of variables in the files, see Appendix B ($NO_2$) and Appendix C (HCHO).

### 3.2.1 Level 2 key variables

Tables 3.1 and 3.2 give the key variables in the Level 2 trace gas files.

**Table 3.1.** TEMPO Level 2 $NO_2$ key variables

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| geolocation | latitude | float (mirror_step, xtrack) | degrees north | Pixel center latitude |
| geolocation | latitude_bounds | float (mirror_step, xtrack, corner) | degrees north | Pixel corner latitude (SW, SE, NE, NW) |
| geolocation | longitude | float (mirror_step, xtrack) | degrees east | Pixel center longitude |
| geolocation | longitude_bounds | float (mirror_step, xtrack, corner) | degrees east | Pixel corner longitude (SW, SE, NE, NW) |
| geolocation | time | double (mirror_step) | seconds since 1980-01-06T00:00:00Z | Radiance exposure start time |
| product | main_data_quality_flag | short (mirror_step, xtrack) | none | Main data quality flag |
| product | vertical_column_troposphere | double (mirror_step, xtrack) | molecules/$cm^2$ | Troposphere $NO_2$ vertical column |
| product | vertical_column_troposphere_uncertainty | double (mirror_step, xtrack) | molecules/$cm^2$ | Troposphere $NO_2$ vertical column uncertainty |
| product | vertical_column_stratosphere | double (mirror_step, xtrack) | molecules/$cm^2$ | Stratosphere $NO_2$ vertical column |

**Table 3.2.** TEMPO Level 2 HCHO key variables

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| geolocation | latitude | float (mirror_step, xtrack) | degrees north | Pixel center latitude |
| geolocation | latitude_bounds | float (mirror_step, xtrack, corner) | degrees north | Pixel corner latitude (SW, SE, NE, NW) |

| | | | | |
|---|---|---|---|---|
| geolocation | longitude | float (mirror_step, xtrack) | degrees east | Pixel center longitude |
| geolocation | longitude_bounds | float (mirror_step, xtrack, corner) | degrees east | Pixel corner longitude (SW, SE, NE, NW) |
| geolocation | time | double (mirror_step) | seconds since 1980-01-06T00:00:00Z | Radiance exposure start time |
| product | main_data_quality_flag | short (mirror_step, xtrack) | none | Main data quality flag |
| product | vertical_column | double (mirror_step, xtrack) | molecules/cm$^2$ | HCHO vertical column |
| product | vertical_column_uncertainty | double (mirror_step, xtrack) | molecules/cm$^2$ | HCHO vertical column uncertainty |

Most users should use the *main_data_quality_flag* in the *product* group to ensure the use of high-quality data. This flag reports the overall quality of the retrieval using three values described in Table 3.3. Further filtering suggestions are given in Section 3.5.

The VCD in pixels classified as bad should never be used. A pixel could be categorized as suspect for multiple reasons; advanced users of the products could use these pixels for specific investigations with extra caution. When the *main_data_quality_flag* is a fill value, this indicates that the retrieval was not attempted, likely because of missing geolocation information or detector saturation flagged in the L1B radiances.

**Table 3.3.** Definition of *main_data_quality_flag* values. $AMF_g$ is the geometric AMF, where $AMF_g$ = sec(*solar_zenith_angle*) + sec(*viewing_zenith_angle*). For $NO_2$, $VCD_{min}$ = -1e19 and $VCD_{max}$ = 1e19 molecules/cm$^2$. For HCHO, $VCD_{min}$ = -5e17 and $VCD_{max}$ = 5e17 molecules/cm$^2$.

| Value | Meaning | Description |
|---|---|---|
| 0 | good | fit_convergence_flag = 1 *AND*<br>$VCD_{min} <= VCD_{total} <= VCD_{max}$ *AND*<br>$(SCD + 2*SCD_{uncert}) >= 0$ *AND*<br>$AMF_g <= 6$ *AND*<br>$AMF_{total} >= 0.1$ |
| 1 | suspect | fit_convergence_flag = 0 *OR*<br>$(SCD + 2*SCD_{uncert}) < 0$ *AND* $(SCD + 3*SCD_{uncert}) >= 0$ *OR*<br>$VCD_{total} < VCD_{min}$ *OR*<br>$VCD_{total} > VCD_{max}$ *OR*<br>$AMF_g > 6$ *OR*<br>$AMF_{total} < 0.1$ |
| 2 | bad | fit_convergence_flag < 0 *OR*<br>$(SCD + 3*SCD_{uncert}) < 0$ *OR*<br>amf_diagnostic_flag(bit1) = 1 |

In addition, the *support_data* and *qa_statistics* groups contain multiple variables useful to diagnose and filter $NO_2$ and HCHO retrievals according to their quality.

- *fit_convergence_flag* in *qa_statistics* group reports the status of the SCD fitting algorithm result with the following values:
  - -2: failed
  - -1: maximum number of iterations exceeded
  - 0: the result of the fitting is suspect (for example, at noise level)
  - 1: good fitting

  The fit convergence flag is considered in the main data quality flag, but can be used separately to investigate the quality of the SCD retrieval itself.

- *amf_diagnostic_flag* in *support_data* group is a 16-bit integer used to provide information about AMF calculations in the form of a bitwise flag. See Table 3.4 for bit definitions.
- *ground_pixel_quality_flag* in *support_data* is copied from L1B files. Please refer to the L1B user guide for information. In the trace gas algorithms, only the “land” flag from this variable is used (to determine the source of GLER information).
- *eff_cloud_fraction* and *amf_cloud_fraction* in the *support_data* group give the effective cloud fraction (read from the CLDO4 *cloud_fraction* variable) and the *amf_cloud_fraction*, which is the radiance cloud fraction at either 440 nm ($NO_2$) or 340 nm (HCHO).
- *snow_ice_fraction* in the *support_data* group gives the snow/ice fraction for a pixel.

**Table 3.4.** Meaning of each bit of *amf_diagnostic_flag* in Level 2 $NO_2$ and HCHO files

| Bit | Bit Meaning |
|---|---|
| bit0 | Good AMF |
| bit1 | Bad AMF / no AMF calculation performed |
| bit2 | Warning: pixel affected by glint |
| bit3 | Warning: climatological cloud pressure information used |
| bit4 | Warning: adjusted surface pressure |
| bit5 | Warning: adjusted cloud pressure |
| bit6 | Not used / reserved for future use |
| bit7 | Not used / reserved for future use |
| bit8 | Not used / reserved for future use |
| bit9 | Not used / reserved for future use |
| bit10 | Error: no albedo information |
| bit11 | Error: no cloud information |
| bit12 | Error: no trace gas profile information |
| bit13 | Error: no scattering weight calculation |
| bit14 | Error: no geolocation information available |
| bit15 | Not used / reserved for future use |

### 3.2.2 Level 3 key variables

Tables 3.5 and 3.6 give the key variables in the $NO_2$ and HCHO Level 3 trace gas files. Note that the Level 3 main data quality flag is only assigned "good" when all pixels used in that grid box are "good" in the Level 2 file. Please see Appendix B ($NO_2$) and C (HCHO) for a complete list of variables in the Level 3 files.

**Table 3.5.** TEMPO Level 3 $NO_2$ key variables

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | latitude | float (latitude) | degrees north | Latitude at grid box center |
| | longitude | float (longitude) | degrees east | Longitude at grid box center |
| | time | double (1) | seconds since 1980-01-06T00:00:00Z | Scan start time |
| product | main_data_quality_flag | short | none | Main data quality flag |

| | | | | |
|---|---|---|---|---|
| | | (time, latitude, longitude) | | |
| product | vertical_column_troposphere | double (time, latitude, longitude) | molecules/cm$^2$ | Troposphere $NO_2$ vertical column |
| product | vertical_column_troposphere_uncertainty | double (time, latitude, longitude) | molecules/cm$^2$ | Troposphere $NO_2$ vertical column uncertainty |
| product | vertical_column_stratosphere | double (time, latitude, longitude) | molecules/cm$^2$ | Stratosphere $NO_2$ vertical column |

**Table 3.6.** TEMPO Level 3 HCHO key variables

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | latitude | float (latitude) | degrees north | Latitude at grid box center |
| | longitude | float (longitude) | degrees east | Longitude at grid box center |
| | time | double (1) | seconds since 1980-01-06T00:00:00Z | Scan start time |
| product | main_data_quality_flag | short (time, latitude, longitude) | none | Main data quality flag |
| product | vertical_column | double (time, latitude, longitude) | molecules/cm$^2$ | HCHO vertical column |
| product | vertical_column_uncertainty | double (time, latitude, longitude) | molecules/cm$^2$ | HCHO vertical column uncertainty |

## 3.3 Usage recommendations

This section provides suggestions for filtering and best use. Note that these recommendations apply to both Level 2 and Level 3 files, as Level 3 files contain unfiltered Level 2 information.

### 3.3.1 Data filtering

Users of the trace gas products should at minimum apply filtering of the data using the main data quality flag and cloud fraction. At large cloud fractions, clouds obscure the lower atmosphere, leading to less sensitivity to the surface and larger uncertainties.

The *main_data_quality_flag* variable in the *product* group of the Level 2 and Level 3 files provides a high-level approximation to product quality. Table 3.3 provides the definition of the *main_data_quality_flag* for $NO_2$ and HCHO retrievals.

Retrievals of the highest quality have a *main_data_quality_flag* equal to “0”. This flag considers the value of the VCDs to detect outliers, the viewing geometry for each pixel and the availability of a successful AMF calculation. Owing to increased uncertainties in the spectral fitting and AMF calculations, pixels with geometric $AMF_g$ larger than 6 (SZA > ~70° and VZA > ~70°) are categorized as “suspect” with main_data_quality_flag equal to “1”. Those pixels identified as outliers or without a successful AMF calculation are categorized as bad with a *main_data_quality_flag* value equal to “2”. Pixels categorized as suspect carry useful information, but their interpretation requires further analysis. Fitting uncertainties in early and late hours of the day increase, and the sensitivity of TEMPO to lower tropospheric $NO_2$ and HCHO is reduced.

The overall precisions of TEMPO $NO_2$ and HCHO (d)SCD retrievals meet the TEMPO pre-launch science requirements. However, in V04 spectral fitting degrades over partly cloudy pixels due to inhomogeneous illumination of the instrument slit, resulting in large uncertainties and, in the case of HCHO, positive biases in the retrieved dSCD associated with straylight contamination. In consequence, the recommendation to use only the highest quality retrievals is to limit analysis to pixels with effective cloud fraction (*support_data/eff_cloud_fraction*) < 0.1. Less strict cloud fraction criterion (e.g., < 0.2) will retain more data, though the retained data will include more pixels with high uncertainties. Pixels with high cloud fractions (> 0.5) are less affected by this issue, and may be used with more confidence for specific applications (like $NO_2$ cloud slicing). Users are thus advised to adjust based on their tolerance. This cloud fraction recommendation is for the current data version, and the cloud filter recommendation may change for future data releases.

GLER look-up-table accuracy is difficult to assess, particularly over snow and ice, bright surfaces, and hours distant from the MODIS overpass times (10:30 AM and 1:30 PM local equatorial crossing time). Thus, we recommend using the *snow_ice_fraction* in the *support_data* group to identify pixels covered by snow and ice and treat them with care.

The *geolocation* and *support_data* groups contain variables necessary to interpret the observations. The *support_data* group contains the variable *amf_diagnostic_flag*, a 16-bit bitwise flag indicating different assumptions/issues in the air mass factor calculation, which advanced users may wish to consult for further insight.

### 3.3.2 Calculation of vertical pressure levels

In Level 2 files, the vertical pressure levels (layer edges) are defined on the same levels as the GEOS-CF model used as input to the retrievals. These pressure levels can be regenerated for each ground pixel using the variable *support_data/surface_pressure* ($p_s$) and the *eta_a* and *eta_b* values which are attributes of the surface pressure variable. Each pressure level (lower edge of layer *z*) can be calculated using:

$$p_{edge}(z) = eta\_a(z) + eta\_b(z) * p_s$$

### 3.3.3 Using alternative AMF

Advanced users may wish to recalculate AMFs using their own a priori gas profiles and the information contained in the Level 2 scattering weight variable *support_data/scattering_weights*. Users may use their own a priori atmospheric temperature profile or the one provided in the variable *support_data/temperature_profile* (from GEOS-CF, and used in the TEMPO retrieval).

First, it is necessary to interpolate the scattering weights, a priori gas profile and temperature profile to the same vertical grid. The scattering weights are provided on the pressure grid described above and include the effects of viewing geometry, surface reflectance, Rayleigh scattering and clouds. (Note that errors in trace gas profiles only account for one part of the AMF error. The other source is from the scattering weights – primarily from errors in clouds and surface albedo. Furthermore, scattering weight errors from clouds and albedo are not independent, as surface albedo significantly influences cloud retrievals.) AMFs can be calculated using the following equation:

$$AMF = \int W(z)S(z)c(z)dz$$

in which *W*, *S* and *c* represent the scattering weight, the shape factor and the temperature correction (which accounts for the use of a single temperature cross section in the spectral fit at altitude *z*). The shape factor is the normalized trace gas profile over the altitude range of interest. It is calculated from the partial columns of the trace gas at each layer, *n(z)*, using

$$S(z) = \frac{n(z)}{\int n(z)dz}$$

For the tropospheric column, the integration goes from the surface to the tropopause. For the stratosphere, it goes from the tropopause to the top of the atmosphere. In the case of the total column, the integration is from the surface to top of the atmosphere.

In the case of HCHO, which has minimal temperature-dependent absorption, users can apply *c(z)* = 1 throughout the profile. For $NO_2$, the temperature correction at each layer can be determined using the following empirical relationship [van Geffen et al., 2022]:

$$c(z) = 1 - a[T(z) - T_{\sigma}] + b[T(z) - T_{\sigma}]^2$$

with values $a = 0.00316$, $b = 3.39\times10^{-6}$, $T_{\sigma} = 220$ K and $T(z)$ = temperature at layer $z$.

### 3.3.4 Use of $NO_2$ total columns

Most users who wish to use the total $NO_2$ column (for instance, for comparisons with total $NO_2$ measured by ground-based Pandora instruments) should use the sum of *vertical_column_troposphere* + *vertical_column_stratosphere* in the main *product* group. The *vertical_column_total* in the *support_data* group is determined using the full profile of scattering weights and full model GEOS-CF profile and is significantly influenced by the relative model amounts of the stratospheric and tropospheric profile, and its use by most users is discouraged.

### 3.3.5 Use of negative trace gas columns

Negative trace gas columns of $NO_2$ and HCHO are expected from the retrieval and should not be excluded from analyses. The trace gas column is a differential column derived from the ratio of a radiance spectrum to a reference spectrum, and negative values may result from both systematic uncertainties and noise in the measurement. The exclusion of negative values could result in a high bias in both temporal and spatial averages, and users are strongly advised to not exclude negative columns. (Unphysical negative values from a bad retrieval are generally flagged in the main data quality flag and can be removed using this flag.)

### 3.3.6 Uncertainties

Error estimates for $NO_2$ and HCHO are limited to fitting uncertainties from the slant column retrieval. For an offline estimate of other uncertainty sources, such as those from the AMF, please consult the relevant ATBD document.

## 3.4 Changes from previous version

(1) Improvements to Level 1 calibration and stray light impact clouds and trace gases. In particular, the overestimation of the cloud fraction has been reduced which impacts trace gas retrievals.
(2) The stray light correction in Level 1 data reduces the high bias in HCHO over cloudy scenes.
(3) Destriping is implemented to mitigate $NO_2$ SCD stripes. New variables *destriping_correction* and *fitted_slant_column_uncorrected* (the original SCD before destriping) have been added to the *support_data* group.
(4) A bug in Version 3, where surface pressure indices were offset by one GEOS-CF grid box, has been corrected. The bug mainly affected AMFs in regions with large topography variations.
(5) A bug in reading the scattering weights from the cloudy-sky look-up table has been corrected. This mainly affected AMFs under partly cloudy conditions with high altitude clouds.
(6) New GLER tables for land are used which better represent diurnal variability of the surface reflectance.
(7) The GEOS-CF model used to provide meteorological inputs and a priori trace gas profiles is updated from v1 to v2.
(8) Clear-sky AMFs and scattering weight variables have been added. These are the AMF and scattering weights if the scene is assumed to be cloud-free.
(9) A planetary boundary layer (PBL) height variable derived from the GEOS-CF model has been added. The PBL height is not used in the trace gas retrieval but may help users interpret observations.
(10) *January 2026 minor update*: The SZA and VZA limits used in the $NO_2$ destriping correction calculation have been reduced from 85° to 80° to reduce the likelihood of spurious destriping values when single scans are used for destriping.
(11) *January 2026 minor update*: The cloud fraction limit for the HCHO radiance reference calculation has been reduced from 0.5 to 0.3 to reduce striping in V04 HCHO.

## 3.5 Known issues in trace gas products

### 3.5.1 Known issues affecting both $NO_2$ and HCHO

(1) The precision and accuracy of the retrieval deteriorates in scenes where the instrument slit is filled in an inhomogeneous way. This is typically associated with partly cloudy pixels. This effect typically starts to appear at effective cloud fractions greater than 0.1,

peaks between 0.2-0.5 and reduces its influence at larger effective cloud fractions. These pixels show markedly higher uncertainties and fitting RMS values.

(2) When cloud pressure is unavailable from CLDO4, it is replaced with a climatology which may result in large uncertainties in the AMF. This is flagged in the *amf_diagnostic_flag* variable.

(3) The accuracy of the retrieval deteriorates at large solar zenith angles (SZA) for both trace gases and clouds. For now, we recommend limiting the analysis to SZA < 70° or to treat pixels with SZA > 70° with great caution. Investigating the performance of TEMPO early morning/late afternoon observations is an ongoing effort.

(4) Calculated GLER values constrained by MODIS but distant from MODIS overpass times may have higher uncertainties. These uncertainties propagate to the $NO_2$ and HCHO AMFs calculations both directly though surface albedo and indirectly through clouds, and may result in artificial spatial variability and/or biases in the cloud information and trace gas VCD.

(5) There are discontinuities in surface reflectance between granules at extremely high SZA.

(6) TEMPO spectra saturate over very bright clouds. This affects the cloud retrieval and the $NO_2$ and HCHO retrievals at high cloud fractions. Most are flagged in the main data quality flag as a fill value (retrieval not attempted) or "bad" but there may be some erroneously flagged as "good" or "suspect", particularly on the edges of clouds.

(7) The effects of aerosols are not currently considered in the cloud or trace gas retrievals. Smoke from biomass burning may be detected as a moderate cloud fraction, and the application of a cloud filter to these data can result in users missing biomass burning plumes.

(8) Retrievals over snow and ice have low accuracy due to uncertainties in the surface reflectance over snow/ice and the poor performance of the cloud fraction retrieval in these regions.

### 3.5.2 Known issues affecting only $NO_2$

(1) *17 September 2025 – 5 January 2026*: Spurious stripes exist in some early morning scans on the day immediately after the weekly irradiance is collected (currently Thursdays).

(2) On rare occasions, large sharply-defined $NO_2$ "clouds" suddenly appear at ~15:00-18:00 UTC and last for 2 to 4 scans. These artifacts are most common in the northern part of the field of regard and in summer. They appear to be correlated with saturated cloudy pixels in the southern half of the field, possibly indicating uncorrected internal mirroring of stray light within the spectrometer. Investigation of this issue in L1B radiances is ongoing.

### 3.5.3 Known issues affecting only HCHO

(1) The radiance reference used for the retrieval of the dSCDs does not take into consideration the presence of large HCHO signals. As a result, dSCDs are biased low and the background correction may be subject to large uncertainties. A fix for this will be implemented in a future version.
(2) Some scans may have a limited number of available observations for radiance reference construction, resulting in stripes of missing data and artifacts in the North-South direction, particularly at Northern latitudes.
(3) Residual straylight in the HCHO fitting window results in vertical column positive bias in cloudy pixels and bright smoke plumes.
(4) *17 September 2025 – 5 January 2026*: The HCHO retrieval shows significant stripes during the central hours of the day, particularly in the southern part of the field of regard. This was due to using too high of a cloud fraction threshold after the reduction of the cloud fraction bias between V03 and V04.

# 4 Cloud product (CLDO4)

## 4.1 Algorithm overview

The TEMPO CLDO4 algorithm is composed of two parts: (1) $O_2$-$O_2$ Slant Column Density (SCD) retrieval, and (2) cloud information retrieval. The SCD retrieval uses SAO's fitting code [González Abad et al., 2015] with $O_2$-$O_2$ fitting configuration. The code for the cloud information retrieval is originally adapted from that used for the OMI cloud product (OMCDO2N) generated by NASA Goddard and is further developed at SAO for TEMPO application.

The TEMPO $O_2$-$O_2$ SCD retrieval is based on direct fitting of the measured spectrum within 439 – 488 nm using $O_2$-$O_2$, $NO_2$, $O_3$, $H_2O$, liquid water, Ring, vibrational Raman scattering (water Ring), wavelength shift, under-sampling and 3$^{rd}$ order baseline and multiplicative closure polynomials. The gas reference spectra are convoluted with the instrument slit function derived during the on-line solar irradiance fitting.

TEMPO cloud information retrieval is mainly based on Vasilkov et al. [2018] which assumes Lambertian clouds with an albedo of a=0.8. It greatly reduces the computational cost while providing reasonable results, thus, it is practical for processing the large volume of TEMPO data. This assumption is consistently adopted by TEMPO clouds and trace gases.

Effective Cloud Fraction (ECF) is derived from the sun-normalized spectral radiance at 466 nm. Cloud Optical Centroid Pressure (OCP) is derived using effective cloud fraction, $O_2$-$O_2$ SCD, Look-Up Tables (LUTs), and ancillary data. The pre-computed LUTs of sun normalized radiances and AMFs at the relevant wavelengths are based on the US standard air and pre-flight instrument slit function.

## 4.2 Key variables

This section lists the key variables in cloud Level 2 and 3 files from the *geolocation* and *product* groups. Many other useful variables can be found in the *support_data* and *qa_statistics* groups.

For a comprehensive list of variables in the files, see Appendix D.

### 4.2.1 Level 2 key variables

**Table 4.1.** TEMPO Level 2 CLDO4 key variables

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| geolocation | latitude | float (mirror_step, xtrack) | degrees north | Pixel center latitude |
| geolocation | latitude_bounds | float (mirror_step, xtrack, corner) | degrees north | Pixel corner latitude (SW, SE, NE, NW) |
| geolocation | longitude | float (mirror_step, xtrack) | degrees east | Pixel center longitude |
| geolocation | longitude_bounds | float (mirror_step, xtrack, corner) | degrees east | Pixel corner longitude (SW, SE, NE, NW) |
| geolocation | time | double (mirror_step) | seconds since 1980-01-06T00:00:00Z | Radiance exposure start time |
| product | cloud_fraction | float (mirror_step, xtrack) | none | Effective Cloud Fraction (ECF) |
| product | cloud_pressure | float (mirror_step, xtrack) | hPa | Effective Optical Centroid Pressure (OCP) |

The Level 2 CLDO4 files contain the following quality flags. Note that there are currently no quality flags for *cloud_fraction* and *cloud_pressure*.

- *fit_convergence_flag* in *qa_statistics* group is for $O_2$-$O_2$ SCD fitting and carries the same meaning as that in $NO_2$ and HCHO.
- *ground_pixel_quality_flag* in *support_data* is copied from L1B files. Please refer to the L1B user guide for information.
- *SCD_MainDataQualityFlag* in *support_data* is for $O_2$-$O_2$ SCD fitting. Note it is not for cloud information. Possible values are: 0 (normal), 1 (suspicious), and 2 (bad), where normal means that the retrieved SCD is in the range permitted, within 2 standard deviations of the fitting uncertainty and with solar zenith angle SZA < 89 degree.

- *processing_quality_flag* in *product* is a 16-bit integer used to indicate error, warning, or information during CLDO4 processing. All 16 bits are used, thus, all values represented by the 16-bit integer are possible, fill_value should be ignored, and bitwise tests should be used. The meaning of each bit is listed in Table 4.2.

**Table 4.2.** Meaning of each bit of *processing_quality_flag* in Level 2 CLDO4

| Bit | Bit Meaning |
|---|---|
| bit0 | Error due to lack of geolocation |
| bit1 | Warning for invalid cloud radiance fraction (CRF) at 466 nm |
| bit2 | Warning: OCP is replaced by scene pressure because ECF<0.05 |
| bit3 | Error due to invalid input surface pressure or input GLER |
| bit4 | Warning: OCP is replaced by scene pressure because snow_ice_fraction>0.05 |
| bit5 | Warning: $O_2$-$O_2$ SCD temperature correction exceeds maximum iteration allowed (20) |
| bit6 | Error for OCP due to invalid or bad $O_2$-$O_2$ SCD |
| bit7 | Information: bad Irradiance (IRR) or Radiance (RAD) at 440 nm |
| bit8 | Error due to bad IRR or RAD at 466 nm |
| bit9 | Warning: ECF is beyond normal range and is thus truncated |
| bit10 | Information: possible error for scene albedo/scene pressure at surface (internal use) |
| bit11 | Information: possible error for scene albedo/scene pressure at cloud (internal use) |
| bit12 | Error: ECF calculation is skipped due to any problem in calculating ECF |
| bit13 | Error: OCP calculation is skipped due to any problem in calculating OCP |
| bit14 | Warning: calculated OCP is out of normal range and thus clipped |
| bit15 | Information: scene pressure / scene albedo is skipped due to any problem |

### 4.2.2 Level 3 key variables

**Table 4.3.** TEMPO Level 3 CLDO4 key variables

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | latitude | float (latitude) | degrees north | latitude at grid box center |
| | longitude | float (longitude) | degrees east | longitude at grid box center |
| | time | double (1) | seconds since 1980-01-06T00:00:00Z | scan start time |
| product | cloud_fraction | float (latitude, longitude) | none | Effective Cloud Fraction (ECF) |

| | | | | |
|---|---|---|---|---|
| product | cloud_pressure | float (latitude, longitude) | hPa | Effective Optical Centroid Pressure (OCP) |

## 4.3 Usage Information

Due to the complex spatial, temporal, and spectral dependence of cloud information, as well as the differences in the assumptions made, the TEMPO cloud information is not expected to be directly comparable with the cloud products from other instruments. Instead, the TEMPO "cloud fraction" is an Effective Cloud Fraction (ECF) and the TEMPO "cloud pressure" is an effective Optical Centroid Pressure (OCP) that are specific for each TEMPO scene calculated with the adopted assumption and a priori data. In a sense, the complex cloud type and distribution for each TEMPO scene is condensed and transferred to a couple of cloud parameters to provide a simple way of self-consistently calculating TEMPO trace gas AMFs. ECF and OCP work best in pairs, as OCP derivation depends on ECF as an input.

Currently, only the "cloud_fraction" (i.e., ECF) and "cloud_pressure" (i.e., OCP) variables in the CLDO4 output are directly used by the SDPC for trace gas AMF calculations. Other variables include various input variables used for cloud retrieval and supplementary output variables that are potentially useful but remain to be tested.

Cloud ECF retrieval is highly sensitive to instrument radiance/irradiance calibration. A calibration bias at 466 nm in the L1B data will lead to a systematic error in cloud fraction retrieval which will subsequently affect the retrieved cloud pressure. A previous high bias of RAD/IRR in Version 3 L1B data has been corrected in Version 4, which leads to improved ECF and OCP in this release.

The cloud retrieval assumes a Lambertian cloud albedo of a = 0.8. In reality, cloud albedo varies widely with cloud type and is also dependent on viewing geometry [Hartmann, 2016]. Stammes et al. [2008] recommends a = 0.8 - 0.9 based on radiative transfer calculations and uses a = 0.8 for backward consistency. A higher cloud albedo will result in a smaller cloud fraction all else being equal, and vice versa. It is possible that a choice other than a = 0.8 could be more appropriate for TEMPO, given the differences in resolution and geometry. A larger ECF implies a higher degree of cloud contamination. Users are advised to adjust their cloud screening criterion based on the caveats described above and their level of tolerance. For filtering TEMPO $NO_2$ and HCHO products, users are advised to refer to the trace gas section of this document.

Biases and errors in *a priori* ancillary data will propagate into the retrieved cloud information which in turn affects the trace gas AMF through the cloudy sky scattering weights. Presumably, the MODIS-based GLER has variable quality over the range of geometries covered by TEMPO.

It remains to be quantified how the GLER works for TEMPO. If the prescribed surface albedo is lower than reality, the cloud algorithm will compensate by increasing ECF, i.e., ECF will appear larger.

$O_2$-$O_2$ SCD contains east-west oriented stripes at ~1-2% level. For context, a 2% increase in SCD can lead to an OCP change of 10 - 40 hPa depending on ECF, with smaller change for larger ECF. The stripes in SCD result from differences in slit functions of the 1024 spatial pixels. The slit functions are derived through a fitting using the solar irradiance IRR measurement. SCDs derived using the same IRR file have a similar stripe pattern. However, whenever a new IRR file is used for $O_2$-$O_2$ fitting, the stripe pattern changes significantly. In this case, the day right after a new IRR file is not de-striped.

To generate the operational TEMPO product, SDPC uses GEOS-CF data available at the time of processing. GEOS-CF has a coarser spatial resolution than a TEMPO pixel and the errors in meteorological fields can affect cloud retrieval result, and particularly cloud pressure. In short, multiple sources of error may compensate or amplify each other, the result of which requires further study.

When the measured reflectance is outside the range of clear-sky and overcast reflectance calculated using the LUTs and a priori data, the calculated ECF will wander outside the nominal range of [0.0, 1.0]. This occurs because the assumed cloud albedo a = 0.8 and input *a priori* data lead to theoretical reflectance that is incompatible with the observation. Following the practice employed by the NASA OMI product, the TEMPO ECF is allowed to go somewhat beyond the nominal range to [-1.0, 2.0] and the ECF values are clipped at 0.0 and 1.0 before output. If the calculated ECF is beyond the relaxed range of [-1.0, 2.0], the ECF retrieval fails as the calculated and measured values are irreconcilable under the retrieval setup, and the corresponding cloud result will be missing.

## 4.4 Changes from previous version

(1) Version 3 ECF histogram peaked near 0.05 instead of 0, suggesting an overestimate of cloud fraction. The shift in ECF histogram has been fixed in Version 4. The improvement is mainly due to L1B and secondarily to land GLER.

(2) Destriping is implemented to mitigate $O_2$-$O_2$ SCD stripes. For forward processing, SDPC currently implements a traditional statistics based method for the standard product, where a vector of destriping factor is derived from the previous day’s granules and applied on the current day’s granules. The NRT product is generated without destriping. A new variable “scddes” for the destriped SCD is added to the support group of the standard CLDO4 product, and the variable attribute specifies the destriping status and method.

(3) A bug in Version 3, where surface pressure indices were offset by one GEOS-CF grid box, has been corrected. The bug mainly affected OCP in regions with large topography variations.
(4) New GLER tables for land are used which better represent diurnal variability of the surface reflectance. These contribute to improvements in the ECF histogram.
(5) The GEOS-CF model used to provide meteorological inputs is updated from v1 to v2.
(6) Improved $O_2$-$O_2$ convergence rate to reduce the number of missing values.
(7) The $O_2$-$O_2$ SCD temperature correction uses the temperature at 0.707 * OCP during iteration in V4, while 0.7937 * OCP was used in Version 3. The factor 0.707 mathematically corresponds to the half mass point of the $O_2$-$O_2$ column above clouds. However, whether it is the best value to use still awaits further investigation.
(8) Restrict the maximum solar wavelength shift and radiance wavelength shift to 0.1 nm (instead of 0.5 nm), as large shift is usually associated with unreliable $O_2$-$O_2$ retrievals.

## 4.5 Known issues in cloud product

(1) The calculated ECF value depends on assumed cloud albedo. Reasonable choices of cloud albedo are 0.8 - 0.9 based on previous studies for polar orbiting satellites. Currently, a = 0.8 is used in both TEMPO CLDO4 and TEMPO trace gases for consistency and backward compatibility. A larger cloud albedo will lead to a lower ECF which in turn changes OCP and trace gas AMF.
(2) ECF retrieval fails when the measured reflectance is outside and far from the range set by the calculated (*i.e.*, expected) clear-sky and overcast reflectance for the same scene under the assumptions made. Failure potentially suggests unrealistic ancillary inputs or invalid assumption of cloud albedo. When the ECF retrieval fails, OCP cannot be retrieved.
(3) In general, cloud retrieval is challenging when the contrast between cloud albedo and surface albedo becomes small and for large $AMF_g$. For instance, ECF retrievals tend to fail over very bright surfaces, e.g., glint or bright snow and ice.
(4) Spectral saturation for the relevant wavelengths can lead to errors in ECF, failed $O_2$-$O_2$ fitting, or large $O_2$-$O_2$ fitting uncertainties. When $O_2$-$O_2$ fitting fails, OCP cannot be derived, and ECF should be used with caution.
(5) Over bright cloud tops, OCP retrieval fails due to spectral saturation (e.g., hurricanes), ECF values appear lower than expected (<1) on at least some occasions, these ECFs may have been affected by saturation so that the measured RAD (and subsequently derived ECF) is not as high as it should be.
(6) GLER may still not be able to fully account for the true surface albedo dependence on geometry. Stationary surface features (over snow/ice, non-snow/ice, mountainous or

coastal areas) could sometimes show up in ECF maps, which suggests possible problems with the GLERs.

(7) $O_2$-$O_2$ fitting tends to have larger uncertainty for larger SZA or for partly cloudy scenes.

(8) Cloud retrievals with large SZA (>70°) have large uncertainties and should be used with caution.

(9) ECF and OCP are not as reliable for very small ECF. In the literature, pixels with small ECF (< 0.1 or 0.05) are sometimes treated as clear in trace gas AMF calculations, although this approach is not currently implemented for TEMPO trace gas products. Similarly, OCP for small ECF is sometimes set to scene pressure as the cloud pressure retrieved under this condition tends to have large uncertainty. TEMPO cloud pressure is set to scene pressure when ECF < 0.05.

(10) Cloud shadows are noticeable when SZA is large, *e.g.*, in morning and evening scans. CLDO4 retrievals do not consider any shadows. There are currently no flags for cloud shadows in the product.

(11) CLDO4 retrievals do not explicitly consider aerosols. Any contribution from aerosols (background or episode) would be aliased into the ECF and OCP. For example, thick smoke from wildfires is frequently aliased as clouds in ECF maps.

(12) High-altitude cloud retrievals are in general less robust than mid- and low-altitude cloud retrievals.

(13) LUTs for clouds were calculated using a single US standard air temperature-pressure (T-P) profile and pre-flight slit function. Corrections for the deviations from these conditions have not been performed.

(14) *A priori* meteorological fields are from GEOS-CF data that are available at the time of processing. Forward processing has to use forecasts which are generally less accurate than reanalysis. Furthermore, downscaling from GEOS-CF to TEMPO pixel resolution is through simple space-time interpolation of instantaneous fields which ignores finer resolution dynamical influence.

(15) Cloud pressure depends on cloud fraction in the CLDO4 algorithm. Therefore, ECF and OCP are closely related and work best in pairs. When OCP is unavailable, TEMPO trace gas algorithm replaces it with a climatology which can lead to large errors.

(16) For forward processing, whenever a new IRR file is used for fitting, SDPC cannot use the previous day’s granules to derive a destriping vector due to changes in stripe pattern. The granules on the day of the IRR change will thus not be destriped using the traditional method. Destriping works well for most stripes, but could miss some occasionally, especially near the Pacific portion during some time of day, as traditional destriping works better for medium SCDs than for small and large SCDs.

# References

Geddes, J. A., et al. (2018). Stratosphere–troposphere separation of nitrogen dioxide columns from the TEMPO geostationary satellite instrument, *Atmospheric Measurement Techniques*, 11, 6271–6287, doi:10.5194/amt-11-6271-2018.

Gonzalez Abad, G. et al. (2015). Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval. *Atmospheric Measurement Techniques*, 8, 19-32, doi:10.5194/amt-8-19-2015.

Hartmann, D. (2016). *Global Physical Climatology*. Australia: Elsevier. pp. 76–78. ISBN 978-0-12-328531-7.

Martin, R. V. et al. (2002). An improved retrieval of tropospheric nitrogen dioxide from GOME. *Journal of Geophysical Research: Atmospheres*, 107, D20, doi:10.1029/2001JD001027.

Nowlan, C. et al. (2025). TEMPO Near Real-Time Data Products: User Guide and ATBD Supplement (Version 1.0). *https://asdc.larc.nasa.gov/documents/tempo/guide/TEMPO_NRT_user_guide_supplement_V1.0.pdf*

Stammes, P., Sneep, M., de Haan, J. F., Veefkind, J. P., Wang, P., and Levelt, P. F. (2008). Effective cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and validation. *Journal of Geophysical Research*, 113, D16S38, doi:10.1029/2007JD008820.

Vasilkov, A. et al. (2018). A cloud algorithm based on the $O_2$-$O_2$ 477 nm absorption band featuring an advanced spectral fitting method and the use of surface geometry-dependent Lambertian-equivalent reflectivity. *Atmospheric Measurement Techniques*, 11, 4093-4107, doi:10.5194/amt-11-4093-2018.

TEMPO Validation Team (2023). Tropospheric Emissions: Monitoring of Pollution (TEMPO) – Level 2 Science Data Product Validation Plan, SAO-DRD-11.

van Geffen, J. H. G. M., Eskes, H. J., Boersma, K. F., and J. P. Veefkind (2022). TROPOMI ATBD of the total and tropospheric $NO_2$ data products, Issue 2.4.0, Document number 5P-KNMI-L2-0005-RP.

Zoogman, P., Liu, X., Suleiman, R., Pennington, W., Flittner, D., Al-Saadi, J. et al. (2017). Tropospheric Emissions: Monitoring of Pollution (TEMPO). *Journal of Quantitative Spectroscopy and Radiative Transfer*, 186, 17–39. doi:10.1016/j.jqsrt.2016.05.008

# Appendix A: List of acronyms

**Table A1.** List of acronyms

| Acronym | Meaning |
|---|---|
| AMF | Air mass factor |
| ASDC | Atmospheric Science Data Center |
| ECF | Effective cloud fraction |
| EOSDIS | Earth Observing System Data and Information System |
| GEOS-CF | Goddard Earth Observing System Composition Forecasting |
| GEMS | Geostationary Environment Monitoring Spectrometer |
| GLER | Geometry dependent Lambertian Equivalent Reflectance |
| GPS | Global Positioning System |
| IMS | Interactive Multisensor Snow and Ice Mapping System |
| L1 | Level 1 |
| L2 | Level 2 |
| L3 | Level 3 |
| LUT | Look-up table |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| NASA | National Aeronautics and Space Administration |
| NE | Northeast |
| NRT | Near real-time |
| NW | Northwest |
| OCP | Optical centroid pressure |
| OMI | Ozone Monitoring Instrument |
| PBL | Planetary boundary layer |
| SAO | Smithsonian Astrophysical Observatory |
| SCD | Slant column density |
| SDPC | Science Data Processing Center |
| SE | Southeast |
| SW | Southwest |
| SZA | Solar zenith angle |
| TEMPO | Tropospheric Emissions: Monitoring of Pollution |
| TROPOMI | TROPOspheric Monitoring Instrument |
| UTC | Coordinated Universal Time |
| VCD | Vertical column density |

# Appendix B: $NO_2$ file description

**Table B1.** Full list of variables in $NO_2$ Level 2 files

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | xtrack | int (xtrack) | none | Pixel index along slit |
| | mirror_step | int (mirror_step) | none | Scan mirror position index |
| product | main_data_quality_flag | short (mirror_step, xtrack) | none | Main data quality flag |
| product | vertical_column_stratosphere | double (mirror_step, xtrack) | molecules/cm$^2$ | Stratosphere $NO_2$ vertical column |
| product | vertical_column_troposphere | double (mirror_step, xtrack) | molecules/cm$^2$ | Troposphere $NO_2$ vertical column |
| product | vertical_column_troposphere_uncertainty | double (mirror_step, xtrack) | molecules/cm$^2$ | Troposphere $NO_2$ vertical column uncertainty |
| geolocation | latitude | float (mirror_step, xtrack) | degrees north | Pixel center latitude |
| geolocation | latitude_bounds | float (mirror_step, xtrack, corner) | degrees north | Pixel corner latitude (SW, SE, NE, NW) |
| geolocation | longitude | float (mirror_step, xtrack) | degrees east | Pixel center longitude |
| geolocation | longitude_bounds | float (mirror_step, xtrack, corner) | degrees east | Pixel corner longitude (SW, SE, NE, NW) |
| geolocation | relative_azimuth_angle | float (mirror_step, xtrack) | degrees | Relative azimuth angle at pixel center |
| geolocation | solar_azimuth_angle | float (mirror_step, xtrack) | degrees | Solar azimuth angle at pixel center |

| | | | | |
|---|---|---|---|---|
| geolocation | solar_zenith_angle | float (mirror_step, xtrack) | degrees | Solar zenith angle at pixel center |
| geolocation | time | double (mirror_step) | seconds since 1980-01-06T00:00:00Z | Radiance exposure start time |
| geolocation | viewing_azimuth_angle | float (mirror_step, xtrack) | degrees | Viewing azimuth angle at pixel center |
| geolocation | viewing_zenith_angle | float (mirror_step, xtrack) | degrees | Viewing zenith angle at pixel center |
| qa_statistics | fit_convergence_flag | short (mirror_step, xtrack) | none | Radiance fit convergence flag |
| qa_statistics | fit_rms_residual | float (mirror_step, xtrack) | none | Radiance fit RMS |
| support_data | albedo | float (mirror_step, xtrack) | none | Surface albedo |
| support_data | amf_cloud_fraction | float (mirror_step, xtrack) | none | Cloud radiance fraction for AMF calculation |
| support_data | amf_cloud_pressure | float (mirror_step, xtrack) | hPa | Cloud pressure for AMF calculation |
| support_data | amf_diagnostic_flag | short (mirror_step, xtrack) | none | Bitwise air mass factor calculation diagnostic flag |
| support_data | amf_stratosphere | float (mirror_step, xtrack) | none | $NO_2$ stratosphere air mass factor |
| support_data | amf_stratosphere_clear_sky | float (mirror_step, xtrack) | none | $NO_2$ clear sky stratosphere air mass factor |

| | | | | |
|---|---|---|---|---|
| support_data | amf_total | float (mirror_step, xtrack) | none | $NO_2$ total air mass factor |
| support_data | amf_total_clear_sky | float (mirror_step, xtrack) | none | $NO_2$ clear sky total air mass factor |
| support_data | amf_troposphere | float (mirror_step, xtrack) | none | $NO_2$ troposphere air mass factor |
| support_data | amf_troposphere_clear_sky | float (mirror_step, xtrack) | none | $NO_2$ clear sky troposphere air mass factor |
| support_data | destriping_correction | float (mirror_step, xtrack) | molecules/cm$^2$ | Across track dependent $NO_2$ slant column destriping correction |
| support_data | eff_cloud_fraction | float (mirror_step, xtrack) | none | Effective cloud fraction |
| support_data | fitted_slant_column | double (mirror_step, xtrack) | molecules/cm$^2$ | $NO_2$ fitted slant column |
| support_data | fitted_slant_column_uncertainty | double (mirror_step, xtrack) | molecules/cm$^2$ | $NO_2$ fitted slant column uncertainty |
| support_data | fitted_slant_column_uncorrected | double (mirror_step, xtrack) | molecules/cm$^2$ | $NO_2$ fitted slant column before destriping correction |
| support_data | gas_profile | float (mirror_step, xtrack, swt_level) | molecules/cm$^2$ | Vertical profile of $NO_2$ partial columns |
| support_data | ground_pixel_quality_flag | int (mirror_step, xtrack) | none | Ground pixel quality flag (ground type) |
| support_data | pbl_height | short (mirror_step, xtrack) | m | Planetary boundary layer height |
| support_data | scattering_weights | float (mirror_step, | none | Vertical profile of |

| | | | | |
|---|---|---|---|---|
| | | xtrack, swt_level) | | scattering weights |
| support_data | scattering_weights_clear_sky | float (mirror_step, xtrack, swt_level) | none | Vertical profile of clear-sky scattering weights |
| support_data | snow_ice_fraction | float (mirror_step, xtrack) | none | Fraction of pixel area covered by snow and/or ice |
| support_data | surface_pressure | float (mirror_step, xtrack) | hPa | Surface pressure |
| support_data | temperature_profile | float (mirror_step, xtrack, swt_level) | K | Vertical profile of air temperature |
| support_data | terrain_height | short (mirror_step, xtrack) | m | Terrain height |
| support_data | tropopause_pressure | float (mirror_step, xtrack) | hPa | Tropopause pressure |
| support_data | vertical_column_total | double (time, latitude, longitude) | molecules/cm$^2$ | $NO_2$ vertical column uncertainty |
| support_data | vertical_column_total_uncertainty | double (time, latitude, longitude) | molecules/cm$^2$ | $NO_2$ vertical column |
| support_data | wind_speed | float (mirror_step, xtrack) | m/s | 2-meter wind speed |

**Table B2:** Full list of variables in $NO_2$ Level 3 files

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | latitude | float (latitude) | degrees north | Latitude at grid box center |
| | longitude | float (longitude) | degrees east | Longitude at grid box center |
| | weight | float (latitude, longitude) | $km^2$ | Sum of Level 2 pixel overlap areas |
| | time | double (time) | seconds since 1980-01-06T00:00:00Z | Scan start time |
| product | main_data_quality_flag | short (time, latitude, longitude) | none | Main data quality flag |
| product | vertical_column_stratosphere | double (time, latitude, longitude) | molecules/$cm^2$ | $NO_2$ stratosphere vertical column uncertainty |
| product | vertical_column_troposphere | double (time, latitude, longitude) | molecules/$cm^2$ | $NO_2$ troposphere vertical column |
| product | vertical_column_troposphere_uncertainty | double (time, latitude, longitude) | molecules/$cm^2$ | $NO_2$ troposphere vertical column uncertainty |
| geolocation | relative_azimuth_angle | float (time, latitude, longitude) | degrees | Relative azimuth angle at pixel center |
| geolocation | solar_zenith_angle | float (time, latitude, longitude) | degrees | Solar zenith angle at pixel center |
| geolocation | viewing_zenith_angle | float (time, latitude, longitude) | degrees | Viewing zenith angle at pixel center |
| support_data | albedo | float (time, latitude, longitude) | none | Surface albedo |

| | | | | |
|---|---|---|---|---|
| support_data | amf_cloud_fraction | float (time, latitude, longitude) | none | Cloud radiance fraction for AMF calculation |
| support_data | amf_cloud_pressure | float (time, latitude, longitude) | hPa | Cloud pressure for AMF calculation |
| support_data | amf_stratosphere | float (time, latitude, longitude) | none | $NO_2$ stratosphere air mass factor |
| support_data | amf_total | float (time, latitude, longitude) | none | $NO_2$ air mass factor |
| support_data | amf_troposphere | float (time, latitude, longitude) | none | $NO_2$ troposphere air mass factor |
| support_data | eff_cloud_fraction | float (time, latitude, longitude) | none | Effective cloud fraction |
| support_data | fitted_slant_column | double (time, latitude, longitude) | molecules/cm$^2$ | $NO_2$ fitted slant column |
| support_data | fitted_slant_column_uncertainty | double (time, latitude, longitude) | molecules/cm$^2$ | $NO_2$ fitted slant column uncertainty |
| support_data | pbl_height | short (mirror_step, xtrack) | m | Planetary boundary layer height |
| support_data | snow_ice_fraction | float (time, latitude, longitude) | none | Fraction of pixel area covered by snow and/or ice |
| support_data | surface_pressure | float (time, latitude, longitude) | hPa | Surface pressure |
| support_data | terrain_height | short (time, latitude, longitude) | m | Terrain height |
| support_data | tropopause_pressure | float (time, latitude, longitude) | hPa | Tropopause pressure |

| | | | | |
|---|---|---|---|---|
| support_data | vertical_column_total | double (time, latitude, longitude) | molecules/cm$^2$ | $NO_2$ vertical column |
| support_data | vertical_column_total_uncertainty | double (time, latitude, longitude) | molecules/cm$^2$ | $NO_2$ vertical column uncertainty |
| qa_statistics | num_vertical_column_troposphere_samples | int (time, latitude, longitude) | none | Number of Level 2 pixel values contributing to Level 3 grid |
| qa_statistics | min_vertical_column_troposphere_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Smallest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | max_vertical_column_troposphere_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Largest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | num_vertical_column_troposphere_uncertainty_samples | int (time, latitude, longitude) | none | Number of Level 2 pixel values contributing to Level 3 grid |
| qa_statistics | min_vertical_column_troposphere_uncertainty_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Smallest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | max_vertical_column_troposphere_uncertainty_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Largest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | num_vertical_column_stratosphere_samples | int (time, latitude, longitude) | none | Number of Level 2 pixel values contributing to Level 3 grid |

| | | | | |
|---|---|---|---|---|
| qa_statistics | min_vertical_column_stratosphere_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Smallest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | max_vertical_column_stratosphere_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Largest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | num_vertical_column_total_samples | int (time, latitude, longitude) | none | Number of Level 2 pixel values contributing to Level 3 grid |
| qa_statistics | min_vertical_column_total_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Smallest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | max_vertical_column_total_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Largest Level 2 pixel value contributing to Level 3 grid |

# Appendix C: HCHO file description

**Table C1.** Full list of variables in HCHO Level 2 files

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | xtrack | int (xtrack) | none | Pixel index along slit |
| | mirror_step | int (mirror_step) | none | Scan mirror position index |
| product | main_data_quality_flag | short (mirror_step, xtrack) | none | Main data quality flag |
| product | vertical_column | double (mirror_step, xtrack) | molecules/cm$^2$ | HCHO vertical column |
| product | vertical_column_uncertainty | double (mirror_step, xtrack) | molecules/cm$^2$ | HCHO vertical column uncertainty |
| geolocation | latitude | float (mirror_step, xtrack) | degrees north | Pixel center latitude |
| geolocation | latitude_bounds | float (mirror_step, xtrack, corner) | degrees north | Pixel corner latitude (SW, SE, NE, NW) |
| geolocation | longitude | float (mirror_step, xtrack) | degrees east | Pixel center longitude |
| geolocation | longitude_bounds | float (mirror_step, xtrack, corner) | degrees east | Pixel corner longitude (SW, SE, NE, NW) |
| geolocation | relative_azimuth_angle | float (mirror_step, xtrack) | degrees | Relative azimuth angle at pixel center |
| geolocation | solar_azimuth_angle | float (mirror_step, xtrack) | degrees | Solar azimuth angle at pixel center |
| geolocation | solar_zenith_angle | float (mirror_step, xtrack) | degrees | Solar zenith angle at pixel center |

| | | | | |
|---|---|---|---|---|
| geolocation | time | double (mirror_step) | seconds since 1980-01-06T00:00:00Z | Radiance exposure start time |
| geolocation | viewing_azimuth_angle | float (mirror_step, xtrack) | degrees | Viewing azimuth angle at pixel center |
| geolocation | viewing_zenith_angle | float (mirror_step, xtrack) | degrees | Viewing zenith angle at pixel center |
| qa_statistics | fit_convergence_flag | short (mirror_step, xtrack) | none | Radiance fit convergence flag |
| qa_statistics | fit_rms_residual | float (mirror_step, xtrack) | none | Radiance fit RMS |
| support_data | albedo | float (mirror_step, xtrack) | none | Surface albedo |
| support_data | amf | float (mirror_step, xtrack) | none | HCHO air mass factor |
| support_data | amf_clear_sky | float (mirror_step, xtrack) | none | HCHO clear sky air mass factor |
| support_data | amf_cloud_fraction | float (mirror_step, xtrack) | none | Cloud radiance fraction for AMF calculation |
| support_data | amf_cloud_pressure | float (mirror_step, xtrack) | hPa | Cloud pressure for AMF calculation |
| support_data | amf_diagnostic_flag | short (mirror_step, xtrack) | none | Bitwise air mass factor calculation diagnostic flag |
| support_data | background_correction | float (mirror_step, xtrack) | molecules/cm$^2$ | HCHO background correction |
| support_data | eff_cloud_fraction | float (mirror_step, xtrack) | none | Effective cloud fraction |

| | | | | |
|---|---|---|---|---|
| support_data | fitted_slant_column | double (mirror_step, xtrack) | molecules/cm$^2$ | HCHO fitted slant column |
| support_data | fitted_slant_column_uncertainty | double (mirror_step, xtrack) | molecules/cm$^2$ | HCHO fitted slant column uncertainty |
| support_data | gas_profile | float (mirror_step, xtrack, swt_level) | molecules/cm$^2$ | Vertical profile of HCHO partial columns |
| support_data | ground_pixel_quality_flag | int (mirror_step, xtrack) | none | Ground pixel quality flag (ground type) |
| support_data | pbl_height | short (mirror_step, xtrack) | m | Planetary boundary layer height |
| support_data | scattering_weights | float (mirror_step, xtrack, swt_level) | none | Vertical profile of scattering weights |
| support_data | scattering_weights_clear_sky | float (mirror_step, xtrack, swt_level) | none | Vertical profile of clear-sky scattering weights |
| support_data | snow_ice_fraction | float (mirror_step, xtrack) | none | Fraction of pixel area covered by snow and/or ice |
| support_data | surface_pressure | float (mirror_step, xtrack) | hPa | Surface pressure |
| support_data | temperature_profile | float (mirror_step, xtrack, swt_level) | K | Vertical profile of air temperature |
| support_data | terrain_height | short (mirror_step, xtrack) | m | Terrain height |
| support_data | wind_speed | float (mirror_step, xtrack) | m/s | 2-meter wind speed |

**Table C2.** Full list of variables in HCHO Level 3 files

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | longitude | float (longitude) | degrees east | Longitude at grid box center |
| | latitude | float (latitude) | degrees north | Latitude at grid box center |
| | weight | float (latitude, longitude) | $km^2$ | Sum of Level 2 pixel overlap areas |
| | time | double (time) | seconds since 1980-01-06T00:00:00Z | Scan start time |
| product | main_data_quality_flag | short (time, latitude, longitude) | none | Main data quality flag |
| product | vertical_column | double (time, latitude, longitude) | $molecules/cm^2$ | HCHO vertical column |
| product | vertical_column_uncertainty | double (time, latitude, longitude) | $molecules/cm^2$ | HCHO vertical column uncertainty |
| geolocation | solar_zenith_angle | float (time, latitude, longitude) | degrees | Solar zenith angle at pixel center |
| geolocation | viewing_zenith_angle | float (time, latitude, longitude) | degrees | Viewing zenith angle at pixel center |
| geolocation | relative_azimuth_angle | float (time, latitude, longitude) | degrees | Relative azimuth angle at pixel center |
| support_data | albedo | float (time, latitude, longitude) | none | Surface albedo |
| support_data | amf | float (time, latitude, longitude) | none | HCHO air mass factor |
| support_data | amf_cloud_fraction | float (time, latitude, longitude) | none | Cloud radiance fraction for AMF calculation |

| | | | | |
|---|---|---|---|---|
| support_data | amf_cloud_pressure | float (time, latitude, longitude) | hPa | Cloud pressure for AMF calculation |
| support_data | eff_cloud_fraction | float (time, latitude, longitude) | none | Effective cloud fraction |
| support_data | fitted_slant_column | double (time, latitude, longitude) | molecules/cm$^2$ | HCHO fitted slant column |
| support_data | fitted_slant_column_uncertainty | double (time, latitude, longitude) | molecules/cm$^2$ | HCHO fitted slant column uncertainty |
| support_data | pbl_height | short (mirror_step, xtrack) | m | Planetary boundary layer height |
| support_data | snow_ice_fraction | float (time, latitude, longitude) | none | Fraction of pixel area covered by snow and/or ice |
| support_data | surface_pressure | float (time, latitude, longitude) | hPa | Surface pressure |
| support_data | terrain_height | short (time, latitude, longitude) | m | Terrain height |
| qa_statistics | num_vertical_column_samples | int (time, latitude, longitude) | none | Number of Level 2 pixel values contributing to Level 3 grid |
| qa_statistics | min_vertical_column_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Smallest Level 2 pixel value contributing to Level 3 grid |
| qa_statistics | max_vertical_column_sample | double (time, latitude, longitude) | molecules/cm$^2$ | Largest Level 2 pixel value contributing to Level 3 grid |

# Appendix D: CLDO4 file description

**Table D1.** List of variables in CLDO4 Level 2 files

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | xtrack | int (xtrack) | none | Pixel index along slit |
| | mirror_step | int (mirror_step) | none | Scan mirror position index |
| product | cloud_fraction | float (mirror_step, xtrack) | none | Effective cloud fraction at 466 nm |
| product | cloud_pressure | short (mirror_step, xtrack) | hPa | Optical centroid pressure for cloud |
| product | CloudRadianceFraction440 | float (mirror_step, xtrack) | none | Cloud radiance fraction at 440 nm |
| product | CloudRadianceFraction466 | float (mirror_step, xtrack) | none | Cloud radiance fraction at 466 nm |
| product | processing_quality_flag | short (mirror_step, xtrack) | none | 16-bit processing quality flag |
| geolocation | time | double (mirror_step) | seconds since 1980-01-06T00:00:00Z | Radiance exposure start time |
| geolocation | latitude | float (mirror_step, xtrack) | degrees north | Pixel center latitude |
| geolocation | latitude_bounds | float (mirror_step, xtrack, corner) | degrees north | Pixel corner latitude (SW, SE, NE, NW) |
| geolocation | longitude | float (mirror_step, xtrack) | degrees east | Pixel center longitude |
| geolocation | longitude_bounds | float (mirror_step, xtrack, corner) | degrees east | Pixel corner longitude (SW, SE, NE, NW) |
| geolocation | solar_zenith_angle | float (mirror_step, xtrack) | degrees | Solar zenith angle at pixel center |

| | | | | |
|---|---|---|---|---|
| geolocation | solar_azimuth_angle | float (mirror_step, xtrack) | degrees | Solar azimuth angle at pixel center |
| geolocation | viewing_zenith_angle | float (mirror_step, xtrack) | degrees | Viewing zenith angle at pixel center |
| geolocation | viewing_azimuth_angle | float (mirror_step, xtrack) | degrees | Viewing azimuth angle at pixel center |
| geolocation | relative_azimuth_angle | float (mirror_step, xtrack) | degrees | Relative azimuth angle at pixel center |
| support_data | GLER440 | float (mirror_step, xtrack) | none | 440nm surface reflectivity used in calculation |
| support_data | GLER466 | float (mirror_step, xtrack) | none | 466nm surface reflectivity used in calculation |
| support_data | SCD_MainDataQualityFlag | int (mirror_step, xtrack) | none | Main data quality flag for fitted_slant_column |
| support_data | SceneLER440 | float (mirror_step, xtrack) | none | 440nm reflectance calculated at ScenePressure |
| support_data | SceneLER466 | float (mirror_step, xtrack) | none | 466nm reflectance calculated at ScenePressure |
| support_data | ScenePressure | float (mirror_step, xtrack) | hPa | Scene pressure |
| support_data | surface_pressure | float (mirror_step, xtrack) | hPa | Surface pressure |
| support_data | fitted_slant_column | double (mirror_step, xtrack) | molecules/$cm^2$ | Collision induced oxygen complex fitted slant column |
| support_data | fitted_slant_column_uncertainty | double (mirror_step, xtrack) | molecules/$cm^2$ | Collision induced oxygen fitted slant column uncertainty |
| support_data | scddes | double (mirror_step, xtrack) | molecules/$cm^2$ | Destriped fitted_slant_column |
| support_data | snow_ice_fraction | float (mirror_step, xtrack) | none | Fraction of pixel area covered by snow and/or ice |
| support_data | pbl_height | short (mirror_step, xtrack) | m | Planetary boundary layer height |

| | | | | |
|---|---|---|---|---|
| support_data | terrain_height | short (mirror_step, xtrack) | m | Terrain height |
| support_data | ground_pixel_quality_flag | int (mirror_step, xtrack) | none | Ground pixel quality flag (ground type) |
| qa_statistics | fit_rms_residual | float (mirror_step, xtrack) | none | Radiance fit RMS |
| qa_statistics | fit_convergence_flag | short (mirror_step, xtrack) | none | Radiance fit convergence flag |

**Table D2.** List of variables in CLD04 Level 3 files

| Group | Name | Format (dimension) | Unit | Description |
|---|---|---|---|---|
| | longitude | float (longitude) | degrees east | Longitude at grid box center |
| | latitude | float (latitude) | degrees north | Latitude at grid box center |
| | weight | float (latitude, longitude) | $km^2$ | Sum of Level 2 pixel overlap areas |
| | time | double (time) | seconds since 1980-01-06T00:00:00Z | Scan start time |
| product | cloud_fraction | float (time, latitude, longitude) | none | Effective cloud fraction at 466 nm |
| product | cloud_pressure | float (time, latitude, longitude) | hPa | Optical centroid pressure for cloud |
| product | CloudRadianceFraction440 | float (time, latitude, longitude) | none | Cloud radiance fraction at 440 nm |
| product | CloudRadianceFraction466 | float (time, latitude, longitude) | none | Cloud radiance fraction at 466 nm |
| geolocation | solar_zenith_angle | float (time, latitude, longitude) | degrees | Solar zenith angle at pixel center |
| geolocation | viewing_zenith_angle | float (time, latitude, longitude) | degrees | Viewing zenith angle at pixel center |
| geolocation | relative_azimuth_angle | float (time, latitude, longitude) | degrees | Relative azimuth angle at pixel center |

| | | | | |
|---|---|---|---|---|
| support_data | GLER440 | float (time, latitude, longitude) | none | 440nm surface reflectivity used in calculation |
| support_data | GLER466 | float (time, latitude, longitude) | none | 466nm surface reflectivity used in calculation |
| support_data | surface_pressure | float (mirror_step, xtrack) | hPa | Surface pressure |