CENTER FOR ASTROPHYSICS
HARVARD & SMITHSONIAN

VERSION: 1.0
RELEASE DATE: May 20, 2024

TROPOSPHERIC EMISSIONS: MONITORING OF POLLUTION (TEMPO) PROJECT

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

May 20, 2024

Distributed by the Atmospheric Science Data Center
https://asdc.larc.nasa.gov

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

PRODUCT and SCIENCE DATA PROCESSING CENTER PIPELINE VERSION

Product Version Designation	Science Data Processing Center Pipeline Version	Release
V03	4.4	May 20, 2024; first public release of TEMPO Level 2 and Level 3 products

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 monitor air quality over North America during daylight hours at high spatial resolution (2×4.75 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 the NASA Atmospheric Science Data Center (ASDC) for public distribution. 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 filename 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 constitute the first public release of TEMPO Level 2 and 3 products. As of 20 May 2024, these are declared at the Beta maturity level. The TEMPO validation document defines the Beta level as “the product is minimally validated but may still contain significant errors; based on product quick looks using the initial calibration parameters. Publication of research based on Beta maturity products is not recommended and highly discouraged." [TEMPO Validation Team, 2023]. The maturity is anticipated to be declared provisional after validation team approval. Users should consult the relevant data landing pages for current product maturity.

1.4. User guide description

This document provides a user guide for the V03 Level 2 and Level 3 trace gas products nitrogen dioxide (NO2) and formaldehyde (HCHO), and V03 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 Document (ATBD), available on the product's data set landing webpage 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	Beta
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	Beta
HCHO	2	Formaldehyde vertical columns	Native TEMPO ground pixel, one granule	Beta
HCHO	3	Formaldehyde vertical columns on regular grid	All Level 2 granules merged for a single scan, on 0.02° × 0.02° regular grid	Beta
CLDO4	2	Effective cloud fraction (ECF) and cloud optical centroid pressure (OCP)	Native TEMPO ground pixel, one granule	Beta

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	Beta

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 files 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. In some cases, Level 2 trace gas and cloud scans may be missing early granules (e.g., G01) when twilight (city light) measurements 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 resolution 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_V03_{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 3 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_V03_20240510T001504Z_S017G03.nc

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

TEMPO_{GAS}_L3_V03_{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_V03_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc
NO2 Level 3	TEMPO_NO2_L3_V03_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}.nc
HCHO Level 2	TEMPO_HCHO_L2_V03_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc
HCHO Level 3	TEMPO_HCHO_L3_V03_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}.nc
CLDO4 Level 2	TEMPO_CLDO4_L2_V03_{YYYY}{MM}{DD}T{HH}{NN}{SS}Z _S{XXX}G{YY}.nc
CLDO4 Level 3	TEMPO_CLDO4_L3_V03_{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 V03, *swt_level* = 72, corresponding to the number of vertical layers in the GEOS-CF 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.

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 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 three parts: (1) NO_2 SCD retrieval, (2) AMF calculation, and (3) 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 V1 forecasts that are available at the time of SDPC processing. These are instantaneous hourly forecasts sampled at 0.25° longitude × 0.25° latitude (from the model run on a cubed-sphere at c360 horizontal resolution). Typically, forecasts within 24-hour of the TEMPO observations are used. Hourly climatology for each month provides a backup if the forecast is not available, though this rarely occurs.

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 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,	degrees north	Pixel corner latitude (SW,

		xtrack, corner)		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 extreme care. 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, SCD_{min} = -1e19 and SCD_{max} = 1e19 molecules/cm^2. For HCHO, SCD_{min} = -5e17 and SCD_{max} = 5e17 molecules/cm^2.

Value	Meaning	Description
0	good	fit_convergence_flag = 1 *AND* (SCD_{min} <= SCD <= SCD_{max}) *AND* ((VCD_{total} + 2*VCD_{total})*AMF_{total} >= 0) *AND* AMF_g <= 6 *AND* AMF_{total} >= 0.1
1	suspect	fit_convergence_flag = 0 *OR* (((VCD_{total} + 2*VCD_{total})*AMF_{total} < 0) *AND* ((VCD_{total} + 3*VCD_{total})*AMF_{total} >= 0)) *OR* (SCD < SCD_{min}) *OR* (SCD > SCD_{max}) *OR* AMF_g > 6 *OR* AMF_{total} < 0.1
2	bad	fit_convergence_flag < 0 *OR* ((VCD_{total} + 3*VCD_{total})*AMF_{total} < 0) *OR* 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 (time, latitude, longitude)	none	Main data quality flag
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 potential 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 the TEMPO to lower tropospheric concentrations of 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 V03 spectral fitting degrades over partially cloudy pixels due to the inhomogeneous illumination of the instrument slit, resulting in large uncertainties and, in the case of HCHO, large positive biases in the retrieved dSCD. 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*) < 0.2. More strict cloud fraction criterion (e.g., < 0.15) will retain less data, though the retained data will have less cloud influence. 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* and the *EtaA* and *EtaB* values which are attributes of the surface pressure variable. The pressure at each altitude level *z* can be calculated using:

$$P_{edge}(z) = EtaA(z) + EtaB(z) * P_{surface}$$

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 the GEOS-CF forecast, 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, which primarily result from errors in clouds and surface albedo. 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:

$$AMF = \frac{\sum_i W_i \Omega_i \alpha_i}{\sum_i \Omega_i}$$

in which W_i, Ω_i and α_i represent the scattering weight, the trace gas partial column and the temperature correction (which accounts for the use of a single temperature cross section in the spectral fit at atmospheric layer *i*). 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 should apply $\alpha_i = 1$ 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]:

$$\propto_i = 1 - a(T_i - T_0) + b(T_i - T_0)^2$$

with values $a = 0.00316$, $b = 3.39\times10^{-6}$, $T_0 = 220\text{K}$ and T_i = temperature at layer *i*.

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 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 please consult the ATBD document.

3.4 Known issues in trace gas products

3.4.1 Known issues affecting both NO_2 and HCHO

(1) The precision and accuracy of the retrieval deteriorates in the presence of clouds, due to inhomogeneity associated with partially cloudy pixels. This effect typically starts to

appear at effective cloud fractions between 0.1-0.2, peaks between 0.3-0.4 and stabilizes and/or reduces its influence for effective cloud fractions larger than 0.4.

(2) There are several unresolved issues in the cloud retrieval (discussed in the CLDO4 section of this document). Of these, the most significant is a high bias in the cloud fractions for clear or nearly clear sky pixels, possibly mainly due to a combination of a high bias in the TEMPO absolute radiances and low bias in the GLER.

(3) 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.

(4) 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.

(5) The accuracy of GLER look-up-tables derived from MODIS BRDF (v6.1) observations needs further investigation. Calculated GLER values at times of the day distant from the MODIS instrument overpass times may result in higher uncertainties of the GLER. 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 fraction and trace gas VCD.

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

(7) 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 correctly 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.

3.4.2 Known issues affecting only NO_2

(1) There is mild striping along the North-South direction in the SCD. A fix for this will be implemented in a future version.

3.4.3 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 calculation resulting in North-South stripes of missing data and artifacts, particularly at Northern latitudes.

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 3rd 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 proving 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.

Cloud fraction is derived from the sun-normalized spectral radiance at 466 nm. Cloud pressure is derived using cloud fraction, O_2-O_2 SCD, Look-Up Tables (LUTs), and ancillary data. The pre-computed LUTs of sun normalized radiances and AMFs 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. Geolocation information on viewing geometry, latitude, longitude, and pixel footprint bounds can be found in the *geolocation* group.

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 L1B user guide for information.
- *SCD_MainDataQualityFlag* in *support_data* is for O_2-O_2 SCD fitting instead of cloud information. Possible values are: 0 (normal), 1 (suspicious), and 2 (bad), where normal means that the retrieved SCD is in the range permitted and within 2 standard deviations of the fitting uncertainty.
- *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

16-bit integer are possible, fill_value should be ignored, and bitwise test 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 recommendations

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.

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. Preliminary analysis suggests that the RAD/IRR based on current L1B data may have a high bias by a few percent, which will lead to an overestimation of ECF in the CLDO4 product.

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] recommend a = 0.8 - 0.9 based on radiative transfer calculations and used 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 might be possible a different choice than a = 0.8 may be more appropriate for TEMPO given the differences in resolution and geometry. Consistent with the independent pixel approximation, trace gas AMF is composed of a clear-sky part and a cloudy-sky part. 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.

Biases and errors in *a priori* ancillary data will propagate into the retrieved cloud information which in turn affects the AMF through the cloudy sky scattering weights. Presumably, the MODIS based GLER are better constrained at the MODIS overpass time than at other local times, however, TEMPO observations span the whole of daytime hours. It remains to be quantified how the input 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.

To generate operational products that keep up with the continuously collected TEMPO data, SDPC uses GEOS-CF forecasts available at the time of processing. GEOS-CF has a coarser spatial resolution than a TEMPO pixel and the forecast errors in meteorological fields can affect the 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 Known issues in cloud product

(1) The Effective Cloud Fraction (ECF) histogram statistically peaks at 0.05 which deviates from the monotonic decreasing trend from 0 to 1 seen by sensors in polar orbits. The shift from zero could be related to bias in the L1B radiance/irradiance calibration and surface GLER. A small contribution may also come from the background aerosol. Although this issue is also seen in a version of clouds from the GEMS sensor (making similar measurements as TEMPO but over Asia), the shift implies that the TEMPO ECF is biased high. The bias seems to depend on the geometric air mass factor AMF_g (for which solar zenith angle SZA is a proxy).

(2) Preliminary analysis suggests that the 466 nm reflectance calculated from the L1B data may be biased high by a few percent, which can in turn lead to larger ECFs. In addition, if the specified GLER is too low, the cloud algorithm will compensate for it with increased ECF, and vice versa. The ECF histogram also depends on the assumed cloud albedo and reasonable choices are 0.8 - 0.9 based on previous studies. Currently, a = 0.8 is used in both TEMPO CLDO4 and TEMPO trace gases. A larger cloud albedo will lead to a lower ECF but must be applied consistently with the trace gases.

(3) 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. ECF retrievals tend to fail over very bright surfaces, e.g., glint or bright snow and ice. When the ECF retrieval fails, OCP cannot be retrieved. In general, cloud retrieval is challenging when the contrast between cloud albedo and surface albedo becomes small and for large AMF_g.

(4) Spectral saturation at 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.

(5) ECF in the mountainous terrain appears to show some local time dependence. This could be due to the current input GLER not fully able to account for the true surface albedo dependence on geometry. Similar problems could also exist elsewhere.

(6) Stationary surface features (over snow/ice, non-snow/ice, mountainous or coastal areas) sometimes show up in ECF maps, which suggests the GLERs used in the retrieval are too low for the affected regions.

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

(8) 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 AMF calculations, although this approach is not currently implemented for TEMPO trace gas products which always use the provided ECF. 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.

(9) 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.

(10) 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.

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

(12) 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.

(13) *A priori* meteorological fields are from GEOS-CF forecasts that are available each hour. Forecasts 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.

(14) Cloud pressure depends on cloud fraction in the CLDO4 algorithm. Therefore, ECF and OCP are closely related and work best in pairs.

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.org:10.1029/2001JD001027.

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
GEOS-CF	Goddard Earth Observing System Composition Forecasting
GEMS	Geostationary Environment Monitoring Spectrometer
GLER	Geometry dependent Lambertian Equivalent Reflectance
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
NW	Northwest
OCP	Optical centroid pressure
OMI	Ozone Monitoring Instrument
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_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
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	vertical_column_total	double (mirror_step, xtrack)	molecules/cm^2	NO_2 vertical column
support_data	vertical_column_total_uncertainty	double (mirror_step, xtrack)	molecules/cm^2	NO_2 vertical column uncertainty
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	snow_ice_fraction	float (mirror_step, xtrack)	none	Fraction of pixel area covered by snow and/or ice
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)
support_data	surface_pressure	float (mirror_step, xtrack)	hPa	Surface pressure
support_data	tropopause_pressure	float (mirror_step, xtrack)	hPa	Tropopause pressure
support_data	scattering_weights	float (mirror_step, xtrack, swt_level)	none	Vertical profile of scattering weights
support_data	gas_profile	float (mirror_step, xtrack, swt_level)	molecules/cm^2	Vertical profile of NO_2 partial columns

support_data	albedo	float (mirror_step, xtrack)	none	Surface albedo
support_data	temperature_profile	float (mirror_step, xtrack, swt_level)	K	Vertical profile of air temperature
support_data	amf_total	float (mirror_step, xtrack)	none	NO_2 total air mass factor
support_data	amf_diagnostic_flag	short (mirror_step, xtrack)	none	Bitwise air mass factor calculation diagnostic flag
support_data	eff_cloud_fraction	float (mirror_step, xtrack)	none	Effective cloud fraction
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_troposphere	float (mirror_step, xtrack)	none	NO_2 troposphere air mass factor
support_data	amf_stratosphere	float (mirror_step, xtrack)	none	NO_2 stratosphere air mass factor
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 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_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
product	vertical_column_stratosphere	double (time, latitude, longitude)	molecules/cm^2	NO_2 stratosphere 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	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	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	snow_ice_fraction	float (time, latitude, longitude)	none	Fraction of pixel area covered by snow and/or ice
support_data	terrain_height	short (time, latitude, longitude)	m	Terrain height
support_data	surface_pressure	float (time, latitude, longitude)	hPa	Surface pressure
support_data	tropopause_pressure	float (time, latitude, longitude)	hPa	Tropopause pressure
support_data	albedo	float (time, latitude, longitude)	none	Surface albedo
support_data	amf_total	float (time, latitude, longitude)	none	NO_2 air mass factor
support_data	eff_cloud_fraction	float (time, latitude, longitude)	none	Effective cloud fraction
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_troposphere	float (time, latitude, longitude)	none	NO_2 troposphere air mass factor
support_data	amf_stratosphere	float (time, latitude, longitude)	none	NO_2 stratosphere air mass factor
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	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 latitude (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	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	snow_ice_fraction	float (mirror_step, xtrack)	none	Fraction of pixel area covered by snow and/or ice
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)
support_data	surface_pressure	float (mirror_step, xtrack)	hPa	Surface pressure
support_data	scattering_weights	float (mirror_step, xtrack, swt_level)	none	Vertical profile of scattering weights
support_data	gas_profile	float (mirror_step, xtrack, swt_level)	molecules/cm^2	Vertical profile of HCHO partial columns
support_data	albedo	float (mirror_step, xtrack)	none	Surface albedo
support_data	temperature_profile	float (mirror_step, xtrack, swt_level)	K	Vertical profile of air temperature
support_data	amf	float (mirror_step, xtrack)	none	HCHO air mass factor

support_data	amf_diagnostic_flag	short (mirror_step, xtrack)	none	Bitwise air mass factor calculation diagnostic flag
support_data	eff_cloud_fraction	float (mirror_step, xtrack)	none	Effective cloud fraction
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	background_correction	float (mirror_step, xtrack)	molecules/cm^2	HCHO background correction
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 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	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	snow_ice_fraction	float (time, latitude, longitude)	none	Fraction of pixel area covered by snow and/or ice

support_data	terrain_height	short (time, latitude, longitude)	m	Terrain height
support_data	surface_pressure	float (time, latitude, longitude)	hPa	Surface pressure
support_data	albedo	float (time, latitude, longitude)	none	Surface albedo
support_data	amf	float (time, latitude, longitude)	none	HCHO air mass factor
support_data	eff_cloud_fraction	float (time, latitude, longitude)	none	Effective cloud fraction
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
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. Full 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	snow_ice_fraction	float (mirror_step, xtrack)	none	Fraction of pixel area covered by snow and/or ice
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. Full 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