TITLE:

Ozone_cci

Product User Guide

Version/Issue 1/2 (PUG v1.2)

Reference: Ozone_cci_PUG_01_02

Date of issue: 5/12/2015

Distributed to: Ozone_cci Consortium

WP Manager: V. Sofieva
WP Manager Organization: FMI

Other partners:
EOST: DLR-IMF, BIRA-IASP, RAL, KNMI, IUP, LATMOS, FMI, U. Saskatchewan, U. Chalmers, ULB
VALT: AUTH, NKUA, BIRA-IASB
CRG: DLR-PA, KNMI

This work is supported by the European Space Agency

DOCUMENT PROPERTIES

Title	PUG
Reference	Ozone_cci_PUG_01_01
Issue	01
Revision	02
Status	draft
Date of issue	5/12/2015
Document type	Deliverable

	FUNCTION	NAME	DATE	SIGNATURE
LEAD AUTHORS	Scientists	Ronald van der A Melanie Coldewey-Egbers Christophe Lerot Diego Loyola Jacob van Peet Richard Siddans Viktoria Sofieva Nabiz Rahpoe Klaus-Peter Heue Rosa Astoreca		
EDITOR		Viktoria Sofieva		
REVIEWED BY	ESA Technical Officer	Claus Zehner		
ISSUED BY	Project Coordinator	Michel van Roozendael		

DOCUMENT CHANGE RECORD

Issue	Revision	Date	Modified items	Observations
01	01	19/12/2013	All	First draft version
01	02	15/11/2015	Sections 3, 4, new sections 4.2, 6	

CONTENTS

Executive summary

The Product User Guideline (PUG) is a deliverable of the ESA Ozone_cci project (http://www.esa-ozone-cci.org/). The Ozone_cci project is one of twelve projects of ESA's Climate Change Initiative (CCI). The Ozone_cci project will deliver the Essential Climate Variable (ECV) Ozone in line with the "Systematic observation requirements for satellite-based products for climate" as defined by GCOS (Global Climate Observing System)in (GCOS-107 2006): "Product A.7: Profile and total column of ozone".

During the first 2 years of this project, which started 1st Sept 2010, a so-called Round Robin (RR) exercise has been conducted. During this phase several existing retrieval algorithms to produce vertical profiles and total columns of ozone from satellite observations were compared. During the last year of this project Ozone ECVs were generated.

The purpose of this document is to describe the ozone products generated in the framework of Ozone_cci, including a detailed description of the file format.

1 Applicable documents

Ozone_cci SoW
Ozone_cci DARD
Ozone_cci PSD
Ozone_cci URD
Ozone_cci ATBD
ESA CCI Project Guidelines

2 Overview of Ozone_cci products

The Ozone_cci includes data products for total ozone columns, ozone profiles from nadir sensors and stratospheric ozone profiles from limb and occultation sensors. All data sets are reported in NetCDF-4 CF format following CCI and GCOS standards, and are freely available on the Ozone_cci web site (http://www.esa-ozone-cci.org/?q=node/160).

Product identifier	Source/ Processing center	Time periods																			
		95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14
Level-2 Data Sets																					
TC_L2_GOME	BIRA	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			
TC_L2_SCIA	BIRA								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
TC_L2_GOME2A	BIRA													🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
TC_L2_GOME2B	BIRA																			🟩	🟩
TC_L2_OMI	BIRA										🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
NP_L2_GOME	RAL			🟩																	
NP_L2_GOME2	RAL													🟩	🟩						
NP_L2_IASI	ULB								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
LP_L2_SCIA	UBR								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L2_MIPAS	KIT											🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L2_GOMOS	ESA								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L2_OSIRIS	UoS							🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L2_SMR	CHALM							🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L2_ACE	UofT										🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
Level-3 Data Sets																					
TC_L3_MRG	DLR/BIRA	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
NP_L3_MRG	RAL/KNMI			🟩										🟩	🟩						
LP_L3_SCIA	UBR								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L3_MIPAS	KIT											🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L3_GOMOS	FMI								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L3_OSIRIS	UoS							🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L3_SMR	CHALM							🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L3_ACE	UofT										🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
LP_L3_MRG-MZM	FMI													🟩	🟩						
LP_L3_MRG-SMM	FMI													🟩	🟩						
TTC_L3_GOME	DLR	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩	🟩	🟩	🟩	🟩	🟩			
TTC_L3_SCIA	DLR								🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
TTC_L3_OMI	DLR										🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
TTC_L3_GOME2	DLR													🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Level-4 Data Sets																					
NP_L4_MRG	KNMI			🟩										🟩	🟩						

On total ozone, 19 years of harmonized level-2 data records from GOME, SCIAMACHY and GOME-2A, GOME-2B and OMI sensors have been produced using an advanced version of the direct-fitting GODFIT-3 prototype algorithm. This data set includes the Level 2 products for each instrument (over full instrument lifetime) and a merged monthly mean gridded data set using OMI in combination with GOME as long-term stability reference.

For ozone profiles, data set from GOME (for the year 1997) and GOME-2 (for the years 2007-2008) instruments have been generated. Beside the level 2 data sets for the GOME and GOME-2 instruments, monthly mean gridded and assimilated 6 hourly global ozone fields are provided. A Level 2 dataset from IASI for the year 2008 has been produced and is available.

As regards limb sensors, the so-called Harmonized single instruments (HARMOZ) data sets has been generated for the GOMOS, MIPAS, SCIAMACHY, OSIRIS, SMR and ACE-FTS instruments. These data records (covering instrument lifetime except for MIPAS – after 2005 only) include individual profiles with a common pressure grid and concentration unit, auxiliary information for converting into mixing ratio and/or geometric altitude. In addition, for each pair of instruments, drift and bias tables are provided. Beside the single profile data, single instrument zonal mean time series (10° latitude bin) including detailed uncertainty/variability information are also available.

Merged ozone profile data sets covering two contiguous years (2007-2008) have been created from all limb/occultation sensors on board of ENVISAT (GOMOS, MIPAS, SCIAMACHY) as well as from the Third Party Missions OSIRIS, SMR and ACE-FTS. The merged data sets include monthly zonal mean and bi-weekly mean (20° longitude, 10° latitude, bi-weekly) ozone profiles. In addition fine resolution data sets (5°x5°, 3 day time step) covering the years 2007 and 2008 have been generated for MIPAS and SCIAMACHY instruments as they provide very high spatial sampling.

Tropical tropospheric ozone columns from the CCD (convective cloud differential) algorithm are available for the same periods and instruments as the total ozone columns. The retrieved L2 data from the instruments GOME, SCIAMACHY, OMI, and GOME-2A are read by the L3 algorithm to retrieve the L3 tropospheric ozone columns. However for the period July 2003 to August 2006 (tape recorder failed) the data coverage from GOME was not sufficient for the CCD method, the situation improved slightly after August 2007. GOME_2B were not yet analyzed in all details.

3 Total Ozone ECV

3.1 L2 Total Ozone (BIRA-IASB)

Within the Ozone_cci project, the baseline algorithm for total ozone retrieval from backscatter UV sensors is the GOME-type direct-fitting (GODFIT) algorithm jointly developed at BIRA-IASB, DLR-IMF and RT-Solutions for implementation in version 5 of the GOME Data Processor (GDP) operational system. In contrast to previous versions of the GDP which were based on the DOAS method, GODFIT uses a least-squares fitting inverse algorithm including direct multi-spectral radiative transfer simulation of earthshine radiances and Jacobians with respect to total ozone, albedo closure and other ancillary fitting parameters. A detailed description of the GODFIT v4 algorithm can be found in the Ozone_cci ATBD.

3.1.1 Data Processing

Level-2 total ozone column data sets derived from the sensors GOME/ERS-2, SCIAMACHY/ENVISAT, GOME-2/METOP-A, GOME-2/METOP-B and OMI/AURA have been processed with the retrieval algorithm GODFIT v4 developed at BIRA-IASB. The data sets are provided for the complete instrumental time series, under the condition of availability of the input parameters, and are based on the latest level-1 data (see Table 3.1).

GODFIT is a direct-fitting algorithm using a non-linear least-squares adjustment of LIDORT-based spectral simulations of the backscattered earthshine radiance to measured spectra in the 325-335 nm interval. More details on the algorithm itself can be found in the ATBD (Rahpoe et al., 2015) or in Lerot et al. (2014). There is one ozone column measurement per ground pixel observed by the sensor and the level-2 data sets are distributed via Net-CDF files (one file per orbit). For each measurement, geolocation information, auxiliary and additional fitted parameters, quality indicators, a-priori O_3 profile shape and averaging kernels are also provided in the output files.

Figure 3.1 shows an example of total ozone columns retrieved from one day of GOME-2/METOP-A observations.

Table 3.1: Time coverage of the level-2 data sets and level-1 versions used in the processing chains.

Sensor	Time coverage	Level-1 data
GOME/ERS-2	Jul. 1995 – Jun. 2011	ESA L1 v4.00/4.01/4.03
SCIAMACHY/ENVISAT	Aug. 2002 – Apr. 2012	ESA L1 v8.0x
GOME-2/METOP-A	Jan. 2007 – Jul. 2016	EUMETSAT L1 v5.12/6.12
GOME-2/METOP-B	Jan. 2013 – Jul. 2016	EUMETSAT L1 v5.12/6.12
OMI/AURA	Oct. 2004 - Dec. 2016	NASA Collection 3

3.1.2 Quality Control Criteria

The delivered Net-CDF files contain only measurements for which the convergence has been reached with a number of iterations less than 6 (the typical number of iterations is 3-4). No retrieval is performed for pixels with solar zenith angle larger than 89°. The quality of the total ozone measurements following some specific instrumental operations (e.g. decontamination episodes) may be degraded. These measurements are in general easily detectable and have already been filtered out from the delivered level-2 data sets.

An estimation of the random error is associated to each total ozone column given in the product. This value has been derived via propagation of the level-1 radiance and irradiance statistical errors throughout the inversion algorithm. The reduced chi-squared value is a good indicator of the consistency between the fit residuals and the level-1 errors. Assuming perfectly estimated level-1 errors, the reduced chi-squared will be very close to 1 for a fit without any systematic structures in its residuals. In practice, they are generally ranging between 0.3 and 3. The root mean-squared (RMS) of the fit residuals is another indicator for the fit quality, but does not provide any hint on the nature of the residuals (random or systematic).

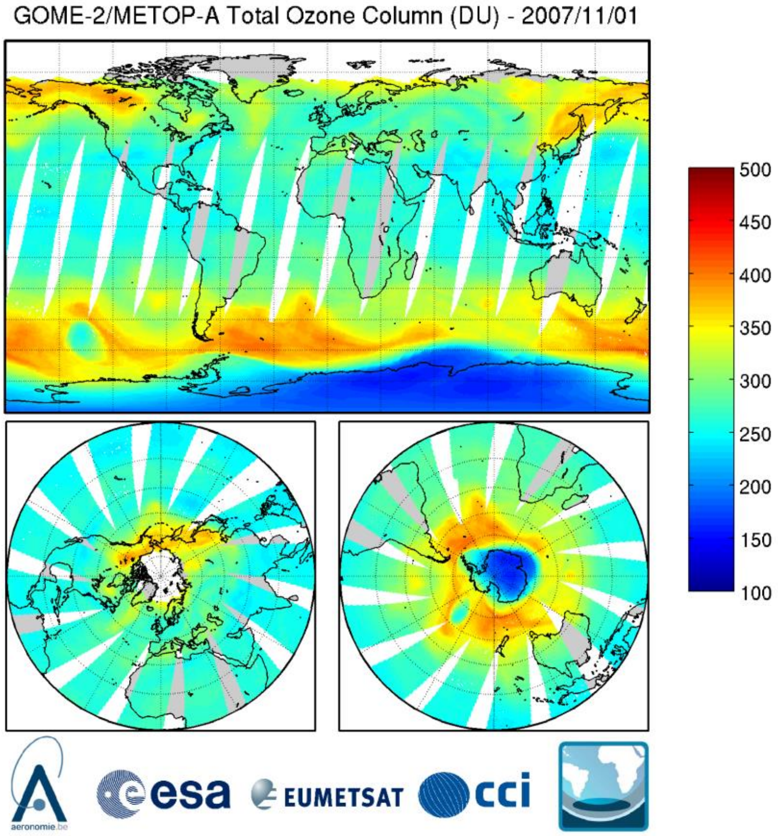

Figure 3.1: Total ozone columns retrieved from GOME-2/METOP-A observations on 1st November 2007

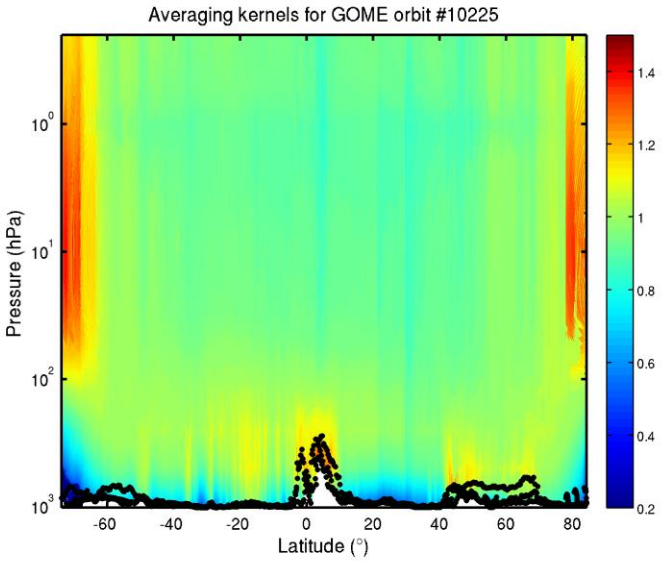

Figure 3.2: Typical averaging kernels of total ozone retrievals for one GOME orbit. The black dots represent the pressure of the effective scenes considered

As mentioned before, the averaging kernels are also provided for all measurements. They represent the sensitivity of the total column retrieval to a real change in the ozone concentration at a given layer, considering both the observation geometry and the algorithmic features. At low and mid-latitudes, these averaging kernels are generally close to 1 in the stratosphere and upper troposphere and decrease for the lowermost layers, depending on the surface albedo and cloud contamination. At higher solar zenith angles, they change more rapidly with the altitude, making the retrieval quality much more dependent on the a priori profile shape information. Typical averaging kernels are illustrated in Fig. 2 for one GOME orbit. The black dots represent the pressure of the effective scene considered for the total ozone retrieval. A smoothing error estimate is also provided in the level-2 files, which represents the impact of the a priori profiles shape on the retrieved column. This is computed using both the averaging kernel and the covariance matrices associated to the a priori profile climatology.

These different parameters can be used by the user to apply additional filtering for an optimal use of these data sets adapted to its own application. Although the total error on the individual measurements is generally within a few percent, it can be much larger in some specific geophysical conditions unaccounted for in the retrieval algorithm like the presence of large aerosol plumes or major volcanic eruptions leading to clouds of SO2 and ashes.

3.1.3 Data format

3.1.3.1 Filename structure

An example of filename for the L2 total ozone column output file of one GOME orbit is:
ESACCI-OZONE-L2P-TC-GOME_ERS2-BIRA_010185-19970401143000-fv0300.nc
where:

- “GOME_ERS2” indicates the instrument and platform. Alternatively, it can be “SCIAMACHY_ENVISAT”, “GOME2_METOPA”, “GOME2_METOPB” or “OMI_AURA”.
- “010185” represents the orbit number
- “19970401143000” indicates the date and time of the beginning of the orbit. This is to be interpreted as YYYMMDDhhmmss.
- “fv0300” is the product number. This is to be interpreted as v03.00. This number may vary from a sensor to another. v03.00 corresponds to latest products generated during phase-II (Reprocessing performed in between July 2016 and Feb. 2017).

3.1.3.2 Data content

Table 3.2 describes all variables contained in the level-2 total ozone output NetCDF files.

Table 3.2: Dimension and description of all variables contained in the L2 total ozone NetCDF files. N_p represents the total number of measurements for scanning instruments (GOME, SCIAMACHY, GOME-2) and the number of viewing lines for imager instruments (OMI). N_r is the number of rows for imager instruments (60 for OMI), and is 1 for scanner instruments. N_{sw} is the number of subwindows used in the wavelength calibration procedure applied once per orbit and N_{cal} is the number of fitted parameters during this procedure.

Variable Name	Unit	Dimension	Description
time	Days	N_p x N_r	Time of measurement in days since 1995-1-1 00:00:00
time_of_measurement_string	-	N_p x N_r x 19	String indicating the time of measurement at a glance: YYYYMMDDThhmmss.sss
pixel_number	-	N_p x N_r	Ground pixel number
state_number	-	N_p x N_r	State/MDR/Viewing line number. Only relevant for

			SCIAMACHY, GOME-2 and OMI.
row_number	-	N_p x N_r	Row index number. Only relevant for OMI.
pixel_type	-	N_p x N_r	Pixel type: 0 for forward pixels, 3 for backscan pixels, -1: NA
latitude	degree	N_p x N_r	Latitude of the pixel center
latitude_corner	degree	4 x N_p x N_r	Latitudes of the pixel corners
longitude	degree	N_p x N_r	Longitude of the pixel center
longitude_corner	degree	4 x N_p x N_r	Longitudes of the pixel corners
solar_zenith_angle	degree	N_p x N_r	Solar zenith angle at the pixel center
viewing_zenith_angle	degree	N_p x N_r	Viewing zenith angle at the pixel center.
relative_azimuth_angle	degree	N_p x N_r	Relative azimuth angle at the pixel center
retrieval_mode_flags	-	N_p x N_r	retrieval mode: 0 for normal mode, 1 for snow/ice mode from cloud algorithm
processing_flags	-	N_p x N_r	0: Nominal mode; 1: irregular L1 data - No retrieval; 2: Solar zenith angle larger than 89° - No retrieval; 3: No cloud data - No retrieval; 8: Forward model failure - No retrieval; 9: inversion failure - No retrieval; 21: Pixel affected by row anomaly - No retrieval; 22-24: Pixel might be affected by row anomaly - uncertain output
Total_ozone_column	mol.m-2	N_p x N_r	Retrieved total ozone column
Total_ozone_column_ random_error	mol.m-2	N_p x N_r	Random error associated to the retrieved total column
Total_ozone_column_ smoothing_error	mol.m-2	N_p x N_r	Error due to the a priori profile associated to the retrieved total column
ozone_ghost_column	mol.m-2	N_p x N_r	Partial ozone column comprised between the ground and the effective surface
fitted_ring_coefficient	-	N_p x N_r	Retrieved Ring scaling parameter
fitted_state_vector	Various	8 x N_p x N_r	Full fitted state vector (Total O3, T°-shift, 4 polynomial coefficients, Ring scale factor, Radiance wavelength shift)
effective_temperature	°K	N_p x N_r	Retrieved effective temperature
cloud_fraction	-	N_p x N_r	Effective cloud fraction
cloud_top_pressure	hPa	N_p x N_r	Cloud Top pressure
cloud_albedo	-	N_p x N_r	Effective cloud top albedo provided by
effective_scene_pressure	hPa	N_p x N_r	Pressure at the effective scene used for the retrieval
effective_scene_albedo	-	N_p x N_r	Retrieved effective albedo of the scene
surface albedo	-	N_p x N_r	Minimum surface albedo at 335 nm from OMI LER climatology
surface_altitude	m	N_p x N_r	Surface altitude extracted from GTOPO30
rms	-	N_p x N_r	Root mean square of fit residuals
reduced_chi_squared	-	N_p x N_r	Reduced chi-square of the fit
nb_of_iterations	-	N_p x N_r	Number of iterations before convergence
convergence_flag	-	N_p x N_r	Convergence flag: 0 for failure, 1 for success

atmosphere_ pressure_grid	hPa	15 x N_p x N_r	Pressure at levels defining the layers used in the forward model
averaging_kernels	-	14 x N_p x N_r	Averaging kernels in the layers of the forward model
apriori_ozone_profile	mol.m-2	14 x N_p x N_r	A-priori partial ozone columns in the layers of the forward model
Wavelength_calibration_ parameters	-	N_{cal} x N_{sw} x N_r	Wavelength calibration fitted parameters in each subwindow: 1 wavelength shift and optionally 1 or 2 slit function parameters.
Wavelength_calibration_ rms	-	N_{sw} x N_r	Root mean square of wavelength calibration fit residuals in each subwindow

3.2 L3 Total Ozone (DLR)

Within the second phase of the Ozone_cci project an algorithm has been developed by DLR-IMF for the creation of a level-3 merged monthly mean homogeneous total ozone product combining measurements from the five sensors GOME/ERS-2, SCIAMACHY/ENVISAT, GOME-2/METOP-A, GOME-2/METOP-B, and OMI/AURA. A detailed description of the algorithm is given in the Ozone_cci ATBD and in [*Coldewey-Egbers et al.*, 2015].

3.2.1 Data Processing

Individual GOME, SCIAMACHY, GOME-2, and OMI level-2 total ozone data records, processed with the GODFIT v3.0 retrieval algorithm (see Sect. 3.1), are the input to the level-3 processing. At first 1°x1° daily data are created for each individual sensor. In order to minimize the differences between the individual level-3 products, an inter-satellite calibration approach is used to create the merged total column product. OMI in combination with GOME is used as a long-term reference in which GOME has first been adjusted to OMI based on comparisons during the overlap period from 2004 to 2011. Next, SCIAMACHY and both GOME-2 data records are adjusted to this reference. The correction factors depend on latitude and time. Figure 3.3 shows an example for the merged total ozone product with data from October 2010.

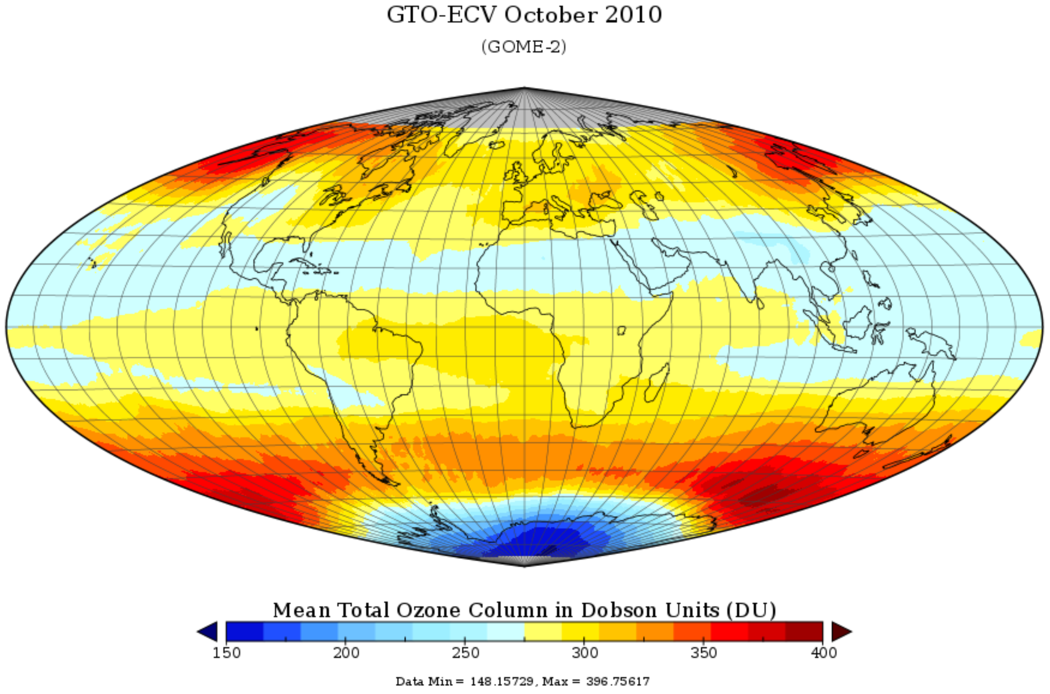

Figure 3.3: GTO-ECV mean total ozone from October 2010 (GOME-2 time period)

The GTO-ECV dataset contains additionally the standard deviation and the standard error for each grid point, see example in Figure 3.4. It is important to note that the standard error quantifies the spatio-temporal sampling errors inherent to the satellite measurements. The larger errors correspond to the time periods where SCIAMACHY (lower sampling with alternating nadir/limb measurements) is used.

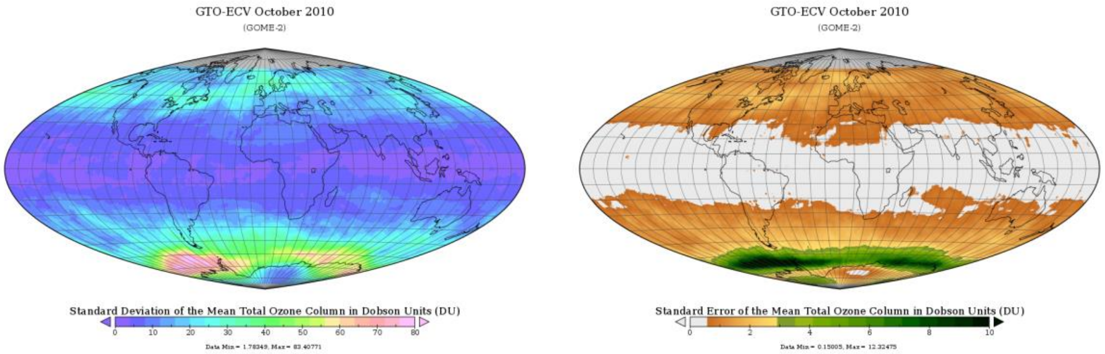

Figure 3.4: GTO-ECV standard deviation and standard error of the mean total ozone from October 2010 (GOME-2 time period)

3.2.2 Quality Control Criteria

GTO-ECV contains only data for the grid points with a representative number of measurements. Cut-off values for latitude as a function of month (see Table 3.3) have been defined in order to provide representative monthly means that contain a sufficient number of measurements equally distributed over time.

Table 3.3: Cut-off values for latitude as a function of month for the level-3 merged monthly mean total ozone product.

Month	Latitudes	Month	Latitudes
January	60.0° N – 90.0° S	July	90.0° N – 57.5° S
February	70.0° N – 90.0° S	August	90.0° N – 62.5° S
March	80.0° N – 80.0° S	September	82.5° N – 72.5° S
April	90.0° N – 65.0° S	October	72.5° N – 85.0° S
May	90.0° N – 60.0° S	November	65.0° N – 90.0° S
June	90.0° N – 57.5° S	December	60.0° N – 90.0° S

3.2.3 NetCDF Output

Table 3.4 describes all variables contained in the level-3 merged monthly mean total ozone output NetCDF files.

Table 3.4: Dimension and description of all variables contained in the L3 merged monthly mean total ozone NetCDF files. N_{lat} = 180 and N_{lon} = 360.

Variable Name	Unit	Dimension	Description
latitude	degree	N_{lat}	Latitude of grid center
longitude	degree	N_{lon}	Longitude of grid center
atmosphere_mole_content_of_ozone	DU	N_{lat} x N_{lon}	Mean Total Ozone Column in Dobson Units
atmosphere_mole_content_of_ozone_standard_deviation	DU	N_{lat} x N_{lon}	Standard Deviation of Mean Total Ozone Column in Dobson Units
atmosphere_mole_content_of_ozone_standard_error	DU	N_{lat} x N_{lon}	Standard Error of Mean Total Ozone Column in Dobson Units
atmosphere_mole_content_of_ozone_number_of_observations	-	N_{lat} x N_{lon}	The Number of Measurements used to derive the Mean Total Ozone in Dobson Units

4 Nadir Profile ECV

4.1 L2 Nadir Profile (RAL)

This note describes the details of this particular ozone profile dataset, including pertinent attributes of the data and algorithm used. For a full technical description of the retrieval algorithm used please refer to the Ozone_cci ATBD.

4.1.1 Data Processing and Parameters

NCDF files are produced by the RAL nadir profile ozone scheme for GOME aboard ERS-2 and GOME-2 aboard MetOp-A. Currently, one year of processed data is provided for GOME (1997) and two years for GOME-2 (2007-2008). The data processed was based on the whole-orbit level 1b data available for these years from the British Atmospheric Data Centre (BADC) as of February 2013 (Product Format Version 4.0).

For GOME-2, the native Band 2 pixels (40 km x 80 km) have been combined to produce pixels with a footprint size of 160 km by 160 km. This is done to improve the signal to noise of the measurement or the purposes of optimizing the retrieval of the tropospheric part of the profile. Whilst the spatial resolution is reduced this results in less 'noisy' retrieved tropospheric ozone sub-columns.

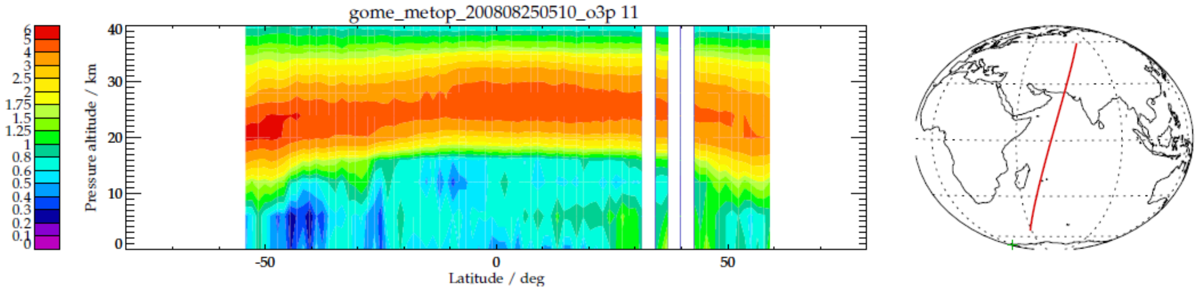

Figure 4.1 : Retrieved number density orbit cross section on 25th August 2008 (nadir pixel)

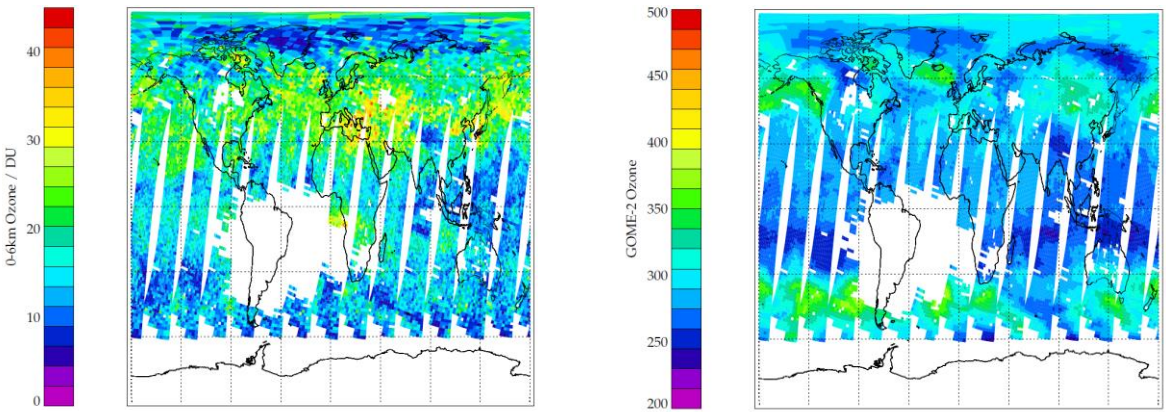

Figure 4.2: Lowest layer retrieved sub-column ozone on 25th August 2008 (left), and Retrieved total column ozone (right) on the same day

The RAL retrieval scheme derives profiles of number density on a set of pressure levels, spaced approximately every 4-6 km in altitude (taken from the SPARC-DI grid). The optimal estimation method is used. Averaging kernels are provided on this basis. It is noted that the vertical resolution of the retrieval is relatively coarse compared to the vertical grid and that tropospheric levels in particular have significant bias towards the assumed *a priori* state. It is therefore important to take account of the characterization of the retrieval provided by the averaging kernels when comparing these results to other data-sets, particularly where those are more highly vertically resolved.

Figure 4.1 shows an example of retrieved number density profiles over 1 orbit. Retrieved ozone and ozone error are also provided on levels in volume mixing ratio, in addition to sub-column and sub-column error estimates. Figure 4.2 shows examples of the lowest retrieved sub-column and total column ozone for 1 day in August 2008. For convenience vertically integrated sub-column amounts between the retrieval levels are also reported.

The algorithm is sequential retrieval. It uses information from GOME-2 Band 1 initially before performing surface albedo retrieval in Band 2 and finally ozone profile retrieval in Band 2 incorporating the information derived from Band 1 as input. The output from both retrievals are included in the product with that from band 1 indicated in the variable name, however it should be highlighted that the output from the Band 1 retrieval is not the algorithm final solution. Other trace gas spectra are fit as part of the ozone retrieval in order to accurately fit the ozone profile (such as CH_2O and BrO) and their column values are included in the output file, but it should be noted that the chosen fitting window is optimized for ozone rather than these trace gases.

4.1.2 Quality Control Criteria

It is recommended that some quality control criteria be applied to the ozone profile data, using parameters also supplied within the NetCDF file:

- ‘ncost’ (the normalised total fit cost) is less than 2
- ‘aconv’ (the convergence flag) is equal to 1
- ‘sza’ (solar zenith angle) is less than 80°
- 1/cos(‘lza’)*’o3_b1_tc’ (the line-of-sight zenith component of the Band 1 retrieved total column amount) is less than 500 during the months of January to May. This is due to very high stratospheric ozone at high Northern latitudes which limits the ability to discern tropospheric ozone beneath.

Tropospheric ozone can have a low bias in the presence of thick cloud (as indicated by ‘cloudf’ (cloud fraction) and/or high cloud ‘cloudp’ (effective cloud top pressure).

4.1.3 NetCDF Output

The format of the Level 2 ozone profile product file from the RAL GOME ozone profile algorithm is NetCDF. The values in all groups are taken from the level 1 or other input data files, or calculated by the program. The file includes output from both the retrieval algorithm and geolocation information, in addition to ancillary information such as surface pressure obtained from ERA-Interim reanalysis. The parameters of netCDF files are collected in Table 4.1

Table 4.1 Main parameters in the netDCF files. N_{prof} and N_{levels} denotes the number of profiles and pressure levels, respectively.

Parameter and unit	Dimension and precision	Description
o3_nd (cm-3)	float, $N_{prof} \times$ n_o3_nd	Ozone molecular number density
o3_vmr	float, $N_{prof} \times$ n_o3_vmr	Ozone volume mixing ratio
o3_error (%)	float, $N_{prof} \times$ n_o3_error	Retrieved ozone uncertainty
o3_ap	float, $N_{prof} \times$ n_o3_ap	Ozone a priori volume mixing ratio
o3_ap_error (%)	float, $N_{prof} \times$ n_o3_ap_error	Ozone a priori error
o3_sub_col (DU)	float, $N_{prof} \times$ n_o3_sub_col	Ozone partial column
o3_sub_col_error (DU)	double, $N_{prof} \times$ n_o3_sub_col_error	Ozone partial column error
o3_ap_sub_col (DU)	double, $N_{prof} \times$ n_o3_sub_col	Ozone a priori partial column
o3_ap_sub_col_error (DU)	double, $N_{prof} \times$ n_o3_ap_sub_col	Ozone a priori partial column error
o3_tc (DU)	Float, $N_{prof} \times 1$	Total column ozone
o3_tc_error (DU)	Float, $N_{prof} \times 1$	Total column ozone error
o3_ap_tc_error (DU)	double, $N_{prof} \times 1$	Ozone a priori total column error
o3_b1_sub_col (DU)	double, $N_{prof} \times$ n_o3_b1_sub_col	Band 1 ozone partial column
o3_b1_sub_col_error (DU)	double, $N_{prof} \times$ n_o3_b1_sub_col_error	Band 1 ozone partial column error

Parameter and unit	**Dimension and precision**	**Description**
o3_b1_tc (DU)	double, $N_{prof}\times 1$	Band 1 total ozone column
o3_b1_tc_error (DU)	double, $N_{prof}\times 1$	Band 1 total ozone column error
nit	long int, $N_{prof}\times 1$	Number of Iterations
b1nit	long int, $N_{prof}\times 1$	Band 1 number of iterations
cost	float, $N_{prof}\times 1$	Final cost function value
ncost	float, $N_{prof}\times 1$	Normalized final cost function value
b1cost	float, $N_{prof}\times 1$	Band 1 cost function value
aconv	int. array, $N_{prof}\times 1$	Convergence flag
b1conv	Long int array, $N_{prof}\times 1$	Band 1 convergence flag
achi	Long int array, $N_{prof}\times 1$	Chi squared flag
spres (hPa)	Float, $N_{prof}\times 1$	Surface Pressure
levs (hPa)	Float, $N_{levels}\times 1$	Pressure levels of retrieved ozone profiles
lat (degrees north)	Float, $N_{prof}\times 1$	Latitude of ground pixel center
lon (degrees east)	Float, $N_{prof}\times 1$	Longitude of ground pixel center
ll (degrees north/degrees east)	Float, $N_{prof}\times 8$	Latitude and longitude of ground pixel corners. [lat1,lon1,lat2,lon2,lat3,lon3,lat4,lon4]
Pixno	Long Int, $N_{prof}\times 1$	Orbit ground pixel number ([scan line number * 100]+cross track scan position index)
sza (degrees)	float, $N_{prof}\times 1$	Solar zenith angle
lza (degrees)	float, $N_{prof}\times 1$	Line-of-sight zenith angle
time (hours)	float, $N_{prof}\times 11$	Hours since 00:00.00hrs on date
scp	short, $N_{prof}\times 1$	Across track scan index
cloudf	double, $N_{prof}\times 1$	FRESCO effective cloud fraction
cloudp (hpa)	double, $N_{prof}\times 1$	FRESCO cloud top pressure
clouda	double, $N_{prof}\times 1$	FRESCO cloud albedo
cloud_ffail	short, $N_{prof}\times 1$	FRESCO cloud fit fail indication
cloud_mode	short, $N_{prof}\times 1$	FRESCO cloud fit mode
cloud_s6	double, $N_{prof}\times 1$	Expected scaling of 0-6km sub column due to cloud
cloud_s12	double, $N_{prof}\times 1$	Expected scaling of 0-12km sub column due to cloud
salb	float, $N_{prof}\times 1$	Retrieved surface albedo
ring	float, $N_{prof}\times 1$	Retrieved ring spectrum scaling parameter
xsect	float, $N_{prof}\times 1$	Retrieved wavelength shift of absorptions cross sections
bro	float, $N_{prof}\times 1$	BrO column average volume mixing ratio
bro_err	float, $N_{prof}\times 1$	BrO column average volume mixing ratio error
no2	float, $N_{prof}\times 1$	NO_2 column average volume mixing ratio
no2_err	float, $N_{prof}\times 1$	NO_2 column average volume mixing ratio error
ch2o	float, $N_{prof}\times 1$	CH_2O column average volume mixing ratio
ch2o_err	float, $N_{prof}\times 1$	CH_2O column average volume mixing ratio error
rsf	float, $N_{prof}\times$n_misr	Residual spectral pattern scaling factor
slit	float, $N_{prof}\times 1$	Slit function FWHM scaling parameter
misr (nm)	float, $N_{prof}\times$ n_misr	Wavelength shift between radiance and irradiance spectra
tsurf (K)	float, $N_{prof}\times$n_tsurf	Effective surface temperature
sx (cm-6)	float, $N_{prof}\times$ n_sx_1× n_sx_0	Ozone molecular number density solution covariance matrix
sn (cm-6)	float, $N_{prof}\times$ n_sx_1× n_sx_0	Ozone molecular number density measurement noise covariance matrix
ak	dloat, $N_{prof}\times$ n_ak_1×	Ozone molecular number density averaging kernel

Parameter and unit	Dimension and precision	Description
	n_ak_0	matrix

4.2 *IASI L2 nadir profile (ULB)*

This section describes the details of the IASI ozone profile dataset, including attributes of the data and algorithm used. For a full technical description of the retrieval algorithm used please refer to the Ozone_cci ATBD.

4.2.1 Data processing and parameters

The IASI ozone profile data product is a new product of Ozone_cci Phase-II. It is based on the FORLI (Fast Optimal/Operational Retrieval on Layers for IASI) algorithm. FORLI is a fast radiative transfer model capable of processing in near-real-time the numerous radiance measurements made by the high-spatial and high-spectral resolution IASI, with the objective to provide global concentration distributions of atmospheric trace gases.

Currently one year (2008) of processed data has been generated and is provided to users. The retrieval is performed in partial column on altitude levels: the ozone product from FORLI-O3 is a profile retrieved on 41 layers between the surface and 40 km, with an extra layer from 40 to 60 km, the top of the atmosphere. It is provided along with averaging kernels and relative total error profile associated, on the same vertical grid. When the first levels are not available (because of the orography), the value is set to -999. The first layer is between the altitude of the surface and the kilometer just above. The next ones have a thickness of 1 km. The last layer is from 40 to 60 km.

It is important to take account of the characterization of the retrieval provided by the averaging kernels when comparing these results to other data-sets, in particular those that are more highly vertically resolved such as ozonesonde measurements. One should apply the averaging kernels to the highly vertically resolved profile, using the following equation [*Rodgers*, 2000]:

$\boldsymbol{x}_S = \mathbf{A} \times \boldsymbol{x}_r + (\mathbf{I}-\mathbf{A})*\mathbf{x}_a$

with $\boldsymbol{x}_r$ the profile (in partial columns) to be smoothed (for example a sonde profile), $\mathbf{A}$ the IASI averaging kernel matrix, $\mathbf{x}_a$ the IASI a priori profile (in partial columns) and $\boldsymbol{x}_s$ the smoothed profile (in partial column).

4.2.2 Quality control criteria

The variable ‘Retrieval quality flag’ is not implemented for the moment.
The data provided have been filtered using the FORLI quality flags. Below is the list of flags used to discard the data.

Quality input flags:

-Missing T, Q, Cloud input values

-Negative surface altitudes

-Unrealistic skin temperature

Quality processing flags:

-Convergence not reached after maximum number of iterations

-Too high values for Chi Square

-No retrieval done (due to incorrect inputs or other reasons).

-Residuals “biased” or “sloped” or large RMS values

-Fit diverged

-Unrealistic averaging kernels

-Total error covariance matrix ill conditioned

-Unrealistic partial columns

To assess the quality of the profile, users could use the vertical profile of total retrieval error. It is an absolute error (ratio error on observation).

4.2.3 NetCDF output

The format of the Level 2 ozone profile product file from the FORLI algorithm is NetCDF. The FORLI algorithm for IASI operates with multiplication factors, with the a priori as reference, and the profile is adjusted in layer partial columns. The original output profile is in partial columns but is provided here in the units needed to follow the general convention. The values in all groups are taken from the level 1 or other input data files, or calculated by the program. Further details, including on ancillary data, can be found in the ATBD. The main parameters of the netCDF-4 files are collected in Table 4.2.

The netcdf file format for the profiles and the averaging kernels is as follows: the vertical profiles are stored from the surface to 40 km of altitude, each kilometer, with an extra layer from 40 km to the top of the atmosphere (TOA). The first data corresponds to the layer between Earth's surface and the kilometer just above. The averaging kernels are stored row by row, from the surface to the TOA. The missing values are set to -999.

Table 4.2: The variables in the NetCDF files. N_{alt} denotes the number of vertical layers and N_{obs} denotes the number of observations in the day.

Parameter and unit	**Dimension and precision**	**Description**
o3_sub_col (DU)	Float, $N_{alt}\times N_{obs}$	Ozone partial column vertical profile
o3_sub_col_error (DU)	Float, $N_{alt}\times N_{obs}$	Vertical profile of total retrieved error
o3_ap_sub_col (DU)	Float, $N_{alt}\times N_{obs}$	Ozone a priori partial columns vertical profile
o3_tc (DU)	Float, $1\times N_{obs}$	total column ozone
Ak (molec cm^{-2}/molec cm^{-2})	Float, $N_{alt}\times N_{alt}\times N_{obs}$	Averaging kernels
sza (degrees)	Float, $1\times N_{obs}$	solar zenith angle
Cloudf (%)	Float, $1\times N_{obs}$	EUMETSAT Cloud coverage in the pixel
time (hhmmss)	Int array, $1\times N_{obs}$	Hour in the day
lat (degrees)	Float, $1\times N_{obs}$	latitude of the ground pixel
lon (degrees)	Float, $1\times N_{obs}$	longitude of the ground pixel
dofs	Float, $1\times N_{obs}$	Degrees Of Freedom of the Signal
ret_flag	Int, $N_{alt}\times N_{obs}$	Retrieval quality flag
tropo_alt	Float, $1\times N_{obs}$	tropopause altitude (from Eumetsat IASI L2 atmospheric profile with WMO definition)

therm_contrast (K)	Float, 1× N_{obs}	thermal contrast (defined as difference between Eumetsat skin temperature and Eumetsat atmospheric temperature at the first level, just above the surface)

4.3 L3 Nadir Profile (KNMI)

This section gives a short description of the algorithm that calculates averaged ozone fields on a regular latitude-longitude grid and gives a description of its output files. Input that should be provided are L2 satellite measurements, output is in NetCDF format complying with the CF 1.6 metadata conventions.

4.3.1 Algorithm

The pixels in the satellite data (L2) are assumed to be ordered as indicated in Figure 4.1. If this is not the case, the reading routine should provide the appropriate transformation. **A** is the first corner in the longitude and latitude arrays, **B** the second etc. The across track direction is given by the lines the lines **A**-**D** and **B**-**C**, while the along track direction is given by the lines **A**-**B** and **D**-**C**. Note that corners **C** and **D** are reversed with respect to the GOME/GOME-2 convention.

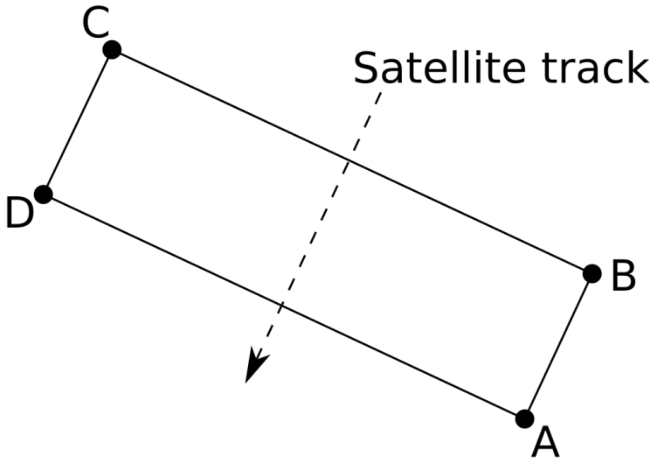

Figure 4.1: Pixel layout assumed in the nadir L3 algorithm.

The along track pixel edges **AB** and **DC** and cross track pixel edges **AD** and **BC** (see Figure 4.1) are divided into a number of points. The first point on **AB** and the first on **DC** form a line which is divided into the same number of points as **AD**. Each of these points is assigned to a gridcell, see for example Figure 4.2.

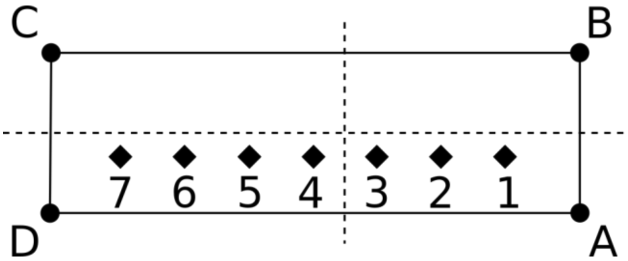

Figure 4.2: A L2 pixel is divided into subpixels (diamonds 1-7). Each subpixel is assigned to a TM5 gridcell (dashed) and the average and standard deviation are calculated (see text).

Suppose that ABCD in Figure 4.2 is the pixel of interest and that the horizontal line marked with the diamonds are the subpixels (numbered 1 to 7). Furthermore, the two dashed lines denote the gridcell boundaries which are numbered the same way as the pixel corners (i.e. gridcell A is the lower right cell). In this case, subpixels 1 $\sim$ 3 are added to gridcell A, and the counter for gridcell A is increased by 3. Subpixels 4 $\sim$ 7 are added to gridcell D and the counter for gridcell D is increased by 4. The pixel values are weighted by $1/\sigma^2$ before adding, so the weighted mean gridcell value and the corresponding standard deviation are given by:

$$mean = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}$$

$$sdev = \sqrt{\frac{1}{\sum_i \frac{1}{\sigma_i^2}}}$$

These values are provided for partial columns in the L3 files on a layer-by-layer basis and for the total column. An example is shown in Figure 4.1 for January 2008, based on the L2 dataset provided in phase 1 of the ozone CCI project.

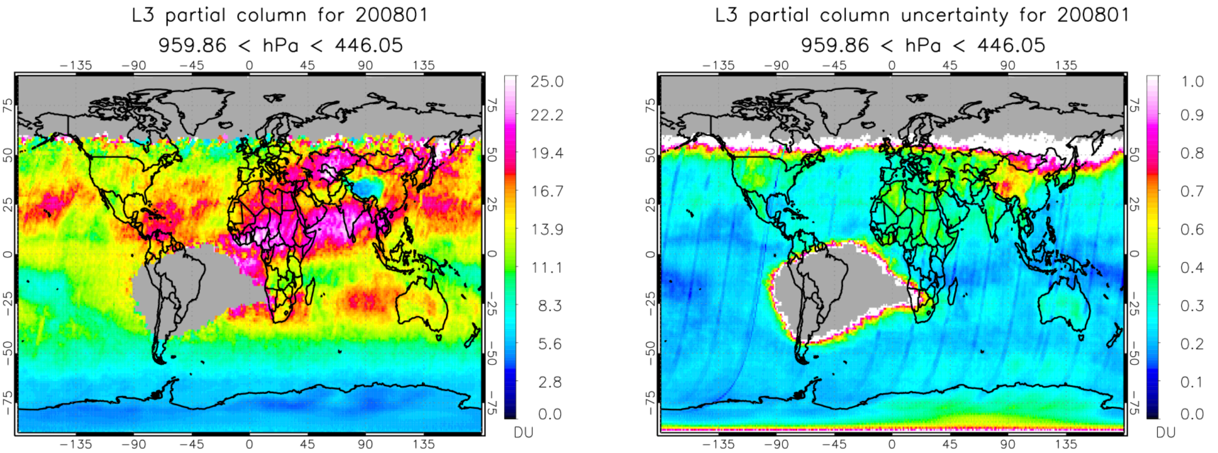

Figure 4.3: mean partial ozone column (left) and its uncertainty (right) for January 2008, based on L2 data provided in the first phase of the Ozone-CCI project.

4.3.2 NetCDF Output

Common datasets for all NetCDF output files are time, lat, lon, surface pressure and air pressure. Missing values in the dataset are indicated with the IEEE 'NaN' values.
Time is given in seconds since some reference time. Since the L3 fields are monthly averages, the time is equal to the reference time, which has been set to the first day of the month. The fields lat and lon give the latitude and longitude of the L3 gridcell centers. Latitude varies between -90 and +90 and longitude between -180 and +180.

The surface pressure and air pressure fields are given in hPa and to obtain the full 3D pressure field, one should extend the surface pressure field in the third dimension with the air pressure field. The first entry from the air pressure field should not be used, since it is only a dummy entry for the surface pressure.

The NetCDF (partial) column datasets are O3 du, O3e du, O3 du tot and O3e du tot, which are the profile in partial columns and its associated error (both in DU/layer) and the total column and its associated error (both in DU). If the original L2 data was given in number density, the weighted mean number density and its error and the volume mixing ratio and its error are also given as O3 ndens, O3e ndens, O3 vmr and O3e vmr. The partial column datasets have been calculated for the layers between the number density levels. The full list with NetCDF variables in can be found in Table 4.3.

Table 4.3: The variables in the NetCDF files. N_{time}, N_{layer}, N_p, N_{lat} and N_{lon} are number of time, layers, pressures levels, latitude and longitude zones, respectively.

Parameter and unit	**Dimension and precision**	**Description**
lon(degree East)	float, $N_{lon}\times 1$	longitude, from -180 (west) to +180 (east) given at gridcell centers. NetCDF dimension
lat(degree North)	float, $N_{lat}\times 1$	latitude, from -90 (south) to +90 (north) given at gridcell centers. NetCDF dimension
layers	integer, $N_{layer}\times 1$	layer number, starting at 1. NetCDF dimension.
air_pressure (hPa)	float, $N_p\times 1$	air pressure at layer boundaries, replace the first element from this array with the corresponding surface pressure element. NetCDF dimension.
time	integer, $N_{time}\times 1$	seconds since reference time, usually the start of the month. NetCDF dimension.
surface_pressure	float, $N_{time}\times N_{lat}\times N_{lon}$	pressure at the bottom of the atmosphere
O3_du (DU)	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	weighted average of the partial ozone columns (DU/layer)
O3e_du (DU)	float $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	uncertainty in the weighted average of the partial ozone columns (DU/layer)
O3_du_tot (DU)	float, $N_{time}\times N_{lat}\times N_{lon}$	total column: vertically integrated O3_du dataset
O3e_du_tot (DU)	float, $N_{time}\times N_{lat}\times N_{lon}$	total column uncertainty: quadratically added o3e_du
O3_vmr (ppmv)	float $N_{time}\times N_p\times N_{lat}\times N_{lon}$	weighted average of the volume mixing ratio
O3e_vmr (ppmv)	float, $N_{time}\times N_p\times N_{lat}\times N_{lon}$	uncertainty in the weighted average of the volume mixing ratio
O3_ndens (molec cm^{-3})	float $N_{time}\times N_p\times N_{lat}\times N_{lon}$	weighted average of the number density (#molecules/cm3) with dimensions (time, air_pressure, lat, lon).

O3e_ndens (molec cm^{-3})	float $N_{time} \times N_p \times N_{lat} \times N_{lon}$	uncertainty in the weighted average of the number density

4.4 *L4 Nadir Profile (KNMI)*

The data assimilation algorithm will take the level-2 data produced by the merged retrieval algorithm as input. Besides the profiles themselves, other important data that have to be provided in the level-2 product are the averaging kernel (AK) and the covariance matrices. The data are assimilated using the Kalman filter technique that is outlined in the next section. It is basically a form of optimal interpolation to find the weighted average between model results and measurements. Required for this approach are a model and its associated uncertainties (covariance matrix) and the measurements with uncertainties and the averaging kernel. The used model to assimilate the ozone profiles is TM5.

4.4.1 Algorithm

The equations for the state vector $\boldsymbol{x}$ and the measurement vector $\boldsymbol{y}$ are given by:

$$\boldsymbol{x}_{i+1} = M(\boldsymbol{x}_i) + \mathbf{w}_i,\ \ \mathbf{w}_i \sim N(\mathbf{0}, \mathbf{Q}_i)$$
$$\boldsymbol{y}_i = H(\boldsymbol{x}_i) + \mathbf{v}_i,\ \ \mathbf{v}_i \sim N(\mathbf{0}, \mathbf{R}_i)$$

where M is the model that propagates the statevector in time. It has an associated uncertainty $\boldsymbol{w}$, which is assumed to be normally distributed with zero mean and covariance matrix $\mathbf{Q}$. The observation operator H gives the relation between $\boldsymbol{x}$ and $\boldsymbol{y}$. The uncertainty is given by $\boldsymbol{v}$, which is also assumed to have zero mean and covariance matrix $\mathbf{R}$. In matrix notation, the propagation of the state vector and its covariance matrix ($\mathbf{P}$) are given by:

$$\boldsymbol{x}_{i+1}^{f} = \mathbf{M}(\boldsymbol{x}_i^{a})$$
$$\mathbf{P}_{i+1}^{f} = \mathbf{M}\mathbf{P}_i^{a}\mathbf{M}^{T} + \mathbf{Q}_i$$

where $\boldsymbol{x}^a$ is the statevecctor at time $t = i$, after assimilation of the observations. The observations are assimilated according to:

$$\boldsymbol{x}_i^{a} = \boldsymbol{x}_i^{f} + \mathbf{K}_i\left(\boldsymbol{y}_i - \mathbf{H}_i\boldsymbol{x}_i^{f}\right)$$
$$\mathbf{P}_i^{a} = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i^{f}$$
$$\mathbf{K}_i = \mathbf{P}_i^{f}\mathbf{H}_i^{T}\left(\mathbf{H}_i\mathbf{P}_i^{f}\mathbf{H}_i^{T} + \mathbf{R}_i\right)^{-1}$$

where $\mathbf{K}$ is called the Kalman gain matrix.

The covariance matrix $\mathbf{P}$ is too large to handle, it's size is the number of elements in the state vector squared. For the 44-layer 2° × 3° (latitude × longitude) TM5 grid, this amounts to $(475200)^2$ elements. To reduce $\mathbf{P}$ to something more manageable we parameterize it into a time dependent standard deviation field and a constant correlation field.

We cannot apply the forecast equation for the covariance matrix directly because of two problems. First, because you have to add $\mathbf{Q}$, the original parameterization is not conserved and $\mathbf{P}$ will "fill up". Eventually, $\mathbf{P}$ will become too large to handle. Second, errors in the ozone chemistry should also be taken into account. Therefore, the Kalman covariance propagation is replaced by an approach where we first apply the model's advection operator to the standard deviation field, and then model the error growth.

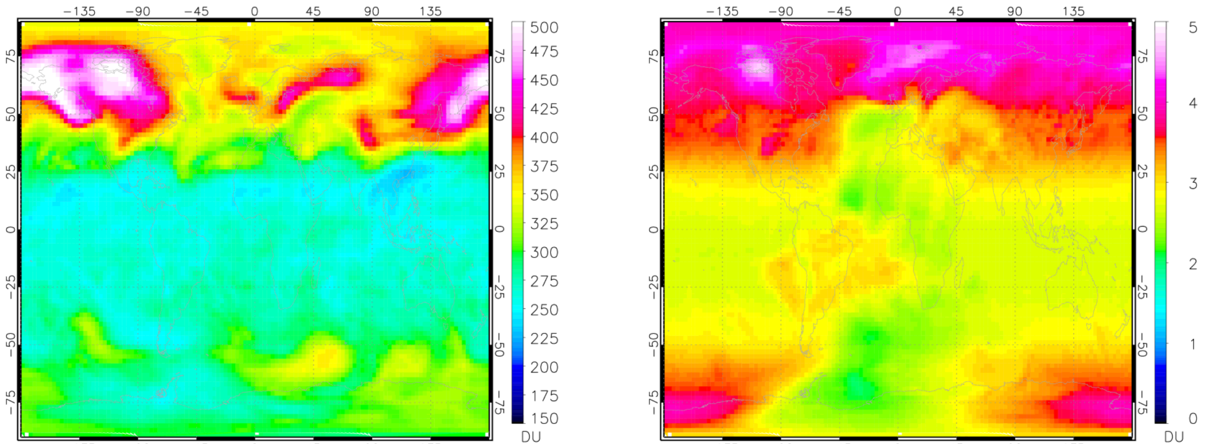

Figure 4.4: Assimilated total ozone column (left) and the corresponding error for 12 UTC January 31st, 2008, based on L2 data provided in the first phase of the Ozone-CCI project.

In the analysis equations, the number of elements in an ozone profile is generally much larger than the degrees of freedom (about 5 to 6). We therefore reduce the number of data points per profile by taking the singular value decomposition of the AK, and transform the profiles accordingly. Finally, we use an eigenvalue decomposition to calculate the $\mathbf{H}_i\mathbf{P}_i^f\mathbf{H}_i^T$ matrix inverse in the Kalman filter equation. We truncate it at a number of eigenvalues representing about 98% of the original trace.

In the L4 files, the ozone concentrations are given as both column densities and volume mixing ratios. The associated uncertainties are given by the time dependent standard deviation field mentioned above. In Figure 4.4 an example plot is shown for 12 UTC January 31st, 2008, based on the data provided in the first phase of the Ozone-CCI project. The left plot shows the total column, while the right plot shows the uncertainty on the total column, calculated as: $\sigma_{tot} = \sqrt{\sum(\sigma_i)^2}$

4.4.2 NetCDF Output

The assimilation output is given on a 2°×3° (latitude × longitude) grid of 44 layers. Time is given in seconds since some reference time, and has been set to zero, i.e. the reference time is the time of the ozone field. The fields lat and lon give the latitude and longitude of the L3 gridcell centers. Latitude varies between -90 and +90 and longitude between -180 and +180. Pressure (Pa) is given on hybrid levels, and to reconstruct the 3D pressure field, one should take each cell in the Psurf field, multiply it by the “Hybride coef b” vector and add the “Hybride coef a” vector. The temperature field is given on the layer centers in K.

Ozone is given in both column density (“O3 dens”, molecules / m2) and volume mixing ratio (“O3 vmr”, ppv) and their standard errors (“O3s dens” and “O3s vmr” respectively). The full list with NetCDF variables in can be found in Table 4.4.

Table 4.4: The variables in the NetCDF files. N_{time}, N_{layer}, N_{level}, N_{lat} and N_{lon} are number of time, layers, pressures levels, latitude and longitude zones, respectively.

Parameter and unit	Dimensions and precision	Description
lon (degree East)	float, N_{lon}×1	longitude from -180 (west) to +180 (east) given at gridcell

Parameter and unit	Dimensions and precision	Description
		centers. NetCDF dimension.
lat (degree North)	float, $N_{lat}\times1$	latitude from -90 (south) to +90 (north) given at gridcell centers. NetCDF dimension.
layers	integer, $N_{layer}\times1$	layer number, starting at 1. NetCDF dimension.
levels	float, $N_{level}\times1$	levels = layers boundaries, starting at 1. NetCDF dimension.
time (hours)	integer, $N_{time}\times1$	hours since reference time, usually midnight. NetCDF dimension.
Psurf (Pa)	float, $N_{time}\times N_{lat}\times N_{lon}$	surface air pressure
Temperature (K)	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	air temperature at layer centers
Gph (m)	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	geopotential height at layer centers
O3_vmr (ppv)	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	volume mixing ratio
O3s_vmr (ppv)	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	uncertainty in the volume mixing ratio
O3_dens (molec m^{-2})	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	column density of ozone in
O3s_dens (molec m^{-2})	float, $N_{time}\times N_{layer}\times N_{lat}\times N_{lon}$	uncertainty in the column density of ozone
Hybride_coef_a (Pa)	float, $N_{level}\times 1$	Hybride half levels : p(k) = hyb_a(k) + hyb_b(k) * ps [Pa]. Surface first.
Hybride_coef_b	float, $N_{level}\times 1$	Hybride half levels : p(k) = hyb_a(k) + hyb_b(k) * ps [Pa]. Surface first.
Hybride_coef_da (Pa)	float, $N_{layer}\times 1$	surface first
Hybride_coef_db	float, $N_{layer}\times 1$	surface first
Hybride_coef_fa (Pa)	float, $N_{layer}\times 1$	surface first
Hybride_coef_fb	float, $N_{layer}\times 1$	surface first
Cell_area (m^2)	float, $N_{lat}\times1$	cell area per latitude

5 Limb Profile ECV

5.1 L2 HARMonized dataset of Ozone profiles (HARMOZ)

5.1.1 Overview of the Dataset

The HARMonized dataset of OZone profiles (HARMOZ) is based on limb and occultation measurements from Envisat (GOMOS, MIPAS and SCIAMACHY), Odin (OSIRIS, SMR) and SCISAT (ACE-FTS) satellite instruments. HARMOZ consists of original retrieved ozone profiles from each instrument, which are screened for invalid data by the instrument teams. While the original ozone profiles are presented in different units and on different vertical grids, the harmonized dataset is given on a common pressure grid in NetCDF-4 format. The Ozone_cci pressure grid corresponds to vertical sampling of ~1 km below 20 km and 2-3 km above 20 km. The vertical range of the ozone profiles is specific for each instrument, thus all information contained in the original data is preserved. Provided altitude and temperature profiles allow the representation of ozone profiles in number density or mixing ratio on a pressure or altitude vertical grids. Geolocation, uncertainty estimates and vertical resolution are provided for each profile. For each instrument, optional parameters, which are related to the data quality, are also included.

For convenience of users, tables of biases between each pair of instruments for each month, as well as bias uncertainties, are provided. These tables characterize the data consistency and can be used in various bias and drift analyses, which are needed, for instance, for combining several datasets to obtain a long-term climate dataset.

The detailed description of the HARMOZ data can be found in [*Sofieva et al.*, 2013]. The dataset is available at http://www.esa-ozone-cci.org/?q=node/161 or at dx.doi.org/10.5270/esa-ozone_cci-limb_occultation_profiles-2001_2012-v_1-201308.

5.1.2 NetCDF Output

HARMOZ ozone profiles are structured in folders corresponding to each instrument. Each folder contains monthly data files with self-explanatory names: ESACCI-OZONE-L2-LP-IIII_SSSS-PP_VV-YYYYMM-Z.nc, where L2=Level 2 , LP= Level2, IIII= instrument, SSSS=satellite, PP=processing center, VV= processor version, YYYY= year, MM=month, Z=file version. For example, the file ESACCI-OZONE-L2-LP-GOMOS_ENVISAT-IPF_V6-200801-fv0004.nc contains GOMOS ozone profiles for January 2008.

Each file contains the mandatory parameters, which are the same for all instruments (Table 5.1). The files contain also optional instrument-specific parameters (Table 5.2), which might be related to might be related to the data quality.

Table 5.1 Mandatory parameters in the HARMOZ NetCDF files. N_{alt} and N_{prof} denote the number of pressure levels and the number of profiles, respectively.

Parameter and unit	*Dimensions*	*Description*
time(days since 1900-01-01 00:00:00)	$N_{prof} \times 1$	The parameter to index the profiles
air_pressure (hPa)	$N_{alt} \times 1$	The vertical coordinate
altitude (km)	$N_{alt} \times N_{prof}$	The geometric altitude above the mean sea-level
latitude (degree_north)	$N_{prof} \times 1$	Latitude of each profile
longitude (degree_east)	$N_{prof} \times 1$	Longitude of each profile
mole_concentration_of_ ozone_in_air (mol/cm^3)	$N_{alt} \times N_{prof}$	Vertical profiles of ozone. Number density (cm^{-3}) is acquired by multiplying the variable with Avogadro constant N_A=6.02214e23 mol^{-1}
mole_concentration_of_ozone_ in_air standard_error (mol/cm^3)	$N_{alt} \times N_{prof}$	Uncertainty (random error) associated with the ozone profiles
vertical_resolution (km)	$N_{alt} \times N_{prof}$ or $N_{alt} \times 1$	FWHM of the averaging kernel
air_temperature (K)	$N_{alt} \times N_{prof}$	Temperature profiles at the locations of measurements, for conversion from concentration to mixing ratio

Table 5.2 Optional parameters in HARMOZ NetCDF files N_{alt} and N_{prof} denote the number of pressure levels and the number of profiles, respectively.

	Parameter and unit	*Dimensions*	*Description/comment*
GOMOS	orbit_number	$N_{\text{prof}} \times 1$	Envisat orbit number
	star_number	$N_{\text{prof}} \times 1$	Star number in GOMOS catalogue
	star_magnitude	$N_{\text{prof}} \times 1$	Star visual magnitude
	star_temperature (K)	$N_{\text{prof}} \times 1$	Star effective temperature
	obliquity (deg)	$N_{\text{prof}} \times 1$	Obliquity of occultation: the angle between the orbital plane and the line of sight
	sza (deg)	$N_{\text{prof}} \times 1$	solar zenith angle at tangent point
	Chi2	$N_{\text{alt}} \times N_{\text{prof}}$	Profiles of normalized χ^2- statistics. Usually close to 1. Large values indicate problems with retrievals
	illumination_condition_flag	$N_{\text{prof}} \times 1$	0-full dark, 3-straylight, 2- twilight, 4- straylight&twilight.
	SAA_flag	$N_{\text{prof}} \times 1$	The indicator showing that the data might be affected by the Southern Atlantic Anomaly (cosmic rays); 0- no, 1- yes
SCIAMACHY	orbit_number	$N_{\text{prof}} \times 1$	Envisat orbit number
	state_id	$N_{\text{prof}} \times 1$	State ID of the SCIA measurement
	height_sat (km)	$N_{\text{prof}} \times 1$	Satellite altitude above the sea-level, for each profile
	radius_earth (km)	$N_{\text{prof}} \times 1$	The Earth radius at locations above the tangent points
	sza_tanpnt (deg)	$N_{\text{prof}} \times 1$	solar zenith angle at tangent point
	pixel_lat (degree_north)	$N_{\text{prof}} \times 4$	the ground latitudes of the four corners of the limb scan pixel
	pixel_lon (degree)	$N_{\text{prof}} \times 4$	the ground longitude of the four corners of the limb scan pixel
	total_ozone_column (mm)	$N_{\text{prof}} \times 1$	Total ozone column for each profile; 1mm=100 DU (Dobson Unit)
	systematic_error (%)	$N_{\text{alt}} \times N_{\text{prof}}$	Systematic errors derived from parameter deviation simulation (see ozone-CCI ATBD)
MIPAS	apriori_temperature (K)	$N_{\text{alt}} \times N_{\text{prof}}$	temperature profiles at locations of measurements based on ECMWF and MSIS data
	geo_id	$N_{\text{prof}} \times 22$	MIPAS geolocation identifier formatted as XXXXX_YYYYMMDDThhmmssZ where XXXXX=orbit, YYYY=year, MM=month, DD=day, hh=hour, mm=minute, ss=second
	orbit_number	$N_{\text{prof}} \times 1$	Envisat orbit number
	sza(deg)	$N_{\text{prof}} \times 1$	Solar zenith angle
	chi2	$N_{\text{prof}} \times 1$	Normalized χ^2- value of retrievals
	dof	$N_{\text{prof}} \times 1$	degrees of freedom of target retrieval
	rms (nW/cm/sr)	$N_{\text{prof}} \times 1$	root mean square of residual spectra
OSIRIS	scan_number	$N_{\text{prof}} \times 1$	OSIRIS scan number
	albedo	$N_{\text{prof}} \times 1$	Retrieved albedo
	ssa (deg)	$N_{\text{prof}} \times 1$	Solar scattering angle
	sza(deg)	$N_{\text{prof}} \times 1$	Solar zenith angle
	optics_temperature (K)	$N_{\text{prof}} \times 1$	Average optics box temperature
SMR	quality	$N_{\text{prof}} \times 1$	Quality flag 0: best quality, 4: tolerable
	solar_zenith_angle (deg)	$N_{\text{prof}} \times 1$	
	local_solar_time (h)	$N_{\text{prof}} \times 1$	
	measurement_response	$N_{\text{alt}} \times N_{\text{prof}}$	Proportion of measurement; measurements with weak influence of a priori have measurement response close to 1.
	scaled_potential_vorticity (K m^2 kg^{-1} s^{-1})	$N_{\text{alt}} \times N_{\text{prof}}$	Profiles of potential vorticity (Lait, 1994) scaled at 475 K potential temperature level
	equivalent_latitude (deg)	$N_{\text{alt}} \times N_{\text{prof}}$	Profiles of equivalent latitude at locations of measurements

ACE-FTS	beta_angle (deg)	$N_{\mathrm{prof}} \times 1$	β-angle is defined as the angle between the orbit plane of ACE-FTS and the vector from the Sun. It is a proxy for vertical resolution.

5.1.3 Data Agreement Tables (bias tables)

In addition, the tables of biases between each pair of instruments for each month, as well as the bias uncertainties, are provided. The bias tables are computed using the collocated measurements with the following restrictions on time difference Δt, distance between tangent points Δd, and latitude difference $\Delta\theta$:

(i) standard collocation: $|\Delta t| \le 24\ \mathrm{h}$, $|\Delta d| \le 1000\ \mathrm{km}$, $|\Delta\theta| \le 2°$.

(ii) tight collocation: $|\Delta t| \le 4\ \mathrm{h}$, $|\Delta d| \le 400\ \mathrm{km}$.

The bias b is calculated as:

$$b = 2\frac{\langle x_1 - x_2 \rangle}{\langle x_1 \rangle + \langle x_2 \rangle}, \qquad (1)$$

where x_1 and x_2 are collocated measurements from two instruments at a given altitude and $\langle . \rangle$ denotes mean/median estimates (both are provided). The relative uncertainty of b is estimated as:

$$\sigma_b = \frac{2}{\langle x_1 \rangle + \langle x_2 \rangle} \cdot \frac{\sigma_{(x_1 - x_2)}}{\sqrt{N}} \qquad (2)$$

where $\sigma_{(x_1 - x_2)}$ is the sample standard deviation of the difference distribution computed in a standard or in a robust way as $\sigma = \frac{1}{2}(P_{84} - P_{16})$, P_{84} and P_{16} are 84th and 16th percentiles, respectively, and N is the number of collocated measurements. In the tables, both parameters b and σ_b are presented in %.

The bias is evaluated for each month in 20° latitude bins from 90°S to 90°N. The bias tables are structured in 15 folders corresponding to the instrument pairs, e.g., “GOMOS_OSIRIS”. The folders contain bias tables corresponding to each month in NetCDF format. The file names contain information about the year and the month, as well as the instruments. For example, the file “ESACCI-OZONE-AgreementTable_GOMOS_OSIRIS_200801.nc” contains the bias table between GOMOS (x_1) and OSIRIS (x_2) for January 2008, for the standard collocation criterion. The files for tight collocation criterion are ended with “_tight.nc”. The parameters included in NetCDF file are presented in Table 5.3.

Table 5.3 Main parameters of bias tables in the NetCDF format

Parameter and unit	*Dimensions*	*Description/comment*
air_pressure (hPa)	$N_{\mathrm{alt}} \times 1$	The vertical coordinate
approximate_altitude (km)	$N_{\mathrm{alt}} \times 1$	Approximate altitude at pressure levels computed as $z = 16\ \log_{10}(1013 / P)$, P is pressure in hPa
latitude_centers (degree_north)	$N_{\mathrm{lat}} \times 1$	Centers of latitude bins: 80S, 60S, 40S, 20S, 0S, 20N, 40N, 60N, 80N
bias (%)	$N_{\mathrm{lat}} \times N_{\mathrm{alt}}$	Bias between instrument#1 and instrument#2 estimated as the

		mean of differences , Eq. (1)
robust_bias (%)	$N_{lat} \times N_{alt}$	As “bias”, but the median estimates are used
bias_uncertainty (%)	$N_{lat} \times N_{alt}$	Uncertainty of the bias estimated using the standard sample std of differences, Eq.(2)
robust_bias _uncertainty (%)	$N_{lat} \times N_{alt}$	Uncertainty of the bias estimated using the robust sample std of differences, Eq.(2)
number_of _collocated_data	$N_{lat} \times N_{alt}$	number of collocated data in each latitude bin and at each pressure level

Sample scripts to read the NetCDF files with MATLAB, IDL and IGOR Pro are also available on http://www.esa-ozone-cci.org/?q=node/161 or dx.doi.org/10.5270/esa-ozone_cci-limb_occultation_profiles-2001_2012-v_1-201308.

5.1.4 Relative drifts and biases between limb-profile datasets

For the calculation of pairwise relative bias and drift, the harmonized limb ozone profiles (HARMOZ) created during the first phase of Ozone CCI (Climate Change Initiative) have been used. Involved in this comparison are ACE-FTS, GOMOS, MIPAS, OSIRIS, SCIAMACHY, and SMR datasets. The ozone profiles in the HARMOZ format are generated using a common CCI pressure grid with corresponding number density, altitude, pressure, and temperature information as described in [*Sofieva et al.*, 2013].

In addition to the Agreement Tables (Chapter 5.1.3), an alternative a multiple regression analysis has been performed to derive the relative bias β and, in addition the relative drift α from the zonal mean monthly mean difference time series *b(t,z)* of the collocated pairs of profiles [*Rahpoe et al.*, 2015]. The following multi-regression has been used:

$$b(t,z) = \alpha(t,z)\cdot(t - t') + \beta(z) + \sum_{i=1,2}\left[\kappa_i(z)\sin(\omega_i\,\mathrm{t}) + \lambda_i(z)\cos(\omega_i\,\mathrm{t})\right] + R(t,z), \qquad (3)$$

where κ_i, λ_i and ω_i are amplitude and frequency of harmonic components with the periods of 6 and 12 months. For this analysis February 2005 is chosen as the reference time t' for all pairs. The noise term $R(t,z)$ is assumed to be autoregressive function with lag one, AR(1). We used the methods described in [*Weatherhead et al.*, 1998] and [*Gebhardt et al.*, 2014] to derive autocorrelation, white noise, σ_α , and σ_β for each pair of instruments.

The results of the linear regression are stored in ASCII and NetCDF files with corresponding overview plots that can be found viewed at the BIRA ftp server in the directory Limb_Profiles/REL_DRIFT_BIAS/.

Table 5.4 Main parameters of relative drift and relative bias data in the NetCDF format

Parameter and unit	*Dimensions*	*Description/comment*
air_pressure (hPa)	$N_{alt}\times 1$	The vertical coordinate
approximate_altitude (km)	$N_{alt}\times 1$	Approximate altitude at pressure levels computed as $z = 16\ \log_{10}(1013/P)$, *P* is pressure in hPa
latitude_centers (degree_north)	$N_{lat}\times 1$	Centers of latitude bins: 80S, 60S, 40S, 20S, 0S, 20N, 40N, 60N, 80N
Drift (%/decade)	$N_{lat} \times N_{alt}$	Drift between instrument#1 and instrument#2 estimated from multi-regression model

2_sigma_drift (%/decade)	$N_{lat} \times N_{alt}$	Uncertainty of the drift estimated using the covariances from the autocorrelation method
Bias (%)	$N_{lat} \times N_{alt}$	Relative bias for the reference time t' derived from the multi-regression model
2_sigma_bias (%)	$N_{lat} \times N_{alt}$	Uncertainty of the bias estimated using the covariances from the autocorrelation method
number_of _collocated_data	$N_{lat} \times N_{alt}$	number of collocated data in each latitude bin and at each pressure level

5.2 L3 Limb Profile Datasets

5.2.1 Monthly Zonal Mean ozone profiles from individual instruments (MZM)

5.2.1.1 Overview of the Dataset

The monthly zonal mean data in 10° latitude zones from 90°S to 90°N are created for all Ozone_cci limb and occultation instruments. The HARMOZ data [*Sofieva et al.*, 2013] are used as an input.

For all sensors, the monthly zonal average is computed as the mean of ozone profiles. x_k :

$$\bar{x} = \frac{1}{N}\sum x_k \ , \qquad (4)$$

where *N* is the number of measurements. MZM ozone profiles are presented in two forms: as mixing ratio and mole concentration on the Ozone_cci pressure grid. The uncertainty of the monthly mean σ^2_{mean} is estimated as the standard error of the mean:

$$\sigma^2_{mean} = \frac{s^2}{N}, \qquad (5)$$

where $s^2 = \left\langle (x_k - \bar{x})^2 \right\rangle$ is the sample variance. Both sample standard deviation *s* and the standard error of the mean σ_{mean} are stored in the MZM files. For SMR, only the data having the measurement response larger than 0.75 are used.

The mean of individual error estimates e_k :

$$\bar{e} = \frac{1}{N}\sum e_k \ , \qquad (6)$$

are also provided in the MZM data files.

In order to characterize the non-uniformity of sampling, we provide inhomogeneity measures in latitude, H_{lat} , and in time, H_{time} . The definition of this measures and details of the related analyses can be found in ATBD and the dedicated Technical Note. Each inhomogeneity measure ranges from 0 to 1 (the more homogeneous, the smaller *H*).

5.2.1.2 NetCDF Output

The monthly zonal mean data are structured into yearly NetCDF files, for each instrument separately. The self-explaining name indicates the instrument and the year. For example, the file “ESACCI-OZONE-L3-LP-GOMOS_ENVISAT-MZM-2008.nc” contains monthly zonal mean data for GOMOS in 2008. The variable that are included into NetCDF files are collected in Table 5.5.

Table 5.5 The variables in MZM NetCDF files. N_{month}, N_{alt}, N_{lat} are number of months, pressures levels and latitude zones, respectively.

Parameter and unit	*Dimensions*	*Description*
Time	$N_{month}\times 1$	The parameter to index the months. The time is assigned to the middle of month and presented in “'days since 1900-01-01 00:00:00”
air_pressure (hPa)	$N_{alt}\times 1$	The vertical coordinate
approximate_altitude (km)	$N_{alt}\times 1$	Approximate altitude at pressure levels computed as $z = 16\ \log_{10}(1013/P)$, P is pressure in hPa
latitude_centers (degree_north)	$N_{lat}\times 1$	Centers of latitude bins: -85°: 10°:85°
ozone_mixing_ratio	$N_{lat}\times N_{alt}\times N_{month}$	Monthly zonal mean ozone mixing ratio vertical profiles
ozone_mole_ concentration (mol/cm^3)	$N_{lat}\times N_{alt}\times N_{month}$	Monthly zonal mean ozone mole concentration vertical profiles
standard_error_of_the_ mean (%)	$N_{lat}\times N_{alt}\times N_{month}$	Uncertainty of the monthly zonal mean, σ_{mean}, Eq. (5)
sample_standard _deviation (%)	$N_{lat}\times N_{alt}\times N_{month}$	Sample standard deviation in 1 month ×10° spatio-temporal bins, for each pressure level
mean_uncertainty_ estimate (%)	$N_{lat}\times N_{alt}\times N_{month}$	Monthly zonal mean of error estimates, Eq.(6)
inhomogeneity_in_time	$N_{lat}\times N_{alt}\times N_{month}$	Inhomogeneity measure in time
inhomogneity_in_latitude	$N_{lat}\times N_{alt}\times N_{month}$	Inhomogeneity measure in latitude

5.2.2 Merged Monthly Zonal Mean ozone profiles (MMZM)

5.2.2.1 Overview of the Dataset

The merged monthly zonal mean data (MMZM hereafter) include merged ozone profiles in 10° latitude zones for each month in years 2007-2008, at ozone-CCI pressure grid from 250 hPa to 1 hPa, and the parameters, which characterize the uncertainty of the merged profiles. MMZM is the weighted mean of the monthly zonal mean profiles from individual instruments. The weights are inversely proportional to the total errors of MZM:

$$\sigma^2 = \sigma^2_{mean} + \sigma^2_{sampling}, \qquad (7)$$

where σ^2_{mean} is the standard error of the mean (Eq. (5)) and $\sigma^2_{sampling}$ is the sampling uncertainty variance, which is related to potentially non-uniform sampling by measurements in space and in time. It is parameterized as

$$\sigma_{sampling} = \frac{1}{2}\left(H_{lat} + H_{time}\right)\cdot \sigma_{nat}, \qquad (8)$$

where H_{lat} and H_{time} are inhomogeneity measures in latitude and in time, respectively, and σ_{nat} is the profile of natural variability taken from LLM climatology [*McPeters et al.*, 2007], for each month and each latitude bin.

5.2.2.2 NetCDF Output

The merged monthly zonal mean data are structured into monthly NetCDF files with self-explanatory names. For example, the file "ESACCI-OZONE-L3-LP-MERGED-MZM-200801-fv0002.nc" contains merged monthly zonal mean data for January 2008. In addition to the variables of the merged data, the profiles from individual instruments with their uncertainty parameters are also included (for the altitude range 250-1 hPa used in data merging). The variables included into NetCDF files are collected in Table 5.6.

Table 5.6. The variables in MMZM NetCDF files

	Parameter and unit	*Dimensions*	*Description*
General parameters	air pressure (hPa)	$N_{alt}\times 1$	The vertical coordinate
	approximate_altitude (km)	$N_{alt}\times 1$	Approximate altitude at pressure levels computed as $z = 16\ \log_{10}(1013/P)$, P is pressure in hPa
	latitude_centers (degrees_north)	$N_{lat}\times 1$	Centers of latitude bins: -85°: 10°:85°
	instruments	$N_{instru}\times 1$	A dimension for individual datasets, instrument order 1-GOMOS, 2-MIPAS, 3-SCIAMACHY, 4-OSIRIS, 5-ACE-FTS , 6-SMR
Merged data	merged_ozone_vmr	$N_{lat}\times N_{alt}$	Merged monthly zonal mean ozone mixing ratio vertical profiles
	merged_ozone_concentration (mol/cm^3)	$N_{lat}\times N_{alt}$	Vertical profiles of merged monthly zonal mean ozone mole concentration. Number density (cm^{-3}) is acquired by multiplying the variable with Avogadro constant N_A=6.02214e23 mol^{-1}
	uncertainty_of_merged_ozone (%)	$N_{lat}\times N_{alt}$	Uncertainty σ_{merged} of the merged data
Individual datasets	ozone_vmr	$N_{lat}\times N_{alt}\times N_{instru}$	Monthly zonal mean ozone mixing ratio vertical profiles for individual instruments
	ozone_mole_concentration (mol/cm^3)	$N_{lat}\times N_{alt}\times N_{instru}$	Monthly zonal mean ozone mole concentration vertical profiles for individual instruments.
	standard_error_of_the_mean (%)	$N_{lat}\times N_{alt}\times N_{instru}$	Uncertainty of the monthly zonal mean for individual datasets, σ_{mean}, Eq. (5)
	sampling_error (%)	$N_{lat}\times N_{alt}\times N_{instru}$	Sampling error $\sigma_{sampling}$ for individual datasets characterized using (8).
	total_ error (%)	$N_{lat}\times N_{alt}\times N_{instru}$	Total uncertainty of monthly zonal mean data from individual instruments, see Eq.(7)

5.2.3 Semi-Monthly Mean ozone profiles with resolved longitudinal structure (SMM)

5.2.3.1 Overview of the Dataset

The general approach of computing semi-monthly mean data is the same as for creating monthly zonal mean ozone profiles: first semi-monthly mean data (SMM) from individual instruments are created, and then the weighted mean of SMM data is used as merged semi-monthly mean ozone profiles (MSMM).

For the SMM dataset, ozone profiles from individual HARMOZ datasets [*Sofieva et al.*, 2013] are averaged in 10° latitude × 20° longitude zones, twice per month. The data averaging and characterization is performed in the same way as for monthly zonal mean data described above, i.e., via computing the mean of ozone profiles.

SMM ozone profiles are characterized by:

- the standard error of the mean, Eq.(5):
- inhomogeneity in latitude, longitude and in time.

The data merging is performed in full analogy with creating monthly zonal mean data, as described in the previous section. The weights are inversely proportional to total uncertainties, Eq.(7). The sampling error is estimated in the same way as for the monthly zonal mean data, Eq.(8).

5.2.3.2 NetCDF Output

The merged semi-monthly mean ozone profiles are structured into yearly NetCDF files with self-explanatory names. For example, the file “ESACCI-OZONE-L3-LP-SMM-2008-fv0002.nc” contains the semi-monthly mean ozone profiles for January 2008. In addition to the variables of the merged data, the profiles from individual instruments with their uncertainty parameters are also included (for the altitude range 250-1 hPa used in data merging). The variables included into NetCDF files are collected in Table 5.7.

Table 5.7. The variables in MSMM NetCDF files. N_{alt} is number of pressure levels, N_{lat} and N_{lon} are numbers of latitude and longitude bins, respectively, N_{time} is number if temporal intervals and N_{instru} =6 is number of instruments.

	Parameter and unit	*Dimensions*	*Description*
General parameters	air pressure (hPa)	$N_{alt}\times 1$	The vertical coordinate
	approximate_altitude (km)	$N_{alt}\times 1$	Approximate altitude at pressure levels computed as $z = 16\log_{10}(1013/P)$, *P* is pressure in hPa
	latitude_centers (degree_north)	$N_{lat}\times 1$	Centers of latitude bins: -85°: 10°:85°
	longitude_centers (degree_east)	$N_{lon}\times 1$	Centers of longitude bins: -170°:20°:170°
	time	$N_{time}\times 1$ (24×1)	Central date for each half of month, expressed as days since
	instruments	$N_{instru}\times 1$	A dimension for individual datasets, instrument order 1-GOMOS, 2-MIPAS, 3-SCIAMACHY, 4-OSIRIS, 5-ACE-FTS , 6-SMR
Merg-	merged_ozone_vmr	$N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Merged semi-monthly zonal mean ozone mixing ratio vertical profiles
	merged_ozone_con-	$N_{lat}\times N_{lon}\times N_{alt}\times$	Vertical profiles of merged semi-monthly zonal

	centration (mol/cm^3)	N_{time}	mean ozone mole concentration. Number density (cm^{-3}) is acquired by multiplying the variable with Avogadro constant N_A=6.02214×10^{23} mol^{-1}
	uncertainty_of_ merged_ozone (%)	$N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Uncertainty σ_{merged} of the merged data
Individual datasets	ozone_vmr	$N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Semi-monthly zonal mean ozone mixing ratio vertical profiles for individual instruments
	ozone_mole_con-centration (mol/cm^3)	$N_{instru}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Semi-monthly zonal mean ozone mole concentration vertical profiles for individual instruments.
	standard_error_of_ the_mean (%)	$N_{instru}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Uncertainty of the semi-monthly zonal mean for individual datasets, σ_{mean}, Eq. (5)
	sampling_error (%)	$N_{instru}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Sampling error $\sigma_{sampling}$ for individual datasets characterized using (8).
	total_ error (%)	$N_{time}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{instru}$	Total uncertainty of semi-monthly zonal mean data from individual instruments, see Eq.(7)
	inhomogeneity_in_ longitude	$N_{instru}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Inhomogeneity measure in latitude
	inhomogeneity_in_ latitude	$N_{instru}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Inhomogeneity measure in latitude
	inhomogeneity_in_ time	$N_{instru}\times N_{lat}\times N_{lon}\times N_{alt}\times N_{time}$	Inhomogeneity measure in time

6 Tropospheric ozone

6.1 *Level 3 convective cloud differential algorithm*

6.1.1 Data processing

This section gives an introduction to the calculation of the tropospheric ozone column based on the convective cloud differential (CCD) algorithm more details are described in the ATBD. The algorithm is based on level 2 total column products as described in section 3.1. The data are monthly averaged and gridded, whereby only the position of the center coordinate is considered.

With the CCD method, the tropospheric column is calculated as the difference between the stratospheric column and the total column. The stratospheric column is estimated as the column above high reaching convective clouds (cloud cover >0.8 and cloud top height >8 km). The above cloud ozone column might be influenced by up draught of tropospheric pollution; therefore a relative clean reference region with strong convective activity is used (70°E to 170°W), which is assumed to be representative for the respective latitude band. For the total column only the cloud-free observations (cloud cover less than 10%) are considered.

The assumptions that the stratospheric ozone is constant throughout one month and for one latitude band limit the CCD algorithm to the tropics (20°S to 20°N).

6.1.2 NetCDF output

The results are stored in the netCDF-4 format, each variable carries the attributes “standard_name” and “long_name”. The standard names and units follow the CF conventions when possible. However, for some variables no standard_name is specified, e.g. stratospheric_ozone_column.

The dimensions latitude and longitude are defined at the top level. No dimension for the time is given as it is constant within one data file. The data are collected in several groups:

- “PRODUCT”, contains the tropospheric ozone data
    - “SUPPORT_DATA/DETAILED_RESULTS”, contains subgroups for more detailed results:
        - STRATOSPHERIC_OZONE, lists the reference and the stratospheric data per grid cell
        - TOTAL_OZONE, contains the average ozone column for cloud free pixels
        - CLOUD_PARAMETERS includes the average and standard deviation of the cloud data used to calculate the above cloud ozone column
        - SURFACE_PROPERTIES gives the average surface data per grid cell.
- The METADATA data are stored in separate group at the top level domain.

Table 6.1: Overview of the file structure

Variable Name	Unit	Dimension	Description
latitude	degree	N_{lat}	Latitude of grid center
longitude	degree	N_{lon}	Longitude of grid center
PRODUCT			Group containing the results and “SUPPORT_DATA/DETAILED_RESULTS”
METADATA			Group containig the file attributes in a subgroup called “O3CCI_METADATA”

Table 6.2: The data in the “PRODUCT” group

Variable Name	Unit	Dimension	Description
tropospheric_O3	DU	N_{lat} x N_{lon}	Mean tropospheric ozone column in Dobson Units
tropospheric_O3_std	DU	N_{lat} x N_{lon}	Standard Deviation of mean tropospheric ozone column in Dobson Units
tropospheric_O3_mixingratio	ppb[1]	N_{lat} x N_{lon}	Average mixing ratio in the tropospheric column
tropospheric_O3_mixingratio_std	ppb	N_{lat} x N_{lon}	Standard deviation of average mixing ratio in the tropospheric column
tropospheric_O3_number	1	N_{lat} x N_{lon}	The number of measurements used to derive the mean tropospheric Ozone

Table 6.3: the data in the “STRATOSPHERIC_OZONE” subgroup of “SUPPORT_DATA/DETAILED_RESULTS”

Variable Name	Unit	Dimension	Description
stratospheric_O3	DU	N_{lat} x N_{lon}	Mean stratospheric ozone column

---

[1] Units = 1, scaling_factor=1e-9

stratospheric_O3_std	DU	N_{lat} x N_{lon}	Standard Deviation of mean stratospheric ozone column
stratospheric _O3_number	1	N_{lat} x N_{lon}	The number of measurements used to derive the mean stratospheric ozone
stratospheric_O3_reference	DU	N_{lat}	Mean stratospheric ozone column in the reference area
stratospheric_O3_reference_std	DU	N_{lat}	Standard deviation the mean stratospheric ozone column in the reference area
stratospheric_O3 _reference_number	1	N_{lat}	The number of measurements used to derive the mean stratospheric ozone in the reference area
stratospheric_O3_reference_flag	1	N_{lat}	Quality flag of the stratospheric ozone in the reference area (Table 6.7).

Table 6.4: "TOTAL_OZONE" subgroup of "SUPPORT_DATA/DETAILED_RESULTS"

Variable Name	**Unit**	**Dimension**	**Description**
total_O3	DU	N_{lat} x N_{lon}	Mean total ozone column
total_O3_std	DU	N_{lat} x N_{lon}	Standard Deviation of mean total ozone column
total_O3_number	1	N_{lat} x N_{lon}	The number of measurements used to derive the mean total ozone

Table 6.5: “CLOUD_PARAMETERS” in "SUPPORT_DATA/DETAILED_RESULTS"

Variable Name	**Unit**	**Dimension**	**Description**
cloud_albedo	1	N_{lat} x N_{lon}	Mean cloud albedo used for the stratospheric ozone column
cloud_albedo_std	1	N_{lat} x N_{lon}	Standard Deviation of mean cloud albedo
cloud_height	km	N_{lat} x N_{lon}	Mean cloud height used for the stratospheric ozone column
cloud_height_std	km	N_{lat} x N_{lon}	Standard deviation of cloud height
cloud_fraction	1	N_{lat} x N_{lon}	Mean cloud fraction used for the stratospheric ozone column
cloud_fraction	1	N_{lat} x N_{lon}	Standard deviation for the cloud fraction

Table 6.6: "SURFACE_PROPERTIES" supgroup in “SUPPORT_DATA/DETAILED_RESULTS”, in contrast to the other data here no standard deviation of the data is given.

Variable Name	**Unit**	**Dimension**	**Description**
Surface_albedo	1	N_{lat} x N_{lon}	Mean surface albedo
Surface_height	km	N_{lat} x N_{lon}	Mean surface _altitude above mean sea level

Before subtracting the stratospheric reference column from the total column the stratospheric data are quality checked. The stratospheric reference column might be classified as invalid for the reasons given in Table 6.7. In these cases no tropospheric O_3 column is calculated for the entire latitude band. However, there are two exceptions focusing on outlier cases where one stratospheric reference column is classified different from the two neighboring ones:

- If the data number of data is low, but the stratospheric reference column agrees well with the neighboring reference columns and they are not classified as invalid, then the tropospheric column will be given anyway.
- If two stratospheric reference columns are classified as invalid data and only one column in between is classified as valid, no tropospheric columns are given for the single remaining band. In case data are flagged for more than one reason, the flag values are added.

Table 6.7: Quality flag for the stratospheric reference data

Flag value	Description	Threshold
0	valid stratospheric reference data	
1	stratospheric ozone column out of range	< 200 DU
2	number of individual observations too low	< 8 measurements per latitude band
4	standard deviation too high	> 10 DU
8	latitudinal gradient in stratospheric ozone too large	> 8 DU difference between two neighboring latitude bands, both will be flagged

7 References

Coldewey-Egbers, M. et al. (2015), The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative, *Atmos. Meas. Tech.*, *8*(9), 3923–3940, doi:10.5194/amt-8-3923-2015. [online] Available from: http://www.atmos-meas-tech.net/8/3923/2015/

Gebhardt, C., A. Rozanov, R. Hommel, M. Weber, H. Bovensmann, J. P. Burrows, D. Degenstein, L. Froidevaux, and A. M. Thompson (2014), Stratospheric ozone trends and variability as seen by SCIAMACHY from 2002 to 2012, *Atmos. Chem. Phys.*, *14*(2), 831–846, doi:10.5194/acp-14-831-2014. [online] Available from: http://www.atmos-chem-phys.net/14/831/2014/

McPeters, R. D., G. J. Labow, and J. A. Logan (2007), Ozone climatological profiles for satellite retrieval algorithms, *J. Geophys. Res.*, *112*(D5), D05308, doi:10.1029/2005JD006823.

Rahpoe, N. et al. (2015), Relative drifts and biases between six ozone limb satellite measurements from the last decade, *Atmos. Meas. Tech.*, *8*(10), 4369–4381, doi:10.5194/amt-8-4369-2015. [online] Available from: http://www.atmos-meas-tech.net/8/4369/2015/

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric sounding: Theory and Practice*, World Scientific, Singapore.

Sofieva, V. F. et al. (2013), Harmonized dataset of ozone profiles from satellite limb and occultation measurements, *Earth Syst. Sci. Data*, *5*(2), 349–363, doi:10.5194/essd-5-349-2013. [online] Available from: http://www.earth-syst-sci-data.net/5/349/2013/

Weatherhead, E. C. et al. (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res. Atmos.*, *103*(D14), 17149–17161, doi:10.1029/98JD00995. [online] Available from: http://dx.doi.org/10.1029/98JD00995