QA4ECV

Quality Assurance for Essential Climate Variables

Project Number 607405

Deliverable:	Deliverable 4.6 – Product Specification Document for HCHO
Responsible Partner:	BIRA-IASB
Delivery date:	10-02-2017

HCHO Product

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

Product User Guide for HCHO

(Version 1.0)

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

1: BIRA-IASB, 2: KNMI, 3: MPI-C, 4: IUP-B, 5: WUR, 6: NPL

Change Record

Issue	Date	Pages Changed	Comments
0.4	16-09-2016		First draft, including description of contents of the netCDF file
0.5	03-01-2017		Second draft including introduction, error and quality flags, uncertainty descriptions and homogenisation with NO2 PSD.
0.6	01-02-2017		conversion to QA4ECV template
0.64	03-02-2017		Comments included from internal review for NO2 PSD
0.7	09-02-2017		Comments included after internal review for HCHO PSD
1.0	10-02-2017		Version for the website

CONTENTS

TERMS & ACRONYMS

AMF	Air Mass Factor
BIRA	Belgian Institute for Space Aeronomy
DOAS	Differential Optical Absorption Spectroscopy
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
MPIC	Max Planck Institute for Chemistry Mainz
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
RAA	Relative Azimuth Angle
SCD	Slant Column Density
SCIAMACHY	Scanning and Imaging Spectrometer for Atmospheric Chartography
SZA	Solar Zenith Angle
TM5(-MP)	Tracer Model version 5 (Massive Parallel)
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

1. Executive Summary

This is the product user manual for the QA4ECV HCHO data product, version 1.0. The QA4ECV HCHO ECV precursor product monitors the tropospheric HCHO column densities globally for the period 1996-2015. The main product is the tropospheric HCHO tropospheric column density, defined as the vertically integrated number of HCHO molecules between the Earth's surface and the tropopause per unit area of the satellite pixel. The QA4ECV HCHO product is, or will be, available for the sensors OMI, GOME-2(A), SCIAMACHY and GOME.

A consortium consisting of BIRA, IUP Bremen, KNMI, MPI-C, and Wageningen University has set together a community algorithm to produce an 11+ years (October 2004 – December 2015) set of OMI HCHO data based on improved HCHO slant columns, background 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 www.qa4ecv.eu/data/hcho. For details on the QA4ECV HCHO retrieval algorithm, please see the reports provided as deliverable D4.2 [QA4ECV, 2016a] and D4.4 [QA4ECV, 2016b]. Quality assessment of the QA4ECV HCHO algorithm will be available in D5.5, while quality assessment of the product (including validation with reference data) will be available in D5.6.

Section 2 introduces the user to the QA4ECV HCHO ECV precursor product, explaining the main product, applications and its heritage. In section 3, 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.

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 the file name convention, variable names, metadata information, and the data format. Section 4 explains how to obtain the QA4ECV HCHO data.

The document is also 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. This section explains what users should do to visualize and interpret the data, how to use the averaging kernels associated with the retrievals, how to take uncertainties into account, and how to use the quality information in the data files so that the most significant retrievals can be selected for analysis.

2. Introduction

2.1 QA4ECV HCHO ECV precursor product

The QA4ECV HCHO ECV precursor product monitors the tropospheric HCHO column densities globally 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 km2 (GOME) to 24×13 km2 (OMI), and the overpass times are early afternoon for the OMI sensor (~13:40 hrs) and mid-morning for the other instruments.

Daily and monthly averaged HCHO tropospheric vertical columns are also available at the horizontal resolution of 0.25°, in text format. To calculate these averages, satellite observations have been selected based on processing quality flag (see section 5.4).

2.2 Research and applications

Long term satellite observations of tropospheric formaldehyde (HCHO) are essential to support air quality and chemistry-climate related studies from the regional to the global scale. Formaldehyde is an intermediate gas in almost all oxidation chains of non-methane volatile organic compounds (NMVOC), leading eventually to CO_2. NMVOCs are, together with NOx, CO and CH_4, among the most important precursors of tropospheric ozone. NMVOCs also produce secondary organic aerosols and influence the concentrations of OH, the main tropospheric oxidant. The major HCHO source in the remote atmosphere is CH_4 oxidation. Over the continents, the oxidation of higher NMVOCs emitted from vegetation, fires, traffic and industrial sources results in important and localised enhancements of the HCHO levels. Its lifetime being of the order of a few hours, HCHO concentrations in the boundary layer can be directly related to the release of short-lived hydrocarbons, which mostly cannot be observed directly from space. Furthermore, HCHO observations provide information on the chemical oxidation processes in the atmosphere, including CO chemical production from CH_4 and NMVOCs. The seasonal and inter-annual variations of the formaldehyde distribution are principally related to temperature changes and fire events, but also to changes in anthropogenic activities. For these reasons, HCHO satellite observations are used in combination with tropospheric chemistry transport models to constrain NMVOC emission inventories in so-called top-down inversion approaches.

2.3 Data set and evolution

Within the QA4ECV project, a consortium of European groups (BIRA-IASB, IUP Bremen, MPIC, KNMI and WUR), has carefully compared various 'state-of-science' approaches for the main retrieval steps. This exercise has led to recommendations and best practices for the QA4ECV HCHO 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 HCHO algorithm. The comparisons and recommendations for algorithmic and software choices are summarized in QA4ECV Deliverable 4.2 [QA4ECV, 2016a]. A short overview of the set-up of the retrieval software for HCHO, and algorithm choices is given in QA4ECV Deliverable 4.4 [QA4ECV, 2016b].

This version of the PSD relates to version 1 of the QA4ECV HCHO product.

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 HCHO ECV precursor 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.

2.5 Definitions

The general method used for the derivation of HCHO VCDs from UV spectral measurements is the Differential Optical Absorption Spectroscopy method (DOAS) which involves two main steps. First, the effective slant column amount (corresponding to the integrated HCHO concentration along the mean 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. Subsequently, a correction is applied to the slant column values to correct for appearing biases that may be of known or unknown origin. Finally, slant columns are converted into vertical columns by means of air mass factors (M) obtained from suitable radiative transfer calculations, accounting for the presence of clouds, surface properties, and best-guess HCHO vertical profiles.

The main outputs of the algorithm are the slant column density (N_s), the vertical column (N_v), the air mass factor (M), and the values used for the reference sector correction ($N_{s,0}$ and $N_{v,0}$). Complementary product information includes clear sky air mass factor, error on the total column (detailed by N_s, M, $N_{v,0}$ errors), averaging kernel, and quality flags.

$$N_v = \frac{N_s - N_{s,0}}{M} + N_{v,0} = \frac{\Delta N_s}{M} + N_{v,0} \qquad 2\text{-}1$$

$$N_{v,0} = \frac{M_0}{M} N_{v,0,CTM} \qquad 2\text{-}2$$

Table 2-1: Main measured quantity

Quantity [unit]	Field name	Symbol
HCHO tropospheric vertical column [molec./cm2]	PRODUCT/tropospheric_hcho_vertical_column	N_v
HCHO tropospheric amf [n.u.]	PRODUCT/amf_trop	M
HCHO slant column [molec./cm2]	PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/ scd_hcho	N_s
HCHO slant column correction [molec./cm2]	PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/ scd_hcho_correction	$N_{s,0}$
HCHO vertical column correction [molec./cm2]	PRODUCT/SUPPORT_DATA/DETAILED_RESULTS/ vcd_hcho_correction	$N_{v,0}$

3. Available Products

The main product in the QA4ECV HCHO data record is the tropospheric HCHO column density, which is defined as the vertically integrated number of HCHO molecules between the Earth's surface and the tropopause per unit area of the satellite pixel. The unit of all HCHO column density products is molecules cm^{-2}.

The output level-2 file of QA4ECV is one netCDF file for each orbit. The netCDF layout for QA4ECV is based on the TROPOMI format, though with a few changes to adapt the format to the QA4ECV needs.

3.1 Output file name convention

The name of the QA4ECV 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 largest 328.5-359 nm fit window, "fitB" for the 328.5-346 nm fit window, and "fitC" for the 328.5-346 nm fit window, with BrO pre-fitted in fit A. For example:

- QA4ECV_L2_HCHO_OMI_20050202T151200_o02947_fitA_v1.nc
- QA4ECV_L2_HCHO_GOME2A_20130202T174159_o32648_fitC_v1.nc

The product version and fit window definition are also given in the global attributes.

3.2 Intermediate Products

The overall structure of the cdl file is as follows:

- global attributes
- group: METADATA
- group: PRODUCT
    - dimensions
    - variables
    - group: SUPPORT_DATA
        - group: GEOLOCATIONS
            - variables
        - group: DETAILED_RESULTS
            - variables
        - group: INPUT_DATA
            - variables

For the groups in this list, an overview of the contents is given below.

3.3.1 Global attributes

- Conventions = CF-1.7
- background_correction_processor_name = QA4ECV_HCHO_BC
- background_correction_processor_version = 1.2

- cdm_data_type = SWATH
- date_created = 2016-10-05T13:54:55Z
- equator_crossing_longitude = 166.36728
- equator_crossing_time = 2006-01-01T02:23:29.997000Z
- geospatial_lat_max = 90.0
- geospatial_lat_min = -90.0
- geospatial_lon_max = -180.0
- geospatial_lon_min = 180.0
- id = QA4ECV_L2_HCHO_OMI_20060101T0127_o07788_fitA_v1.nc
- institution = BIRA-IASB & IUPB & KNMI & MPIC & WUR
- irradiance_file = OMI-Aura_L1-GLOBAL-OMTMIRRYA_2005m0101t0000-syear-rPDS01_v003-2007m0716t145802.he4
- keywords = 0315 Biosphere/atmosphere interactions (0426, 1610); 0345 Pollution: Urban and Regional; 0365 Troposphere: Composition and Chemistry; 0368 Troposphere: Constituent Transport and Chemistry; 3360 Remote Sensing; 1631 Land/atmosphere interactions (1218, 1843, 3322); 1632 Land cover change
- keywords_vocabulary = AGU index terms, http://publications.agu.org/author-resource-center/index-terms/
- level1b_file = OMI-Aura_L1-OML1BRUG_2005m0101t1331-o02480_v003-2011m0119t174146-p1.he4
- naming_authority = be.aeronomy
- orbit = 7788
- processing_status = Offline QA4ECV processing
- product_version = 1
- project = QA4ECV
- reference = http://www.qa4ecv.eu/
- slant_column_fit_window = [328.5-359] nm
- slant_column_processor_name = QDOAS
- slant_column_processor_version = 2.112.1 - 7 June 2016
- source = OMI / EOS-Aura
- standard_name_vocabulary = NetCDF Climate and Forecast Metadata Conventions Standard Name Table (v29, 08 July 2015), http://cfconventions.org/standard-names.html
- time_coverage_end = 2006-01-01T02:44:54Z
- time_coverage_start = 2006-01-01T01:50:07Z
- time_reference = 2006-01-01T00:00:00Z
- time_reference_days_since_1950 = 20454
- time_reference_julian_day = 2453372.0
- time_reference_seconds_since_1970 = 1104537622
- title = QA4ECV formaldehyde (HCHO) column data
- tracking_id = 21f75eca-0876-413a-8e8e-e97b7023d756
- vertical_column_processor_name = qa4ecv_hcho_omi_amf_hcho
- vertical_column_processor_version = 2.5

3.3.2 Group: METADATA

METADATA/ALGORITHM_SETTINGS/SLANT_COLUMN_RETRIEVAL

- analysis_method = optical density fitting
- convergence_criterion = 0.001
- fit_polynomial_degree = 5 (6 coefficients)
- fit_slit_function = yes
- fit_window = [328.50, 346.00] nm
- instrument_slit_function = OMI_SF_offical_uv2_row30.slf
- 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 = 325.00, 360.00 nm
- least_squares_fit_weighting = no
- linfit_dampening =
- linfit_preshift =
- lv1_calibration_options =
- lv1_extractor_version =
- maximum_number_of_iterations = 3
- processing_algorithm = 2.112.1 - 7 June 2016
- radiance_calibration_shift = true
- radiance_calibration_stretch = 1st order
- reference_spectrum_bro = bro_Fleischmann(2004)_223K_300-385nm.xs
- reference_spectrum_ch2o = ch2o_MellerMoortgat(2000)_298K_224.56-376.00nm(0.01nm).xs
- reference_spectrum_no2 = no2_VANDAELE_1998_220K.xs
- reference_spectrum_o3223 = O3223_Serdyuchenko(2014)_223K_213-1100nm(2013 version).xs, I0 correcitonSC=1e20molec cm-2
- reference_spectrum_o3243 = O3243_Serdyuchenko(2014)_243K_213-1100nm(2013 version).xs, I0 correcitonSC=1e20molec cm-2
- reference_spectrum_o3lambda = o3lambda_serdyunchenko_223.xs
- reference_spectrum_o3squared = o3squared_serdyunchenko_223.xs
- reference_spectrum_o4 = o4_thalman_volkamer_293K.xs
- reference_spectrum_ring = ring_sao2010_hr_norm.xs
- reflectance_reference_spectrum = daily mean earthshine spectra selected per row in the reference sector
- solar_reference = sao2010_solref_vac.dat
- spike_tolerance_factor = 5.0
- sza_limits = 0.0, 80.0 degrees
- traceability_chain = http://www.qa4ecv.eu/ecv/hcho/main/doas
- undersampling_correction = no

METADATA/ALGORITHM_SETTINGS/ BACKGROUND_CORRECTION

- data_selection = CF < 0.5, RMS < 3*mean(RMS) , SZA < 80
- fit_model = polynomial through latitude bins
- hcho_vertical_column_in_the_reference_sector = TM5-MP model
- latitude_bins_for_zonal_correction = 5 degrees
- reference_sector_for_destriping_and_global_offset_correction = [-5 5] lat, [180 240] long
- reference_sector_for_zonal_correction = [-90 90] lat, [180 240] long
- time_step = day
- traceability_chain = http://www.qa4ecv.eu/ecv/hcho/main/background

METADATA/ALGORITHM_SETTINGS/VERTICAL_COLUMN_RETRIEVAL

- amf_a_priori_profile = TM5-MP daily profiles, 1degree x 1degree
- amf_a_priori_profile_uncertainty_profile_height = 100hPa
- amf_cloud_fraction_uncertainty = 0.05
- amf_cloud_pressure_uncertainty = 50hPa
- amf_cloud_product = OMI O2-O2 cloud
- amf_interpolation_method = linear
- amf_surface_albedo = OMI minimum LER 342nm (2005-2010)
- amf_surface_albedo_uncertainty = 0.02
- amf_temperature_correction = no
- amf_vcd_hcho_correction.uncertainty = latitude dependent standard deviation of TM5 - IMAGES over Pacific region
- lut_a_priori_profile.uncertainty = 0.99, 0.9, 0.8, 0.6, 0.4, 0.1, 0.01
- lut_cloud_fraction.uncertainty = 0.0, 0.1, 0.2, 0.4, 0.7, 1.0
- lut_cloud_pressure.uncertainty = 1.0, 0.95, 0.9, 0.85, 0.8, 0.7, 0.6, 0.4, 0.2, 0.0
- lut_ozone_cross_section = Serdyuchenko, 243K, 2013
- lut_profile = US standard atmosphere profile, 80 layers
- lut_relative_azimuth_angle = 0.0, 45.0, 90.0, 135.0, 180.0
- lut_rtm = VLIDORT 2.7
- lut_rtm_polarization = on
- lut_solar_zenith_angle = 1.0, 10.0, 20.0, 30.0, 40.0, 45.0, 50.0, 55.0, 60.0, 65.0, 70.0, 72.0, 74.0, 76.0, 78.0, 80.0, 85.0
- lut_solar_zenith_angle.uncertainty = 1.0, 30.0, 60.0, 70.0, 80.0, 85.0
- lut_sphericity = on
- lut_surface_albedo = 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.6, 0.8, 1.0
- lut_surface_albedo.uncertainty = 0.0, 0.05, 0.1, 0.4, 0.8, 1.0
- lut_surface_pressure = 1063.1, 1037.9, 1013.3, 989.3, 965.8, 920.6, 877.0, 835.0, 795.0, 701.2, 616.6, 540.5, 411.1, 308.0, 227.0, 165.8, 121.1
- lut_surface_pressure.uncertainty = 1063.1, 989.3, 877.0, 701.2, 411.1
- lut_viewing_zenith_angle = 1.0, 10.0, 20.0, 30.0, 40.0, 50.0, 60.0, 65.0, 70.0, 75.0
- lut_viewing_zenith_angle.uncertainty = 0.0, 30.0, 60.0, 70.0, 75.0

- lut_wavelength = 341.0
- traceability_chain = http://www.qa4ecv.eu/ecv/hcho/main/amf

METADATA/ GRANULE_DESCRIPTION

- GranuleEnd = 2006-01-01T02:44:54Z
- GranuleStart = 2006-01-01T01:50:07Z
- InstrumentName = OMI
- LongitudeOfDaysideNadirEquatorCrossing = 166.36728
- MissionName = EOS-Aura
- ProcessingMode = Reprocessing
- ProductName = QA4ECV formaldehyde (HCHO) column data
- ProductShortName = QA4ECV_HCHO_OMI
- ProductVersion = 1

3.3.3 Group: PRODUCT

Dimensions

These set the dimensions of the variables in the whole product; they are known from the current group downwards. Any variable associated with an individual ground pixel (for example the vertical HCHO 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).
bc_lat_bin = 36	Number of latitude bins to calculate de background correction
layer = 34	Layers in the TM5 model: 0,1,…,layer-1 (starting at low pressure going down).
nv = 2	sets the indices of the upper and lower bounds of the TM5 layers, i.e. 0 and 1

Variables

Field name	**Quantity**	**[unit]**	**Symbol**
latitude	pixel center latitude	degree	
longitude	pixel center longitude	degree	
delta_time	offset from reference start time of	mseconds	

	measurement		
time_utc	Time of observation as ISO 8601 date-time string	string	
processing_error_flag	0 for successful processing, 1 in case the processing failed somewhere. See also the "processing_quality_flags" dataset in PRODUCT/ SUPPORT_DATA/ DETAILED_RESULTS.	1	
tropospheric_hcho_vertical_column	formaldehyde vertical column	molecules cm-2	N_v
tropospheric_hcho_vertical_column_ uncertainty_random	random uncertainty on formaldehyde vertical column	molecules cm-2	$\sigma_{N_V}^{rand}$
tropospheric_hcho_vertical_column_ uncertainty_systematic	systematic uncertainty on formaldehyde vertical column	molecules cm-2	$\sigma_{N_V}^{syst}$
averaging_kernel	averaging kernel	1	A
amf_trop	tropospheric air mass factor	1	M
tm5_pressure_level_a	TM5 hybrid A coefficient at interface levels	Pa	
tm5_pressure_level_b	TM5 hybrid B coefficient at interface levels	1	
tm5_surface_pressure	Surface pressure of the ground pixel in PRODUCT/SUPPORT_DATA/INPUT_DATA is used in the cloud retrieval. This pressure is adjusted based on the TM5 surface pressure to take meteorological conditions into account. It is this tm5_surface_pressure that is needed by users of the averaging kernel.		

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-1	1	
solar_zenith_angle	Solar zenith angle at the ground pixel location on the reference ellipsoid. Angle is measured away from the vertical.	degree	
viewing_zenith_angle	Viewing zenith angle of the satellite at the ground pixel location on the reference ellipsoid. Angle is measured away from the vertical.	degree	
relative_azimuth_angle	The relative azimuth angle (raa) is computed from the	degree	

	solar azimuth angle (saa) and viewing azimuth angle (vaa).		
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.	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	

GROUP: PRODUCT/SUPPORT_DATA/DETAILED_RESULTS

Field name	**Quantity**	**[unit]**	**Symbol**
intensity_offset_a	fit coefficient A of the intensity offset	1	
intensity_offset_a_precision	precision of fit coefficient A of the intensity offset	1	
intensity_offset_b		1	
intensity_offset_b_precision		1	
irradiance_calibration_offset	The calibrated irradiance wavelength is: lambda_cal = lambda + offset + (lambda - lambda_center) * stretch	1	
irradiance_calibration_stretch		1	
irradiance_calibration_wavelength		nm	
number_of_spectral_points_in_retrieval	number of spectral points used in the retrieval	1	
polynomial_coefficients	polynomial coefficients of the DOAS fit	1	
polynomial_coefficients_precision	precision of the polynomial coefficients of the DOAS fit	1	
radiance_calibration_offset	The calibrated radiance wavelength is: lambda_cal = lambda + offset + (lambda - lambda_center) * stretch	1	
radiance_calibration_offset_precision		1	
radiance_calibration_stretch		1	
radiance_calibration_stretch_precision		1	
radiance_calibration_wavelength		nm	
ring_coefficient		1	
ring_coefficient_precision		1	
rms_fit		1	
scd_bro		molec.cm-2	
scd_bro_precision		molec.cm-2	
scd_no2		molec.cm-2	
scd_no2_precision		molec.cm-2	
scd_o3_223		molec.cm-2	
scd_o3_223_precision		molec.cm-2	
scd_o3_243		molec.cm-2	

scd_o3_243_precision		molec.cm-2	
scd_o3_lambda		molec.cm-2	
scd_o3_lambda_precision		molec.cm-2	
scd_o3_squared		molec.cm-2	
scd_o3_squared_precision		molec.cm-2	
scd_o4		molec.cm-2	
scd_o4_precision		molec.cm-2	
scd_resol		molec.cm-2	
scd_resol_precision		molec.cm-2	
scd_hcho	hcho slant column	molec.cm-2	N_s
scd_hcho_uncertainty_random	precision of hcho column	molec.cm-2	$\sigma_{N_s}^{rand}$
scd_hcho_uncertainty_systematic	systematic uncertainty of hcho slant columns	molec.cm-2	$\sigma_{N_s}^{syst}$
scd_hcho_corrected	background corrected hcho slant column	molec.cm-2	ΔN_s
scd_hcho_correction	correction for hcho slant column	molec.cm-2	$N_{s,0}$
vcd_hcho_correction	correction for hcho vertical column background	molec.cm-2	$N_{v,0}$
vcd_hcho_correction_uncertainty	accuracy of background correction for hcho vertical column	molec.cm-2	$\sigma_{N_{v,0}}$
tm5_vcd_hcho_background	tm5 hcho background column used for the background correction of the hcho tropospheric columns	molecules cm-2	$N_{v,0,CTM}$
amf_clear	clear-sky air mass factor	1	
averaging_kernel_clear	clear-sky averaging kernel	1	
cloud_radiance_fraction_hcho	cloud radiance fraction at 341 nm for HCHO retrieval	1	
amf_uncertainty	amf_uncertainty	1	σ_M
amf_albedo_uncertainty	AMF uncertainty due to error on albedo	1	
amf_cloud_fraction_uncertainty	amf_cloud_fraction_uncertainty	1	
amf_cloud_pressure_uncertainty	AMF uncertainty due to error on cloud pressure	1	
amf_a_priori_profile_uncertainty	AMF uncertainty due to error on a priori profile	1	
profile_height	profile height of a priori profile	hPa	
processing_quality_flags	Used to indicate errors, filters and warnings, both from the slant column retrieval and the subsequent algorithm steps, where 0 (zero) means success. An overview of the TROPOMI processing_quality_flags that are relevant for HCHO processing is given at the end of this document.	1	
tropospheric_hcho_vertical_column_uncertainty_systematic_scdes	systematic uncertainty on formaldehyde vertical column due to slant column systematic error	molec.cm-2	
tropospheric_hcho_vertical_column_uncertainty_systematic_amfes	systematic uncertainty on formaldehyde vertical column due	molec.cm-2	

	to air mass factor systematic error		
tropospheric_hcho_vertical_column_uncertainty_systematic_vcd0es	systematic uncertainty on formaldehyde vertical column due to background correction systematic error	molec.cm-2	

GROUP: PRODUCT/SUPPORT_DATA/INPUT_DATA

Field name	**Quantity**	**[unit]**	**Symbol**
surface_albedo_hcho	Surface albedo in the HCHO fit window (341 nm). Reference: OMI : Kleipool et al.,2008, http://onlinelibrary.wiley.com/doi/10.1029/2008JD010290/full	1	
surface_altitude	Data from OMCLDO2 and FRESCO. Reference: OMI O2-O2 data: Veefkind et al., 2016; http://www.atmos-meas-tech-discuss.net/amt-2016-48/.	m	
surface_pressure	surface pressure from the cloud product	hPa	
surface_albedo	Surface albedo used in the cloud product (475 nm for OMI)	1	
cloud_fraction	effective cloud fraction from the cloud product OMI O2-O2 data: Veefkind et al., 2016; http://www.atmos-meas-tech-discuss.net/amt-2016-48/.	1	
cloud_pressure	cloud optical centroid pressure from the cloud product	hPa	
scene_pressure	scene pressure from the cloud product	hPa	
scene_albedo	scene albedo in the cloud product (475 nm for OMI)	1	
snow_ice_flag	snow-ice mask, OMI specific variable, Reference: From OMI level 1 files: http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/documents/v003/ RD01_SD467_IODS_Vol_2_issue8.pdf		
surface_classification	Land-water mask from USGS (http://edc2.usgs.gov/glcc/globdoc2_0.php) and NASA SDP toolkit (http://newsroom.gsfc.nasa.gov/sdptoolkit/toolkit.html)	1	
omi_xtrack_flags	OMI specific variable, indicating issues with the so-called row anomaly. From OMI level 1 files	1	
hcho_profile_apriori	TM5 profile in volume mixing ratio for the calculation of AMF.	1	

	Reference: The high-resolution version of TM5-MP for optimised satellite retrievals, Williams et al., 2016, http://www.geosci-model-dev-discuss.net/gmd-2016-125/		

4. Obtaining HCHO

Visit www.qa4ecv.eu. Then:

- click ECV data in the header menu.
- click on the link HCHO in the menu Essential Climate Variables.
- Click on the blue button ‘Data Access’

Alternatively you can go directly to www.qa4ecv.eu/data/hcho. You can directly download individual HCHO Data Files (days). These files contain all the (final and intermediate) products listed in section 2, including a priori profiles.

To obtain 1 year of HCHO data from OMI, use the wget command. TBD

5. Using HCHO

5.1 Processing product data

The tropospheric HCHO column is the principal QA4ECV HCHO precursor product. For this product, we can distinguish different user groups: from users that take our product ‘face value’ (group 3) 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:

User group	User application	Data sets needed
1	Tropospheric chemistry / air quality model evaluation Validation with tropospheric HCHO profile measurements (aircraft, MAX-DOAS) Trend analysis	N_v, σ_{N_v} M, A A_l^{TM5}, B_l^{TM5}, l_{tp}^{TM5}, p_s
2	Tropospheric column comparisons, e.g. with other HCHO column retrievals.	N_v, σ_{N_v}
3	Visualization of the HCHO product, outreach, mapping.	N_v

5.2 How do I read and visualize QA4ECV HCHO data?

An easy option to read and visualize the data is with the panoply tool, available for free 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.

5.3 Processing Uncertainty Data

Single pixel uncertainty

The QA4ECV HCHO ECV precursor product contains an algorithm uncertainty estimate for each individual pixel (tropospheric_hcho_vertical_column_uncertainty_estimate). This uncertainty estimate is calculated 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 HCHO background and the AMF (equation 5-1).

$$\sigma_{N_V}^2 = \frac{1}{M^2}\left({\sigma_{N_S}}^2 + \frac{{\Delta N_S}^2}{M^2}{\sigma_M}^2\right) + {\sigma_{N_{v,0}}}^2 \qquad 5\text{-}1$$

where σ_{N_S}, σ_M and $\sigma_{N_{v,0}}$ are the uncertainties on the slant column, the air mass factor and the reference correction, respectively. These quantities are provided in the detailed results (see Table 5-1).

This uncertainty should be interpreted as the best guess of the retrieval uncertainty for one specific measurement. It contains random and systematic components. Two values are provided for the HCHO tropospheric vertical column uncertainty:

- tropospheric_hcho_vertical_column_uncertainty_random ($\sigma_{N_V}^{rand}$)
- tropospheric_hcho_vertical_column_uncertainty_systematic ($\sigma_{N_V}^{syst}$)

The random uncertainty is by far the dominant HCHO uncertainty, directly related to the signal to noise ratio of the HCHO slant columns retrieved from the level 1 spectra. When averaging over multiple pixels, or averaging over time, random errors will tend to cancel, but (an unknown) part of the total uncertainty is from systematic sources with unknown correlation and will remain even after averaging. In the QA4ECV HCHO product, the selected approach is to consider all the errors as systematic, except for the slant column precision. Furthermore, the different contributions to the AMF uncertainty are provided separately (see Table 5-1).

Uncertainty of a mean of retrievals

When averaging the satellite observations, the uncertainty from random effects can be divided by the square root of the number of observations (N):

$$\sigma_{N_V}^2 = \frac{\sigma_{N_V,rand}^2}{N} + \sigma_{N_V,syst}^2 \qquad 5\text{-}2$$

Table 5-1: Uncertainty field names

Field name of uncertainty	Associated quantity
tropospheric_hcho_vertical_column_ uncertainty_random	Random uncertainty on HCHO VCD
tropospheric_hcho_vertical_column_ uncertainty_systematic	Systematic uncertainty on HCHO VCD
scd_hcho_ uncertainty_random	Random uncertainty on HCHO SCD
scd_hcho_ uncertainty_systematic	Systematic uncertainty on HCHO SCD
vcd_hcho_correction_uncertainty	uncertainty on HCHO VCD correction in the reference sector (mostly systematic)
amf_ uncertainty	AMF uncertainty (systematic and random)
amf_albedo_uncertainty	AMF uncertainty due to error on albedo
amf_cloud_fraction_uncertainty	AMF uncertainty due to error on cloud fraction
amf_cloud_pressure_uncertainty	AMF uncertainty due to error on cloud pressure
amf_a_priori_profile_uncertainty	AMF uncertainty due to error on a priori profile
tropospheric_hcho_vertical_column_ uncertainty_systematic_amfes	Systematic uncertainty on HCHO VCD due to amf uncertainty
tropospheric_hcho_vertical_column_ uncertainty_systematic_scdes	Systematic uncertainty on HCHO VCD due to scd uncertainty
tropospheric_hcho_vertical_column_ uncertainty_systematic_vcd0es	Systematic uncertainty on HCHO VCD due to vcd0 uncertainty

5.4 Product Quality Information

The output for each ground pixel is accompanied by flags indicating the status of the results of the processing.

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/filtering needed, and these can be found in the field processing_quality_flag.
- 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.

In the detailed results, the **processing_quality_flag** can be used for a more advanced selection of the observations, based on several quality criteria. The values of the processing quality flags are taken from the S5P/Tropomi quality flag list. Table 5-2 provides an overview of the processing_quality_flag values (error or filter) that apply to the QA4ECV HCHO 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. Warnings do not require the rejection of the observations. Using processing_quality_flags = 0 or > 255 provides a selection of observations that is considered as optimal. Note that if the processing_quality_flags shows that an error occurred, the processing_error_flag is set to 1, but this is not the case for filters. More advanced users may decide to use HCHO observations (from slant columns to vertical columns) also for scenes that are generally filtered out.

Table 5-2: 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		
7	sza_range_error	Solar zenith angle out of range, maximum value from configuration. SZA > 80°
30	chi2_error	RMS > 1.5e-03
36	cloud_error	No cloud data.
42	generic_exception	Catch all generic error.
48	slant_column_density_error	No slant column
49	airmass_factor_error	No AMF
97	geographic_region_filter	No data in the reference sector, no background correction,
Filters		
5	ler_range_filter	TR_HCHO > 0.3 (surface reflectivity at 342 nm)
70	snow_ice_filter	Observations with permanent_ice (100), snow (101), and mixed_pixels_at_coastlines (103) are discarded (based on NISE classification).
72	cloud_filter	CF > 0.4 (cloud fraction) CRF_HCHO > 0.6 (cloud radiance fraction at 341 nm)

5.5 Application of the averaging kernel

The more advanced users interested in application (1) should be careful in their interpretation of the satellite data product. It is important to realize that the HCHO column retrieval has a non-uniform vertical sensitivity to the HCHO vertical distribution, and that this sensitivity is different for every pixel. The vertical sensitivity is determined by the surface and atmospheric properties and by the viewing geometry at the time of measurement. In general, the measurements that have been taken under relatively cloud-free situations, high overhead sun, and small viewing zenith angles have sensitivity to HCHO down to the surface. For dark surfaces, cloudy scenes, scenes with high aerosol loading, low sun, or extreme viewing angles, sensitivity is much poorer, and potential errors in a priori assumptions propagate strongly in the retrieved column product.

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

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

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

with A total column the averaging kernel, a vector specified at *L* pressure levels, and x_m the vertical distribution of HCHO (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) HCHO profile to the pressure grid of the averaging kernel in order to construct a vertical column $\hat{y}_m$ as would be retrieved by the satellite instrument.

6. Contact Information

http://www.qa4ecv.eu/

For questions about the QA4ECV HCHO ECV precursor product, you can get in touch with the product developers. The first contact is Isabelle De Smedt (BIRA-IASB). Please also check out the QA4ECV User Forum on www.qa4ecv.eu/forum.

For direct questions, you can send an email to:

Isabelle De Smedt	isabelle.desmedt@aeronomie.be	HCHO product
Andreas Richter	richter@iup.physik.uni-bremen.de	HCHO product
Folkert Boersma	boersma@knmi.nl	QA4ECV, NO2
Huan Yu	huan.yu@aeronomie.be	HCHO AMF
Thomas Danckaert	thomas.danckaert@aeronomie.be	DOAS tool
Jos van Geffen	geffen@knmi.nl	NO2, L2 format
Henk Eskes	eskes@knmi.nl	TM5, AK
Steffen Beirle	steffen.beirle@mpic.de	Background correction
Ronald van der A	Ronald.van.der.A@knmi.nl	Accessing data

7. References

Boersma, K. F., H. J. Eskes, and E. J. Brinksma (2004), Error analysis for tropospheric NO2 retrieval from space, J. Geophys. Res., 109, D04311, doi:10.1029/2003JD003962.

Chance, K. V, Palmer, P. I., Martin, R. V, Spurr, R., Kurosu, T. P. and Jacob, D. J.: Satellite observations of formaldehyde over North America from GOME, Geophys. Res. Lett., 27(21), 3461–3464, doi:10.1029/2000GL011857, 2000.

Danielson, J. J., and Gesch, D. B.: Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010), U. S. Geological Survey. Available from http://topotools.cr.usgs.gov/gmted_viewer/, 2011.

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J.-F. and Van Roozendael, M.: Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, Atmos. Chem. Phys., 15(8), 12241–12300, doi:10.5194/acpd-15-12241-2015, 2015.

González Abad, G., Liu, X., Chance, K. V, Wang, H., Kurosu, T. P. and Suleiman, R.: Updated SAO OMI formaldehyde retrieval, Atmos. Meas. Tech. Discuss., 7(1), 1–31, doi:10.5194/amtd-7-1-2014, 2014.

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.

Hewson, W., Bösch, H., Barkley, M. P. and De Smedt, I.: Characterisation of GOME-2 formaldehyde retrieval sensitivity, Atmos. Meas. Tech., 6(2), 371–386, doi:10.5194/amt-6-371-2013, 2013.

Kleipool, Q. L., Dobber, M. R., de Haan, J., & Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data. Journal of Geophysical Research: Atmospheres (1984–2012), 113(D18), 2008.

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 (2016[a]), Recommendations (scientific) on best practices for Land and Atmosphere ECVs, QA4ECV Report / Deliverable n° 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 (2016[b]), Harmonized, sensor-independent retrievals (software) for HCHO and NO_2. Report on MOPITT-IASI total column consistency, QA4ECV Report / Deliverable n° D4.4, A. Lorente (Ed), retrieved from QA4ECV website: restricted access (25-01-2017).

Spurr, R. J.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. J. Quant. Spec. Radiat. Transfer, 102(2), 316-342, 2006.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in Light Scattering Reviews, edited by A. Kokhanovsky, pp. 229–271, Berlin., 2008.

Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F. (2016). Improvements to the OMI O_2–O_2 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.

Wittrock, F., Richter, A., Oetjen, H., Burrows, J. P., Kanakidou, M., Myriokefalitakis, S., Volkamer, R., Beirle, S., Platt, U. and Wagner, T.: Simultaneous global observations of glyoxal and formaldehyde from space, Geophys. Res. Lett., 33(16), 1–5, doi:10.1029/2006GL026310, 2006.