

S5P/TROPOMI HCHO ATBD



documentnumber : S5P- BIRA-L2-400F- AT BD
CI identification : CI-400F-AT BD
issue : 2.2.0
date : 2020 06 15
status : Update for the version 2.1.0 of the
operational processor

Document approval record

	digital signature
prepared:	<div> <div> Isabelle De Smedt Digitally signed by Isabelle De Smedt Date: 2020.07.02 11:25:54 +02'00'</div> <div> Nicolas Theys Digitally signed by Nicolas Theys Date: 2020.07.02 11:28:18 +02'00'</div> <div> Hua n Yu Digitally signed by Huan Yu Date: 2020.07.02 11:29:10 +02'00'</div> </div> <div> <div> Jonas Vlietinck Digitally signed by Jonas Vlietinck Date: 2020.07.02 11:30:19 +02'00'</div> <div> Christophe Lerot Digitally signed by Christophe Lerot Date: 2020.07.02 11:31:10 +02'00'</div> <div> Michel Van Roozenda el Digitally signed by Michel Van Roozenda el Date: 2020.07.02 11:32:06 +02'00'</div> </div>
Checked:	Andreas Hilboll, Andreas Richter, Folkard Wittrock Randall Martin, Michael Barkley Steven Compernelle, Kai-Uwe Eichmann
approved PI:	
approved PM:	

Document change record

issue	date	item	comments
0.0.1	2012-09-25	All	Initial draft version
0.0.2	2012-11-23		Changes after the first internal review
0.1.1	2013-06-04	All	CDR version for internal review
0.5.0	2013-06-21	All	CDR version
0.9.0	2013-11-30		Changes after the first external review
0.10.0	2014-04-15		Consolidated changes after the first external review. Instrument overview section now in separate document. Update figure for the flow Diagram of the L2 HCHO retrieval algorithm
0.11.0	2014-09-30		Minor changes after June 2014 CDR. Harmonized input/output tables between ATBDs. Updated number of observations per orbit.
0.13.0	2015-09-15	5.4.1	Minor changes. Update of altitude-resolved air mass factors LUT table. Update for limited release to S5P validation team.
0.14.0	2015-12-11	All	Update of document for the instrumental binning scheme. Update of figures 1 and 11 and references. Section 5.7 "Error characterization" has been merged with section 7 "Error Analysis".
1.0.0	2016-01-30	All	Update after the second internal review
1.1.0	2016-06-30		Update after S4 ATBD internal review
1.2.0	2017-08-16	SCD, BC	Update after QA4ECV developments
1.3.0	2018-10-15		Description of algorithm updates for operational release (phase E2). Various improvements of the text according to the AMT special issue paper: De Smedt et al., 2018.
2.0.0	2019-07-31		Update for version 2.0.0 of the operational processor New pixel size (5.5x3.5 km ²) from 06 August 2019 and on Update of the background correction of the slant columns in order to minimize sunglint effects. Use of the TM5 tropopause layer index
2.1.0	2020-02-28		Update with pixel size 5.5x3.5 km ² Update of applicable and reference document versions

Contents

Document approval record	2
Document change record	3
Contents	4
List of figures	6
List of tables	7
1 Introduction	8
1.1 Identification	8
1.2 Purpose and objective	8
1.3 Document overview	8
1.4 Acknowledgements	8
2 Applicable and reference documents	9
2.1 Applicable documents	9
2.2 Standard documents	9
2.3 Reference documents	9
2.4 Electronic references	10
3 Terms, definitions and abbreviated terms	11
3.1 Terms and definitions	11
3.2 Acronyms and abbreviations	11
4 Introduction to formaldehyde retrieval	13
4.1 Product description	13
4.2 Heritage	14
4.3 Product requirements	15
4.4 Overview of the retrieval method	16
4.5 Design considerations	17
5 Algorithm description	18
5.1 Overview of the retrieval scheme	18
5.2 DOAS equations for nadir satellite measurements	19
5.3 Formaldehyde slant column retrieval	22
5.3.1 Fitting intervals, absorption cross-sections and retrieval settings	24
5.3.2 Wavelength calibration and convolution to TROPOMI resolution	26
5.3.3 Spike removal algorithm	27
5.4 Tropospheric air mass factor calculation	28
5.4.1 LUT of altitude dependent air mass factors	28
5.4.2 Treatment of partly cloudy scenes	30
5.4.3 Aerosols	32
5.4.4 A priori vertical profile shapes	32
5.5 Across-track and zonal reference sector correction	34
6 Input and output	Error! Bookmark not defined.
6.1 High level data product description	36
6.2 Auxiliary information needs	37
6.2.1 Dynamic information	37
6.2.2 Static information	38
7 Error analyses	40
7.1 Error characterization	40
7.2 Errors on the slant columns	41
7.2.1 Random errors on slant columns	41
7.2.2 Systematic errors on the slant columns	42
7.3 Errors on air mass factors	43

7.3.1	Surface albedo	44
7.3.2	Clouds and aerosols	45
7.3.3	Profile shape	45
7.3.4	AMF wavelength dependence, model atmosphere and LUT interpolation errors	46
7.4	Errors on the reference sector correction.....	47
7.5	HCHO error estimates and product requirements.....	47
7.5.1	Precision	47
7.5.2	Trueness	48
8	Validation	50
9	Conclusions.....	51
A	Description of prototype software.....	52
B.	Description of forward model.....	54
C.	Averaging Kernel	55
	References.....	56

List of figures

Figure 4-1: HCHO vertical column retrieved from TROPOMI during summer 2018 (De Smedt et al., 2018).

.....14

Figure 5-1: Flow Diagram of the L2 HCHO retrieval algorithm.18

Figure 5-2: Sketch of satellite radiation measurement and geometry in a plane parallel atmosphere.20

Figure 5-3: Typical optical densities of O₃, BrO, NO₂, Ring effect (solar lines and molecular), O₂-O₂, and HCHO in the near UV. The slant columns have been taken as 10²⁰ molec.cm⁻² for O₃, 10¹⁴ molec.cm⁻² for BrO, 1x10¹⁶ molec.cm⁻² for NO₂, 0.4x10⁴² molec².cm⁻⁵ for O₂-O₂, and 10¹⁶ molec.cm⁻² for HCHO.24

Figure 5-4: Right panel: Examples of TROPOMI slit functions around 340 nm, for row 1 and row 225. Left panel: TROPOMI spectral resolution in channel 3, as a function of the row and the wavelength, derived from the instrument key data ISFR v2.0.0.27

Figure 5-5: Variation of the altitude dependent air mass factor with: (a) solar zenith angle, (b) viewing zenith angle, (c) relative azimuth angle between the sun and the satellite, (d) surface albedo, (e) surface pressure for a weakly reflecting surface, (f) surface pressure for a highly reflecting surface. Unless specified, the parameters chosen for the radiative transfer simulations are: SZA=30°, VZA=0°, RAA=0°, albedo=0.05, surface pressure=1063hPa, λ=340nm.29

Figure 5-6: Altitude dependent AMF for a cloud free pixel (in blue), a pixel with a cloud radiance fraction of 0.5 (in orange) and a fully cloudy pixel (in grey). The cloud pressure is 540hPa, and the cloud albedo is 0.8. In all cases, we consider a nadir view, a SZA of 30° and a ground albedo of 0.05.32

Figure 5-7: Daily map of TROPOMI tropospheric air mass factors at 340 nm. A priori HCHO profiles from high-resolution TM5-MP model runs are used. The right panel shows air mass factors non corrected for clouds (*M_{clear}*), while on the right panel, the IPA cloud correction is applied for effective cloud fractions *f_{eff}* larger than 10% (*M*). Observations with *f_{eff}* larger than 30% have been filtered out. The operational CRB cloud product is used.34

Figure 7-1: First panel: TM5-MP HCHO profiles extracted in June over the equatorial Pacific ocean (blue) and over Beijing (red). Those profiles have been used to calculate the tropospheric air mass factors shown in the panels a to e, representing the AMF dependence on (a) the surface albedo, (b) the cloud altitude, (c), (d), (e) the cloud fraction. In all cases, we consider a nadir view and a solar zenith angle of 30°. In (a) the pixel is cloud free, in (b) the albedo is 0.02 and the effective cloud fraction is 0.5, in (c), (d), (e) the ground albedo is 0.02 and the cloud pressure is respectively 966, 795 and 540 hPa.44

Figure 7-2: Precision on the TROPOMI HCHO columns, in several NMVOC emission regions, and at different spatial and temporal scales (from individual pixels to monthly averages in 20x20 km² grids). These estimated TROPOMI measurements in 2019.48

Figure 7-3: Regional and monthly average of the relative systematic vertical column AMF-related uncertainties in several NMVOC emission regions, for the period 2005-2014. The 5 contributions to the systematic air mass factor error uncertainty are shown: structural (green), a priori profile (pink), albedo (olive), cloud fraction (blue) and cloud altitude (cyan).49

List of tables

Table 1: HCHO data product requirements for TROPOMI, as given in [RD04] and [RD05] for tropospheric HCHO. Where numbers are given as "a / b", the first is the target requirement and the second is the threshold requirement.....	16
Table 2: Wavelength intervals used in previous formaldehyde retrieval studies [nm].	23
Table 3 : Summary of DOAS settings used to retrieve HCHO slant columns from TROPOMI spectra.	25
Table 4: Parameters in the altitude dependent air mass factors lookup table.....	30
Table 5: Prior information datasets used in the air mass factor calculation.	33
Table 6: List of output fields in the TROPOMI HCHO product. scanline x ground_pixel mean the number of pixels in an orbit along track and across track, respectively.	36
Table 7: Dynamic auxiliary information needs in the HCHO retrieval algorithm	37
Table 8: Static auxiliary information needs in the HCHO retrieval algorithm	38
Table 9: Summary of the different error sources considered in the HCHO slant column error budget.	42
Table 10: Number of grid points considered for the altitude dependent air mass factors and the air mass factor sensitivity lookup tables	43
Table 11: Summary of the different error sources considered in the air mass factor error budget.	46
Table 12: Estimated errors on the reference sector correction.	47
Table 13: Estimated HCHO vertical column uncertainty budget for monthly averaged low and elevated columns (higher than 1×10^{16} molec.cm ⁻²). Contributions from the three retrieval steps are provided, as well as input parameter contributions.	49

1 Introduction

1.1 Identification

This document describes the prototype algorithm for the retrieval of formaldehyde columns from TROPOMI measurements in the UV part of the spectrum.

1.2 Purpose and objective

The purpose of the document is to give a high-level description of the algorithm that is used to retrieve formaldehyde columns from TROPOMI measurements, the necessary input and auxiliary data, and the output that is generated. In addition, information about the size of the product, calculation times, and the accuracy are provided.

1.3 Document overview

Chapter 4 gives a brief introduction to formaldehyde retrieval. Chapter 5 provides a high level description of the retrieval algorithm. In Chapter 6 the feasibility is discussed. In Chapter 7 the errors are investigated. In Chapter 8 validation and tools for validation are discussed. In Chapter 9, conclusions are drawn. A description of the prototype software is given in the Appendix A. A description of the TROPOMI instrument and performance, referred to from all ATBDs, can be found in RD14.

1.4 Acknowledgements

The authors would like to thank the following people for useful discussions, information and other contributions:

Andreas Richter, Maarten Sneep, Pepijn Veefkind, Jos van Geffen, Folkert Boersma, Henk Eskes, Ronald van der A, Mattia Pedernana, Nan Hao, Zhibin Cheng, Diego Loyola, Pieter Valks, Rob Spurr.

2 Applicable and reference documents

2.1 Applicable documents

- [AD01] GMES Sentinel-5 Precursor – S5P System Requirement Document (SRD); source: ESA/ESTEC; ref: S5P-RS-ESA-SY-0002; issue: 4.1; date: 2011-04-29
- [AD02] Sentinel-5P Level 2 Processor Development – Statement of Work -; source: ESA; ref: S5P-SW-ESA-GS-053; issue: 1.1; date: 2012-05-21
- [AD03] S5P Level 2 Processor Development – Level 2 Processor Requirements Specifications source: ESA; ref: S5P-SW-ESA-GS-054; issue 1.2 draft; date: 2014-09-15
- [AD04] S5P/TROPOMI Level 2 Product Development Plan, source: KNMI, ref: S5P-KNMI-L2CO-0010-PL issue: 1.1.0 date: 2014-06-02
- [AD05] S5P – Tailoring of ECSS Standards for the Level 2 Processor Development; source: ESA; ref: SP-RS-ESA-GS-055; issue 1.1dr; date: 2012-10-31
- [AD06] Sentinel-5P Level 2 Processor Development: Coordination Tasks; source: ESA; ref: S5P-SW-ESA-GS-081; issue: 1; date: 2012-06-27

2.2 Standard documents

There are no standard documents

2.3 Reference documents

- [RD01] Terms, definitions and abbreviations for TROPOMI L01b data processor; source: KNMI; ref: S5P-KNMI-L01B-0004-LI; issue: 3.0.0; date: 2013-11-08
- [RD02] Terms, and symbols in the TROPOMI Algorithm Team; source: KNMI; ref: SN-TROPOMI-KNMI-L2-049-MA; issue: 1.0.0; date: 2015-07-16
- [RD03] Science Requirements Document for TROPOMI. Volume 1; source: KNMI & SRON; ref: RS-TROPOMI-KNMI-017; issue: 2.0; date: 2008-10-30.
- [RD04] GMES Sentinels-4 and-5 Mission Requirements Document (MRD); source: ESA; ref: EO-SMA-/1507/JL; issue: 3; date: 2011-09-21
- [RD05] Report Of The Review Of User Requirements For Sentinels-4/-5; source: ESA; ref: EO-SMA-/1507/JL; issue: 2.1; date: 2011-12-21
- [RD06] CAPACITY: Operational Atmospheric Chemistry Monitoring Missions – Final report; source: KNMI; ref: CAPACITY; date: Oct. 2005.
- [RD07] CAMELOT: Observation Techniques and Mission Concepts for Atmospheric Chemistry; source: KNMI; ref: RP-CAM-KNMI-050; date: Nov. 2009.
- [RD08] TRAQ: Performance Analysis and Requirements Consolidation - Final Report; source: KNMI; ref: RP-ONTRAQ-KNMI-051; date: Jan. 2010.
- [RD09] S5P/TROPOMI ATBD of the Aerosol data products; source: KNMI; ref: S5P-KNMI-L2-0008-RP-TROPOMI_ATBD_UVAI; issue: 1.1.0; date: 2018-06-15.
- [RD10] S5P/TROPOMI ATBD of the Cloud data products; source: DLR; ref: S5P-L2-DLR-ATBD-400I_Clouds; issue: 2.2.0; date: 2020-06-15.
- [RD11] S5P/TROPOMI ATBD of the total and tropospheric NO₂ data products; source: KNMI; ref: S5P-KNMI-L2-0005-RP-ATBD_NO2_data_products; issue: 1.4.0; date: 2019-02-06.
- [RD12] S5P/TROPOMI ATBD of the SO₂ data products; source: BIRA-IASB; ref: S5P-L2- BIRA-ATBD-SO2-400E; issue: 2.2.0; date: 2020-06-15.

- [RD13] S5P/TROPOMI Static input for Level 2 processors; source: KNMI; ref: S5P-KNMI-L2CO-0004-SD; issue: 3.0.0; date: 2015-02-27
- [RD14] TROPOMI Instrument and Performance Overview; source: KNMI; ref: S5P-KNMI-L2-0010-RP; issue: 0.10.0; date: 2014-03-15.
- [RD15] Sentinel-5 Precursor Level 2 UPAS Processor Input / Output Definition Document; source: DLR; ref: S5P-L2-DLR-IODD-3002; issue: 3.5.0; date: 2019-08-09.
- [RD16] Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Formaldehyde HCHO; source: DLR; ref: S5P-L2-DLR-PUM-400F; issue: 2.2.0; date: 2020-06-15.
- [RD17] Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products; ref: S5P-MPC-IASB-ROCVR; issue: 6.0.1; date: 2020-03-30.

2.4 Electronic references

- [URL01] <http://uv-vis.aeronomie.be/software/QDOAS/>
- [URL02] <http://www.projects.science.uu.nl/tm5/>
- [URL03] <http://www.rtslidort.com/>
- [URL04] <http://www.temis.nl/fresco/>

3 Terms, definitions and abbreviated terms

Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L0-1b data processor are described in [RD01]. Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L2 data processors are described in [RD02]. Terms, definitions and abbreviated terms that are specific for this document can be found below.

3.1 Terms and definitions

The most important symbols related to the data product described in this document – some of which are not in [RD02]– are the following:

AK	averaging kernel
m	altitude-resolved air mass factor or weighting function
M	air-mass factor
M^{geo}	geometric air-mass factor
N_s	slant column density
N_v	vertical column density

3.2 Acronyms and abbreviations

AAI	Absorbing Aerosol Index
AMF	Air Mass Factor
BrO	Bromine Monoxide
BRDF	Bidirectional reflectance distribution function
CH ₄	Methane
CO	Carbon Monoxide
CRB	Clouds as Reflecting Boundaries
CTM	Chemistry Transport Model
DOAS	Differential Optical Absorption Spectroscopy
ENVISAT	Environmental Satellite
ERS	European Remote Sensing satellite
FRESCO	Fast Retrieval Scheme for Clouds from the Oxygen A band
GOME-2	Global Ozone Monitoring Experiment–2
HCHO	Formaldehyde
LOS	Line Of Sight
MAG	Mission Advisory Group

MetOp	Meteorological Operational Satellite
NMVOC	Non-Methane Volatile Organic Compound
NO _x	Nitrogen oxides
NRT	near-real time (i.e. processing within 3 hours of measurement)
OCRA	Optical Cloud Recognition Algorithm
OE	Optimal Estimation
OMI	Ozone Monitoring Instrument
O ₃	Ozone
PDGS	Sentinel-5 Precursor Payload Data Ground Segment (at DLR)
RAA	Relative Azimuth Angle
ROCINN	Retrieval of Cloud Information using Neural Networks
SCIAMACHY	Scanning Imaging Absorption spectroMeter for Atmospheric Cartography
SZA	Solar Zenith Angle
TM 4/5	Data assimilation / chemistry transport model (version 4 or 5)
TROPOMI	Tropospheric Monitoring Instrument
VOC	Volatile Organic Compound
VZA	Viewing Zenith Angle

4 Introduction to formaldehyde retrieval

4.1 Product description

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 CO₂ (Seinfeld and Pandis, 2006). NMVOCs are, together with NO_x, CO, and CH₄, among the most important precursors of tropospheric ozone. NMVOCs also produce secondary organic aerosols and influence the concentrations of OH, the main tropospheric oxidant (Hartmann et al., 2013). The major HCHO source in the remote atmosphere is CH₄ 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 (as illustrated in Figure 1, Stavrou et al., 2009a). With its lifetime on the order of a few hours, HCHO concentrations in the boundary layer can be 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 CH₄ and NMVOCs. The seasonal and inter-annual variations of the formaldehyde distribution are principally related to temperature changes (controlling vegetation emissions) and fire events, but also to changes in anthropogenic activities (Stavrou et al., 2009b). For all 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 (e.g. Abbot et al., 2003, Palmer et al., 2006; Fu et al., 2007; Millet et al., 2008; Stavrou et al., 2009a, 2009b, 2012, 2015; Curci et al., 2010; Barkley et al., 2011, 2013; Fortems-Cheiney et al., 2012; Marais et al., 2012; Mahajan et al., 2015; Kaiser et al., 2017).

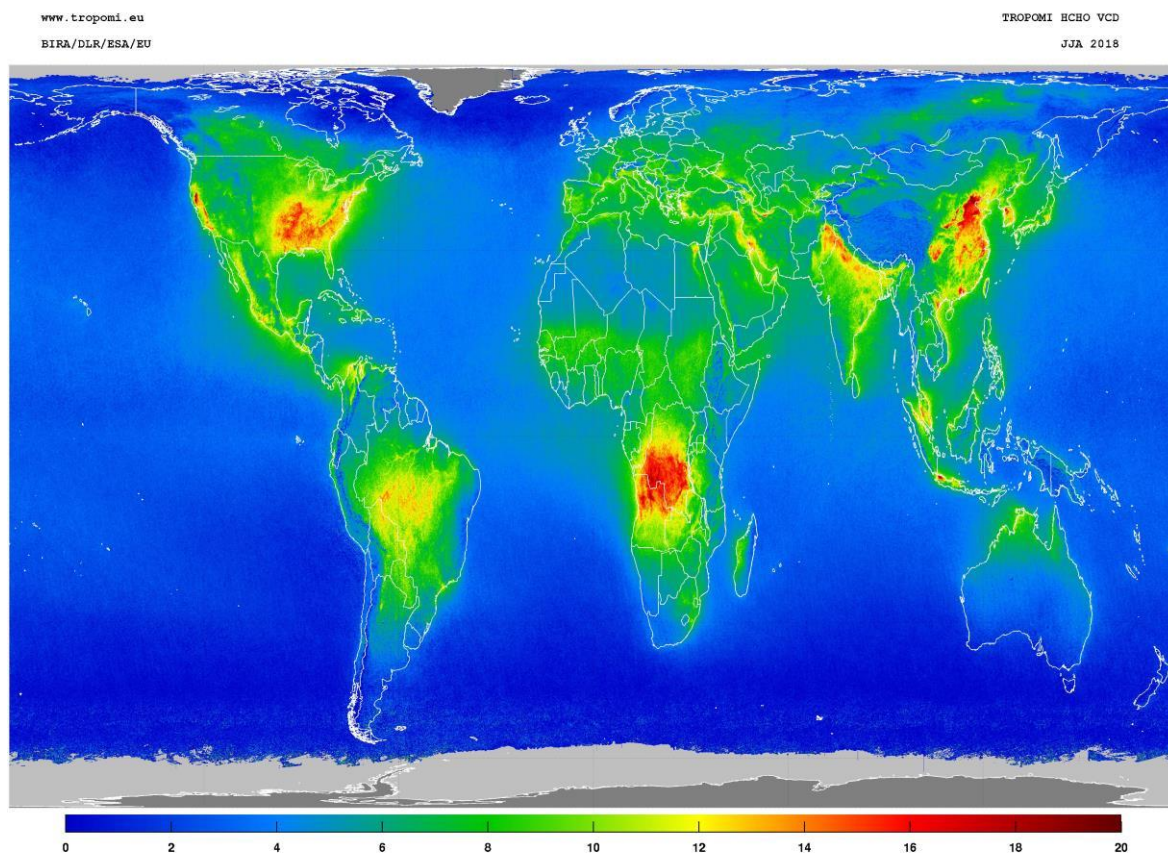


Figure 4-1: HCHO vertical column retrieved from TROPOMI during summer 2018 (De Smedt et al., 2018).

4.2 Heritage

Formaldehyde columns in the boundary layer can only be sounded from space using UV nadir-viewing spectrometers. HCHO tropospheric columns have been successively retrieved from GOME on ERS-2 and from SCIAMACHY on ENVISAT, resulting in a continuous data set covering a period of almost 16 years from 1996 until 2012 (Chance et al., 2000; Palmer et al., 2001; Wittrock et al., 2006; Marbach et al., 2009; De Smedt et al., 2008; 2010). Started in 2007, the measurements made by the three GOME-2 instruments (EUMETSAT METOP-A, B and C) have the potential to extend by more than a decade the successful time-series of global formaldehyde morning observations (Vrekoussis et al., 2010; De Smedt et al., 2012; Hewson et al., 2012). Since its launch in 2004, OMI on the NASA AURA platform has been providing complementary HCHO measurements in early afternoon with a global daily coverage and a better spatial resolution than European sensors (Kurosu et al., 2008; Millet et al., 2008; González Abad et al., 2015; De Smedt et al., 2015) (Figure 4-1). S5P/TROPOMI aims to continue this time series of early afternoon observations, with a daily global Earth coverage, a finer spatial resolution and an equivalent signal to noise ratio compared to OMI, offering an unprecedented view of the spatiotemporal variability of NMVOC emissions.

4.3 Product requirements

The scientific requirements for formaldehyde (HCHO) retrievals are identified from the TROPOMI science objectives RD03, the COPERNICUS requirements for sentinels-4/-5 [RD04], and the S5P MAG report of the review of user requirements for sentinels-4/-5 [RD05]. The requirements for HCHO are summarised in Table 1. Uncertainty requirements include retrieval errors as well as measurement (instrument-related) errors. Absolute requirements (in total column units) relate to background conditions where uncertainties are driven by the signal-to-noise ratio of the measurements. In contrast for elevated columns, retrieval errors are dominated by uncertainties on the air mass factors, and therefore uncertainty requirements are given in percentage value.

These requirements derive from the CAPACITY study [RD06] and have been fine-tuned by the CAMELOT [RD07] and ONTRAQ [RD08] studies. The CAPACITY study has defined three main themes: the ozone layer (A), air quality (B), and climate (C) with further division into sub themes¹. Requirements for HCHO have been specified for a number of these sub themes. With respect to air quality protocol monitoring (B1) which is mostly concerned with trend and variability analysis, the requirements are specified for NMVOC emissions on monthly to annual time scales and for larger region/country [RD05]. Depending on the number of independent observations available over a region and for a given time window, the random component of the uncertainty requirement can be relaxed by the square-root of the number of observations. However, because the random component is only a fraction of the total uncertainty, the relaxation for a certain location in a given time period might be modest. Also aggregation of independent observations over larger regions/countries might relax the threshold uncertainty. In the error analysis section (7), we discuss these requirements and the performances of the HCHO retrieval algorithm.

Note that the coverage requirement for S5P is always global, and that there is no vertical resolution requirement applicable for the formaldehyde column, which is essentially a tropospheric column. Note also that the horizontal resolution, the revisit time and the daily coverage, are driven by the choice of the instrument characteristics and the satellite orbit.

¹ GMES environmental themes and types of user information domains for which HCHO column observation is required:

- Theme B: Air Quality
 - B1 : Air Quality Protocol Monitoring
 - B2: Air Quality Near-Real Time
 - B3: Air Quality Assessment
- Theme C: Climate
 - C3: Climate Assessment

Table 1: HCHO data product requirements for TROPOMI, as given in [RD04] and [RD05] for tropospheric HCHO. Where numbers are given as "a / b", the first is the target requirement and the second is the threshold requirement.

Requirements ¹	Horizontal resolution	Revisit time	Uncertainty ²
B1, B2, B3	5km / 20km	0.5 / 2 hour	20% or 1.3×10^{15} molec.cm ⁻² [RD04]
C3	5km / 50km	6 / 24x3 hour	30% or 1.3×10^{15} molec.cm ⁻² [RD04]
B1			60% or 0.5ppbv (least stringent) [RD05]
B2			30% or 0.5ppbv (least stringent) [RD05]

4.4 Overview of the retrieval method

The general method used for the derivation of HCHO VCDs from UV spectral measurements is the Differential Optical Absorption Spectroscopy method (DOAS; Platt and Stutz, 2008) 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 (AMF) obtained from suitable radiative transfer calculations, accounting for the presence of clouds, surface properties, and best-guess HCHO vertical profiles.

In the UV, the sensitivity to HCHO concentrations in the boundary layer is intrinsically limited from space due to the combined effect of Rayleigh and Mie scattering that limits the radiation fraction scattered back toward the satellite. In addition, ozone absorption reduces the number of photons that reaches the lowest atmospheric layers. Furthermore, the absorption signatures of HCHO are weaker than those of other UV-Vis absorbers, such as e.g. NO₂. As a result, the retrieval of formaldehyde from space is noise sensitive and error prone. While the precision (or random error uncertainty) is driven by the signal to noise ratio of the recorded spectra and by the width of the retrieval interval, the trueness (or systematic error uncertainty) is limited by the current knowledge on the external parameters needed in the different retrieval steps.

The selection of the optimal retrieval interval must maximize the sensitivity of the inversion to the HCHO absorption signatures, while minimizing errors from geophysical and instrument related spectral features. The retrieval interval should be chosen as wide as possible to maximize the number of sampling points, while avoiding overlap with strong atmospheric spectral features from interfering species (mainly O₃, BrO, and O₄). The DOAS algorithm intrinsically assumes that the atmosphere is optically thin so that the optical light path is independent of wavelength within the fitting window. Hence the method is accurate only for modest ozone absorption (i.e., for small to medium solar zenith angles). Generally, the effect of ozone misfit on the retrieval can be handled by introducing wavelength-dependent-AMF in the fit, and by applying appropriate background corrections on the columns. The correlation with BrO absorption can be reduced by using two different

² 1.3×10^{15} molec.cm⁻² = 0.05 DU = 0.4 ppbv mean mixing ratio for a ~1.5 km boundary layer up to 850hPa.

wavelength intervals to fit BrO and HCHO (see section 5.3: Formaldehyde slant column retrieval).

The Sentinel-5P sensor TROPOMI samples the Earth's surface with a revisit time of one day and with an unprecedented spatial resolution of $7 \times 3.5 \text{ km}^2$ ($5.5 \times 3.5 \text{ km}^2$ from 6 August 2019). Furthermore, the signal to noise ratio of TROPOMI is required to be equivalent to OMI in the UV-Visible range. This allows the resolution of fine details and S5P will arguably be a valuable tool to better study NMVOC emissions. Nevertheless, it poses additional constraints on the retrieval code for several reasons:

- Computational speed: the Level 1b data flow delivers spectral measurements for band 3 with a size of 4 gigabytes per orbit (15 orbits daily).
- Precision: given the required signal to noise ratio of measured spectra, the air quality requirements for HCHO [RD04] will not be met for individual ground pixels. Spatial and temporal averages of the HCHO columns will have to be considered in order to provide useful constraints on human and natural NMVOC emissions.
- Trueness: currently, the spatial resolution of global-scale external parameters needed in the AMF calculation (e.g. albedo or a priori HCHO profiles) is much coarser than the resolution of TROPOMI. This introduces errors in the final vertical columns, mostly for coastal regions and localized sources (Heckel et al., 2011)

To fully exploit the potential of satellite data, applications relying on tropospheric HCHO observations require high quality long-term time series, provided with well characterized errors and averaging kernels, and consistently retrieved from the different sensors. Furthermore, as the HCHO observations are aimed to be used synergistically with other species observations (e.g. for air quality applications), it is essential to homogenize as far as possible the retrieval methods as well as the external databases, in order to minimize systematic biases between the observations.

4.5 Design considerations

The design of the TROPOMI HCHO prototype algorithm, developed at BIRA-IASB, has been driven by the experience developed with formaldehyde retrievals from the series of precursor missions OMI, GOME(-2) and SCIAMACHY. Furthermore, within the S5P Level 2 Working Group project (L2WG), a strong component of verification has been developed involving independent retrieval algorithms for each operational prototype algorithm. For HCHO, the University of Bremen (IUP-UB) has been responsible of the algorithm verification. An extensive comparison of the processing chains of the prototype (the retrieval algorithm presented in this paper) and verification algorithm has been conducted. In parallel, within the EU FP7-project Quality Assurance for Essential Climate Variables (QA4ECV, Lorente et al., 2017), a detailed step by step study has been performed, including more scientific algorithms (BIRA-IASB, IUP-UB, MPIC, KNMI and WUR), leading to state-of-the art European products (www.qa4ecv.eu). Those iterative processes led to improvements that have been implemented for the operational release of the HCHO product (phase E2).

5 Algorithm description

5.1 Overview of the retrieval scheme

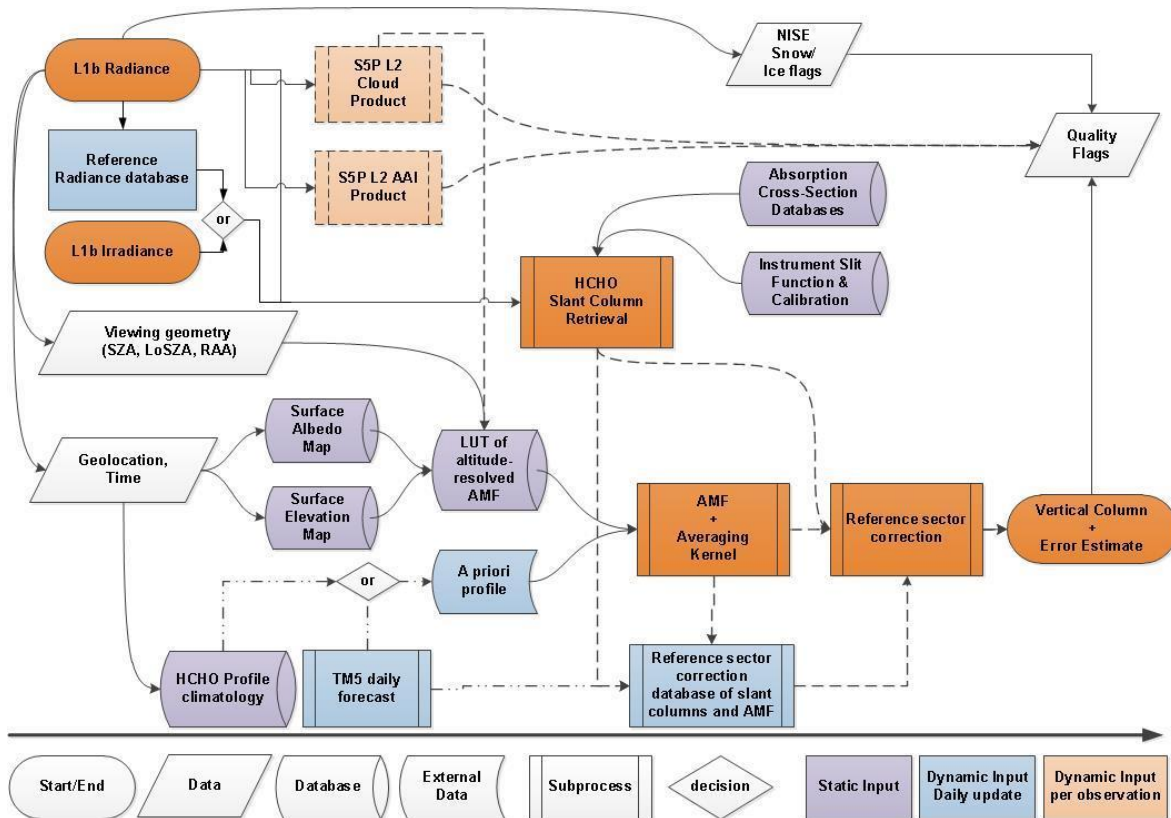


Figure 5-1: Flow Diagram of the L2 HCHO retrieval algorithm.

Figure 5-1 shows the flow diagram of the L2 HCHO retrieval algorithm implemented in the UPAS-2 operational processor RD15. The baseline operation flow scheme is based on a DOAS retrieval algorithm (Platt et al., 1994; Platt and Stutz, 2008; and references therein; Gottwald et al., 2006). The DOAS component is identical to that of the retrieval flow of SO₂ (also developed by BIRA-IASB, see S5P SO₂ ATBD RD12). The interdependencies with auxiliary data and other L2 retrievals, such as clouds, aerosols or NO₂ are also represented, and further described in the feasibility section 6. Here, we will first discuss the principle of the DOAS retrieval before discussing the separate steps of the process in more details.

The DOAS algorithm is a widely-used two-step scheme for obtaining vertical column density from spectral observations. Firstly, the slant column density is derived from the measured reflectance spectrum. Secondly, an AMF is determined for the conversion of the slant column into a vertical column. A third step consisting in background normalization is generally required in the case of weak absorbers such as formaldehyde.

The radiance and irradiance data are read from the L1b file, along with geolocation data as pixel coordinates and observation geometry (sun and viewing angles). The relevant absorption cross section data as well as characteristics of the instrument are used as input for the HCHO slant column density determination (N_s). In parallel to the slant column fit, cloud cover information and AI are obtained from the operational chain, as they are required

as dynamic input. Subsequently, the slant column is converted to a vertical column (N_v) through division by an air mass factor (M) which accounts for the average light path through the atmosphere. For this purpose, several auxiliary data are read from external sources: cloud cover data, topographic information, surface albedo, and the a priori shape of the vertical HCHO profile in the atmosphere. The AMF is derived from interpolation in a look-up-table providing altitude-resolved air mass factors (also used as a basis of the error calculation and retrieval characterization module). This look up table has been created using the VLIDORT 2.6 radiative transfer model (Spurr et al., 2008a) at a single wavelength representative for the retrieval interval (340 nm). It is used to compute the total column averaging kernels (Eskes & Boersma, 2003), which are essential information on the measurement vertical sensitivity and are required for comparison with other types of data (Veefkind et al., 2012). Background normalization of the slant columns is required in the case of weak absorbers such as formaldehyde. Before converting the slant columns into vertical columns, background values of N_s are normalized to compensate for possible systematic offsets (reference sector correction, see below). The tropospheric vertical column end product results therefore from a differential column to which is added the HCHO background due to methane oxidation, estimated using a tropospheric chemistry transport model. ($N_{v,0}$).

The final HCHO vertical column is obtained using the following equation:

$$N_v = \frac{N_s - N_{s,0}}{M} + N_{v,0} \quad 5-1$$

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 the clear sky air mass factor, the uncertainty on the total column, the averaging kernel, and quality flags. Table 6 gives a non-exhaustive set of data fields that are provided in the level 2 data product. A complete description of the level 2 data format is given in the S5P HCHO Product User Manual (Pederghana et al., 2017).

5.2 DOAS equations for nadir satellite measurements

Figure 5-2 sketches the geometry of a nadir satellite radiation measurement, defining the solar and viewing zenith angles (θ_0 and θ , with $\mu_0 = |\cos\theta_0|$ and $\mu = |\cos\theta|$) and the solar and viewing azimuth angles (φ_0 and φ) towards the zenith in a plane parallel atmosphere. Following the definitions adopted in [RD02], the angles are given at the surface ($z = 0\text{m}$).

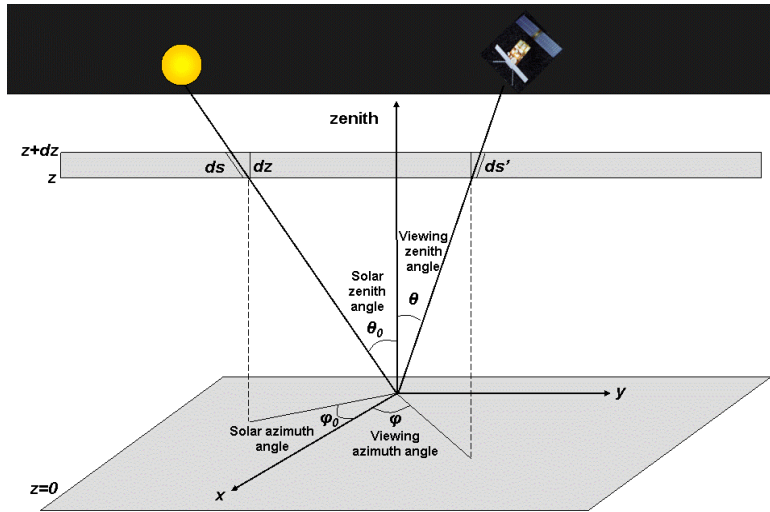


Figure 5-2: Sketch of satellite radiation measurement and geometry in a plane parallel atmosphere.

The DOAS method relies on the application of the Beer-Lambert law to the whole atmosphere. The backscattered earthshine spectrum as measured by the satellite spectrometer contains the strong solar Fraunhofer lines and additional fainter features due to interactions taking place in the Earth atmosphere during the incoming and outgoing paths of the radiation. The basic idea of the DOAS method is to separate broad and narrowband spectral structures of the absorption spectra in order to isolate the narrow trace gas absorption features. In practice, the DOAS approach relies on the following key approximations:

1. A single mean optical path is considered for photons traveling through the atmosphere towards the instrument. This allows expressing light attenuation in term of the Beer-Lambert's law and separating spectroscopic retrievals from radiative transfer calculations.
2. The absorption cross-sections are supposed to be weakly dependent on temperature and independent of pressure: this allows introducing the concept of one effective slant column density for the considered wavelength window.
3. Broadband variations are approximated by a common low-order polynomial to compensate for the effects of loss and gain from scattering and reflections by clouds and/or at the Earth surface.

1. A single "effective" light path is assumed which represents an average of the complex paths of all reflected and scattered solar photons reaching the instrument within the spectral interval used for the retrieval. This simplification is valid if the atmosphere can be considered as optically thin over the considered wavelength interval. The length of the effective light path is dealt with separately using radiative transfer calculations. By expanding Beer-Lambert's law, the various factors that influence the Earth reflectance at the top-of-atmosphere are considered by an equation that includes the effect of all absorbers present over the light path as well as Rayleigh and Mie scattering particles:

$$R(\lambda) = \frac{\pi I(\lambda)}{\mu_0 E_0(\lambda)} = \exp \left(- \int_0^{TOA} (\sum_j \sigma_j(\lambda, s) n_j(s) + \sigma_R(\lambda, s) n_R(s) + \sigma_M(\lambda, s) n_M(s)) ds \right) A(\lambda), \quad 5-2$$

where $I(\lambda)$ and $E_0(\lambda)$ are the measured earthshine backscattered radiance and solar irradiance spectra (respectively in unit of $[\text{W m}^{-2}\text{nm}^{-1}\text{sr}^{-1}]$ and $[\text{W m}^{-2}\text{nm}^{-1}]$). μ_0 is the cosine of the solar zenith angle θ_0 . Note that the formaldehyde retrieval is in practice limited to solar zenith angles smaller than 70° , so that the singularity in the reflectance occurring at θ_0 equal 90° is never met.

The integral is taken along the effective path followed by the photons through the atmosphere. The summation is performed over all relevant species with absorption cross-sections σ_j [$\text{cm}^2\text{molec.}^{-1}$] and densities n_j [molec.cm^{-3}]. σ_R and σ_M are the Rayleigh and Mie extinction cross-sections, while n_R and n_M are the Rayleigh and Mie scattering particle densities, respectively. The densities of the absorbers and scattering particles vary with altitude along the light path. To a lesser extent, it is in principle also the case for the absorption cross-sections because of their dependence on temperature. Furthermore, the factor $A(\lambda)$ describes the spectral dependence of the measured radiance which is a complex function of the scattering efficiency of the Earth/atmosphere system. It also effectively accounts for possible multiplicative errors in the measured reflectance, such as e.g. due to uncorrected instrument degradation effects, uncertainties in the radiometric calibration or possible residual (smooth) polarisation response effects not accounted for in the level 0-1 processing.

2. If it can be assumed that the absorption and scattering cross-sections σ do not depend on altitude, the sum over the individual terms and the integral can be exchanged, allowing the introduction of the slant column density ($N_{s,j} = \int_0^{TOA} n_j(s)ds$) in Equation 5-2:

$$\frac{I(\lambda)}{E_0(\lambda)} = \exp\left(-\left(\sum_j \sigma_j(\lambda) N_{s,j} + \sigma_R(\lambda) N_{s,R} + \sigma_M(\lambda) N_{s,M}\right)\right) A'(\lambda), \quad 5-3$$

where all constants are now incorporated in the parameter $A'(\lambda)$.

This assumption is justified by the fact that the pressure dependence of absorption cross-sections can in most cases be neglected in the UV-Vis region. However, the temperature dependence of absorption cross-sections may be significant (for example that of O_3 and NO_2). This can be corrected in first approximation by introducing correction factors during the AMF calculation (Boersma et al., 2004) or by fitting two absorptions cross-sections at different temperatures assuming linear dependence on temperature (Van Roozendael et al., 2002). The latter approach is applied to the ozone absorption cross-sections during the fit of HCHO.

3. Molecular absorption cross-sections $\sigma_j(\lambda)$ are usually highly structured, while scattering by molecules and particles, as well as reflection at the surface, have broadband spectral dependences that can be approximated by a low order polynomial. The final DOAS equation is obtained by considering the logarithm of (5-3) and including all broadband variations in a single polynomial:

$$\ln \frac{I(\lambda)}{E_0(\lambda)} = - \sum_j \sigma_j(\lambda) N_{s,j} + \sum_p c_p \lambda^p \quad 5-4$$

$$\tau_s^{meas}(\lambda) = \tau_s^{diff}(\lambda, N_{s,j}) + \tau_s^{smooth}(\lambda, c_p), \quad 5-5$$

where the measured optical depth τ_s^{meas} is separated into its highly structured part τ_s^{diff} and its broadband variation τ_s^{smooth} .

Equation (5-4) is a linear equation between the logarithm of the measured quantities (I and E_0), the slant column densities of all relevant absorbers ($N_{s,j}$) and the polynomial coefficients (c_p), at multiple wavelengths. DOAS retrievals consist in solving an over-determined set of linear equations, which can be done by standard methods of linear least squares fit (Platt and Stutz, 2008). The aim of the fitting process is to minimize the chi-square function, i.e. the weighted sum of squares derived from Equation (5-5):

$$X^2 = \sum_{i=1}^k \frac{(\tau_s^{meas}(\lambda_i) - \tau_s^{diff}(\lambda_i, N_{s,j}) - \tau_s^{smooth}(\lambda_i, c_p))^2}{\varepsilon_i^2} \quad 5-6$$

where the summation is made over the individual spectral pixels included in the selected wavelength range (k is the number of spectral pixels in the fitting interval). ε_i is the statistical uncertainty on the measurement at wavelength λ_i . Weighting the residuals by the instrumental uncertainties ε_i is optional. When no measurement uncertainties are used (or no error estimates are available), all uncertainties in Equation (5-6) are set to $\varepsilon_i = 1$, giving all measurement points equal weight in the fit.

Another measure for the goodness of the fit is the so-called root-mean-squared (RMS) differences, which is defined as follows:

$$RMS = \sqrt{\frac{\sum_{i=1}^k (\tau_s^{meas}(\lambda_i) - \tau_s^{diff}(\lambda_i, N_{s,j}) - \tau_s^{smooth}(\lambda_i, c_p))^2}{k}} \quad 5-7$$

In order to optimize the fitting procedure, additional structured spectral effects have to be considered carefully such as the Ring effect (Grainger and Ring, 1962). Furthermore, the linearity of Equation (5-4) may be broken down by instrumental aspects such as small wavelength shifts between I and E_0 , which requires accurate calibration, or possible contamination of measured radiances by spectral straylight, which requires the introduction of an intensity offset parameter.

5.3 Formaldehyde slant column retrieval

Despite the relatively large abundance of formaldehyde in the atmosphere (of the order of 10^{16} molec.cm⁻²) and its well-defined absorption bands, the fitting of HCHO slant columns in earthshine radiances is a challenge because of the low optical density of HCHO compared to other UV-Vis absorbers. As displayed in Figure 5-3, the typical HCHO optical density is one order of magnitude smaller than that of NO₂ and three orders of magnitude smaller than that for O₃. Therefore, the detection of HCHO is limited by the signal to noise ratio of the measured radiance spectra and by possible spectral interferences due to other molecules absorbing in the same fitting interval, mainly ozone and BrO. In general, the correlation between cross-sections decreases if the wavelength interval is extended, but the assumption of a single effective light path defined for the entire wavelength interval may not be fully satisfied, leading to systematic misfit effects that may also introduce biases in the retrieved slant columns.

To optimize DOAS retrieval settings, a trade-off has to be found between these effects and the instrument characteristics. A basic limitation of the classical DOAS technique is the assumption that the atmosphere is optically thin in the wavelength region of interest. At shorter wavelengths, the usable spectral range of DOAS is limited by rapidly increasing Rayleigh scattering and O₃ absorption. The DOAS assumptions start to fail for ozone slant columns larger than 1500 DU (Van Roozendaal et al., 2012). Historically, different wavelength intervals have been selected between 326 and 360 nm for the retrieval of HCHO using previous satellite UV spectrometers (e.g: GOME, Chance et al., 2000; SCIAMACHY, Wittrock et al., 2006, or GOME-2 nm, Vrekoussis et al., 2010). The TEMIS dataset combines HCHO observations from GOME, SCIAMACHY, GOME-2 and OMI measurements retrieved in the same interval (De Smedt et al., 2008; 2012; 2015). The NASA operational and PCA OMI algorithm exploit a larger interval (Kurosu, 2008; González Abad et al., 2015, Li et al., 2015). The latest QA4ECV product uses the largest interval, thanks to the good quality of the OMI level 1 spectra. (De Smedt et al., 2018). A summary of the different wavelength intervals is provided in Table 2.

Table 2: Wavelength intervals used in previous formaldehyde retrieval studies [nm].

	GOME	SCIAMACHY	GOME-2	OMI
Chance et al., 2000	337.5-359			
Wittrock et al., 2006		334-348		
Vrekoussis et al., 2010			337-353	
Kurosu, 2008				327.5-356.5
Hewson et al., 2012			328.5-346	
González Abad et al., 2015				328.5-356.5
TEMIS: De Smedt et al., 2008; 2012; 2015	328.5-346	328.5-346	328.5-346 (BrO in 328.5-359)	328.5-346 (BrO in 328.5-359)
QA4ECV: De Smedt et al., 2018	328.5-359	328.5-359	328.5-359	328.5-359

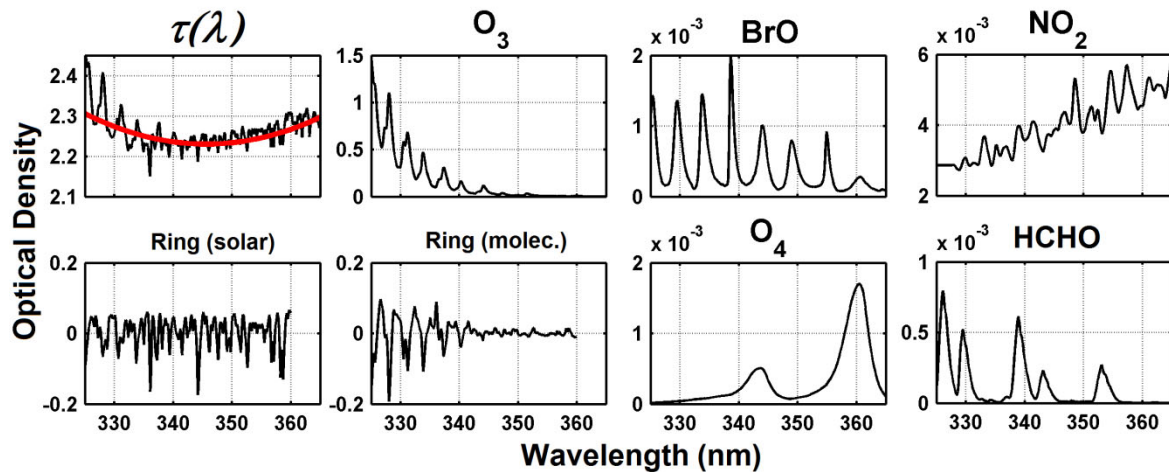


Figure 5-3: Typical optical densities of O_3 , BrO , NO_2 , Ring effect (solar lines and molecular), O_2-O_2 , and $HCHO$ in the near UV. The slant columns have been taken as 10^{20} molec. cm^{-2} for O_3 , 10^{14} molec. cm^{-2} for BrO , 1×10^{16} molec. cm^{-2} for NO_2 , 0.4×10^{42} molec. cm^{-5} for O_2-O_2 , and 10^{16} molec. cm^{-2} for $HCHO$.

5.3.1 Fitting intervals, absorption cross-sections and retrieval settings

As for the TEMIS OMI $HCHO$ product (De Smedt et al., 2015), the TROPOMI L2 $HCHO$ retrieval algorithm includes a two-step DOAS retrieval approach, based on two wavelength intervals:

1. 328.5-359 nm: This interval includes six BrO absorption bands and minimizes the correlation with $HCHO$, allowing a significant reduction of the retrieved slant column noise. Note that this interval includes part of a strong O_4 absorption band around 360 nm, which may introduce geophysical artefacts of $HCHO$ columns over arid soils or high altitude regions.
2. 328.5-346 nm: in a second step, $HCHO$ columns are retrieved in a shorter interval, but using the BrO slant column values determined in the first step. This approach allows to efficiently de-correlate BrO from $HCHO$ absorption while, at the same time, the O_4 -related bias is avoided.

The use of a large fitting interval generally allows for a reduction of the noise on the retrieved slant columns. However, a substantial gain can only be obtained if the level-1b spectra are of sufficiently homogeneous quality over the full spectral range. Indeed, experience with past sensors not equipped with polarization scramblers (e.g. GOME(-2) or SCIAMACHY) has shown that this gain can be partly or totally overruled due to the impact of interfering spectral polarization structures (De Smedt et al., 2012; 2015). During phase E1, recorded TROPOMI spectra appeared to be of excellent quality over the full range of the first fitting interval; therefore, the option of using one single large interval (fitting interval-1) can be applied to TROPOMI, in order to further improve the precision. Results of the retrievals from the two intervals are nevertheless saved in the L2 product files, but the slant columns of the first interval are used to compute the final vertical columns (phase E2 update).

In both intervals, the absorption cross-sections of O_3 at 223K and 243K, NO_2 , BrO and O_4 are included to the fit. The correction for the Ring effect (defined as I_{rrs}/I_{elas} , where I_{rrs} and I_{elas} are the intensities for inelastic (Rotational Raman Scattering; RRS) and elastic scattering processes) is based on the technique outlined by Chance et al. (1997).

Furthermore, in order to better cope with the strong ozone absorption at wavelengths shorter than 336 nm, the method of Puķīte et al. (2010) is implemented. In this method, the variability of the ozone slant column over the fitting window is taken into account. At the first order, the method consists in adding two cross-sections to the fit: $\lambda\sigma_{O_3}$ and $\sigma_{O_3}^2$ (equation 11 in Puķīte et al., 2010), using the O_3 cross-sections at 223K (close to the temperature at ozone maximum in the tropics). It allows a much better treatment of optically thick ozone absorption in the retrieval and therefore to reduce the underestimation of the HCHO slant columns by 50 to 80%, for SZA from 50° to 70°, while keeping the appealing features of DOAS (in terms of performance).

To calculate the reflectance (5-3), the classical satellite-DOAS approach is to use the daily solar irradiance. A more advanced option, is to use daily averaged radiances, selected for each detector row, in the equatorial Pacific (Lat: [-5° 5°], Long: [180° 240°]). The main advantages of this approach are (1) an important reduction of the fit residuals (by up to 40%) mainly due to the cancellation of O_3 absorption and Ring effect present in both spectra; (2) the fitted slant columns are directly corrected for background offsets present in both spectra; (3) possible row-dependent biases (stripes) are greatly reduced by cancellation of small optical mismatches between radiance and irradiance optical channels; and (4) the sensitivity to instrument degradation affecting radiance measurements is reduced because these effects tend to cancel between the analyzed spectra and the references that are used. However, this approach requires the handling of L1 radiances in the reference sector for several days before or around the measurement time. The option of using an equatorial radiance as reference has been activated in the operational processor after the launch of TROPOMI, and is now the algorithm baseline from phase E2.

Table 3 : Summary of DOAS settings used to retrieve HCHO slant columns from TROPOMI spectra.

Fitting interval - 1	328.5-359 nm
Absorption cross-sections	
HCHO	Meller and Moortgat (2000), 298K
Ozone	Serdyuchenko et al. (2013), 223 + 243K
BrO	Fleischmann et al. (2004), 223K
NO ₂	Vandaele et al. (1998), 220K
O ₄ (O ₂ -O ₂)	Thalman et al., (2013), 293K
Fitting interval - 2	328.5-346 nm
Absorption cross-sections	
HCHO	Meller and Moortgat (2000), 298K
Ozone	Serdyuchenko et al. (2013), 223 + 243K
BrO (not fitted)	Fleischmann et al. (2004), 223K
NO ₂	Vandaele et al. (1998), 220K
O ₄ (O ₂ -O ₂)	Thalman et al., (2013), 293K
Ring effect	1 Ring cross-section defined as I_{rrs}/I_{elas} , where I_{rrs} and I_{elas} are the intensities for inelastic (Rotational Raman Scattering; RRS) and elastic scattering processes) is based on the technique outlined by Chance et al. (1997).
Non-linear O₃ absorption effect	2 pseudo-cross sections from the Taylor expansion of the wavelength and the O ₃ optical depth (Puķīte et al., 2010).
Slit function	1 slit function per binned spectrum as a function of wavelength (Pre Flight Model)
Polynomial	5 th order
Intensity offset correction	Linear offset
Reference spectrum (E₀)	Average of radiances, per row, selected in the equatorial Pacific within the last 5 valid days.

5.3.2 Wavelength calibration and convolution to TROPOMI resolution

As described in the TROPOMI instrument description RD14, the spectral registration will differ from ground-pixel to ground-pixel due to thermal variations over the orbit as well as due to inhomogeneous filling of the slit in the flight direction. The Level 1B will contain a spectral assignment, however an accurate fine-tuned spectral calibration based on the Fraunhofer lines needs to be performed by the Level 1-2 algorithms.

Generally speaking, the quality of a DOAS fit critically depends on the accuracy of the alignment between the earthshine radiance spectrum, the reference (solar irradiance) spectrum and the cross sections. The wavelength registration of the reference spectrum can be fine-tuned by means of a calibration procedure making use of the solar Fraunhofer lines. To this end, a reference solar atlas E_s accurate in absolute vacuum wavelength to better than 0.001 nm (Chance and Kurucz, 2010) is degraded to the resolution of the instrument, through convolution by the TROPOMI instrumental slit function (see Figure 5-4).

Using a non-linear least-squares approach, the shift (Δ_i) between the reference solar atlas and the TROPOMI irradiance is determined in a set of equally spaced sub-intervals covering a spectral range large enough to encompass all relevant fitting intervals. The shift is derived according to the following equation:

$$E_0(\lambda) = E_s(\lambda - \Delta_i) \quad 5-8$$

where E_s is the solar spectrum convolved at the resolution of the instrument and Δ_i is the shift in sub-interval i . A polynomial is then fitted through the individual points in order to reconstruct an accurate wavelength calibration $\Delta(\lambda)$ for the complete analysis interval. Note that this approach allows compensating for stretch and shifting errors in the original wavelength assignment.

In the case of TROPOMI, the procedure is complicated by the fact that such calibrations must be performed (and stored) for each separate spectral field on the CCD detector array. Indeed due to the imperfect characteristics of the imaging optics, each row of the TROPOMI instrument must be considered as a separate detector for analysis purposes.

In a subsequent step of the processing, the absorption cross sections of the different trace gases must be convolved with the instrumental slit function. The baseline approach is to use slit functions determined as part of the TROPOMI key data. Slit functions are expected to be delivered for each binned spectrum and as a function of the wavelength. Note that an additional feature of the (prototype) algorithm allows to dynamically fit for an effective slit function of known line shape. This has been used for verification and monitoring purpose during commissioning phase. In brief, the wavelength calibration procedure allows for stretching of slit function left and right wings and the fitted stretch parameters lead to optimized slit functions, which have then be used for all subsequent spectral convolutions. After some iterations, it was shown during the commissioning phase that the optimized slit functions were very close to the latest released TROPOMI instrumental slit function data set.

More specifically, wavelength calibrations are made for each TROPOMI orbit as follows:

1. The irradiances (one for each binned row of the CCD) are calibrated in wavelength over the 325-360 nm wavelength range, using 5 sub-windows.

2. The earthshine radiances are first interpolated on the original L1 irradiance grid. The irradiance calibrated wavelength grid is assigned to those interpolated radiance values.
3. The absorption cross-sections are interpolated (cubic spline interpolation) on the calibrated wavelength grid, prior to the analysis.
4. In the case where averaged radiances are used as reference, an additional step must be performed: the cross-sections are aligned to the reference spectrum by means of shift/stretch values derived from a least-squares fit of the calibrated irradiance towards the averaged reference radiance.
5. During spectral fitting, shift and stretch parameters for the radiance are derived, to align each radiance with cross sections and reference spectrum.

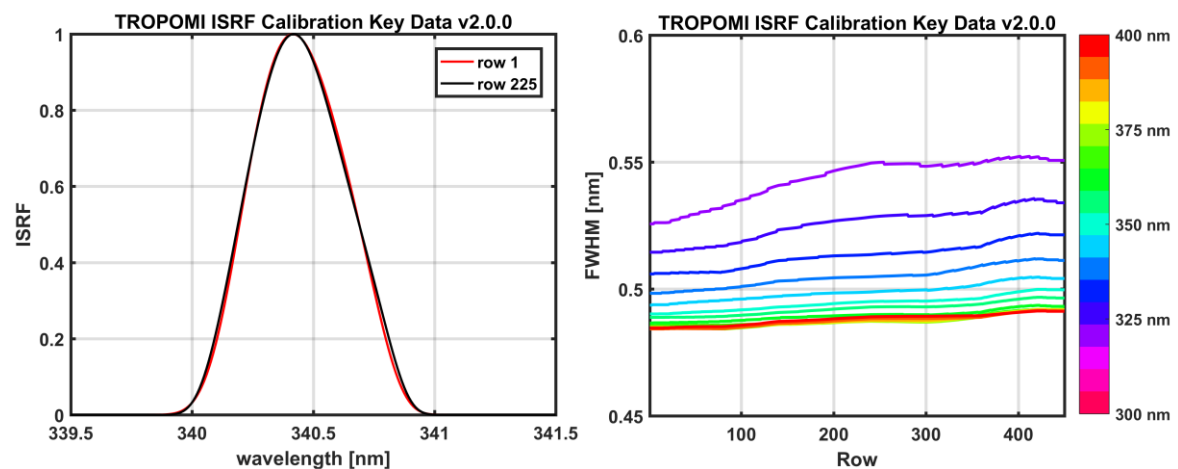


Figure 5-4: Right panel: Examples of TROPOMI slit functions around 340 nm, for row 1 and row 225. Left panel: TROPOMI spectral resolution in channel 3, as a function of the row and the wavelength, derived from the instrument key data ISFR v2.0.0.

5.3.3 Spike removal algorithm

A method to remove individual hot pixels or pixels affected by the South Atlantic Anomaly has been presented for NO₂ retrievals in Richter et al. (2011). Often only a few individual detector pixels are affected and in these cases, it is possible to identify and remove the noisy points from the fit. However, as the amplitude of the distortion is usually only of the order of a few percent or less, it cannot always be found in the highly structured spectra themselves. Higher sensitivity for spikes can be achieved by analysing the residual of the fit where the contribution of the Fraunhofer lines, scattering, and absorption is already removed.

When the residual for a single pixel exceeds the average residual of all pixels by a chosen threshold ratio (the tolerance factor), the pixel is excluded from the analysis, in an iterative process. This procedure is repeated until no further outliers are identified, or until the maximum number of iterations is reached (here fixed to 3). Tests performed with GOME-2 and OMI spectra show that a tolerance factor of 5 improves the HCHO fits. This is especially important to handle the degradation of 2-D detector arrays such as OMI or TROPOMI. However, this improvement of the algorithm has a non-negligible impact on the time of processing (x 1.8). This option is activated in the operational processor.

5.4 Tropospheric air mass factor calculation

In the DOAS approach, an optically thin atmosphere is assumed. The mean optical path of scattered photons can therefore be considered as independent of the wavelength within the relatively small spectral interval selected for the fit and a single wavelength effective air mass factor can be defined as the ratio of the slant column to the vertical column of a particular absorber j :

$$M_j = \frac{N_{s,j}}{N_{v,j}}. \quad 5-9$$

In the troposphere, scattering by air molecules, clouds and aerosols leads to complex altitude-dependent air mass factors. Full multiple scattering calculations are required for the determination of the air mass factors, and the vertical distribution of the absorber has to be assumed *a priori*. For optically thin absorbers, the formulation of Palmer et al. (2001) is conveniently used. It decouples the height-dependent measurement sensitivity from the vertical profile shape of the species of interest, so that the tropospheric AMF (M) can be expressed as the sum of the altitude dependent air mass factors (m_l) weighted by the partial columns (n_{al}) of the *a priori* vertical profile in each vertical layer l , from the surface up to the tropopause index (lt):

$$M = \frac{\sum_{l=1}^{lt} m_l(\lambda, \theta_0, \theta, \varphi, A_s, p_s, f_c, A_{cloud}, p_{cloud}) n_{al}(lat, long, time)}{\sum_{l=1}^{lt} n_{al}(lat, long, time)}, \quad 5-10$$

where A_s is the surface albedo, p_s is the surface pressure, while f_c , A_{cloud} and p_{cloud} are respectively the cloud fraction, cloud albedo and cloud pressure.

The altitude dependent air mass factors represent the sensitivity of the slant column to a change of the partial columns $N_{v,j}$ at a certain level. In a scattering atmosphere, m_l depends on the wavelength, the viewing angles, the surface albedo, and the surface pressure, but not on the partial column amounts or the vertical distribution of the considered absorber (optically thin approximation).

5.4.1 LUT of altitude dependent air mass factors

Generally speaking, m depends on the wavelength, as scattering and absorption processes vary with wavelength. However, the amplitude of the m variation is found to be small (less than 5% for SZA lower than 70°) in the 328.5-359 nm fitting window and a single air mass factor representative for the entire wavelength interval is used (at 340 nm).

Figure 5-5 illustrates the dependency of m with the viewing angles, *i.e.* θ_0 (a), θ (b), and φ (c), and with scene conditions like A_s (d) and p_s for a weakly (e) or highly reflecting surface (f). The decrease of sensitivity in the boundary layer is more important for large solar zenith angles and wide instrumental viewing zenith angles. The relative azimuth angle does have relatively less impact on the measurement sensitivity (note however that aerosols and BRDF effects are not included in those simulations). In the UV, surfaces not covered with snow present albedo lower than 0.1, while snow and clouds generally present albedo larger than 0.8. For a weakly reflecting surface, the sensitivity decreases near the ground because photons are mainly scattered, and scattering can take place at varying altitudes. Larger

values of the surface albedo increase the fraction of reflected compared to scattered photons, increasing measurement sensitivity to tropospheric absorbers near the surface.

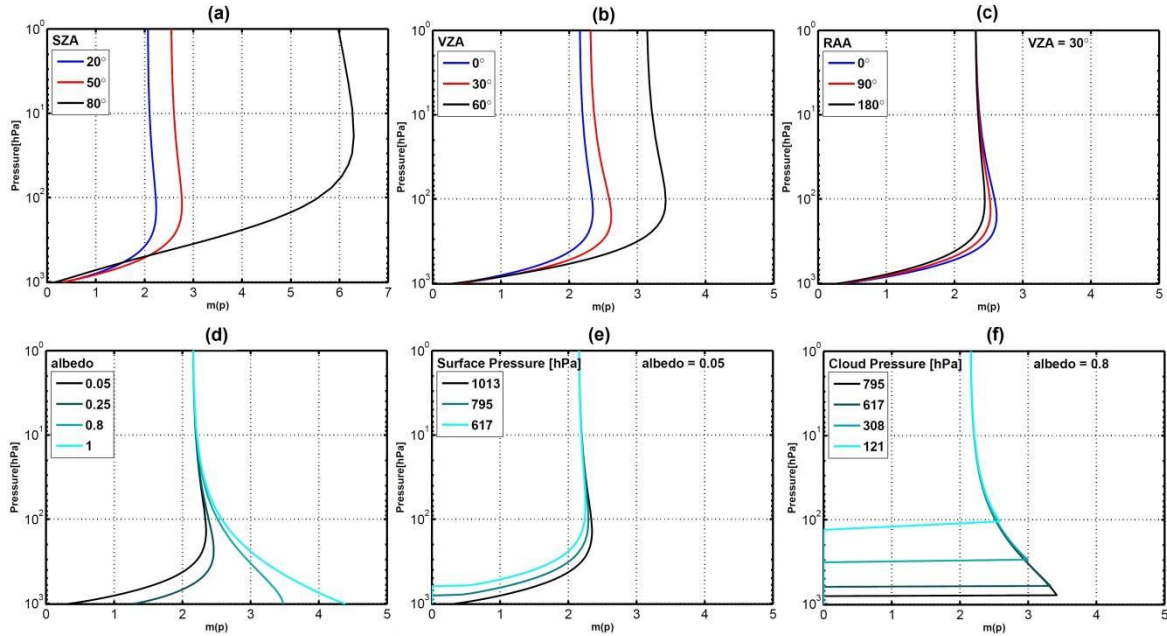


Figure 5-5: Variation of the altitude dependent air mass factor with: (a) solar zenith angle, (b) viewing zenith angle, (c) relative azimuth angle between the sun and the satellite, (d) surface albedo, (e) surface pressure for a weakly reflecting surface, (f) surface pressure for a highly reflecting surface. Unless specified, the parameters chosen for the radiative transfer simulations are: SZA=30°, VZA=0°, RAA=0°, albedo=0.05, surface pressure=1063hPa, $\lambda=340\text{nm}$.

Altitude dependent air mass factors are calculated with the VLIDORT v2.6 radiative transfer model (Spurr, 2008, see B. Description of forward model [URL03]), at 340 nm, using an US standard atmosphere, for a number of representative viewing geometries, surface albedos and surface pressures, and stored in a look-up table. Altitude dependent air mass factors are then interpolated within the lookup table for each particular observation conditions and interpolated vertically on the pressure grid of the a priori profile, defined within the TM5 model (see section 5.4.4). Linear interpolations are performed in $\cos(\theta_0)$, $\cos(\theta)$, relative azimuth angle and surface albedo, while a nearest neighbour interpolation is performed in surface pressure. The parameter values chosen for the look-up table are detailed in Table 4. In particular, the grid of surface pressure is very fine near the ground, in order to minimise interpolation errors caused by the generally low albedo of ground surfaces. Indeed, as illustrated by Figure 5-5 (e) and (f), the variation of the altitude dependent air mass factors is more discontinuous with surface elevation (low reflectivity) than with cloud altitude (high reflectivity). Furthermore, the LUT and model pressures are scaled to their respective surface pressures, in order to avoid extrapolations outside the LUT range.

Table 4: Parameters in the altitude dependent air mass factors lookup table

Parameter name	Nb. of grid points	Grid of values	Symbol
Solar zenith angle [°]	17	0, 10, 20, 30, 40, 45, 50, 55, 60, 65, 70, 72, 74, 76, 78, 80, 85	θ_0
Line of sight zenith angle [°]	10	0, 10, 20, 30, 40, 50, 60, 65, 70, 75	θ
Relative azimuth angle [°]	5	0, 45, 90, 135, 180	φ
Surface albedo	14	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	A_s
Surface pressure [hPa]	17	1063.10, 1037.90, 1013.30, 989.28, 965.83, 920.58, 876.98, 834.99, 795.01, 701.21, 616.60, 540.48, 411.05, 308.00, 226.99, 165.79, 121.11	p_s
Atmospheric pressure [hPa]	64	1056.77, 1044.17, 1031.72, 1019.41, 1007.26, 995.25, 983.38, 971.66, 960.07, 948.62, 937.31, 926.14, 915.09, 904.18, 887.87, 866.35, 845.39, 824.87, 804.88, 785.15, 765.68, 746.70, 728.18, 710.12, 692.31, 674.73, 657.60, 640.90, 624.63, 608.58, 592.75, 577.34, 562.32, 547.70, 522.83, 488.67, 456.36, 425.80, 396.93, 369.66, 343.94, 319.68, 296.84, 275.34, 245.99, 210.49, 179.89, 153.74, 131.40, 104.80, 76.59, 55.98, 40.98, 30.08, 18.73, 8.86, 4.31, 2.18, 1.14, 0.51, 0.14, 0.03, 0.01, 0.001	p_l
Altitude corresponding to the atmospheric pressure, using an US standard atmosphere [km] (for information)	64	-0.35, -0.25, -0.15, -0.05, 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.10, 1.30, 1.50, 1.70, 1.90, 2.10, 2.30, 2.50, 2.70, 2.90, 3.10, 3.30, 3.50, 3.70, 3.90, 4.10, 4.30, 4.50, 4.70, 4.90, 5.25, 5.75, 6.25, 6.75, 7.25, 7.75, 8.25, 8.75, 9.25, 9.75, 10.50, 11.50, 12.50, 13.50, 14.50, 16.00, 18.00, 20.00, 22.00, 24.00, 27.50, 32.50, 37.50, 42.50, 47.50, 55.00, 65.00, 75.00, 85.00, 95.00	z_l

5.4.2 Treatment of partly cloudy scenes

The AMF calculations for TROPOMI use the cloud fraction (f_c), cloud albedo (A_{cloud}) and cloud pressure (p_{cloud}) from the operational cloud retrieval, treating clouds as Lambertian reflectors (OCRA/ROCINN-CRB). The applied cloud correction is based on the independent pixel approximation (Martin et al., 2002 and Boersma et al., 2004), in which a inhomogeneous satellite pixel is considered as made of two independent homogeneous scenes, one completely clear and the other completely cloudy. The intensity measured by the instrument for the entire scene is decomposed into the contributions from the clear-sky and cloudy fractions. Accordingly, for each vertical layer, the altitude dependent air mass factor of a partly cloudy scene is a combination of two air mass factors, calculated respectively for the cloud-free and cloudy fractions of the scene:

$$m_l = (1 - w_c) m_{l_clear}(A_s, p_s) + w_c m_{l_cloud}(A_{cloud}, p_{cloud}) \quad 5-11$$

where m_{l_clear} is the altitude dependent air mass factor for a completely cloud-free pixel, m_{l_cloud} is the altitude dependent air mass factor for a completely cloudy scene, and the cloud radiance fraction w_c is defined as:

$$w_c = \frac{f_c I_{cloud}(A_{cloud}, p_{cloud})}{(1 - f_c) I_{clear}(A_s, p_s) + f_c I_{cloud}(A_{cloud}, p_{cloud})} \quad 5-12$$

I_{clear} and I_{cloud} are respectively the radiance intensities for clear-sky and cloudy scenes whose values are calculated with LIDORT at 340 nm and stored in look-up tables with the same grids as the altitude dependent air mass factors. m_{l_clear} and I_{clear} are evaluated for a surface albedo A_s and a surface pressure p_s , while m_{l_cloud} and I_{cloud} are estimated for a cloud albedo A_{cloud} and at the cloud pressure p_{cloud} . Note that the variations of the cloud albedo are directly related to the cloud optical thickness. Strictly speaking in a Lambertian (reflective) cloud model approach, only thick clouds can be represented. An effective cloud fraction corresponding to an effective cloud albedo of 0.8 ($f_{eff} = f_c \frac{A_c}{0.8}$) can be defined, in order to transform optically thin clouds into equivalent optically thick clouds of reduced extent. In such altitude dependent air mass factor calculations, a single cloud pressure is assumed within a given viewing scene. For low effective cloud fractions (f_{eff} lower than 10%), the cloud top pressure is generally highly unstable and it is therefore reasonable to consider the observation as a clear-sky pixel (i.e. the cloud fraction is set to 0) in order to avoid unnecessary random error propagation through the retrievals. This 10% threshold might be adjusted according to the quality of the cloud product (Loyola et al., 2017).

It should be noted that this formulation of the altitude dependent air mass factor for a partly cloudy pixel implicitly includes a correction for the HCHO column lying below the cloud and therefore not seen by the satellite, the so-called “ghost column”. Indeed, the total AMF calculation as expressed by (5-10) and (5-11) assumes the same a priori vertical profile in both cloudy and clear parts of the pixel and implies an integration of the profile from the top of atmosphere to the ground, for each fraction of the scene. The ghost column information is thus coming from the a priori profiles. For this reason, result analysis should include only observations with moderate cloud fractions (typically f_{eff} lower than 30%).

Figure 5-6 shows three altitude dependent air mass factors calculated respectively for cloud free, half cloudy and totally cloudy pixel, and a cloud albedo of 0.8. For the totally cloudy pixel, the cloud acts as a shield between the satellite and the ground surface and the measurement sensitivity equals zero below the cloud. At the same time, the sensitivity is high above the cloud because of its high reflectivity. The half cloudy altitude dependent air mass factor is a combination of the cloud free and cloudy altitude dependent air mass factors.

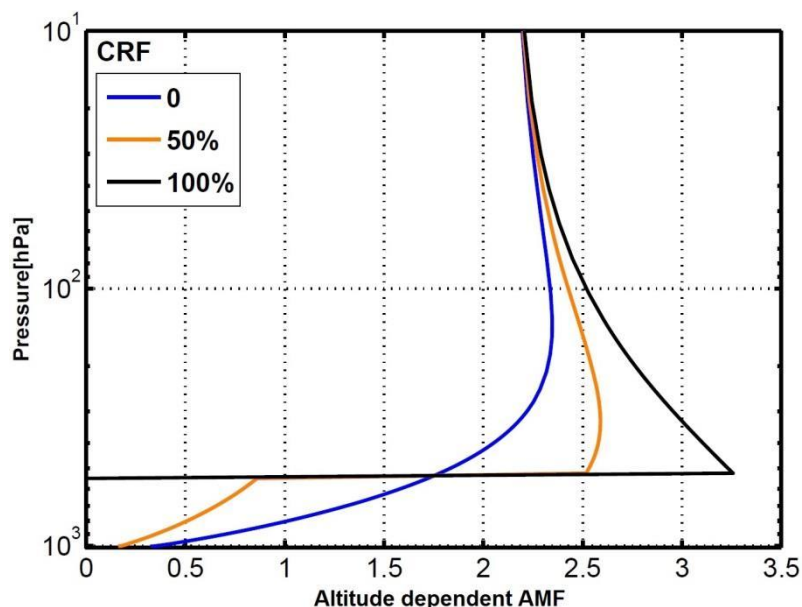


Figure 5-6: Altitude dependent AMF for a cloud free pixel (in blue), a pixel with a cloud radiance fraction of 0.5 (in orange) and a fully cloudy pixel (in grey). The cloud pressure is 540hPa, and the cloud albedo is 0.8. In all cases, we consider a nadir view, a SZA of 30° and a ground albedo of 0.05.

5.4.3 Aerosols

The presence of aerosol in the observed scene may affect the quality of the retrieval. No explicit treatment of aerosols (absorbing or not) is foreseen in the NRT algorithm as there is no general and easy way to treat the aerosols effect on the retrieval. At computing time, the aerosol parameters (extinction profile, single scattering albedo, ...) are unknown. However, the information on the AI [RD09] is included in the L2 files as it gives information to the user on the presence of aerosols and that data should be used and interpreted with care. If, however, reliable information on absorbing aerosol can be obtained from the AI and the aerosol height product, absorbing aerosols might be included in the forward model, in an offline future version of the product.

5.4.4 A priori vertical profile shapes

Formaldehyde concentrations decrease with altitude owing to near-surface sources of short-lived NMVOC precursors, the temperature dependence of CH₄ oxidation, and the altitude dependence of photolysis. The profile shape varies according to local NMHC sources, boundary layer depth, photochemical activity, and other factors.

To resolve this variability in the L2 HCHO product, daily forecasts calculated with the TM5-MP chemical transport model (Huijnen et al., 2010, Williams et al., 2017) will be used to specify the vertical VMR profile shape of the HCHO distribution. TM5-MP will also provide a priori VMR profile shapes for the NO₂, SO₂, and CO retrievals.

TM5-MP is operated with a spatial resolution of 1°x1° in latitude and longitude, and with 34 sigma pressure levels up to 0.1hPa in the vertical direction. TM5-MP uses 3-hourly meteorological fields from the European Centre for Medium Range Weather Forecast (ECMWF) operational model (ERA-Interim reanalysis data for reprocessing, and the operational archive for real time applications and forecasts). These fields include global distributions of wind, temperature, surface pressure, humidity, and (liquid and ice) water content, and precipitation. A more detailed description of the TM5-MP model is given in [URL02] and in the NO₂ ATBD [AD05].

For NRT processing, the daily forecast of the TM5 model (located at KNMI) will be ingested by the UPAS-2 operational processor. For the calculation of the HCHO air mass factors, the profiles are linearly interpolated in space and time, at pixel centre and S5P local overpass time, through a model time step of 30 minutes. To obtain an AMF representative for the troposphere, the integral of the weighting function multiplied by the TM5 profile is performed from the ground to the tropopause.

To reduce the errors associated to topography and the lower spatial resolution of the model compared to the TROPOMI spatial resolution, the a priori profiles need to be rescaled to effective surface elevation of the satellite pixel. Following Zhou et al. (2009) and Boersma et al (2011), the TM5 surface pressure is converted by applying the hypsometric equation and the assumption that temperature changes linearly with height:

$$p_s = p_{TM5} \left(\frac{T_{TM5}}{(T_{TM5} + \Gamma(z_{TM5} - z_s))} \right)^{\frac{g}{R\Gamma}} \quad 5-13$$

Where p_{TM5} and T_{TM5} are the TM5 surface pressure and temperature, $\Gamma = 6.5 \text{Kkm}^{-1}$ the lapse rate, z_{TM5} the TM5 terrain height, and z_s surface elevation for the satellite ground pixel from a digital elevation map at high resolution (see Table 5). The TM5 profile is shifted to start at p_s and scaled so that volume mixing ratios are preserved (see Zhou et al., 2009).

$R=287 \text{ J kg}^{-1} \text{ K}^{-1}$ is the gas constant for dry air, and $g = 9.8 \text{ms}^{-2}$ the gravitational acceleration.

The pressure levels for the a priori HCHO profiles are based on the improved surface pressure level p_s (with $p_i = a_i + b_i p_s$ and a_i, b_i the constants that effectively define the vertical coordinate).

Daily air mass factors obtained using prior information summarized in Table 5, in particular TM5-MP HCHO profiles, are presented in Figure 5-7, in order to give an overview of the tropospheric AMF values and their global regional variations.

Table 5: Prior information datasets used in the air mass factor calculation.

Prior information	Origin of data set	Resolution	Symbol
Surface Albedo	OMI-based monthly minimum LER (update of Kleipool et al., 2008) When available, the TROPOMI-based LER product will be used.	<ul style="list-style-type: none"> month 0.5°x0.5° (lat x long) 342 nm 	A_s
Digital elevation map	GMTED2010 (Danielson et al., 2011)	Average over the ground pixel area.	z_s
Cloud fraction	Operational cloud product based on a Lambertian cloud model (CRB) [RD10].	For each ground pixel.	f_c
Cloud pressure			p_{cloud}
Cloud albedo			A_{cloud}

A priori HCHO profiles	<ul style="list-style-type: none"> • Forecast from TM5 CTM • Backup: last valid day of TM5 profiles 	<ul style="list-style-type: none"> • Daily forecast at overpass time • 1°x1° (lat x long) • 34 sigma pressure levels up to 0.1hPa 	n_a
-------------------------------	---	--	-------

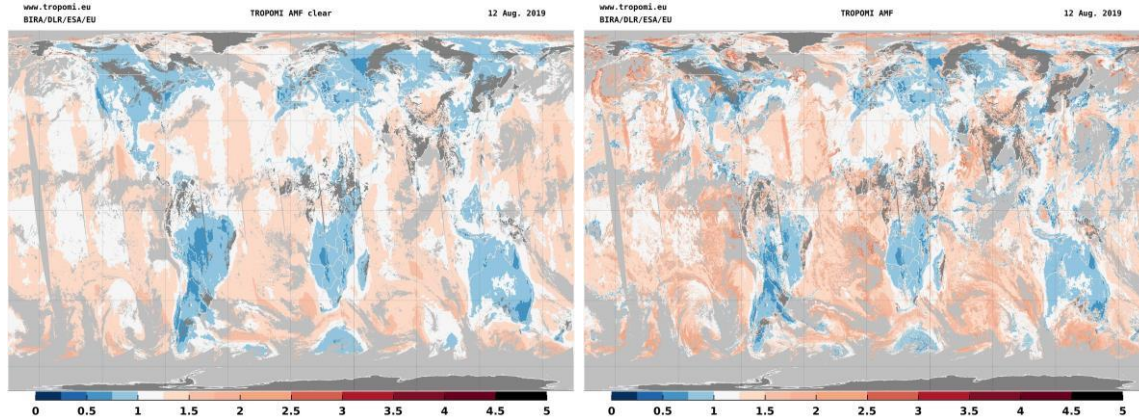


Figure 5-7: Daily map of TROPOMI tropospheric air mass factors at 340 nm. A priori HCHO profiles from high-resolution TM5-MP model runs are used. The left panel shows air mass factors non corrected for clouds (M_{clear}), while on the right panel, the IPA cloud correction is applied for effective cloud fractions f_{eff} larger than 10% (M). Observations with f_{eff} larger than 30% have been filtered out. The operational CRB cloud product is used.

5.5 Across-track and zonal reference sector correction

Residual latitude-dependent biases in the columns, due to unresolved spectral interferences, are known to remain a limiting factor for the retrieval of weak absorbers such as HCHO. Retrieved HCHO slant columns can present large offsets depending on minor changes in the fit settings, and on minor instrumental spectral inaccuracies. Resulting offsets are generally global but also show particular dependencies, mainly with detector row (across-track) and with latitude (along-track). In the case of a 2D-detector array such as OMI or TROPOMI, across-track striping can possibly arise, due to imperfect calibration and different dead/hot pixel masks for the CCD detector regions. Offset corrections are also meant to handle some effects of the time-dependent degradation of the instrument.

A large part of the resulting systematic HCHO slant column uncertainty is reduced by the application of a background correction, which is based on the assumption that the background HCHO column observed over remote oceanic regions (Pacific Ocean) is only due to methane oxidation. The natural background level of HCHO is well estimated from chemistry model simulations of CH₄ oxidation ($N_{v,0,CTM}$). It is ranging from 2 to 4x10¹⁵ molec.cm⁻², depending on the latitude and the season (De Smedt et al., 2008; 2015; González Abad et al., 2015).

For the HCHO retrieval algorithm, we use a 2-steps normalization of the HCHO slant columns:

- Across-track: the mean HCHO slant column is determined for each row in the reference sector around the equator [-5° 5°], [180° 240°]. Data selection is based on the slant column errors from the DOAS fit and on the cloud fraction. Those mean

HCHO values are subtracted from all the slant columns, as a function of the row. The aim is to reduce possible row-dependent offsets. In the case where solar irradiance are used as reference, those offsets can exceed 2×10^{16} molec.cm⁻². They are reduced below 10^{15} molec.cm⁻² by this first step, or when row averaged radiances are used as reference.

- Along-track: the latitudinal dependency of the across-track corrected HCHO N_s is modelled by a polynomial fit through their mean values, all rows combined, in 5° latitude bins in the reference sector ($[-90^\circ 90^\circ]$, $[180^\circ 240^\circ]$). Again, data selection is based on the slant column errors from the DOAS fit and on the cloud fraction.

These two corrections are applied to the global slant columns so that in the reference sector, the mean background corrected slant columns are centered around zero ($\Delta N_s = N_s - N_{s,0}$).

To the corrected slant columns, the background HCHO values from a model have to be added. A latitude-dependent polynomial is fitted daily through 5° latitude bin means of those modelled values in the reference sector. Corresponding values are added to all the columns of the day. Strictly speaking, those background values should be slant columns, derived as the product of air mass factors in the reference sector (M_0) with HCHO vertical columns from the model ($N_{s,0,CTM} = M_{clear,0} N_{v,0,CTM}$) (González Abad et al., 2015). This option requires the storage of the slant columns, the clear air mass factors, and their errors, in a separated database. It is implemented in the operational release of the product. For NRT purpose, the evaluation in the reference sector is made using a moving time window of 5 days. For offline processing, the reference sector correction could still be refined by using daily evaluations. From version 2.0.0 of the operational processor, clear air mass factors ($M_{clear,0}$ instead of M_0) are used to convert the model columns, in order to avoid possible propagation of sun glint effects on the cloud parameters through the AMFs when sun glint appears in the reference sector (see for example right panel of Figure 5-7).

$$N_v = \frac{N_s - N_{s,0}}{M} + N_{v,0} = \frac{\Delta N_s}{M} + \frac{M_{clear,0}}{M} N_{v,0,CTM} \quad 5-14$$

$$= \frac{\Delta N_s - N_{s,0,CTM}}{M}$$

Table 6: 2-steps normalization of the HCHO vertical columns

Correction	Region	Time frame	Column correction	Observation selection
Across-track	Equatorial Pacific Lat: $[-5^\circ 5^\circ]$, Long: $[180^\circ 240^\circ]$	NRT: 5 days moving window	$dN_s(\text{row}) = N_s(\text{row}) - \overline{N_{s,0}(\text{row})}$	$\sigma_{N_s} \leq 3\overline{\sigma_{N_s}}$ $f_c \leq 0.4$
Zonal Along-track	Pacific Lat: $[-90^\circ 90^\circ]$, Long: $[180^\circ 240^\circ]$	Offline: Daily correction	$\Delta N_s(\text{lat}) = dN_s(\text{lat}) - \overline{dN_{s,0}(\text{lat})}$ $N_{s,0,CTM}(\text{lat}) = \overline{M_{clear,0}(\text{lat}) N_{v,0,CTM}(\text{lat})}$ $N_{s,0,CTM}(\text{lat}) = \overline{M_{clear,0}(\text{lat}) N_{v,0,CTM}(\text{lat})}$	$\overline{dN_{s,0}(\text{lat})} \leq 5e16$

6 Input-Output file description

6.1 S5P formaldehyde product description and size

In addition to the main product results, such as HCHO slant column, vertical column and air mass factor, the level 2 data files contain several additional parameters and diagnostic information. Table 6 gives a minimum set of data fields that are present in the Level 2 data. A 1-orbit HCHO column Level 2 file is of about 560 MB (up to 720 MB from 6 August 2019 on), including averaging kernels and a priori profile information. More details about the level 2 data format are provided in the Product User Manual [RD16].

Table 6: List of output fields in the TROPOMI HCHO product. scanline x ground_pixel mean the number of pixels in an orbit along track and across track, respectively.

Symbol	Unit*	Variable name	Number of entries
N_v	mol.m ⁻²	formaldehyde_tropospheric_vertical_column	scanline x ground_pixel
$N_{s,l}$	mol.m ⁻²	fitted_slant_columns_win1	scanline x ground_pixel x number_of_slant_columns_win1
$\sigma_{N,s,1,rand}$	mol.m ⁻²	fitted_slant_columns_precision_win1	scanline x ground_pixel x number_of_slant_columns_win1
$N_s - N_{s,0}$	mol.m ⁻²	formaldehyde_slant_column_corrected	scanline x ground_pixel
$N_{v,0}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_correction	scanline x ground_pixel
M	n.u.	formaldehyde_tropospheric_air_mass_factor	scanline x ground_pixel
M_{clear}	n.u.	formaldehyde_clear_air_mass_factor	scanline x ground_pixel
f_c	n.u.	cloud_fraction_crb	scanline x ground_pixel
w_c	n.u.	cloud_fraction_intensity_weighted	scanline x ground_pixel
p_{cloud}	Pa	cloud_pressure_crb	scanline x ground_pixel
A_{cloud}	n.u.	cloud_albedo_crb	scanline x ground_pixel
A_s	n.u.	surface_albedo	scanline x ground_pixel
z_s	m	surface_altitude	scanline x ground_pixel
$\sigma_{N,v,rand}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_precision	scanline x ground_pixel
$\sigma_{N,v,syst}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_trueness	scanline x ground_pixel
$\sigma_{N,v,syst,kernel}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_kernel_trueness	scanline x ground_pixel
$\sigma_{M,syst}$	n.u.	formaldehyde_tropospheric_air_mass_factor_trueness	scanline x ground_pixel
$\sigma_{M,syst,kernel}$	n.u.	formaldehyde_tropospheric_air_mass_factor_kernel_trueness	scanline x ground_pixel
$\sigma_{M,rand}$	n.u.	formaldehyde_tropospheric_air_mass_factor_precision	scanline x ground_pixel
$\sigma_{N,s,0}$	mol.m ⁻²	formaldehyde_slant_column_corrected_trueness	scanline x ground_pixel

Symbol	Unit*	Variable name	Number of entries
A	n.u.	averaging_kernel	layer x scanline x ground_pixel
n_a	vmr	formaldehyde_profile_apriori	layer x scanline x ground_pixel
p_s	Pa	surface_pressure	scanline x ground_pixel
a_l	Pa	tm5_constant_a	layer
b_l	n.u.	tm5_constant_b	layer
l_t	n.u.	tm5_tropopause_layer_index	scanline x ground_pixel

6.2 Auxiliary information needs

The auxiliary data needs are separated into static and dynamic needs.

6.2.1 Dynamic information

See also the document “Sentinel-5 Precursor Level 2 UPAS Processor Input / Output Definition Document” RD15.

Table 7: Dynamic auxiliary information needs in the HCHO retrieval algorithm

Name/Data	Symbol	Unit	Source	Pre-process needs	Backup if not available
S5P level 1B Earth radiance	I	$\text{mol.s}^{-1}\text{m}^{-2}\text{nm}^{-1}\text{sr}^{-1}$	S5P L1b product	-	No retrieval
S5P level 1B sun irradiance	E_0	$\text{mol.s}^{-1}\text{m}^{-2}\text{nm}^{-1}$	S5P L1b product	Wavelength recalibrated using a high-resolution reference solar spectrum	Use previous measurement
Cloud fraction	f_c	n.u.	S5P operational cloud product based on a Lambertian cloud model [RD10]. UPAS processor.	-	No retrieval
Cloud top pressure	p_{cloud}	Pa			
Cloud top albedo	A_{cloud}	n.u.			
HCHO profile	n_a	vmr	Daily forecast from TM5 CTM run at KNMI.	-	Use last valid day of TM5 CTM
Aerosol index	AI	n.u.	S5P operational AAI product [RD09]. Used for flagging. KNMI processor. In offline product only.	-	Missing information flag.

Snow-ice flag		n.u.	Near real-time global Ice and Snow Extent (NISE) data from NASA.	-	Use snow/ice climatology.
---------------	--	------	--	---	---------------------------

6.2.2 Static information

See also the document “S5P/TROPOMI Static input for Level 2 processors” RD13.

Table 8: Static auxiliary information needs in the HCHO retrieval algorithm

Name/Data	Symbol	Unit	Source	Pre-process needs	Comments
Absorption cross-sections					
HCHO	σ_{HCHO}	cm ² molec. ⁻¹	Meller and Moortgat (2000), 298K	Convolution at the instrumental spectral resolution using the provided slit function	-
Ozone	σ_{03223}	cm ² molec. ⁻¹	Serdyuchenko et al. (2013), 223 + 243K		
	σ_{03243}				
BrO	σ_{BrO}	cm ² molec. ⁻¹	Fleischmann et al. (2004), 223K		
NO ₂	σ_{NO2}	cm ² molec. ⁻¹	Vandaele et al. (1998), 220K		
O ₄ (O ₂ -O ₂)	σ_{O4}	cm ⁵ molec. ⁻²	Thalman et al., 2013, 293K		
High resolution reference solar spectrum	E_s	W m ⁻² nm ⁻¹	Chance and Kurucz, 2010	-	-
Ring effect	$\sigma_{ringev1}$ $\sigma_{ringev2}$	cm ² molec. ⁻¹	2 Ring cross-sections generated internally.	A high-resolution reference solar spectrum and the instrument slit function are needed to generate the data set.	Calculated in an ozone containing atmosphere for low and high SZA, using LIDORT_RRS (Spurr et al., 2008b) and a standard atmosphere (Camelot European Pollution atmospheric profile).
Non-linear O ₃ absorption effect	σ_{o3l} σ_{o3sq}	nm.cm ² molec. ⁻¹ cm ⁴ molec. ⁻²	2 pseudo-cross sections generated internally.	The O ₃ cross-section at 223 K is needed.	Calculated from the Taylor expansion of the wavelength and the O ₃ optical depth (Puķīte et al., 2010).
Instrument slit function	SF	n.u.	Slit Function per wavelength/detector.	-	Values between 325 and 360nm.
Surface Albedo	A_s	n.u.	OMI-based monthly minimum	-	Values at 342nm.

			LER (Kleipool et al., 2008)		
Digital elevation map	z_s	m	GMTED2010 (Danielson et al., 2011)	-	Average over the ground pixel area
HCHO profile	n_a	vmr	Last day of valid TM5 profiles.	-	Used for the reference sector correction and in case of break-up in the NRT processing chain.
Look-up table of altitude-resolved AMFs	m	n.u.	Calculated internally with the LIDORTv 3.3 RTM (Spurr et al., 2008a).	-	-

7 Error analyses

7.1 Error characterization

The total uncertainty (trueness and precision) on the HCHO column products is composed of many sources of error. In part, those are related to the measuring instrument, such as uncertainties due to noise or knowledge of the slit function. In our DOAS-type algorithm, those instrumental errors propagate into the uncertainty on the slant column. Other types of error can be considered as model errors and are related to the representation of the observation physical properties that are not measured. Examples of model errors are uncertainties on the trace gas absorption cross-sections, the treatment of clouds and uncertainties of the a priori profile. Model errors can affect the slant column results or the air mass factor.

A formulation of the error can be derived analytically by error propagation, starting from the equation of the vertical column (5-1) which directly results from the different retrieval steps. As the main algorithm steps are performed independently, they are assumed to be uncorrelated. If we further assume normal probability distributions, the total error on the tropospheric vertical column can be expressed as (Boersma et al., 2004, De Smedt et al., 2008):

$$\sigma_{N,v}^2 = \frac{1}{M^2} \left(\sigma_{N,s}^2 + \frac{(\Delta N_s + M_0 N_{v,0,CTM})^2}{M^2} \sigma_M^2 + \sigma_{N,s,0}^2 + N_{v,0,CTM}^2 \sigma_{M,0}^2 + M_0^2 \sigma_{N,v,0,CTM}^2 \right) \quad 7-1$$

where $\sigma_{N,s}$, σ_M , $\sigma_{N,s,0}$, $\sigma_{M,0}$ and $\sigma_{N,v,0,CTM}$ are respectively the errors on the slant column, the air mass factor, and the slant column correction, the air mass factor, and the model vertical column in the reference sector (indicated by suffix 0). For each of these three categories, several sources of uncertainties and, where possible, their estimated size are respectively summarized in Table 10, Table 11 and Table 12.

A difficulty in the presented error formulation comes from the fact that it assumes the different error sources of the algorithm to be independent and uncorrelated, which is not strictly valid. For example, the background correction is designed to overcome systematic features/deficiencies of the slant column fitting and these two steps cannot be considered as independent. Hence, summing up all the corresponding error estimates would lead to overestimated error bars.

Another important point to note is that, in principle, systematic and random components of each given error source should be discriminated. If so, when deriving the error on the HCHO vertical column, each of the components of (7-1) can be written as:

$$\sigma_k^2 = \frac{\sigma_{k,rand}^2}{N} + \sigma_{k,syst}^2 \quad 7-2$$

where N is the number of ground pixels considered when averaging the observations. It is however very complicated to separate them in practice (see for example Boersma et al., 2004). Our approach is to qualify an error contribution as “random” if its impact on the column tends to average out in space and time. Any other error contribution will be qualified

as “systematic”. This pragmatic classification is based on sensitivity tests performed with GOME-2 and OMI HCHO retrievals.

The error analysis is complemented by the total column averaging kernel (AK) formulation for DOAS-type retrievals (Eskes and Boersma, 2003):

$$AK(p) = \frac{m(p)}{M} \quad 7-3$$

Where m and M are respectively the altitude-resolved air mass factor, and the column air mass factor. Column averaging kernels are particularly useful when comparing measured columns with e.g. model simulations or correlative validation data sets, because they allow removing the effect of the a priori profile shape information used in the retrieval (see C: Averaging Kernel, Boersma et al., 2004).

At the end of this section, Figure 7-2 and Figure 7-3 present respectively the precision (random error uncertainty) and the accuracy (systematic error uncertainty) of the TROPOMI HCHO vertical columns. Precision and trueness of the HCHO product are discussed along with the user requirements that have been summarised in the Introduction to formaldehyde retrieval (4).

7.2 Errors on the slant columns

Error sources that contribute to the total uncertainty on the slant column originate both from instrument characteristics and from uncertainties in the DOAS slant column fitting procedure itself.

7.2.1 Random errors on slant columns

The retrieval noise for individual observations is limited by the SNR of the spectrometer measurements. A good estimate of the random error variance of the reflectance (which results from the combined noise of radiance and reference spectra) is given by the reduced χ^2 of the fit, which is defined as the sum of squares divided by the number of degrees of freedom in the fit. The covariance matrix (Σ) of the linear least squares parameter estimate is then:

$$\Sigma = \frac{\chi^2}{(k-n)} (A^T A)^{-1} \quad 7-4$$

where k is the number of spectral pixels in the fitting interval, n is the number of parameters to fit and the matrix $A(j \times k)$ is formed by the absorption cross-sections. For each absorber j , the value $\sigma_{N_s,j}$ is usually called the slant column error (SCE or $\sigma_{N_s,rand}$).

$$\sigma_{N_s,j}^2 = \frac{\chi^2}{(k-n)} (A^T A)^{-1}_{j,j} \quad 7-5$$

For individual satellite ground pixels, the random error on the slant columns is the most important source of error on the total vertical column. It can be reduced by averaging the observations, but of course to the expense of a loss in time and/or spatial resolution.

7.2.2 Systematic errors on the slant columns

Equation (7-4) does not take into account systematic errors, that are mainly dominated by slit function and wavelength calibration uncertainties, absorption cross-section uncertainties, by interferences with other species (O_3 , BrO and O_4), or by stray light corrections. Those contributions to the slant column errors are estimated from sensitivity tests (see Table 9). The choice of the retrieval interval can have a significant impact on the retrieved HCHO slant columns. However, related zonal systematic offsets are corrected by the reference sector correction (see also the Errors on the reference sector correction). Localised fitting interval effects are estimated to account for a remaining error of 10%.

Non-linear effects due to O_3 absorption are to a large extent accounted for using the Taylor expansion of the O_3 optical depth (Pukite et al., 2010). Again, remaining systematic offsets and zonal biases are removed using the reference sector correction; hence residual systematic features are believed to be small.

All effects summed in quadrature, the various contributions are estimated to account for a systematic error uncertainty of 20% of the background-corrected slant column:

$$(\sigma_{N_s, syst} = 0.2 * \Delta N_s).$$

7-6

Table 9: Summary of the different error sources considered in the HCHO slant column error budget.

Error source	Type	Parameter uncertainty	Averaged uncertainty on HCHO SCD	Evaluation method - reference
Measurement noise	random	S/N=800-1000	$6 \times 10^{16} \text{ molec.cm}^{-2}$	Standard deviation of tropomi HCHO slant columns in remote regions.
HCHO	systematic	Based on alternative cross-section datasets, offset and polynomial orders.	9%	Sensitivity tests using GOME-2 and OMI data. De Smedt et al., 2008; De Smedt et al., 2011; Hewson et al., 2013 Pinardi et al., 2013
O_3			5%	
BrO			5%	
NO_2			3%	
O_4			2%	
Ring			5%	
Offset order			7%	
Polynomial order			7%	
Instrumental slit function and wavelength calibration	systematic	Based on alternative calibrations	10%	Sensitivity tests using GOME-2 and OMI data.
Wavelength interval	systematic	Based on alternative wavelength intervals	10%	Sensitivity tests using GOME-2 and OMI data. Hewson et al., 2013
Temperature dependence of the HCHO XS	systematic	0.05%/°K	2%	Sensitivity tests based on Meller and Moorgat (2000)

7.3 Errors on air mass factors

The uncertainties on the air mass factor depend on input parameter uncertainties and on the sensitivity of the air mass factor to each of them. This contribution is broken down into the squared sum (Boersma et al., 2004, De Smedt et al., 2008):

$$\sigma_M^2 = \left(\frac{\partial M}{\partial A_s} \cdot \sigma_{A_s} \right)^2 + \left(\frac{\partial M}{\partial f_c} \cdot \sigma_{f_c} \right)^2 + \left(\frac{\partial M}{\partial p_{cloud}} \cdot \sigma_{p_{cloud}} \right)^2 + \left(\frac{\partial M}{\partial s} \cdot \sigma_s \right)^2 + (0.15M)^2 \quad 7-7$$

The contribution of each parameter to the total air mass factor uncertainty depends on the observation conditions. The air mass factor sensitivities ($\frac{\partial M}{\partial parameter}$), i.e. the air mass factor derivatives with respect to the different input parameters, can be derived for any particular condition of observation using the altitude-dependent AMF LUT, created with LIDORTv3.3, and using the model profile shapes. In practice, a LUT of AMF sensitivities has been created using coarser grids than the AMF LUT, and one parameter describing the shape of the profile: the profile height, i.e. the altitude (pressure) below which resides 75% of the integrated HCHO profile. $\frac{\partial M}{\partial s}$ is approached by $\frac{\partial M}{\partial s_h}$ where s_h is half of the profile height. Relatively small variations of this parameter have a strong impact on the total air mass factors, because altitude-resolved air mass factors decrease quickly in the lower troposphere, where the HCHO profiles peak (see 7.3.3).

The uncertainties $\sigma_{A_s}, \sigma_{f_c}, \sigma_{p_{cloud}}, \sigma_{s_h}$ are typical uncertainties on the surface albedo, cloud fraction, cloud top pressure and profile shape, respectively. They are estimated from the literature or derived from comparisons with independent data (see Table 11). Together with the sensitivity coefficients, these give the first four contributions on the right of equation (7-7). The fifth term on the right of equation (7-7) represents the uncertainty contribution due to possible errors in the AMF model itself (Lorente et al., 2017). We estimate this contribution to 15% of the air mass factor.

Estimates of the air mass factor errors and of their impact on the vertical columns are listed in Table 11. They are based on the application of equation 7-7 to HCHO columns retrieved from OMI measurements. In expression 7-7, the impact of possible correlations between errors on parameters is not considered, like for example the surface albedo and the cloud top pressure. We implicitly make the assumption that such errors are random in their relative behaviour. Note also that errors on the solar angles, the viewing angles and the surface pressure are supposed to be negligible, which is not totally true in practice, since equation 5-13 does not yield the true surface pressure but only a good approximation.

Table 10: Number of grid points considered for the altitude dependent air mass factors and the air mass factor sensitivity lookup tables

Parameter name	Nb. of grid points in AMF LUT	Nb. of grid points in AMF sensitivity LUT	Symbol
Solar zenith angle [°]	17	6	θ_0
Line of sight zenith angle [°]	10	5	θ

Relative azimuth angle [°]	5	1	φ
Surface albedo	14	6	A_s
Surface pressure [hPa]	17	5	p_s
Profile shape height	-	7	s_h
Cloud fraction	-	6	f_c
Cloud pressure	-	10	p_{cloud}

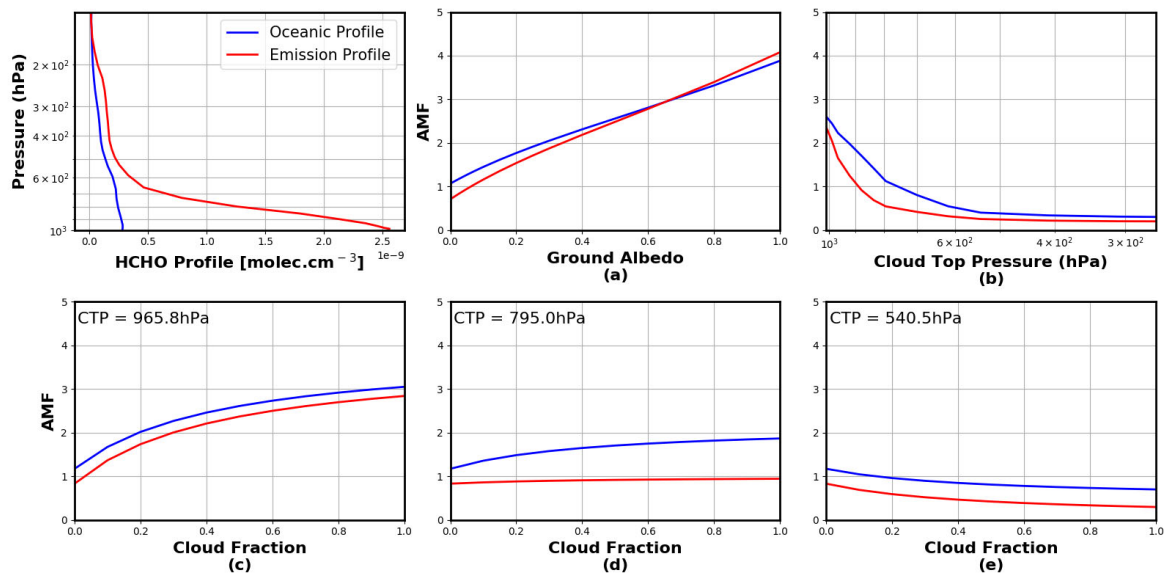


Figure 7-1: First panel: TM5-MP HCHO profiles extracted in June over the equatorial Pacific ocean (blue) and over Beijing (red). Those profiles have been used to calculate the tropospheric air mass factors shown in the panels a to e, representing the AMF dependence on (a) the surface albedo, (b) the cloud altitude, (c), (d), (e) the cloud fraction. In all cases, we consider a nadir view and a solar zenith angle of 30°. In (a) the pixel is cloud free, in (b) the albedo is 0.02 and the effective cloud fraction is 0.5, in (c), (d), (e) the ground albedo is 0.02 and the cloud pressure is respectively 966, 795 and 540 hPa.

7.3.1 Surface albedo

A reasonable uncertainty on the albedo is of 0.02 (Kleipool et al., 2008). This translates to an uncertainty on the air mass factor using the slope of the air mass factor as a function of the albedo and can be evaluated for each satellite pixel (equation 7-7). As an illustration, (a) shows the air mass factor dependence on the ground albedo for two typical HCHO profile shapes (in blue: remote profile, in red: emission profile). The AMF sensitivity (the slope), is almost constant with albedo, being only slightly higher for low albedo values. As expected, the AMF sensitivity to albedo is higher for an emission profile peaking near the surface than for a background profile more spread in altitude.

More substantial errors can be introduced if the real albedo differs considerably from what is expected, for example in the case of the sudden snowfall or ice cover. Snow/ice cover maps will therefore be used for flagging such cases.

7.3.2 Clouds and aerosols

An uncertainty on the cloud fraction of 0.05 is considered, while an uncertainty on the cloud top pressure of 50hPa is taken. (b) shows the air mass factor variation with cloud pressure. The AMF is very sensitive to the cloud altitude (the slope is steepest) when the cloud height is located below or at the altitude of the formaldehyde peak. For higher clouds, the sensitivity of the air mass factor to any change in cloud altitude is very weak. As illustrated in (c), (d) and (e), for which a cloud top pressure of 966, 795 and 540 hPa is respectively considered, the sensitivity to the cloud fraction is mostly significant when the cloud lies below the HCHO layer.

The effect of aerosols on the air mass factors is not explicitly considered in the HCHO retrieval algorithm. To a large extent, however, the effect of the non-absorbing part of the aerosol extinction is implicitly included in the cloud correction (Boersma et al., 2004). Indeed, in the presence of aerosols, the cloud detection algorithm is expected to overestimate the cloud fraction. Since non-absorbing aerosols and clouds have similar effects on the radiation in the UV-visible range, the omission of aerosols is partly compensated by the overestimation of the cloud fraction, and the resulting error on air mass factor is small, typically below 15% (Boersma et al., 2004; Millet et al., 2006). In some cases however, the effect of clouds and aerosols will be different. For example, when the cloud height is significantly above the aerosol layer, clouds will have a shielding effect while the aerosol amplifies the signal through multiple scattering. This will result in an underestimation of the AMF. Absorbing aerosols have also a different effect on the air mass factors, since they tend to decrease the sensitivity to HCHO concentration. In this case, the resulting error on the air mass factor can be as high as 30% (Palmer et al., 2001; Martin et al., 2002). This may, for example, affect significantly the derivation of HCHO columns in regions dominated by biomass burning as well as over heavily industrialized regions. In the TROPOMI HCHO product, observations with an elevated absorbing aerosol index will be flagged, to be used with caution.

Based on sensitivity studies, errors related to cloud and aerosol corrections are considered as random, because their impact on regional/monthly averaged HCHO columns is very limited (less than 10%, when a cloud filtering of 30% is applied on the cloud effective fractions). However, those effects cannot be neglected when looking at daily observations at a finer spatial scale (Leitao et al., 2010; Gonzi et al., 2011; De Smedt et al., 2012; Barkley et al., 2012).

7.3.3 Profile shape

Errors related to a priori profile shape uncertainties are considered as systematic because they do not average out in space or time. For this reason, this contribution to the total AMF error is the largest when considering monthly averaged observations (see Table 11). This is supported by validation results using MAX-DOAS profiles measured around Beijing (see also the Validation report [RD17]).

TM5 is the model of choice for its high spatial resolution, its forecast mode and for consistency across the different air quality trace gas retrievals. However, taking into account

the averaging kernels allows to remove from the comparison the error related to the a priori profiles, when validating the results against other modelled or measured profiles (see C: Averaging Kernel). This is why the systematic error uncertainty on the AMF and on the vertical columns are also provided without this contribution ($\sigma_{M,syst, kernel}$ and $\sigma_{N,v,syst, kernel}$).

7.3.4 AMF wavelength dependence, model atmosphere and LUT interpolation errors

Radiative transfer simulations show that AMF calculations at 340 nm provide representative results for the entire 328.5-346 nm retrieval interval (variations are generally smaller than 5% over the entire wavelength interval). For solar zenith angles larger than 70°, ozone absorption leads to tropospheric air mass factor variations up to 10%. A limit for solar zenith angles smaller than 70° will therefore be applied to the quality flag of the vertical columns.

Errors related to uncertainties in the atmospheric profiles used as input to LIDORT for the altitude-dependent air mass factors LUT calculations (O₃ profile, standard atmosphere pressure and temperature profiles) are considered to be small in the case of tropospheric formaldehyde retrievals.

Errors related to interpolations within the grids of the LUT are minimized by a careful selection of the grids for each parameter, and by refining the table where it is needed, *i.e.* where the variability of the AMF is the largest. In particular, the altitude grid of the lowest layers is known to have the largest impact on the tropospheric AMF sensitivity. These errors are therefore expected to be small (less than 10%).

Table 11: Summary of the different error sources considered in the air mass factor error budget.

Error source	Type	Parameter uncertainty	Averaged uncertainty on HCHO VCD	Evaluation method - reference
Surface albedo	systematic	0.02	10-20%	Equation 7-7
Cloud fraction	random	0.05	0-15%	Equation 7-7 Sensitivity tests on OMI data.
Cloud height	random	50hPa	10-20%	
Profile shape height	systematic	50hPa	20-60%	Validation of model profiles Sensitivity tests using MAXDOAS profiles.
AMF wavelength dependency	systematic	15%	5-10%	RTM simulations
LUT interp. errors	systematic		<5%	
Model atmosphere	systematic		<5%	
Cloud model/cloud correction	random	5 %	<10% on monthly averaged data	

7.4 Errors on the reference sector correction

$$\sigma_{N,v,0}^2 = \frac{1}{M^2} (\sigma_{N,s,0}^2 + N_{v,0,CTM}^2 \sigma_{M,0}^2 + M_0^2 \sigma_{N,v,0,CTM}^2) \quad 7-8$$

This error includes contributions from the model background vertical column, from the error on the air mass factor in the reference sector, and from the amplitude of the normalization applied to the HCHO columns. As mentioned in 7.2, we consider that $\sigma_{N,s,0}$ is taken into account in Equation (7-6). From sensitivity studies with previous satellite retrievals, $\sigma_{N,s,0}$ is taken proportional (20%) to the amplitude of the normalization that needs to be applied to the HCHO columns ($N_{s,0}$). It ranges from 0 (if no normalisation was needed) to 4×10^{15} molec.cm⁻² (for larger spectral interference effects or instrumental degradation effects). The uncertainty on the air mass factor in the reference sector $\sigma_{M,0}$ is calculated as in Equation (7-7) and saved during the background correction step. Uncertainty on the model background has been estimated from the monthly averaged differences between two different CTM simulations in the reference sector: IMAGES (Stavrakou et al., 2009a) and TM5 (Huijnen et al., 2010). The differences range between 0.5 and 1.5×10^{15} molec.cm⁻².

Table 12: Estimated errors on the reference sector correction.

Error source	Type	Uncertainty on HCHO VCD	Evaluation method – reference
Model background	systematic	0.5 and 1.5×10^{15} molec.cm ⁻²	Difference between IMAGES and TM model
Amplitude of the column normalisation	systematic	0 to 4×10^{15} molec.cm ⁻²	Sensitivity tests using GOME-2 and OMI data.

7.5 HCHO error estimates and product requirements

From equations 7-4, 7-7 and 7-8, the total error on the vertical column can be calculated for each satellite observation. Results are discussed for different NMVOC emission regions, separately for the random and systematic parts of the total error, considered respectively as the precision and the trueness of the HCHO product.

7.5.1 Precision

When considering individual pixels, the total error is dominated by the random error on the slant columns. Our simulations and tests on real satellite measurements show that the precision by which the HCHO can be measured is well defined by the instrument signal-to-noise level. For the actual TROPOMI SNR level, that is better than the requirements in the UV band 3 (RD14), the observed precision of single-pixel slant column measurements is 6×10^{15} molec.cm⁻² ([RD07], De Smedt et al., 2018, Vigouroux et al., 2020). This slant column precision is 30% better than the precision obtained with OMI HCHO retrievals (De Smedt et al., 2015), but with a ground pixel size of 7×3.5 km² (5.5×3.5 km² from 6 August 2019).

The precision of the vertical columns is dependent on the air mass factors, and therefore on the observation conditions and on the cloud statistics. Figure 7-2 shows the precision of the TROPOMI HCHO vertical columns, based on the provided individual pixel precision ($\sigma_{N,v,rand}$), which is then divided by the square root of the number of measurements taken in the mean. Results are shown in several regions, and at different spatial and temporal scales

(from individual pixels to monthly averaged column in $20 \times 20 \text{ km}^2$ grids). The product requirements for HCHO measurements state a precision of $1.3 \times 10^{15} \text{ molec.cm}^{-2}$. This particular requirement is not possible to be achieved with TROPOMI individual observations (or with any other existing satellite UV-Vis instrument). The precision can however be increased by averaging the satellite observations in space or in time. As represented on Figure 7-2, the requirement can be approached using daily observations at the spatial resolution of $20 \times 20 \text{ km}^2$ or using monthly averaged columns at the TROPOMI resolution. The precision can be brought below $1 \times 10^{15} \text{ molec.cm}^{-2}$ if a spatial resolution of $20 \times 20 \text{ km}^2$ is considered for monthly averaged columns.

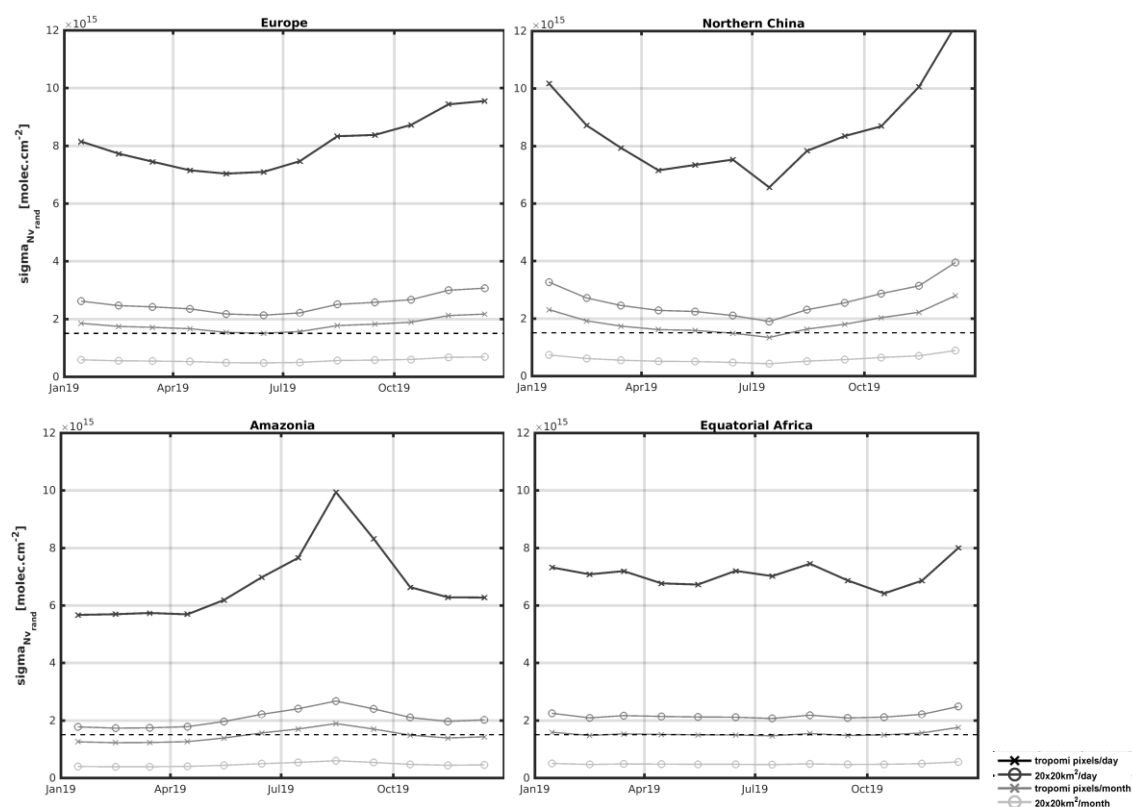


Figure 7-2: Precision on the TROPOMI HCHO columns, in several NMVOC emission regions, and at different spatial and temporal scales (from individual pixels to monthly averages in $20 \times 20 \text{ km}^2$ grids). These estimated TROPOMI measurements in 2019.

7.5.2 Trueness

In this section, we present monthly averaged values of the systematic error uncertainties on the vertical columns estimated for OMI retrievals between 2005 and 2014. The contribution of the air mass factor uncertainties is the largest contribution to the vertical column systematic error uncertainties (see also Table 13). Figure 7-3 presents the VCD uncertainties due to AMF errors, and the five considered contributions. The largest contributions are from the a priori profile uncertainty and from the structural uncertainty (taken as 15% of the AMF). In the case where the satellite averaging kernels are used for comparisons with external HCHO columns, the a priori profile contribution can be removed

from the comparison uncertainty budget, leading to a total uncertainty in the range of 25% to 50%. Table 13 wraps up the relative contributions to the HCHO vertical column uncertainty, in the case of monthly averaged columns for typical low and high columns.

Considering these estimates of the HCHO column trueness, the requirements for HCHO product (30%) are achievable in regions of high emissions and for certain times of the year. In any case, observations need to be averaged to reduce random error uncertainties at a level comparable or smaller than systematic error uncertainties.

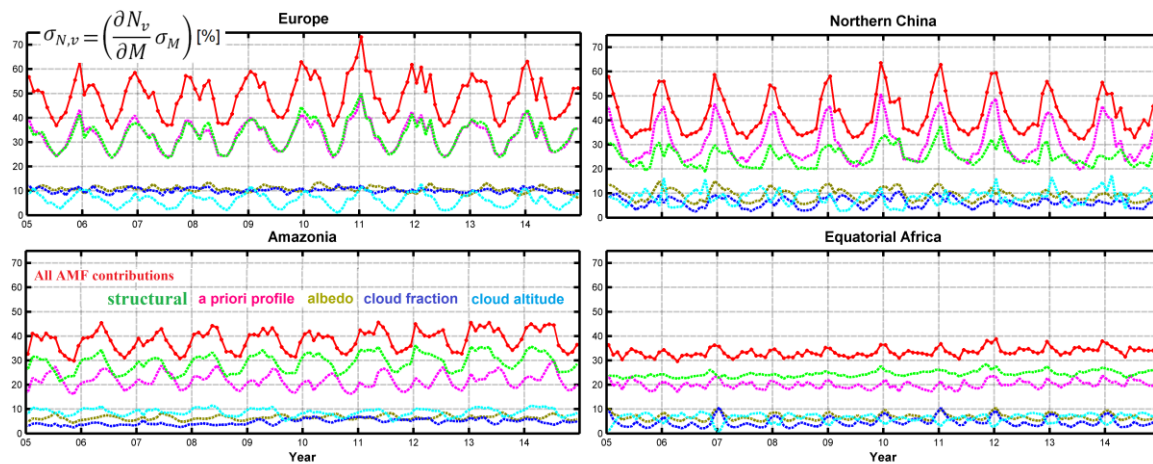


Figure 7-3: Regional and monthly average of the relative systematic vertical column AMF-related uncertainties in several NMVOC emission regions, for the period 2005-2014. The 5 contributions to the systematic air mass factor error uncertainty are shown: structural (green), a priori profile (pink), albedo (olive), cloud fraction (blue) and cloud altitude (cyan).

Table 13: Estimated HCHO vertical column uncertainty budget for monthly averaged low and elevated columns (higher than 1×10^{16} molec.cm⁻²). Contributions from the three retrieval steps are provided, as well as input parameter contributions.

HCHO vertical error uncertainty	Remote regions / low columns	Elevated column regions / periods
Contribution from systematic slant columns uncertainties	25%	15%
Contribution from air mass factors uncertainties	75%	30%
<ul style="list-style-type: none"> from a priori profile errors from model errors from albedo errors from cloud top pressure errors from cloud fraction errors 	<ul style="list-style-type: none"> 60% 35% 20% 20% 15% 	<ul style="list-style-type: none"> 20% 15% 10% 10% 05%
Contribution from background correction uncertainties	40%	10%
Total	90%	35%
Total without smoothing error	50%	25%

8 Validation

Please refer to the Mission Performance Centre (MPC) validation reports that are regularly updated on <http://mpc-vdaf.tropomi.eu/index.php/formaldehyde> [RD17].

9 Conclusions

The retrieval algorithm for the TROPOMI formaldehyde product generation is based on the heritage from algorithms successfully developed for the GOME, SCIAMACHY, GOME-2, and OMI sensors. As the quality of the first recorded spectra of TROPOMI allows it, one large fitting interval (328.5-359 nm) has been selected to retrieve the HCHO slant columns of the operational product. Earthshine radiance from the previous days averaged in the remote Pacific are used as reference for the DOAS fit. Spectral outlier screening is applied during the fitting procedure (spike removal algorithm). The HCHO retrieval algorithm also includes a post-processing across-track reference sector correction, to minimize striping effects, and a more accurate background correction scheme (as developed for the QA4ECV product).

A detailed uncertainty budget is provided for every satellite observation. The precision and trueness of the HCHO tropospheric column can come close to the COPERNICUS product requirements in regions of high emissions and, at mid-latitude, for summer (high sun) conditions. To improve the precision, spatial and temporal averaging is needed. The trueness of the vertical columns is also improved, owing to the use of daily forecasts for the estimation of HCHO vertical profile shapes, that is provided by a new version of the TM5-MP model, running at the spatial resolution of 1x1 degree in latitude and longitude.

The validation of satellite retrievals in the lower troposphere is known to be challenging. Ground-based measurements, where available, often sample the atmosphere at different spatial and temporal scales than the satellite measurements, which leads to ambiguous comparisons. Additional correlative measurements are needed over a variety of regions, in particular in the Tropics and at the sub-urban level in mid-latitudes. These aspects are covered by a number of projects developed in the framework of the TROPOMI validation plan (Fehr, 2016; Vigouroux et al., 2020).

A. Description of prototype software

The prototype algorithm for the retrieval of HCHO slant columns is developed using the QDOAS software, a multi-purpose DOAS analysis software developed at BIRA-IASB since early nineties (Fayt and Van Roozendaal, 2001; Danckaert et al., 2012) [URL01].

This program has been extensively validated in a number of intercomparison exercises (e.g. Roscoe et al., 1999, Van Daele et al., 2004; Roscoe et al., 2011). It has also been used worldwide and for many different DOAS applications.

The QDOAS fitting algorithm uses a combination of Singular Value Decomposition (SVD) and the Levenberg-Marquard methods to solve the linear and non-linear parts of the DOAS equation.

The main QDOAS features include:

- Analysis
 - DOAS/intensity fitting modes;
 - shift/stretch fully configurable for any spectral item (cross-section or spectrum);
 - possibility to filter spectra and cross-sections before analysis (supported filters include Kaiser, gaussian, boxcar, Savitsky Golay...);
 - possibility to define gaps within fitting intervals (e.g. to eliminate bad pixels);
 - possibility to fit an instrumental offset;
 - possibility to define several configurations of spectral windows under a project ;
 - non-linear parameters (offset, shift and stretch) can be fitted using wavelength polynomials up to the second order.
- Calibration and slit function characterization
 - wavelength calibration and instrumental slit function characterization using a non-linear least-squares (NLLS) fitting approach where measured intensities are fitted to a high resolution solar spectrum degraded to the resolution of the instrument. The fitting method (DOAS or intensity fitting) can be different from the method used in the analysis;
 - possibility to correct for atmospheric absorption and Ring effect;
 - supports different analytical line shapes.
- Cross-sections handling
 - possibility to calculate differential absorption cross-sections (by orthogonalization or high-pass filtering);
 - possibility to correct cross-sections using wavelength dependent AMF;
 - possibility to fix the column density of any selected species;
 - possibility to convolve cross-sections in real time using a user defined slit function or the information on calibration and slit function provided by the wavelength calibration procedure;
 - possibility to handle differences in resolutions between measured and control spectra;
 - standard and I_0 -corrected convolutions are supported;

- calculates Ring effect cross-sections (Rotational Raman Scattering approach);

QDOAS has been developed in collaboration with the Dutch company S[&]T. The graphical user interface is built on the Open-Source version of the QT-4 toolkit, a cross-platform application framework, and QWT libraries. QDOAS is a free software distributed under the terms of the GNU General Public License; it is open source and the code is available on request by contacting authors.

Errors on slant column densities in QDOAS

The slant column density is fit linearly according to Equation (5-6). Uncertainties on the retrieved slant columns N_s depend on the sensitivity of the sum of squares (5-6) with respect to variations of the slant column around the fitted minimum and on the noise on the measurements. We use the following expression for the covariance matrix $\sum \sigma_{N_{s,j}}$ of the weighted linear least squares parameter estimate:

$$\sum \sigma_{N_{s,j}} = (A^T E^2 A)^{-1} \quad 9-1$$

where the matrix $A(n \times k)$ is formed by the absorption cross-sections, and the diagonal matrix $E(k \times k)$ contains the measurement errors (with k the number of spectral pixels in the fitting interval):

$$E_{ij} = \frac{1}{\varepsilon_i} \quad \text{if } i = j \quad 9-2$$

If the measurement errors ε_i are not used in the fit, all ε_i are set to 1 and the random error on the measurements ε^2 is estimated by the reduced χ^2 , e.g. the sum of squares (5-6) divided by the number of degrees of freedom in the fit ($k-n$), where n is the number of parameters to fit. Equation (9-1) then becomes:

$$\sum \sigma_{N_{s,j}} = \frac{\chi^2}{(k-n)} (A^T A)^{-1} \quad 9-3$$

This estimate for the uncertainties assumes that the errors on the measurements at each spectral pixel are independent and normally distributed. When measurement errors are correlated or contain systematic components, the uncertainties calculated with (9-1) will underestimate the true error on the slant column density. Remaining residual structures after the fit, or a value of $\frac{\chi^2}{(k-n)} \gg 1$, are an indication of such a bad fit.

Equations (9-1) or (9-3) do not take into account the uncertainties on the cross sections, or other systematic errors in the fit parameters.

B. Description of forward model

The linearized discrete ordinate radiative transfer code LIDORT is the radiative transfer model applied in the calculation of the weighting function look-up tables, used in HCHO retrieval for the determination of the air mass factor. LIDORT is being developed and maintained by the company RT Solutions, Inc in Cambridge, MA in the United States. Many aspects of LIDORT environments have been realized through intensive cooperation between RT Solutions, Inc. and BIRA-IASB. As an example, LIDORT has been the core of the radiative transfer calculations in the direct fitting algorithm GODFIT (Lerot et al., 2010), developed at BIRA-IASB for the derivation of improved total ozone columns from measurements of the GOME instrument onboard the ERS-2 platform. During that project, LIDORT was extended with the capability of calculating analytical weighting functions for total gas columns, which plays a key role in the GODFIT O3 column derivation. Later versions of GODFIT benefited from many improvements in LIDORT, both regarding performance and the treatment of physics. Currently GODFIT uses LIDORT v3.3 (Spurr, 2008a).

C. Averaging Kernel

Retrieved satellite quantities always represent a weighted average over all parts of the atmosphere that contribute to the signal observed by the satellite instrument. The DOAS total column retrieval is implicitly dependant on the a priori tracer profile n_a . Indeed, the radiative transfer calculation accounts for the sensitivity of the measurement to the HCHO concentrations at all altitudes and the sensitivities are weighted with the assumed a priori profile shape to produce the vertical column. The averaging kernel (AK) is proportional to this measurement sensitivity profile, and provides the relation between the retrieved column N_v and the true tracer profile x (Rodgers, 2000; Rodgers and Connor, 2002):

$$N_v - N_{v,a} = AK \cdot (x^{pc} - n_a^{pc}) \quad 9-4$$

where the profiles are expressed in partial columns (pc). For total column observations of optically thin absorbers DOAS averaging kernels are calculated as follows (*Eskes and Boersma, 2003*): $AK(z) = \frac{m(z)}{M}$, where $m(z)$ is the altitude-resolved air mass factor (calculated from equation 5-10) and M is the tropospheric air mass factor (calculated from equation 5-10). The air mass factor and therefore the retrieved vertical column, depends on the a priori profile shape, in contrast to the altitude-resolved air mass factor which describes the sensitivity of the slant column to changes in trace gas concentrations at a given altitude and does not depends on the a priori profile in an optically thin atmosphere. From the definition of AK , we have $N_{v,a} = AK \cdot n_a^{pc}$ and Equation 9-4 simplifies to:

$$N_v = AK \cdot x^{pc} \quad 9-5$$

The averaging kernel varies with the observation conditions. In the HCHO retrieval product, the AK is provided together with the uncertainty budget for each individual pixel. The provided HCHO vertical columns can be used in two ways, each with its own associated error (*Boersma et al., 2004*):

1. For independent study and/or comparison with other independent measurements of total column amounts. In this case, the total error related to the column consists of slant column measurement errors, reference sector correction errors, and air mass factor errors. The latter consists of errors related to uncertainties in the assumed profile n_a and errors related to the $m(z)$ parameters as developed in section 5.4: Tropospheric air mass factor calculation.
2. For comparisons with chemistry transport models or validation with independent profile measurements, if the averaging kernel information is used, the a priori profile shape error no longer contributes to the total error. Indeed, the relative difference between the retrieved column N_v and an independent profile x_i is:

$$\delta = \frac{N_v - AK \cdot x_i^{pc}}{N_v} \quad 9-6$$

The total AMF M cancels since it appears as the denominator of both N_v and AK . Because only the total AMF depends on the a priori tracer profile n_a , the comparison using the averaging kernel is not influenced by the chosen a priori profile shape. The a priori profile error does not influence the comparison, but of course, it still does influence the error on the retrieved vertical column.

References

- Abbot, D. S., Palmer, P. I., Martin, R. V., Chance, K. V., Jacob, D. J. and Guenther, A.: Seasonal and interannual variability of North American isoprene emissions as determined by formaldehyde column measurements from space, *Geophys. Res. Lett.*, 30(17), 1886, 2003.
- Barkley, M. P., Palmer, P. I., Ganzeveld, L., Arneth, A., Hagberg, D., Karl, T., Guenther, A., Paulot, F., Wennberg, P. O., Mao, J., Kurosu, T. P., et al.: Can a state of the art chemistry transport model simulate Amazonian tropospheric chemistry?, *J. Geophys. Res.*, 116(D16), D16302, doi:10.1029/2011JD015893, 2011.
- Barkley, M. P., Kurosu, T. P., Chance, K., Smedt, I. De, Van Roozendaal, M., Arneth, A., Hagberg, D., Guenther, A. and De Smedt, I.: Assessing sources of uncertainty in formaldehyde air mass factors over tropical South America: Implications for top-down isoprene emission estimates, *J. Geophys. Res.*, 117(D13), D13304, doi:10.1029/2011JD016827, 2012.
- Barkley, M. P., De Smedt, I., Van Roozendaal, M., Kurosu, T. P., Chance, K. V., Arneth, A., Hagberg, D., Guenther, A. B., Paulot, F., Marais, E. A., others, et al.: Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI formaldehyde columns, *J. Geophys. Res. Atmos.*, 118(12), n/a–n/a, doi:10.1002/jgrd.50552, 2013.
- Boersma, K. F., Eskes, H. J. and Brinksma, E. J.: Error analysis for tropospheric NO₂ retrieval from space, *J. Geophys. Res.*, 109(D4), doi:10.1029/2003JD003962, 2004.
- Brion, J., et al.: Absorption spectra measurements for the ozone molecule in the 350–830 nm region, *J. Atmos. Chem.*, 30, 291–299, 1998.
- Chance, K. and R. J. Spurr: Ring effect studies: Rayleigh scattering including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, *Applied Optics*, 36, 5224–5230, 1997.
- Chance, K. V., Palmer, P. I., Martin, R. V., Spurr, R. J. D., Kurosu, T. P. and Jacob, D. J.: Satellite observations of formaldehyde over North America from GOME, *Geophysical Research Letters*, 27(21), 3461–3464, doi:10.1029/2000GL011857, 2000.
- Chance, K. and Kurucz, R. L.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, *J. Quant. Spectrosc. Radiat. Transf.*, 111(9), 1289–1295, 2010.
- Clémer, K., Van Roozendaal, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and De Mazière, M.: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, *Atmos. Meas. Tech.*, 3, 863–878, 2010.
- Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K. and Visconti, G.: Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, *Atmos. Chem. Phys.*, 10(23), 11501–11517, 2010.
- Eskes, H. J. and K. F. Boersma, Averaging kernels for DOAS total-column satellite retrievals, *Atmos. Chem. Phys.*, 3, 1285–1291, 2003.
- Danielson, J.J., and Gesch, D.B.: Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, 26 p, 2011.
- Daumont, M., Brion, J., Charbonnier, J., and Malicet, J.: Ozone UV spectroscopy, I: Absorption cross-sections at room temperature, *J. Atmos. Chem.*, 15, 145–155, 1992.
- De Smedt, I., Müller, J.-F., Stavrou, T., van der A, R., Eskes, H. and Van Roozendaal, M.: Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, *Atmos. Chem. Phys.*, 8(16), 4947–4963, 2008.
- De Smedt, I., Stavrou, T., Müller, J. F., van Der A, R. J. and Van Roozendaal, M.: Trend detection in satellite observations of formaldehyde tropospheric columns, *Geophys. Res. Lett.*, 37(18), L18808, doi:10.1029/2010GL044245, 2010.
- De Smedt, I.: Long-Term Global Observations of Tropospheric Formaldehyde Retrieved from Spaceborne Nadir UV Sensors, Ph.D. thesis, Université Libre De Bruxelles, Laboratoire de Chimie Quantique et Photophysique, Faculté de Sciences Appliquées, 2011.
- De Smedt, I., Van Roozendaal, M., Stavrou, T., Müller, J.-F., Lerot, C., Theys, N., Valks, P., Hao, N., and van der A, R.: Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A

- addressing noise reduction and instrumental degradation issues, *Atmos. Meas. Tech. Discuss.*, 5, 5571–5616, doi:10.5194/amtd-5-5571-2012, Special Issue: GOME-2: calibration, algorithms, data products and validation, 2012.
- De Smedt, I., Stavrou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., They s, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., et al.: Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, *Atmos. Chem. Phys. Discuss.*, 15(8), 12241–12300, doi:10.5194/acpd-15-12241-2015, 2015.
- De Smedt, I., They s, N., Yu, H., Danckaert, T., Lerot, C., Compernelle, S., Van Roozendaal, M., Richter, A., Hilboll, A., Peters, E., Pedernana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F. and Veefkind, P.: Algorithm Theoretical Baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, *Atmos. Meas. Tech.*, (11), 2395–2426, doi:10.5194/amt-2017-393, 2018.
- Dufour, G., F. Wittrock, M. Camredon, M. Beekmann, A. Richter, B. Aumont, and J. P. Burrows, SCIAMACHY formaldehyde observations: constraint for isoprene emission estimates over Europe?, *Atmos. Chem. Phys.*, 9(5), 1647–1664, 2009.
- Fayt, C. and M. Van Roozendaal: Windoas 2.1, Software User Manual, BIRA-IASB, 2001.
- Danckaert, T., Fayt, C., Van Roozendaal, M., De Smedt, I., Letocart, V., Merlaud, A., Pinardi, G: Qdoas Software User Manual, Version 2.1, http://uv-vis.aeronomie.be/software/QDOAS/QDOAS_manual_2.1_201212.pdf, 2012.
- Fehr, T.: Sentinel-5 Precursor Scientific Validation Implementation Plan, EOP-SM/2993/TF-tf, 1.0, <http://doi.org/10.5281/zenodo.165739>, 2016.
- Fleischmann, O. C., et al. : New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy, *J. Photochem. Photobiol. A*, 168, 117–132, 2004.
- Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Saunois, M., Szopa, S., Cressot, C., Kurosu, T. P., Chance, K. and Fried, A.: The formaldehyde budget as seen by a global-scale multi-constraint and multi-species inversion system, *Atmos. Chem. Phys. Discuss.*, 12(3), 6909–6955, doi:10.5194/acpd-12-6909-2012, 2012.
- Fu, T.-M., Jacob, D. J., Palmer, P. I., Chance, K. V., Wang, Y. X., Barletta, B., Blake, D. R., Stanton, J. C. and Pilling, M. J.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone, *J. Geophys. Res.*, 112(D6), D06312, 2007.
- González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T. P. and Suleiman, R.: Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval, *Atmos. Meas. Tech.*, 8(1), 19–32, doi:10.5194/amt-8-19-2015, 2015.
- Gonzi, S., Palmer, P. I., Barkley, M. P., De Smedt, I. and Van Roozendaal, M.: Biomass burning emission estimates inferred from satellite column measurements of HCHO: Sensitivity to co-emitted aerosol and injection height, *Geophys. Res. Lett.*, 38(14), L14807, doi:10.1029/2011GL047890, 2011.
- Gottwald, M., Bovensmann, H. et al.: SCIAMACHY, Monitoring the Changing Earth's Atmosphere, DLR, Institut für Methodik der Fernerkundung (IMF), 2006.
- Grainger, J. F. and J. Ring: Anomalous Fraunhofer line profiles, *Nature*, 193, 762, 1962.
- Greenblatt, G. D., Orlando, J. J., Burkholder, J. B., and Ravishankara, A. R.: Absorption measurements of oxygen between 330 and 1140 nm, *J. Geophys. Res.*, 95(D11), 18 577–18 582, doi:10.1029/90JD01375, 1990.
- Heckel, A., Kim, S.-W., Frost, G. J., Richter, A., Trainer, M. and Burrows, J. P.: Influence of low spatial resolution a priori data on tropospheric NO₂ satellite retrievals, *Atmos. Meas. Tech.*, 4(9), 1805–1820, doi:10.5194/amt-4-1805-2011, 2011.
- Hewson, W., Bösch, H., Barkley, M. P. and De Smedt, I.: Characterisation of GOME-2 formaldehyde retrieval sensitivity, *Atmospheric Measurement Techniques*, 6(2), 371–386, doi:10.5194/amt-6-371-2013, 2013.
- Hilboll, A., Richter, A. and Burrows, J. P.: Long-term changes of tropospheric NO₂ over megacities derived from multiple satellite instruments, *Atmospheric Chemistry and Physics*, 13(8), 4145–4169, doi:10.5194/acp-13-4145-2013, 2013.

- Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H.-W., The global chemistry transport model tm5: description and evaluation of the tropospheric chemistry version 3.0., *Geoscientific Model Development*, 3(2):445-473, 2010.
- Kaiser, J., Jacob, D. J., Zhu, L., Travis, K. R., Fisher, J. A., González Abad, G., Zhang, L., Zhang, X., Fried, A., Crounse, J. D., St. Clair, J. M., and Wisthaler, A.: High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US, *Atmos. Chem. Phys.*, 18, 5483-5497, <https://doi.org/10.5194/acp-18-5483-2018>, 2018.
- Kleipool, Q. L., Dobber, M. R., de Haan, J. F. and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, *J. Geophys. Res.*, 113(D18), D18308, doi:10.1029/2008JD010290, 2008.
- Koelemeijer, R. B. A., Stammes, P., Hovenier, J. W. and de Haan, J. F.: A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment, *J. Geophys. Res.*, 106(D4), 3475-3490, doi:10.1029/2000JD900657 2001.
- Khokhar, M. F.: Spatio-Temporal Analyses of Formaldehyde over Pakistan by Using SCIAMACHY and GOME-2 Observations, *Aerosol Air Qual. Res.*, 1-14, doi:10.4209/aaqr.2014.12.0339, 2015.
- Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications., *Atmos. Chem. Phys.*, 5(2):417-432, 2005.
- Kurosu, T. P., OMHCHO README FILE,
http://www.cfa.harvard.edu/tkurosu/SatelliteInstruments/OMI/PGEReleases/READMEs/OMHCHO_README.pdf, last access: 14/08/2012, 2008.
- Leitao, J., Richter, A., Vrekoussis, M., Kokhanovsky, A., Zhang, Q.J., Beekmann, M., and Burrows, J. P.: On the improvement of NO₂ satellite retrievals – aerosol impact on the airmass factors, *Atmos. Meas. Tech.*, 3, 475-493, doi:10.5194/amt-3-475-2010, 2010.
- Leue, C.: Detektion der troposphärischen NO₂ Daten anhand von GOME. Ph.D. thesis, Univ. Heidelberg, Heidelberg, Germany, 1999.
- Li, C., Joiner, J., Krotkov, N. A. and Dunlap, L.: A New Method for Global Retrievals of HCHO Total Columns from the Suomi National Polar-orbiting Partnership Ozone Monitoring and Profiler Suite, *Geophys. Res. Lett.*, doi:10.1002/2015GL063204, 2015.
- Loyola, D. G., Gimeno García, S., Lutz, R., Romahn, F., Spurr, R. J. D., Pedernana, M., Doicu, A., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2017-128>, in review, 2017
- Mahajan, A. S., De Smedt, I., Biswas, M. S., Ghude, S., Fadnavis, S., Roy, C. and van Roozendael, M.: Inter-annual variations in satellite observations of nitrogen dioxide and formaldehyde over India, *Atmos. Environ.*, 116, 194-201, doi:10.1016/j.atmosenv.2015.06.004, 2015.
- Malicet, C., Daumont, D., Charbonnier, J., Parisse, C., Chakir, A., and Brion, J.: Ozone UV spectroscopy, II: Absorption cross-sections and temperature dependence, *J. Atmos. Chem.*, 21, 263-273, 1995.
- Marais, E. A., Jacob, D. J., Kurosu, T. P., Chance, K., Murphy, J. G., Reeves, C., Mills, G., Casadio, S., Millet, D. B., Barkley, M. P., Paulot, F., et al.: Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns, *Atmos. Chem. Phys. Discuss.*, 12(3), 7475-7520, doi:10.5194/acpd-12-7475-2012, 2012.
- Marbach, T., Beirle, S., Platt, U., Hoor, P., Wittrock, F., Richter, A., Vrekoussis, M., Grzegorski, M., Burrows, J. P. and Wagner, T.: Satellite measurements of formaldehyde linked to shipping emissions, *Atmos. Chem. Phys.*, 9(21), 2009.
- Martin, R. V, Chance, K. V, Jacob, D. J., Kurosu, T. P., Spurr, R. J. D., Bucsela, E. J., Gleason, J., Palmer, P. I., Bey, I., Fiore, A. M., Li, Q., et al.: An improved retrieval of tropospheric nitrogen dioxide from GOME, *J. Geophys. Res.*, 107(D20), doi:10.1029/2001JD001027, 2002.
- Meller, R., and Moortgat, G. K.: Temperature dependence of the absorption cross section of HCHO between 223 and 323K in the wavelength range 225-375 nm, *J. Geophys. Res.*, 105(D6), 7089-7102, doi:10.1029/1999JD901074, 2000.
- Millet, D. B., Jacob, D. J., Boersma, K. F., Fu, T.-M., Kurosu, T. P., Chance, K. V., Heald, C. L. and Guenther, A.: Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, *Journal of Geophysical Research*, 113(D2), 1-18, doi:10.1029/2007JD008950, 2008.

- Palmer, P. I., Jacob, D. J., Chance, K. V., Martin, R. V., D, R. J., Kurosu, T. P., Bey, I., Yantosca, R. and Fiore, A.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, *Journal of Geophysical Research*, 106(D13), 14539-14550, doi:10.1029/2000JD900772, 2001.
- Palmer, P. I., Abbot, D. S., Fu, T.-M., Jacob, D. J., Chance, K. V., Kurosu, T. P., Guenther, A., Wiedinmyer, C., Stanton, J. C., Pilling, M. J., Pressley, S. N., et al.: Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, *Journal of Geophysical Research*, 111(D12), 1-14, doi:10.1029/2005JD006689, 2006.
- Pedernghana, M., Loyola, D., Apituley, A., Sneep, M., Veefkind, J. P.: Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Formaldehyde HCHO, S5P-L2-DLR-PUM-400F, 0.11.4, http://www.tropomi.eu/sites/default/files/files/S5P-L2-DLR-PUM-400F-Product_User_Manual_for_the_Sentinel_5_precursor_Formaldehyde_HCHO-00.11.04-20170601_signed.pdf, 2017.
- Pinardi, G., Van Roozendaal, M., Abuhassan, N., Adams, C., Cede, a., Clémer, K., Fayt, C., Frieß, U., Gil, M., Herman, J., Hermans, C., et al.: MAX-DOAS formaldehyde slant column measurements during CINDI: intercomparison and analysis improvement, *Atmospheric Measurement Techniques*, 6(1), 167–185, doi:10.5194/amt-6-167-2013, 2013.
- Platt, U.: Differential optical absorption spectroscopy (DOAS), in *Air Monitoring by Spectroscopic Techniques*, M.W. Sigrist ed., Chemical Analysis Series, Wiley, New York, 127, 27-84, 1994.
- Platt, U and Stutz, J.: *Differential Optical Absorption Spectroscopy: Principles and Applications (Physics of Earth and Space Environments)*, Springer-Verlag, Berlin, Heidelberg, ISBN 978-3540211938, 2008.
- Pukite, J., Kühl, S., Deutschmann, T., Platt, U., and Wagner, T.: Extending differential optical absorption spectroscopy for limb measurements in the UV, *Atmos. Meas. Tech.*, 3, 631-653, 2010.
- Richter, A., Begoin, M., Hilboll, A. and Burrows, J. P.: An improved NO₂ retrieval for the GOME-2 satellite instrument, *Atmos. Meas. Tech.*, 4(6), 213-246, doi:10.5194/amt-4-1147-2011, 2011.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding, Theory and Practice*, World Scientific Publishing, Singapore-New-Jersey-London-Hong Kong, 2000.
- Rodgers, C. D., and B. J. Connor: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, doi:10.1029/2002JD002299, 2003.
- Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., and Burrows, J. P.: High spectral resolution ozone absorption cross-sections – Part 2: Temperature dependence, *Atmos. Meas. Tech.*, 7, 625-636, doi:10.5194/amt-7-625-2014, 2014.
- Spurr, R. J. D.: 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, 2008a.
- Spurr, R. J. D., J. de Haan, R. van Oss, and A. Vasilkov, Discrete ordinate radiative transfer in a stratified medium with first-order rotational Raman scattering, *J.Q.S.R.T* 109, Iss. 3, 404425, 2008b.
- Stavrakou, T., Müller, J. F., De Smedt, I., Van Roozendaal, M., van der Werf, G. R., Giglio, L. and Guenther, A.: Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006, *Atmos. Chem. Phys.*, 9(3), 1037-1060, 2009a.
- Stavrakou, T., Smedt, I. D., Roozendaal, M. V., Vrekoussis, M., Wittrock, F., Burrows, J., Building, M., Lane, B., Gifford, C. and Kingdom, U.: The continental source of glyoxal estimated by the synergistic use of spaceborne measurements and inverse modelling, 2009b.
- Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Van Roozendaal, M., Guenther, a., Wild, M. and Xia, X.: Isoprene emissions over Asia 1979–2012: impact of climate and land-use changes, *Atmos. Chem. Phys.*, 14(9), 4587–4605, doi:10.5194/acp-14-4587-2014, 2014.
- Stavrakou, T., Müller, J., Bauwens, M., Smedt, I. De and Roozendaal, M. Van: How consistent are top-down hydrocarbon emissions based on formaldehyde observations from GOME-2 and OMI ?, *Atmos. Chem. Phys. Discuss.*, 12007–12067, doi:10.5194/acpd-15-12007-2015, 2015.
- Tanskanen, A. Lambertian Surface Albedo Climatology at 360 nm from TOMS Data Using Moving Time-Window Technique. In: *Proceedings of the XX Quadrennial Ozone Symposium*, 1-8 June 2004, Kos, Greece.

- Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O₂-O₂ collision pairs between 340 and 630 nm and at atmospherically relevant pressure., *Phys. Chem. Chem. Phys.*, 15(37), 15371–81, doi:10.1039/c3cp50968k, 2013.
- U.S. Standard Atmosphere, U.S. Government Printing Office, Washington, D.C., 1976.
- van der A, R.J., H.J. Eskes, K.F. Boersma, T.P. van Noije, et al., Trends, seasonal variability and dominant NO_x source derived from a ten year record of NO₂ measured from space, *J. Geophys. Res.*, 113, D04302, doi: 10.1029/2007JD009021, 2008.
- Vandaele A.C., C. Hermans, P.C. Simon, M. Carleer, R. Colin, S. Fally, M.F. Mérienne, A. Jenouvrier, and B. Coquart, Measurements of the NO₂ absorption cross-section from 42000 cm⁻¹ to 10000 cm⁻¹ (238-1000 nm) at 220 K and 294 K, *J.Q.S.R.T.*, 59, 171-184, 1998.
- Van Roozendael, M., V. Soebijanta, C. Fayt, and J.-C. Lambert: Investigation of DOAS Issues Affecting the Accuracy of the GDP Version 3.0 Total Ozone Product, in ERS-2 GOME GDP 3.0 Implementation and Delta Validation, ERSE-DTEX-EOD-TN-02-0006, ESA/ESRIN, Frascati, Italy, Chap.6, pp.97-129, 2002.
- Van Roozendael, M., Spurr, R., Loyola, D., Lerot, C., Balis, D., Lambert, J.-C., Zimmer, W., Van Gent, J., Van Geffen, J., Koukoulis, M., Granville, J., Doicu, A., Fayt, C. & Zehner, C.: Sixteen Years Of GOME/ERS-2 Total Ozone Data: The New Direct-Fitting GOME Data Processor (Gdp) Version 5 - Algorithm Description, *J. geophys. Res.*, 117, D03305, Doi: 10.1029/2011jd016471, 2012.
- Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., et al.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sensing of Environment*, 120(0), 70-83, 2012.
- Vigouroux, C., F. Hendrick, T. Stavrou, B. Dils, I. De Smedt, C. Hermans, A. Merlaud, F. Scolas, C. Senten, G. Vanhalewyn, S. Fally, M. Carleer, J.-M. Metzger, J.-F. Müller, M. Van Roozendael, and M. De Mazière, Ground-based FTIR and MAX-DOAS observations of formaldehyde at Réunion Island and comparisons with satellite and model data, *Atmos. Chem. Phys.*, 9, 9523-9544, doi:10.5194/acp-9-9523-2009.
- Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J., Jones, N., Kivi, R., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI/S5P formaldehyde validation using an extensive network of ground-based FTIR stations, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2020-30>, in review, 2020.
- Vountas, M., Rozanov, V. V. and Burrows, J. P.: Ring effect: impact of rotational Raman scattering on radiative transfer in earth's atmosphere, *J. of Quant. Spec. and Rad. Trans.*, 60(6), 943-961, 1998.
- Vrekoussis, M., Wittrock, F., Richter, A. and Burrows, J. P.: GOME-2 observations of oxygenated VOCs: what can we learn from the ratio glyoxal to formaldehyde on a global scale?, *Atmos. Chem. Phys.*, 10(21), 10145-10160, 2010.
- Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, *Geosci. Model Dev.*, 10, 721-750, doi:10.5194/gmd-10-721-2017, 2017.
- 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, *Geophysical Research Letters*, 33(16), 1-5, doi:10.1029/2006GL026310, 2006.
- Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio, M. P., De Smedt, I., Gonzalez Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J. W., Hanisco, T. F., Richter, D., Scarino, A. J., Walega, J., Weibring, P. and Wolfe, G. M.: Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC4RS aircraft observations over the Southeast US, *Atmos. Chem. Phys.*, 0, 1–24, doi:10.5194/acp-2016-162, 2016.
- Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P.: An improved tropospheric NO₂ retrieval for OMI observations in the vicinity of mountainous terrain, *Atmos. Meas. Tech.*, 2, 401-416, doi:10.5194/amt-2-401-2009, 2009.