QA4ECV

Quality Assurance for Essential Climate Variables

Project Number 607405

Deliverable:	D4.6 – Product Specification Document for NO_2 ECV precursor product
Responsible Partner:	KNMI
Delivery date:	10-02-2017

QA4ECV NO_2 ECV precursor

A joint product by: KNMI, BIRA-IASB, University of Bremen, MPI-C, and Wageningen University

Product Specification Document for the QA4ECV NO_2 ECV precursor product

(Version 1.1)

K. Folkert Boersma[1,5], Jos van Geffen[1], Henk Eskes[1], Ronald van der A[1], Isabelle De Smedt[2], Michel Van Roozendael[2], Huan Yu[2], Andreas Richter[3], Enno Peters[3], Steffen Beirle[4], Thomas Wagner[4], Alba Lorente[5], Tracy Scanlon[6], Steven Compernolle[2], Jean-Christopher Lambert[2]

1: KNMI, 2:BIRA-IASB, 3: IUP Bremen, 4: MPIC, 5: WUR, 6: NPL

Change Record

Issue	Date	Pages Changed	Comments
Issue and draft number, i.e. 1A	Date changes approved	List of pages changed	Comments on why elements have been changed
0.61	15-01-2017		First draft after conversion to QA4ECV template
0.62	19-01-2017		Finished user guidance criteria
0.63	25-01-2017		Abbreviations included
0.64	26-01-2017		References included
0.7	01-02-2017		Comments Isabelle De Smedt included
0.75	02-02-2017		Comments Jos van Geffen and Alba Lorente included
0.77	03-02-2017		Formatting for equivalence with HCHO
0.8	03-02-2017		Comments Andreas Richter and Steven Compernolle included
0.9	08-02-2017		Comments from Steffen Beirle and Thomas Wagner included
1.0	10-02-2017		Review comments by Jean-Christopher Lambert included
1.1	06-05-2017	120	Website link corrected following Technical Review Report

CONTENTS

TERMS & ACRONYMS

AMF	Air Mass Factor
AGU	American Geophysical Union
BIRA-IASB	Belgian Institute for Space Aeronomy
DAK	Doubling Adding KNMI radiative transfer model
DOAS	Differential Optical Absorption Spectroscopy
DOMINO	Dutch OMI NO2 retrieval system
ECV	Essential Climate Variable
GOME	Global Ozone Monitoring Experiment
IUP Bremen	Institute for Environmental Physics Bremen
KNMI	Royal Netherlands Meteorological Institute
Level-1b	Geophysical radiances and irradiance measured by satellite sensor at top-of-atmosphere
MAX-DOAS	Multi-Axis DOAS
MPI-C	Max Planck Institute for Chemistry Mainz
NDACC	Network for the Detection of Atmospheric Composition Change
NetCDF	Network Common Data Form (www.unidata.ucar.edu/software/netcdf/)
NISE	Near-real-time Ice and Snow Extent
OMCLDO2	OMI/Aura Cloud Pressure and Fraction (O_2-O_2 absorption)
OMI	Ozone Monitoring Instrument
QA4ECV	Quality Assurance for Essential Climate Variables
RAA	Relative Azimuth Angle
SCD	Slant Column Density
SCIAMACHY	Scanning and Imaging Spectrometer for Atmospheric Chartography
STREAM	STRatospheric Estimation Algorithm from Mainz
SZA	Solar Zenith Angle
TM5 (-MP)	Tracer Model version 5 (Massive Parallel)
USGS	United States Geological Survey
UUID	Universally Unique Identifier
UV/Vis	Ultraviolet/Visible
VCD	Vertical Column Density
VZA	Viewing Zenith Angle
WGS84	World Geodetic System reference coordinate system (est. 1984)
WUR	Wageningen University and Research
ZSL	Zenith-scattered-light DOAS spectrometer

1. Executive Summary

This document is the product user manual for the QA4ECV NO_2 ECV precursor data product. This is the first document a user should read to get acquainted with the QA4ECV NO_2 data product. The QA4ECV NO_2 ECV precursor data product provides the tropospheric, stratospheric, and total NO2 column densities on the global scale for the period 1996-2015. The main product is the tropospheric NO_2 column density, which is defined as the vertically integrated number of NO_2 molecules between the Earth's surface and the tropopause (WMO-criterion [WMO, 1992]: tropopause defined as the lowest level at which the lapse rate decreases to 2 K/km or less) per unit area of the satellite ground pixel. The QA4ECV NO_2 product is, or will be, available for the sensors OMI, GOME-2(A), SCIAMACHY and GOME.

This document specifies the QA4ECV NO_2 data product, version 1.0, available since October 2016. A consortium consisting of BIRA-IASB, IUP Bremen, KNMI, MPI-C, and WUR has set together a community algorithm to produce an 11+ years (October 2004 – December 2015) record of OMI NO2 data based on improved NO2 slant columns, stratospheric correction methods, and air mass factor calculations within the framework of the EU FP7 QA4ECV project. The product is publicly available as data and images through http://www.qa4ecv.eu/ecv/no2-pre/data. For details on the QA4ECV NO_2 retrieval algorithm, please see the reports provided as deliverable D4.2 [QA4ECV, 2016a] and D4.4 [QA4ECV, 2016b] in the QA4ECV project. More information on the QA4ECV NO_2 AMF approach can be found in Lorente et al. [2016]. Quality assessment of the QA4ECV NO_2 algorithm will be available in D5.5, while quality assessment of the product (including validation with reference data) will be available in D5.6.

The document is partly a Product User Guide, i.e. it introduces the users to the product. Section 5 provides practical guidance on how to use the data. We distinguish different user groups, in order to provide tailor-made guidance to users since not all users need to interpret the data at the same level of scrutiny. This section explains what users should do to visualize and interpret the data, how to use the averaging kernels associated with the retrievals, and how to take associated uncertainties into account for specific applications. Moreover it presents advice on how not to use the data. Important practical hints are provided for how to use the quality information in the data files so that suspect retrievals can be excluded from analysis. The user can get an answer to the question 'How can I be sure I'm interpreting valid QA4ECV NO_2 data'?

Section 2 explains the file name convention. Section 3 of this document can be read as a Product Specification Document, i.e. it provides an overview of all the technical details of the product. It explains variable names, metadata information, and the data format. Section 4 explains how to obtain the QA4ECV NO_2 data. More details on the theoretical grounds of the ECV data production, the ex-ante assessment of the product quality and of its associated uncertainties, and the ex-post geophysical validation of the data using external means can be found in ad hoc documents, namely, the Algorithm Theoretical Basis Document (ATBD), the Quality Assessment Report of Atmosphere ECV Retrieval Algorithm (QA4ECV D5.5) and the Quality Assessment Report of the Atmosphere ECV data Products (QA4ECV D5.6).

2. Introduction

2.1 QA4ECV NO_2 ECV precursor product

The QA4ECV NO_2 ECV precursor product provides the tropospheric, stratospheric, and total NO_2 column densities on the global scale for the period 1996-2015. Harmonized, daily level 2 data are produced with one consistent retrieval algorithm from the UV/Vis spectrometers GOME (1996-2003), SCIAMACHY (2002-2011), OMI (2004-2015), and GOME-2(A) (2007-2015). The spatial resolution of the (nadir) pixels varies from 320×40 km^2 (GOME) to 24×13 km^2 (OMI), and the overpass times are early afternoon for the OMI sensor (~13:40 hrs) and mid-morning (09:30-10:30 hrs) for the other instruments.

The main product(s) of the QA4ECV NO_2 ECV precursor product are the tropospheric, stratospheric, and total NO_2 column densities on the global scale for the period 1996-2015. The main product (detailed in section 3.4), the tropospheric NO_2 column density, is defined as the vertically integrated number of NO_2 molecules between the Earth's surface and the tropopause [WMO, 1992] per unit area of the satellite pixel. Another frequently used data product is the stratospheric NO_2 column density, defined as the vertically integrated number of NO_2 molecules between the tropopause and the top-of-atmosphere (NO_2 in the mesosphere is negligible) per unit area of the satellite ground pixel. The unit of all NO_2 column density products is molecules cm^{-2}.

The name of the QA4ECV NO_2 ECV precursor output level-2 (hence the "L2") file has the following name convention:

QA4ECV_L2_<product>_<instrument>_<starttime>_<orbitnumber>_<fitwindow>_<version>.nc

with the <starttime> taken from the original level-1b filename and <fitwindow> is "fitA" for the 425-450 nm fit window, "fitB" for the 405-465 nm fit window. For example:

- QA4ECV_L2_NO2_GOME2A_20130202T174159_o32648_fitA_v1.nc
- QA4ECV_L2_NO2_OMI_20050101T002000_o02472_fitB_v1.nc

2.2 Research and applications

Nitrogen dioxide (NO_2) is a toxic pollutant close to the Earth's surface and plays a major role in the catalytic cycles controlling influences the ozone layer in the stratosphere. NO_2 also plays a critical role in the formation of tropospheric aerosols and ozone, both important air pollutants and climate agents at the local to regional scale. In large quantities it can also play a moderate role in the radiation budget at local scale. The satellite NO_2 products are used to monitor tropospheric and stratospheric NO_2, and to investigate the burden of tropospheric pollution from anthropogenic (e.g. traffic, power plants, ships) and natural emissions (e.g. soils, biomass burning, lightning). In combination with other pieces of information, tropospheric NO_2 columns retrieved from satellite measurements are used to determine human exposure to pollution, nitrogen-deposition, ozone formation, and the chemical regime of the atmosphere.

2.3 Data set and evolution

With the launch of the GOME sensor on board of ESA's ERS-2 in 1995, it became possible to detect tropospheric columns of NO_2 on a global scale. Algorithms have been developed since to retrieve NO_2 slant columns from backscatter reflectance spectra. To estimate the stratospheric contribution to those slant columns, sophisticated schemes have been set up that identify the NO_2 slant columns as mostly stratospheric in regions with very small amounts of tropospheric NO_2, such as over the Pacific Ocean. This allows the separation of the stratospheric and tropospheric contributions to the NO_2 slant columns. Subsequently, air mass factors are calculated that convert the NO_2 slant columns into vertical columns, by taking into account ancillary information on cloud cover, surface albedo, surface pressure, and vertical NO_2 and temperature vertical profiles. The main retrieved parameters are: (1) tropospheric NO_2 vertical column density, (2) stratospheric NO_2 vertical column density, and (3) total NO_2 vertical column density.

Within the FP7 QA4ECV project, a consortium of European groups (BIRA-IASB, IUP Bremen, MPIC, and WUR) led by KNMI, has carefully compared various 'state-of-science' approaches for the main retrieval steps. This exercise (with contributions from Peking University, Leicester University, and NASA GSFC) has led to recommendations and best practices for the QA4ECV NO_2 retrieval algorithm, and also resulted in more insight in the nature of the retrieval uncertainties. The results and findings of these detailed investigations have been adopted in the QA4ECV NO_2 algorithm. The comparisons and recommendations for algorithmic and software choices are summarized in QA4ECV Deliverable 4.2 [QA4ECV, 2016[a]]. A short overview of the set-up of the retrieval software for NO_2, and basic algorithm choices is given in QA4ECV Deliverable 4.4 [QA4ECV, 2016[b]].

2.4 Purpose and Scope of Document.

The purpose of this guide is to provide users with a basic understanding of the architecture and contents of the QA4ECV NO_2 ECV precursor data product(s), enabling users to interpret and use the products. The guide assumes that the user has a basic knowledge of the construction and operation of the UV/Vis spectrometers for which the QA4ECV products are generated.

This guide includes information and explanations that should enhance a user's understanding of the products. It includes descriptions and explanations of characteristics and quality of the product and the online guide has links (or future links) to imagery and graphics exemplifying those characteristics. This section (2) introduces the user to the QA4ECV NO_2 ECV precursor data product, explaining the main product, applications and its heritage. In section 2, the output data files and contents are detailed, including a description of the output file's metadata content of all settings and ancillary data used in the retrieval, ensuring full traceability. A description of intermediate and final data products is included.

2.5 Definitions

The general method used for the derivation of NO_2 vertical column densities (VCDs) from UV/Vis spectral measurements is the Differential Optical Absorption Spectroscopy method (DOAS), which involves 3 steps. First, the effective slant column density (SCD) (corresponding to the integrated NO_2 concentration along the effective atmospheric optical path: N_s) is derived through a least-squares fit of the measured Earth reflectance spectrum by laboratory absorption cross-sections and a low order polynomial. Then, the stratospheric contribution to the NO_2 SCD is estimated from a stratospheric correction scheme. Finally the tropospheric and stratospheric slant column densities are converted into VCDs by means of air mass factors (M) obtained from radiative transfer calculations, accounting for the presence of clouds, surface properties, and best-guess NO_2 vertical profiles.

The main outputs of the algorithm are the tropospheric vertical columns ($N_{v,,trop}$), defined as the vertically integrated number of NO_2 molecules between the Earth's surface and the tropopause per unit area of the satellite ground pixel, the tropospheric AMF (M_{trop}), and the averaging kernel. Other output products are the stratospheric vertical column density ($N_{v,,strat}$), the slant column density (N_s), and the total vertical column (N_v). Complementary product information includes clear sky air mass factors, uncertainties on the tropospheric column (detailed by N_s, M, $N_{v,0}$ errors) and quality flags.

Abbreviation	Description	Symbols
SAA	Solar Azimuth Angle at WGS84 ellipsoid for center ground pixel, defined East-of-North	ϕ_0
VAA	Viewing Azimuth Angle at WGS84 ellipsoid for center ground pixel, defined East-of-North	ϕ
RAA	Relative Azimuth Angle defined as difference between Solar and Viewing Azimuth angles *For OMI and GOME-2:*	$\lvert 180° - (\phi_0 - \phi) \rvert$
	For GOME and SCIAMACHY:	$\lvert \phi - \phi_0 \rvert$
SZA	Angle between the zenith and the line of sight to the Sun at WGS84 ellipsoid for center ground pixel	θ_0 $\cos\theta_0 = \mu_0$
VZA	Angle between the zenith and the line of sight to the satellite at WGS84 ellipsoid for center ground pixel	θ $\cos\theta = \mu$

The definitions of relative azimuth angle are different between the instruments because the definitons of SAA and VAA differ between different level 1 formats.

3. Available Products

3.1 Intermediate Products

The overall structure of the QA4ECV NO_2 ECV precursor cdl file is as follows:

- Global Attributes
- group: PRODUCT
    - dimensions
    - variables
    - group: SUPPORT_DATA
        - group: GEOLOCATIONS
            - variables
        - group: DETAILED_RESULTS
            - variables
        - group: INPUT_DATA
            - variables
- group: METADATA

For the groups in this list, an overview of the contents is given below using an OMI file as example.

3.2 Global attributes

Name in file	Explanation
Conventions = "CF-1.7"	The file follows the Climate and Forecast metadata conventions
date_created = "2016-03-14T14:49:37Z"	The date/time of the final step in the level-1-to-2 data processing chain, when the file was created.
equator_crossing_time = “2005-01-01T01:16:07Z”	Time (UTC) when spacecraft passed the Equator
equator_crossing_longitude = -173.78	Longitude (in degrees) where spacecraft passed the Equator
institution = "BIRA-IASB & IUPB & KNMI & MPIC & WUR"	The QA4ECV institutes responsible for this data product, given in alphabetic order
product_version = "1"	The main version number of the QA4ECV NO2 product appearing in the filename
project = "QA4ECV”	Name of the FP7 Space project generating this product
reference = “http://www.qa4ecv.eu/"	Website where data and information on the NO2 ECV precursor data can be found
source = "OMI / EOS-Aura"	Instrument and satellite
slant_column_fit_window = "[405-465] nm"	Spectral fitting window used in retrieval
slant_column_processor_name = "QDOAS"	Name of the spectral fitting processor used to retrieve SCDs
slant_column_processor_version = "v2.111"	Version of the spectral fitting processor used to retrieve SCDs
title = " QA4ECV nitrogen dioxide (NO2)	Main title of the data product

column data"	
vertical_column_processor_name = "TM5-MP v1.0 DOMINO v3.2"	Version of the stratospheric correction and AMF processor used to retrieve the final product
vertical_column_processor_version = "3.2"	The version of the data assimilation procedure is DOMINO v3.2
keywords_vocabulary = "AGU index terms, http://publications.agu.org/author-resource-center/index-terms/"	The "keywords" have been selected from the keywords_vocabulary list to be: • 0345 Pollution: urban and regional • 0365 Troposphere: composition and chemistry • 0368 Troposphere: constituent transport and chemistry • 3360 Remote sensing • 3363 Stratospheric Dynamics
keywords = "0345 Pollution, Urban and Regional; 0365 Troposphere, Composition and …"	
standard_name_vocabulary = "NetCDF Climate and Forecast Metadata Conventions Standard …."	
cdm_data_type = "SWATH"	Type of the data
naming_authority = "nl.knmi"	The 'tracking_id' holds a unique UUID number to identify the file, and the 'id' is the also unique logical filename (i.e. without extention), linking the L2 file to the beginning of the processing chain; both are connected to the 'naming_authority'. To harmonise this, KNMI will generate these id's for all output files.
tracking_id = "df9c9c9a-4f33-11e6-a9d6-901b0e192aa6"	
id = "QA4ECV_L2_NO2_OMI_20050101T002000_o02472_fitB_v1"	Filename without extension
orbit = 2472	The number of the orbit since EOS-Aura's launch.
level1b_file = "OMI-Aura_L1-OML1BRVG_2005m0101t0020-o02472_v003-2007m0327t164446.he4"	The name of the level-1b file processed in the spectral fit to retrieve the SCD
time_reference = "2005-01-01T00:00:00Z"	Date of the start of the orbit; variable delta_time gives the offset w.r.t. this reference
time_reference_days_since_1950 = 20089L	Number of days passed since 01-01-1950
time_coverage_start = "2005-01-01T00:42:52Z"	The begin time (UTC) of the first ground pixel in the orbit and the end time of the last ground pixel in the file. The start time as given in the level-1b name may be different.
time_coverage_end = "2005-01-01T01:37:41Z"	
time_reference_julian_day = 2453371.5	
time_reference_seconds_since_1970 = 1104537600L	Number of seconds passed since 01-01-1970 (Unix Reference Time)
geospatial_lat_min = -90.	Definitions for users to produce world maps
geospatial_lat_max = 90.	
geospatial_lon_min = -180.	
geospatial_lon_max = 180.	
processing_status = "Offline QA4ECV	QA4ECV algorithm Status at the end of the

processing, DOMINO v3.2 complete"	processing chain
chemistry_transport_model_name = TM5-MP	
chemistry_transport_model_version = 1.0, 26 September 2016	
vertical_column_processor_name = DOMINO v3	
vertical_column_processor_version = 3.2.0, 5 October 2016	

Note that the order of these attributes is not necessarily how your netcdf software lists them.

3.3 Group: product

Dimensions

Any variable associated with an individual ground pixel (for example the vertical NO_2 column) will have dimensions (e.g. time, scanline, ground_pixel). For each of the dimensions a variable of the same name is defined, as shown below.

scanline = UNLIMITED	The along-track dimension, which varies from orbit to orbit; it will be automatically set if the variables are filled. Note: the word "scanline" originates from (TROP)OMI, for which it reflects the individual swaths, each of which has "ground_pixel" across-track pixels. For GOME-2, SCIAMACHY and GOME-1 "scanline" refers to individual pixels (scans), since the dimension "ground_pixel" is 1.
ground_pixel = 60	The across-track dimension: for OMI this is 60, while for the other instruments it is 1, so that it refers to individual pixels (scans).
corner = 4	The number of corners per ground pixel.
time = 1	The 'time' dimension refers to the 'time slice' of the data; for qa4ecv it always has the value '1'.
polynomial_exponents = 5	Number of coefficients in the DOAS polynomial, for the polynomial_coefficients dataset (see below).
layer = 34	Layers in the TM5 model: 0,1,…,33.
nv = 2	sets the indices of the upper and lower bounds of the TM5 layers, i.e. 0 and 1

Variables

Field name PRODUCT/*	Quantity	[unit]	Symbol
latitude	pixel center latitude	degree	
longitude	pixel center longitude	degree	
delta_time	offset from reference start time of measurement	seconds	
time_utc	Time of observation as ISO 8601 date-time string		
processing_error_flag	0 for successful processing, 1 in case the processing failed somewhere. See also the "processing_quality_flags" dataset in	1	

	PRODUCT/ SUPPORT_DATA/ DETAILED_RESULTS.		
tropospheric_no2_vertical_column	tropospheric NO2 vertical column density (per area unit)	molecules cm^{-2}	$N_{v,trop}$
tropospheric_no2_vertical_column_uncertainty	uncertainty on tropospheric NO2 vertical column density (per area unit)	molecules cm^{-2}	$\sigma_{N_{v,trop}}$
tropospheric_no2_vertical_column_uncertainty_kernel	uncertainty on tropospheric NO2 vertical column density (per area unit) when averaging kernel is applied	molecules cm^{-2}	$\sigma_{N_{v,trop,k}}$
averaging_kernel	(vertical) averaging kernel (34 layers)	1	$\mathbf{A}$
amf_trop	Tropospheric air mass factor	1	M_{trop}
amf_total	Total air mass factor	1	M
tm5_tropopause_layer_index	Index of TM5 vertical layer where tropopause is located	1	l_{tp}^{TM5}
tm5_pressure_level_a	TM5 hybrid A coefficient at interface levels	Pa	A_l^{TM5}
tm5_pressure_level_b	TM5 hybrid B coefficient at interface levels	1	B_l^{TM5}
tm5_surface_pressure	It is this tm5_surface_pressure that is needed by users of the kernels. It is based on the climatological surface pressure of the ground pixel (surface_pressure, from the cloud product) and corrected for meteorological conditions (available from TM5-MP).	Pa	p_s

3.4 Group: PRODUCT/SUPPORT_DATA/GEOLOCATIONS

Field name	Quantity	[unit]	Symbol
pixel_type	• 0 = regular pixel: GOME-2, SCIA, GOME-1 forward & OMI nominal • 1 = backscan pixel: GOME-2, SCIA, GOME-1 • 2 = narrow swath pixel: GOME-2, GOME-1 forward & OMI zoom mode • 3 = narrow swath backscan pixel: GOME-2, GOME-1 • 4 = reduced swath forward pixel: GOME-2 • 5 = reduced swath backscan pixel: GOME-2 • 8 = nadir static pixel: GOME-2 • 16 = polar view pixel: GOME		
solar_zenith_angle	Solar zenith angle of the satellite measurement at the centre of the	degree	θ_0

	ground pixel location on the reference ellipsoid. Angle is measured away from the vertical.		
viewing_zenith_angle	Viewing zenith angle of the satellite measurement at the centre of the ground pixel location on the reference ellipsoid. Angle is measured away from the vertical.	degree	θ
relative_azimuth_angle	The relative azimuth angle (raa) at the centre of the ground pixel on the reference ellipsoid is computed from the solar azimuth angle (saa) and viewing azimuth angle (vaa) in the following way, where the “definition” attribute of the two raa variables is set by the SCD processing institute: a) for GOME-1 and SCIAMACHY: raa = | vaa – saa | ; if (raa > 180) then raa = 360 – raa b) for OMI and GOME-2: raa = | 180 – (saa - vaa) | ; if (raa > 180) then raa = 360 – raa	degree	$\varphi - \varphi_0$
solar_zenith_angle_sat	Solar zenith angle at the satellite. Angle is measured away from the vertical.	degree	
viewing_zenith_angle_sat	Viewing zenith angle of the satellite at the satellite. Angle is measured away from the vertical.	degree	
relative_azimuth_angle_sat	Relative azimuth angle at satellite.	degree	
latitude_bounds	According to the CF standard, the ground pixel corner coordinates should be given in a counter-clockwise order, i.e. not the usual SCIAMACHY or GOME order.	degree	
longitude_bounds		degree	
satellite_latitude	Latitude of the geodetic sub satellite point on the WGS84 reference ellipsoid.	degree	
satellite_longitude	Longitude of the geodetic sub satellite point on the WGS84 reference ellipsoid.	degree	
satellite_altitude	The altitude of the satellite with respect to the geodetic sub satellite point on the WGS84 reference ellipsoid.	m	

3.5 Group: PRODUCT/SUPPORT_DATA/DETAILED_RESULTS

Data sets holding the intermediate slant column results and support data for advanced users

Field name	Quantity	[unit]	Symbol
processing_quality_flags	used to indicate errors, filters and warnings, both from the SCD retrieval and the subsequent algorithm steps, where 0 means success. In the SCD retrieval step of QA4ECV, limited to 0 for success and 42 ("generic exception") for SCD retrieval failure. In case of failure and if more information about the cause of the failure is available, one of the other values from the list of TROPOMI flag values can be used. An overview of the processing_quality_flags is given at the end of this document.	1	
tropospheric_no2_vertical_column_stream	tropospheric NO2 vertical column density (per area unit) with STREAM stratospheric NO2 correction	molecules cm^{-2}	$N_{v,trop,str}$
scd_no2	NO2 slant column density	molecules cm^{-2}	N_s
scd_no2_uncertainty	uncertainty of the NO2 slant column density (per area unit)	molecules cm^{-2}	σ_{N_s}
scd_o3	O3 slant column density (per area unit)	molecules cm^{-2}	
scd_o3_uncertainty	uncertainty of the O3 slant column density (per area unit)	molecules cm^{-2}	
scd_o2o2	O2-O2 slant column density (per area unit)	molecules cm^{-2}	
scd_o2o2_uncertainty	uncertainty of the O2-O2 slant column density (per area unit)	molecules cm^{-2}	
scd_h2o_vapor	H2O slant column density (per area unit)	molecules cm^{-2}	
scd_h2o_vapor_uncertainty	uncertainty of the H2O slant column density (per area unit)	molecules cm^{-2}	
scd_h2o	liquid water fitting coefficient	m	
scd_h2o_liquid_uncertainty	uncertainty of the liquid water fit coefficient	m	
ring_coefficient	fit coefficient for the Ring spectrum	1	
ring_coefficient_uncertainty	uncertainty of the Ring fit coefficient	1	
intensity_offset_a	fit coefficient A of the intensity offset	1	
intensity_offset_a_uncertainty	uncertainty of fit coefficient A of the intensity offset	1	
intensity_offset_b		nm^{-1}	
intensity_offset_b_uncertainty		nm^{-1}	
polynomial_coefficients	polynomial coefficients of DOAS fit	1	
polynomial_coefficients_uncertainty	uncertainty of the polynomial coefficients of DOAS fit	1	
rms_fit	Root-mean-square residual of the DOAS fit	1	

number_of_spectral_points_in_retrieval	number of spectral points used in the retrieval	1	
radiance_calibration_offset	Wavelength calibration parameters of the Earth spectral radiance	nm	
radiance_calibration_stretch		1	
radiance_calibration_wavelength		nm	
irradiance_calibration_offset	Wavelength calibration parameters of the solar spectral irradiance	nm	
irradiance_calibration_stretch		1	
irradiance_calibration_wavelength		nm	
stratospheric_no2_vertical_column	stratospheric NO2 vertical column density (per area unit)	molecules cm-2	$N_{v,strat}$
stratospheric_no2_vertical_column_uncertainty	uncertainty of the stratospheric NO2 vertical column density (per area unit)	molecules cm-2	$\sigma_{N_{v,strat}}$
total_no2_vertical_column	total NO2 vertical column density (per area unit) (calculated as N_s/M)	molecules cm-2	N_v
total_no2_vertical_column_uncertainty	uncertainty of the total NO2 vertical column density (per area unit)	molecules cm-2	σ_{N_v}
summed_no2_total_vertical_column	summed NO2 vertical column (calculated as $N_{v,trop}+N_{v,strat}$)	molecules cm-2	$N_{v,sum}$
summed_no2_total_vertical_column_uncertainty	uncertainty of summed NO2 vertical column density (per area unit)	molecules cm-2	$\sigma_{N_{v,sum}}$
stratospheric_no2_vertical_column_stream	stratospheric NO2 vertical column density (per area unit) estimated with STREAM method	molecules cm-2	$N_{v,stream}$
stratospheric_no2_vertical_column_stream_uncertainty	uncertainty in STREAM stratospheric NO2 vertical column density (per area unit)	molecules cm-2	$\sigma_{N_{v,stream}}$
amf_strat	stratospheric air mass factor	1	M_{strat}
amf_geo	geometric air mass factor	1	M_{geo}
amf_clear	clear-sky tropospheric air mass factor	1	M_{clear}
ghost_column	below-cloud NO2 column density (per area unit) according to TM5 simulation	molecules cm-2	$N_{v,ghost}$
cloud_radiance_fraction_no2	fraction of the radiance coming from the cloudy part of the satellite pixel	1	
tropospheric_no2_vertical_column_uncertainty_scd	uncertainty in the tropospheric NO2 vertical column density (per area unit) due to SCD uncertainty	molecules cm-2	
tropospheric_no2_vertical_column_uncertainty_stratosphere	uncertainty in the tropospheric NO2 vertical column density (per area unit) due to stratospheric correction uncertainty	molecules cm-2	
tropospheric_no2_vertical_column_uncertainty_amftrop	uncertainty in the tropospheric NO2 vertical column due to uncertainty in tropospheric air mass factor	molecules cm-2	
tropospheric_no2_vertical_column_uncertainty_amftrop_albedo	uncertainty in the tropospheric NO2 vertical column density (per area unit) due to uncertainty in the assumed surface albedo	molecules cm-2	
tropospheric_no2_vertical_column_uncertainty_amftrop_cloud_fraction	uncertainty in the tropospheric NO2 vertical column density (per area unit) due to uncertainty in the observed cloud fraction	molecules cm-2	
tropospheric_no2_vertical_column	uncertainty in the tropospheric NO2	molecules	

n_uncertainty_amftrop_cloud_pressure	vertical column density (per area unit) due to uncertainty in the observed cloud pressure	cm-2	
tropospheric_no2_vertical_column_uncertainty_amftrop_tm5_profile	uncertainty in the tropospheric NO2 vertical column density (per area unit) due to uncertainty in the a priori TM5 NO2 profile	molecules cm-2	
scd_no2_stripe_correction_omi	stripe correction for OMI NO2 slant column density (for 60 OMI rows)	molecules cm-2	

3.6 Group: PRODUCT/SUPPORT_DATA/INPUT_DATA

The input data (ancillary data) used in the retrieval of NO_2 columns.

Field name	Quantity	[unit]	Symbol
surface_altitude	Pixel-average terrain height used in OMCLDO2 and FRESCO cloud algorithms. OMI O2-O2 data: Veefkind et al., 2016; http://www.atmos-meas-tech-discuss.net/amt-2016-48/	m	
surface_altitude_uncertainty	Uncertainty in the pixel-average terrain height	m	
surface_classification	For later use	1	
surface_pressure	Pixel surface pressure from the cloud product calculated from surface_altitude and mid-latitude standard atmosphere profile.	Pa	p_s
surface_albedo_no2	surface albedo in the NO2 fit window (440 nm, mode). Reference: OMI: Kleipool et al., 2008, http://onlinelibrary.wiley.com/doi/10.1029/2008JD010290/full	1	a_s
surface_albedo	Surface albedo used in the cloud product (471 nm for OMI)	1	
cloud_fraction	effective cloud fraction from the cloud data product. OMI O2-O2 data: Veefkind et al., 2016; http://www.atmos-meas-tech-discuss.net/amt-2016-48/.		f_{cl}
cloud_fraction_uncertainty	uncertainty in the cloud fraction	1	$\sigma_{f_{cl}}$
cloud_pressure	cloud optical centroid pressure from the cloud data product	Pa	p_{cl}
cloud_pressure_uncertainty	uncertainty in the cloud pressure	Pa	$\sigma_{p_{cl}}$
cloud_albedo	cloud albedo from the cloud data product	1	
cloud_albedo_uncertainty	uncertainty in the cloud albedo	1	
scene_pressure	scene pressure from the cloud data product (retrieved in situations with high surface reflectivity, i.e. snow/ice)	Pa	
scene_pressure_uncertainty	uncertainty in the retrieved scene pressure	Pa	
scene_albedo	scene albedo in the cloud data product (retrieved in situation with high surface	1	

	reflectivity, i.e. snow/ice; at 471 nm for OMI)		
scene_albedo_uncertainty	Uncertainty in the retrieved scene albedo		
snow_ice_flag	flag indicating presence of snow or ice at center of ground pixel using NISE flag values	1	
omi_xtrack_flags	OMI specific variable, indicating issues with the so-called row anomaly. From OMI level 1 files		

Group: METADATA

The information on settings and ancillary inputs used in the retrieval is given in the metadata group of the output file. The interested user should investigate these settings to find out which settings and corrections were used exactly in the retrieval of SCDs and VCDs. The following shows the metadata of an OMI file as example:

METADATA/ALGORITHM_SETTINGS/SLANT_COLUMN_RETRIEVAL/

analysis_method = optical density fitting
convergence_criterion = 1.0e-3
fit_polynomial_degree = 4 (5 coefficients)
fit_slit_function = no
fit_window = 405.00, 465.00 nm
instrument_slit_function = row_12-47_omi_vis_slit_aver.dat
intensity_offset_coefficients = 1
intensity_offset_i_0 = true
interpolation_method = spline
irradiance_calibration_analysis_method = optical density fitting
irradiance_calibration_shift = true
irradiance_calibration_stretch = 1st order
irradiance_calibration_subwindows = 1
irradiance_calibration_window_limits = 405.00, 465.00 nm
irradiance_file = OMI-Aura_L1-GLOBAL-OMTMIRRYA_2005m0101t0000-syear-rPDS01_v003-2007m0716t145802.he4
least_squares_fit_weighting = no
linfit_dampening =
linfit_preshift =
lv1_calibration_options =
lv1_extractor_version =
maximum_number_of_iterations = 3
processing_algorithm = QDOAS version Globalcalib6 – 14 March 2016
qdoas_converter_version = 1.30
radiance_calibration_shift = true
radiance_calibration_stretch = 1st order
radiance_file = OMI-Aura_L1-OML1BRVG_2011m0505t1638-o36194_v003-2011m0506t000412.he4
reference_spectrum_h2o_liquid = water_DONNEELITTERA_6_POPEFREY_HIGHERRES_380-727_smoothed_new_OMIconv60_m-1.xs
reference_spectrum_h2o_vapor = h2o_HITRAN_2013_293K_OMIcon60.xs
reference_spectrum_no2 = no2_VANDAELE_1998_220K_OMIconv60.xs

reference_spectrum_o3 = O3_SERDYUCHENKO_OMIconv60.xs
reference_spectrum_o2o2 = o4_THALMAN_VOLKAMER_2-13_293_EDITED.VAC_OMIconv60.xs
reference_spectrum_ring = ring_sao2010_hr_norm_OMIconv60.xs
solar_reference = sao2010_solref_vac_OMIconv60.dat
spike_tolerance_factor = 5
sza_limits = 0.0, 110.0 degrees
traceability_chain: http://www.qa4ecv.eu/ecv/no2-pre/main/doas
undersampling_correction = no

METADATA/ALGORITHM_SETTINGS/VERTICAL_COLUMN_RETRIEVAL/
amf_lookup_table_filename = S5P_OPER_LUT_NO2AMF_00000000T000000_99999999T999999_20160527T173500.nc

assimilation_correlation_concentration_scale = 30.0, 1/e reduction for this percentage concentration difference
assimilation_correlation_length_scale = 600 km
assimilation_maximum_cloud_fraction = 1.01
assimilation_maximum_omf_difference = 10.0e15 molec/cm^2
assimilation_maximum_sza = 85.0 degree
assimilation_minimum_cloud_fraction = -0.01
assimilation_stratospheric_column_error = 0.2e15 molec/cm^2
assimilation_tropospheric_column_error = 6.0e15 molec/cm^2
chemistry_transport_model_name = TM5-MP
chemistry_transport_model_version = 1.0, 26 September 2016
fit_window_centre = 437.5 nm
minimum_tropospheric_amf = 0.02
processing_algorithm = QA4ECV
processing_algorithm_version = 1
slant_column_noise_scaling_factor = 0.814
stripe_correction_averaging_time = 7.0 days
stripe_correction_pacific_longitude_maximum = -137.0 degree
stripe_correction_pacific_longitude_minimum = -163.0 degree
temperature_correction_factor = 1.0 - 0.00316 (T-Tdoasfit) + 3.39e-6 (T-Tdoasfit)^2; Tdoasfit = 220.0
uncertainty_model_stratosphere_verical_column = 0.2e15 molec/cm2
uncertainty_o2o2_cloud_fraction = 0.025
uncertainty_o2o2_cloud_top = 50.0 hPa
uncertainty_surface_albedo = 0.015
vertical_column_processor_name = DOMINO v3
vertical_column_processor_version = 3.2.0, 5 October 2016
pre_vcd_processor_version = 1.00
post_vcd_processor_version = 0.95

4. Obtaining QA4ECV NO_2 ECV precursor product

4.1 How to obtain the QA4ECV NO_2 ECV precursor data

Visit www.qa4ecv.eu, then:

- click ECV data in the header menu.
- click on the link NO_2 in the menu Essential Climate Variables.
- click on the blue button 'Data Access'

You can directly download individual NO_2 Data Files (daily files). These files contain all the (final and intermediate) products listed in section 3.

To obtain 1 year of NO_2 data from OMI, you can use the (linux) wget command. As a preliminary example (link will be updated):

wget -A tar -r -l1 -nH -N -a "logfile" http://temis.nl/qa4ecv/no2col/data/omi/v1/2007/

4.2 How to obtain a priori NO_2 profiles used in the retrieval

For the AMF calculation in the QA4ECV NO_2 ECV precursor product, NO_2 a priori vertical profiles are used as simulated by the TM5-MP chemistry transport model [Williams et al., 2016]. These vertical profiles are sometimes requested by users to verify the assumptions in the retrieval, or to verify or reproduce their own AMF calculations. Because of the large additional volume of data in the QA4ECV data product, it was decided to not include the TM5 NO_2 profiles in the output data file.

To obtain the TM5 NO_2 profiles, one can send a request to Henk Eskes (KNMI, eskes@knmi.nl) on how to transfer the TM5-MP simulation's output. The TM5-MP files are in netcdf, and contain the 3-dimensional fields (1°×1° gridded, 3-hourly) of mixing ratios of NO_2, O_3, SO_2, HCHO, and temperature. At a later stage, and pending dissemination facilities, these data will be made publicly available as well.

5. Using the QA4ECV NO_2 ECV precursor product

5.1 Processing Product Data

The tropospheric NO_2 column is the principal QA4ECV NO_2 precursor product. For this product, we can distinguish different user groups: from users that take our product 'face value' (group 6) to more advanced users working on extensive scientific projects doing model-to-measurement comparisons and/or satellite validation studies (group 1). This product is useful for the following applications (see 5.3 below on how to use the data):

User group	User application	Data sets needed[1]
1	Tropospheric chemistry / air quality model evaluation Validation with tropospheric NO_2 profile measurements (e.g. aircraft, balloon, MAX-DOAS) Trend analysis	$N_{v,trop}$, $\sigma_{N_{v,trop}}$ M_{trop}, M, $\mathbf{A}$ A_l^{TM5}, B_l^{TM5}, l_{tp}^{TM5}, p_s
2	Tropospheric column comparisons, e.g. with other NO_2 column retrievals Emission/lifetime estimates (without explicit use of models)	$N_{v,trop}$, $\sigma_{N_{v,trop}}$
3	Stratospheric chemistry model evaluation and data assimilation Validation with stratospheric NO_2 profile measurements (limb/occultation satellite measurements)	$N_{v,strat}$, $\sigma_{N_{v,strat}}$ M_{strat}, M, $\mathbf{A}$ A_l^{TM5}, B_l^{TM5}, l_{tp}^{TM5}, p_s
4	Stratospheric column comparisons, e.g. with ground-based remote sensing instruments (MAX-DOAS, NDACC ZSL-DOAS)	$N_{v,strat}$, $\sigma_{N_{v,strat}}$
5	Whole atmosphere (troposphere+stratosphere) data assimilation systems	N_v, σ_{N_v} $\mathbf{A}$ A_l^{TM5}, B_l^{TM5}, l_{tp}^{TM5}, p_s
6	Visualization of the NO_2 product, outreach, mapping.	$N_{v,trop}$, $N_{v,strat}$, $N_{v,sum}$

5.2 How do I read and visualize QA4ECV NO_2 data?

An easy option to read and visualize the data is with the Panoply tool, available publicly from https://www.giss.nasa.gov/tools/panoply/. Make sure you install the version 4.6.2 of Panoply, released 2016-10-31, or a newer version.

Other tools, including read in software in python, IDL, and Fortran90 are under development and will be made available in due time.

[1] Along with these data sets, the user will have to obtain geolocation information (centre and corner coordinates of the piexels), pressure grids, and information on the clouds provided in the product.

5.3 Usage of the QA4ECV NO_2 data

The more advanced users interested in applications (1), (3), and (5), which require a deep understanding of the satellite data product, should engage sufficient effort to ensure that they seize appropriate understanding of both the capabilities and limitations of the satellite data product. It is important to realize that the satellite radiance measurement and the NO_2 column retrieval have a non-uniform vertical sensitivity to the NO_2 vertical distribution, and that this sensitivity is different for every ground pixel. The vertical sensitivity is determined by the surface and atmospheric properties, and by the viewing geometry at the time of measurement. In general, measurements taken under relatively cloud-free situations (typically cloud fraction below 20% of the ground pixel), high overhead sun, and small viewing zenith angles have good sensitivity to NO_2 down to the surface. For dark surfaces, cloudy scenes, scenes with high aerosol loading, low sun, and/or extreme viewing angles, sensitivity is much poorer, and potential errors in a priori assumptions propagate strongly in the retrieved column product. For very bright surfaces, especially the assumptions on the vertical NO_2 distribution also have the potential to introduce uncertainties in the retrieved column product.

Clouds

Cloud information is essential in the retrieval of tropospheric NO_2 columns because clouds shield the boundary layer, but enhance the visibility for absorbers at the cloud top due to multiple scattering and above due to the high albedo. For cloud fractions exceeding approximately 0.2 (or cloud radiance fractions >0.5), most of the satellite-observed signal originates from the cloudy part of the retrieval scene, and such scenes should not be used for applications aimed at emissions, deposition, surface NO2, and boundary layer chemistry. However, cloudy retrievals are useful for assessments of above-cloud NO_2 columns. The excellent sensitivity to above-cloud NO_2 has been demonstrated in a number of studies, for instance on lightning NO_x estimates (Beirle et al. [2006]; Boersma et al. [2005]) and cloud-slicing (Choi et al. [2014]; Belmonte Rivas et al. [2015]).

5.4 Application of the averaging kernel

To account for the differences in the vertical sensitivity and dependency on assumptions, users are strongly encouraged to make use of the averaging kernels provided along with the data product. The total column averaging kernels represent the relationship between the retrieved NO_2 column and the actual, true NO_2 distribution in the troposphere. Using the kernels is especially important in application where independent vertically resolved data (from models, or aircraft or MAX-DOAS measurements) is compared to the QA4ECV NO_2 ECV precursor.

For example, users interested in a model – satellite comparison may want to map the modelled NO_2 profiles via the averaging kernel [Eskes and Boersma, 2003] to what the sensor would retrieve ($\hat{y}_m$ is the 'retrieved' quantity) as follows:

$$\hat{y}_m = \mathbf{A} \cdot x_m = \sum_{l=1}^{L} A_l \, S_l x_{m,l} \qquad (1)$$

with **A** the column averaging kernel, a vector specified at L pressure levels, and x_m the vertical distribution of NO_2 (in partial subcolumns) from a chemistry-transport model (or from collocated validation measurements) at the same L pressure levels. The user thus needs to convert his or her vertical (subcolumn) NO_2 profile to the pressure grid of the averaging kernel in order to construct a vertical column $\hat{y}_m$ as would be retrieved by OMI. In Eq. (1), S_l are the components at the l^{th} layer of an operator that executes a mass-conserving vertical interpolation or integration followed by a conversion to sub-columns (molecules cm^{-2}) should the model vertical distribution $x_{m,l}$ not yet be given in those units (but e.g. in mixing ratio). The operator ensures that the NO_2 column after vertical interpolation has the exact same value to the column before interpolation.

Users will often be interested in the tropospheric NO_2 column. For tropospheric retrievals (with $\hat{y}_{m,trop}$ now the tropospheric column), equation (1) becomes:

$$\hat{y}_{m,trop} = \mathbf{A_{trop}} \cdot x_{m,trop} \quad (2)$$

with $\mathbf{A_{trop}}$ the column averaging kernel for tropospheric retrievals, defined as:

$$\mathbf{A_{trop}} = \mathbf{A} \cdot \frac{M}{M_{trop}} \quad (3)$$

and $x_{m,trop}$ the profile shape for tropospheric levels (levels up to level number tm5_tropopause_layer_index (l_{tp}^{TM5}) as specified in the group PRODUCT). The pressure at level tm5_tropopause_layer_index does not necessarily correspond to the tropopause pressure but rather gives the pressure of the layer in which the tropopause occurs according to the WMO 1985 tropopause criterion [WMO, 1992] (tropopause defined as the lowest level at which the lapse rate decreases to 2 K/km or less, provided the average lapse rate between this level and all higher levels within 2 km does not exceed 2 K/km).

5.5 Spatial averaging and gridding

To obtain a spatially representative average of retrieved tropospheric NO_2 columns (for instance for gridding to a regular grid), we recommend computing the weighted average of all individual retrievals within a certain grid cell over the entire area covered by all (valid) retrievals, where the weight is given by the pixel area w_i (in km^2):

$$\hat{y}_o = \frac{\sum_i w_i \hat{y}_i^o}{\sum_i w_i} \quad (4)$$

5.6 How not to use the data

When interpreting the QA4ECV NO_2 data, users should realize that the retrievals have a distinct temporal characteristic, i.e. they hold for the day and time of measurement (e.g. mid-morning for GOME-2 or early afternoon for OMI). In the troposphere, NO_2 has a relatively short lifetime (~hours), and its emissions have distinct diurnal variation [Boersma et al., 2008]. In the stratosphere, NO_2 lives longer but there is a photochemically driven diurnal cycle. When interpreting the data against models and when interpreting data records from multiple satellites, users should take this temporal representativeness into account. This can be done following the recipe of Boersma et al. [2016]. Users should also realize that satellite UV-visible sensors cannot see under the clouds and, consequently, that satellite data sets of tropospheric NO_2 are biased towards cloud-free conditions.

What should be avoided at all costs is:

a) Interpreting the total column N_v as a diagnostic of tropospheric pollution. While the total column (N_v) will have a contribution from tropospheric NO_2, the accuracy of the total column retrieval has not been optimized for tropospheric pollution, and pollution data users should rather work with $N_{v,trop}$.
b) a comparison of mean model fields (24-hr or 31-day) to averaged satellite data (representative of overpass time and relatively cloud-free conditions only)
c) interpreting cloudy tropospheric NO_2 retrievals as valid down to the Earth's surface.

Scenes with substantial cloud cover (e.g. cloud fraction>0.2) screen the NO_2 pollution below the clouds from the instrument's field of view. While these retrievals may still be useful for above-cloud NO_2 column estimates (cloud-slicing technique, e.g. Choi et al. [2014]) they should not be used for total tropospheric column evaluations, e.g. for emissions or surface NO_2 concentration estimates.

d) a direct comparison of mid-morning NO_2 columns (e.g. GOME-2) to early afternoon NO_2 columns (OMI). NO_2 varies substantially within a timeframe of 4 hours, and this time difference (and pixel location and size differences) should be taken into account when comparing data sets from two different sensors.

e) Trend analysis based on all available data irrespective of sampling conditions. When doing trend analyses, users should avoid sampling biases introduced for instance in OMI because of a time-dependent row anomaly (http://projects.knmi.nl/omi/research/calibration/instrument_status_v3/index.html). For long time series, users should ensure that satellite data is sampled in a consistent manner throughout the time period analyzed.

5.7 Processing Uncertainty Data

Single pixel uncertainty

The QA4ECV NO_2 ECV precursor product contains an ex-ante algorithm uncertainty estimate asscoaited with each individual pixel (tropospheric_no2_vertical_column_uncertainty_estimate). This uncertainty estimate is calculated theoretically via uncertainty propagation based on the principal retrieval equation, uncertainties in level-1 data and subsequent spectral fitting uncertainties, and contributions from uncertainties in a priori and ancillary data required to calculate the stratospheric NO_2 background and the AMF. For more detail see Boersma et al. [2004]. This uncertainty should be interpreted as the best guess of the retrieval uncertainty for one specific measurement. It contains random and systematic components. When averaging data over multiple pixels, or averaging data over time, parts of the uncertainty will cance out or be smoothed, but (an unknown) part of the uncertainty is systematic and will remain even after averaging.

Uncertainty of a spatial mean of retrievals

When comparing model and satellite NO_2 columns (user groups 1 and 2), and for inverse modeling we recommend using the below equation to calculate the uncertainty for spatially averaged (gridded) data (σ_o) by first calculating the area-weighted (statistical) retrieval uncertainty σ, and then accounting for a partial correlation in the errors between pixels as in Eskes et al. [2003].

$$\sigma_o = \sigma\sqrt{\frac{1-c}{n} + c} \tag{5}$$

with c the error correlation between the n retrievals. In Miyazaki et al. [2012], and Boersma et al. [2016], c=0.15 is proposed based on the consideration that errors in clouds, albedo, a priori profile, and aerosols are typically correlated in space accounting for a partial correlation in the errors between pixels as in Eskes et al. [2003]:

Uncertainty of a temporal mean of retrievals

In model-column comparisons and in trend analysis studies, e.g. it is often important to work with an uncertainty estimate for temporally averaged satellite retrievals, for instance on the monthly mean basis. The temporal variability in tropospheric NO_2 is typically strong (e.g.

diurnal cycle, day-to-day variability, weekly cycle, seasonal cycle, …), and this implies also considerable variability in the day-to-day uncertainty estimates. To obtain the uncertainty of a monthly mean tropospheric NO_2 column over a certain region, we recommend taking whichever is largest:

(A) the temporally averaged values for σ_o, or

(B) the standard deviation of the daily tropospheric NO_2 columns.

If there is strong temporal variability (e.g. day-to-day changes in photochemistry, transport events, weekend reductions in emissions) the standard deviation of the mean will likely dominate and be a good representation of the uncertainty in the monthly mean, but over regions with background NO_2 (oceans), the temporally averaged uncertainty (B) will likely be higher.

Retrieval uncertainty when using the averaging kernel

For (tropospheric) applications using the averaging kernel, the error in $\hat{y}_{m,trop}$ will reduce to tropospheric_no2_vertical_column_uncertainty_kernel ($\sigma_{N_{v,trop,k}}$) since uncertainties on the a priori vertical NO_2 profile no longer contribute. A user should be aware that he or she should then no longer use tropospheric_no2_vertical_column_uncertainty_estimate, because this uncertainty includes the profile uncertainty contribution which may now be discarded.

5.8 Product Quality Information

The output for each ground pixel will be accompanied by four flags indicating the status of the results of the processing. In order of hierarchy:

1. The processing_error_flag is included in the main PRODUCT and has value 0 (retrieval processing has succeeded) or 1 (retrieval failure).

   If the processing_error_flag = 0, there may still be warnings about the retrieval result, and these can be found in the field processing_quality_flag. Such warnings include for instance situations with a successful retrieval over the South Atlantic Anomaly, or over ocean scenes with Sun glint.

   If the processing_error_flag = 1, the processing_quality_flag can be checked for detailed information on individual event(s) that led to processing failure.

2. The snow_ice_flag indicates the presence of snow or ice at the Earth's surface in the pixel area, as derived from independent data from the NISE, and from the cloud retrieval data used as input to calculate the AMF.

How can I be sure I'm interpreting valid QA4ECV NO_2 data?

1. processing_error_flag = 0

Ensure that processing_error_flag = 0 for a pixel. If the value instead equals 1, this means that the algorithm has encountered a fatal error (irrespective of the exact algorithm step in which this occurred) and has not produced a valid retrieval for this pixel. Usually the large majority of the pixels in an orbit should have processing_error_flag = 0, but near the terminator (large solar zenith angles), values of 1 occur. Note that the type of the flag is byte. The flag is located here: PRODUCT/processing_error_flag.

2. solar_zenith_angle < 80

Even though the QA4ECV NO_2 retrieval scheme accepts NO_2 slant column densities for solar zenith angles up until 85°, the data assimilation scheme is facing difficulties close to the terminator in wintertime. Stratospheric NO_2 in polar nocturnal darkness is not well constrained in the TM5-MP model because the absence of sunlight precludes any measurements from being ingested by the assimilation procedure. Close to the terminator, the stratospheric NO_2 estimates are therefore of questionable quality, and we recommend not to use retrievals over these areas. The problem occurs mainly in the winter hemisphere and is strongest in mid- to late winter. The SZA is found here: PRODUCT/SUPPORT_DATA/GEOLOCATIONS/solar_zenith_angle.

3. snow_ice_flag < 10 or snow_ice_flag = 255

The value for snow_ice_flag=0 corresponds to "snow free land", and the value 255 to "ice free ocean", and 001-010 to a sea-ice coverage from 1% to 10%. In principle, good retrievals are possible over snow and ice-covered areas, but the quality of the cloud retrievals over such areas is not well known. Pending further research into the issue, we recommend to discard snow and ice-covered pixels using the above filter. This takes out pixels that have been classified by NISE as snow or ice covered, and also any ocean pixels with an ice cover that exceeds 10%. Note that the type of the snow_ice_flag is byte. PRODUCT/SUPPORT_DATA/INPUT_DATA/snow_ice_flag.

Applying criteria 1, 2, and 3 removes the high tropospheric column values close to the terminator. This becomes apparent from Figure 1, where after applying criterion 1 the left panel shows a small band of high tropospheric NO_2 column along 65°N. This band is effectively removed by applying criteria 1-3 as was done in the right panel of Figure 1.

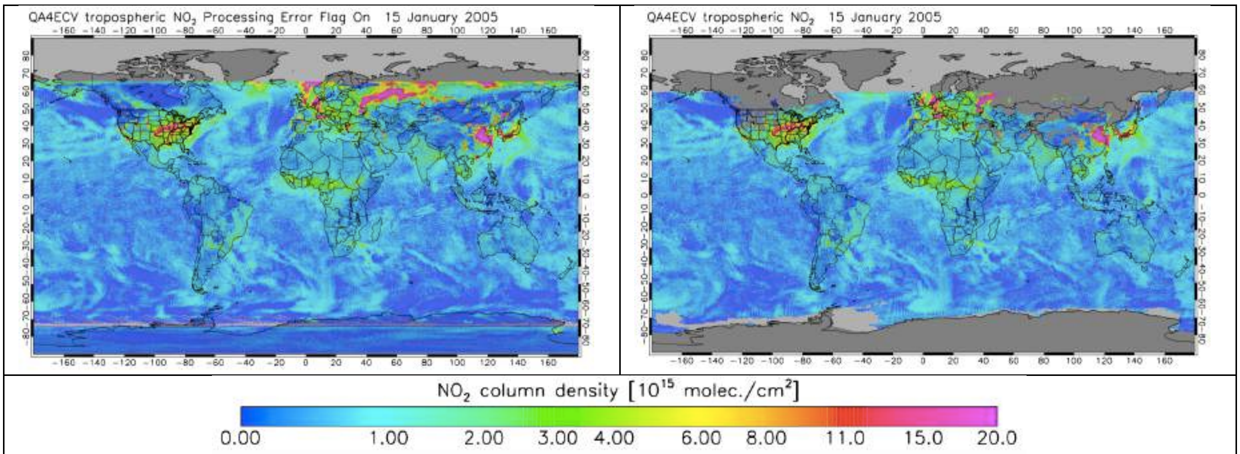

Figure 1. QA4ECV tropospheric NO_2 columns on 15 January 2005 sampled according to criterion 1 only: processing_error_flag=0 (left panel), and sampled according to criteria 1-3, corresponding to processing_error_flag=0, modest solar zenith angles (< 80°), and very little snow/ice (right panel).

4.amf_trop/amf_geo > 0.2

We recommend this setting to avoid situations in which the retrieval is based on very low (relative) tropospheric air mass factors. Such situations typically occur when the TM5 model predicts large amounts of NO_2 close to the surface in combination with aerosols or clouds effectively screening this NO_2 from detection. This selection criterion ensures that the tropospheric AMF cannot multiply the residual tropospheric NO_2 slant column in the retrieval by more than a factor of 2. The scenes rejected by criterion 4 are mostly situated over polluted areas, and their associated column values are often extreme. The right panel of Figure 2 shows that the normalized probability distribution of the rejected ensemble (38401 gridded scenes, solid line) is showing a much larger probability of negative (< -3 10^{15} molec.cm^{-2}) and strongly positive (>10 10^{15} molec.cm^{-2}) column values than the ensemble after application of criteria 1-4 (666485 scenes, dashed line). PRODUCT/amf_trop; PRODUCT/SUPPORT_DATA/DETAILED_RESults/amf_geo.

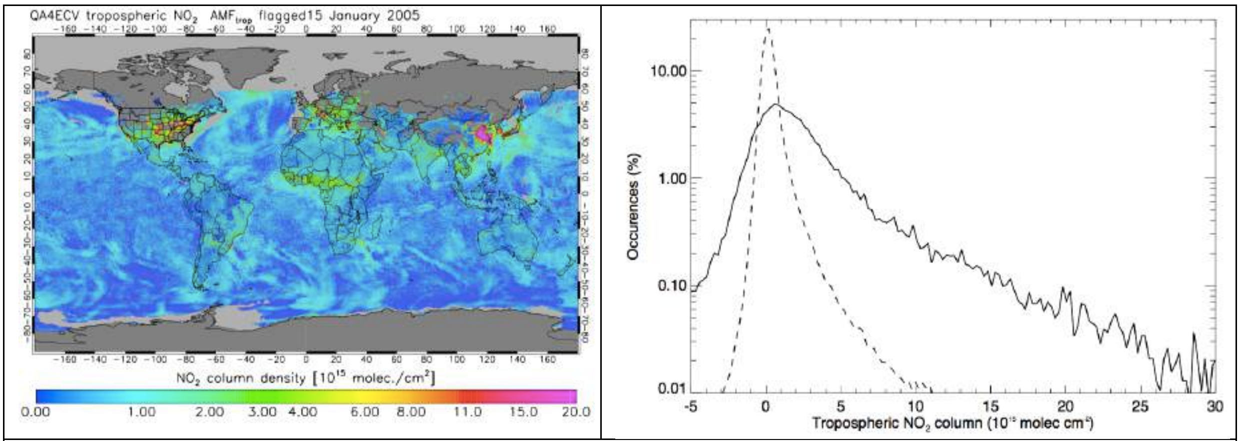

Figure 2. QA4ECV tropospheric NO_2 columns on 15 January 2005 sampled according to criteria 1-4 (left panel). Normalized probability distribution function tropospheric NO_2 columns from the rejected subset (solid line) and the subset passing criteria 1-4 (dashed line).

5. cloud_radiance_fraction_no2 ≤ 0.5

For user groups (1), (2), and (6) we recommend to filter out retrievals of tropospheric NO_2 columns done for cloud radiance fractions >0.5. For these retrievals, there is limited sensitivity to NO_2 below the cloud level, where most NO_2 pollution is situated. For retrievals with cloud radiance fraction <0.5, most of the radiance originates from the clear-sky part of the pixel, which limits the influence (and errors) of a priori assumptions made in the AMF calculation. The left panel of Figure 3 shows the global distribution of NO_2 columns after filtering for cloud radiance fractions >0.5. Regions with large-scale cloudiness are obviously masked out after applying the cloud radiance fraction criterion. In the right panel we see that the probability distribution function is shifted to less negative and somewhat higher values when applying criteria 1-5 compared to applying criteria 1-4. PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/cloud_radiance_fraction_no2.

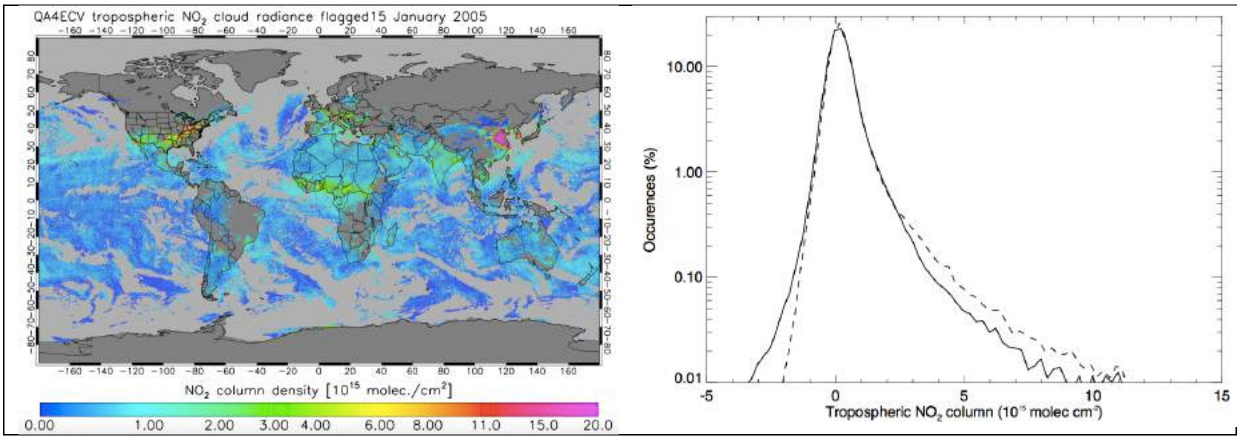

Figure 3. Tropospheric NO_2 columns on 15 January 2005 sampled according to criteria 1-5 (left panel). Normalized probability distribution function of tropospheric NO_2 columns from the rejected subset (solid line) and the subset passing criteria 1-5 (dashed line).

How can I filter more strictly or learn more about possibly suspect retrievals?

It is possible for advanced users to explore the flags in QA4ECV NO_2 product to learn more about the reasons for retrieval failure, or for stricter filtering than recommended here.

Table 1 below provides an overview of the processing_quality_flag values (error, filter) that apply to the QA4ECV NO_2 data product. Note that the flag is filled in a bit-wise manner, with the last two bits the errors and filters, and the other bits for warnings. Note that if the processing_quality_flags shows that a failure or filter occurred (i.e. a non-zero value set in bits 0-7), the processing_error_flag is set to 1 (one).

Table 1: Possible values of the processing_quality_flag and their meaning

value	***short name***	***description***
0	success	No failures, output contains value. Warnings still possible.
Errors		
1	radiance_missing	The number of spectral pixels in the radiance due to flagging is too small to perform the fitting.
2	irradiance_missing	The number of spectral pixels in the irradiance due to flagging is too small to perform the fitting.
3	input_spectrum_missing	The reflectance spectrum does not contain enough points to perform the retrieval. This is different from (ir)radiance_missing in that the missing points may not be aligned.
7	sza_range_error	Solar zenith angle out of range, maximum value from configuration.
9	lut_range_error	Extrapolation in lookup table (AMF, cloud radiances).
11	wavelength_offset_error	Wavelength offset exceeds maximum from configuration.
12	initialization_error	Error occurred during the processing of the pixel, no output generated. The following errors raise this flag: Mismatch between irradiance and radiance wavelengths; on-ground distance between band 1 and band 2 ground pixels exceeds threshold set in the configuration; derived a-priori information not valid, processing impossible.
19	convergence_error	The main algorithm did not converge.
24	geolocation_error	Geolocation out of range
36	cloud_error	No cloud data.
41	generic_range_error	Generic range error.
42	generic_exception	Catch all generic error.
43	input_spectrum_alignment_error	Input Earth radiance and solar irradiance spectra not aligned correctly.
44	abort_error	Not processed because processor aborted prematurely (time out or user abort).
45	wrong_input_type_error	Wrong input type error, mismatch between expectation and received data.
46	wavelength_calibration_error	An error in the wavelength calibration of this pixel.
Filters		
64	solar_eclipse_filter	Solar eclipse.

95	time_range_filter	Time is out of the range that is to be processed.
96	pixel_or_scanline_index_filter	Not processed because pixel index does not match general selection criteria.
97	geographic_region_filter	Pixel falls outside the specified regions of interest.

Criterion 3 above rejects retrievals taken over snow or ice. It is worthwhile to point out that users may want to investigate the quality of retrievals over snow and ice in more detail. To do so, they would have to change (relax) criterion 3. In the criterion, the snow_ice_flag threshold is currently set at 10, which implies that pixels with more than 10% sea ice, and all snow-covered scenes are excluded.

However, if more than 50% of the pixel is covered by snow or by sea ice, the cloud algorithm is switched to 'scene mode'. For these bright scenes, corresponding to snow_ice_flag values between 51-150, the NO_2 retrieval then is switched to using the effective scene albedo and scene pressure rather than the cloud fraction and cloud pressure retrieved from the cloud algorithm. This means that the NO_2 AMFs are calculated differently, i.e. for a clear-sky situation with the scene albedo, and scene pressure as inputs. It is well possible that these retrievals are useful, but since they have not been evaluated or validated, we recommend rejection for now, but encourage further investigations into the issue.

Table 2 below shows the possible values for the snow_ice_flag, whicgh follows the NISE database, given in the SnowIceExtent dataset (downloaded and used by the OMCLDO2 algorithm [Veefkind et al., 2016]).

Table 2: Possible values of the processing_quality_flag and their meaning

snow_ice_flag		***processing_quality_flags***
value	***meaning***	
0	snow-free land	0
001-100	sea-ice concentration (percent)	0
101	permanent ice	0
103	snow (dry or wet)	0
252	mixed pixels at coastline	36
253	suspect ice value	36
254	corners (undefined) == FillValue	36
255	ocean	0
--	error	36

6 Contact Information

For questions about the QA4ECV NO_2 ECV precursor product, you can get in touch with the product developers. For questions on the retrieval of the stratospheric correction please contact Henk Eskes and Steffen Beirle. For questions related to the AMF please contact Folkert Boersma. For questions on the retrieval of slant columns, please consult Andreas Richter (GOME-2, SCIAMACHY, GOME), Isabelle De Smedt (GOME-2, OMI), or Jos van Geffen (OMI). Please also check out the QA4ECV User Forum on www.qa4ecv.eu/forum/267.

For direct questions, you can send an email to:

Henk Eskes: eskes@knmi.nl

Folkert Boersma: boersma@knmi.nl

Jos van Geffen geffen@knmi.nl

Isabelle De Smedt isabelle.desmedt@aeronomie.be

Andreas Richter richter@iup.physik.uni-bremen.de

Steffen Beirle steffen.beirle@mpic.de

http://www.qa4ecv.eu/

7 References

Beirle, S., Spichtinger, N., Stohl, A., Cummins, K. L., Turner, T., Boccippio, D., Cooper, O. R., Wenig, M., Grzegorski, M., Platt, U., and Wagner, T. (2006), Estimating the NO_x produced by lightning from GOME and NLDN data: a case study in the Gulf of Mexico, Atmos. Chem. Phys., 6, 1075-1089, doi:10.5194/acp-6-1075-2006.

Belmonte Rivas, M., Veefkind, P., Eskes, H., and Levelt, P. (2015), OMI tropospheric NO_2 profiles from cloud slicing: constraints on surface emissions, convective transport and lightning NO*x*, Atmos. Chem. Phys., 15, 13519-13553, doi:10.5194/acp-15-13519-2015.

Boersma, K. F., H. J. Eskes, and E. J. Brinksma (2004), Error analysis for tropospheric NO_2 retrieval from space, J. Geophys. Res., 109, D04311, doi:10.1029/2003JD003962.

Boersma, K. F., Eskes, H. J., Meijer, E. W., and Kelder, H. M. (2005), Estimates of lightning NO_x production from GOME satellite observations, Atmos. Chem. Phys., 5, 2311-2331, doi:10.5194/acp-5-2311-2005.

Boersma, K. F., D. J. Jacob, H. J. Eskes, R. W. Pinder, J. Wang, and R. J. van der A (2008), Intercomparison of SCIAMACHY and OMI tropospheric NO_2 columns: Observing the diurnal evolution of chemistry and emissions from space, J. Geophys. Res., 113, D16S26, doi:10.1029/2007JD008816.

Boersma, K. F., Vinken, G. C. M., & Eskes, H. J. (2016). Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV–Vis tropospheric column retrievals. Geoscientific Model Development, 9(2), 875-898.

Choi, S., Joiner, J., Choi, Y., Duncan, B. N., Vasilkov, A., Krotkov, N., and Bucsela, E. (2014), First estimates of global free-tropospheric NO_2 abundances derived using a cloud-slicing technique applied to satellite observations from the Aura Ozone Monitoring Instrument (OMI), Atmos. Chem. Phys., 14, 10565-10588, doi:10.5194/acp-14-10565-2014.

Eskes, H. J., Van Velthoven, P. F. J., Valks, P. J. M., & Kelder, H. M. (2003). Assimilation of GOME total-ozone satellite observations in a three-dimensional tracer-transport model. Quarterly Journal of the Royal Meteorological Society, 129(590), 1663-1682.

Eskes, H. J. and Boersma, K. F. (2003). Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3, 1285-1291, doi:10.5194/acp-3-1285-2003.

Lorente, A., Boersma, K. F., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J. T., Kroktov, N., Stammes, P., Wang, P., Eskes, H. J., and Krol, M. (2016). Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-306, in review.

QA4ECV (2016a), Recommendations (scientific) on best practices for Land and Atmosphere ECVs, QA4ECV Report / Deliverable no D4.2, K. F. Boersma, A. Lorente, and J.-P. Muller (Eds.), retrieved from QA4ECV website: http://www.qa4ecv.eu/sites/default/files/D4.2_v1_final_1736.pdf.

QA4ECV (2016b), Harmonized, sensor-independent retrievals (software) for HCHO and NO2. Report on MOPITT-IASI total column consistency, QA4ECV Report / Deliverable no D4.4, A. Lorente (Ed), retrieved from QA4ECV website: restricted access (25-01-2017).

Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F. (2016). Improvements to the OMI O2–O2 operational cloud algorithm and comparisons with ground-based radar–lidar observations, Atmos. Meas. Tech., 9, 6035-6049, doi:10.5194/amt-9-6035-2016.

Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W. (2016). The high-resolution version of TM5-MP for optimised satellite retrievals: Description and Validation, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-125, in review.

WMO (1992) International meteorological vocabulary (2nd edn). Secretariat of the World Meteorological Organization, Geneva, p 636. ISBN 92-63-02182-1.