Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

BIRA·IASB

S5P/TROPOMI Total Ozone ATBD

document number	:	S5P-L2-DLR-ATBD-400A
authors	:	Rob Spurr, Diego Loyola, Michel Van Roozendael, Christophe Lerot, Klaus-Peter Heue, Jian Xu
CI identification	:	CI-400A-ATBD
issue	:	2.2
date	:	2020-06-15
status	:	Released

Document approval record

	digital signature	
prepared:	Klaus-Peter Heue Digitally signed by Klaus-Peter Heue DN: C=DE, S=Bavaria, O=DLR, OU=IMF-ATP, CN=Klaus-Peter Heue, E=Klaus-Peter.Heue@DLR.de Location: Oberpfaffenhofen Date: 2020-06-22 15:34:18 Foxit PhantomPDF Version: 9.3.0	Christophe Lerot Signature numérique de Christophe Lerot Date : 2020.07.03 13:57:19 +02'00'
	Michel Van Roozendael (Signature) Digitally signed by Michel Van Roozendael (Signature) Date: 2020.07.03 09:09:57 +02'00'	Diego Loyola Digitally signed by Diego Loyola DN: C=DE, O=DLR, CN=Diego Loyola, E=Diego.Loyola@dlr.de Reason: I am the author of this document Date: 2020-07-08 09:09:22
	Jian Xu Digitally signed by Jian Xu DN: C=DE, OU=DLR, O=IMF-ATP, CN=Jian Xu, E=jian.xu@dlr.de Reason: I am the author of this document Location: Oberpfaffenhofen, Germany Date: 2020-06-22 15:56:43 Foxit PhantomPDF Version: 9.3.0	
checked:		
approved PI:	Diego Loyola Digitally signed by Diego Loyola DN: C=DE, O=DLR, CN=Diego Loyola, E=Diego.Loyola@dlr.de Reason: I am approving this document Date: 2020-07-08 09:11:00	
approved PM:		
approved CM:		

Document change record

issue	Date	item	comments
0.1.0	2012-10-04	All	Initial draft version
0.2.0	2012-11-23	All	Updated version following the internal L2-WG review
0.3.0	2013-06-04	All	Document revised according to SRR/PDR review of February 2013
0.5.0	2013-06-21	All	Updated version following the internal L2-WG review
0.9.0	2013-11-29	All	Updated version following the external review Added recently published papers on GODFIT
0.10.0	2014-04-15	All	Remove instrument description (former chapter 4) and minor changes in chapter 7.
0.11.0	2014-09-30	All Section 6	Harmonization of symbols across ATBDs. Input/output tables added.
0.13.0	2015-09		Minor changes for preliminary release to S5P validation teams
0.14.0	2015-12	4.1, 5.2, 10 6.1	Recent references added Estimates updated with the new binning scheme
1.0	2016-02-01	All	Included comments from L2-WG reviewers First public release
1.5	2018-04-30	All	Updates following Commissioning phase
1.6	2018-10-17	All	Updates following reviews before OFFL product release
2.0	2019-06-30	5.2.3 5.2.7 5.4.6	include surface albedo fit for NRTI clarify destriping algorithm update QA value for OFFL and NRTI
2.1	2020-02-28	8 all	replaced validation (sect 8) by a reference to validation reports small updates
2.2	2020-06-15	all	small updates with respect of the operational status of TROPOMI / S5P

Contents

List of <TBD>’s

1 Introduction

1.1 Purpose of the ATBD

Ozone is of crucial importance for the equilibrium of the Earth atmosphere. In the stratosphere, the ozone layer shields the biosphere from dangerous solar ultraviolet radiation. In the troposphere, it acts as an efficient cleansing agent, but at high concentration it also becomes harmful to human and animal health and vegetation. Ozone is an important greenhouse-gas contributor to ongoing climate change. Since the discovery of the Antarctic ozone hole in the mid-eighties and the subsequent Montreal protocol that regulated the production of chlorine-containing ozone-depleting substances, ozone has been routinely monitored from the ground and from space.

The TROPOMI (Tropospheric Monitoring Instrument) is the payload instrument for the Sentinel 5 Precursor (S5P) Mission. The S5P platform was launched into a sun-synchronous low-earth orbit in October 13th, 2017. TROPOMI is a nadir-viewing atmospheric chemistry instrument measuring at moderate spectral resolution from the UV to the near infrared [RD7]. TROPOMI/S5P takes global measurements of a number of trace species, including ozone - there are separate products for ozone profiles and total column amounts.

There are two algorithms that will deliver total ozone column amounts to be retrieved from S5P/TROPOMI. The first algorithm (herewith called S5P_TO3_DOAS) is based on the "DOAS-style" GOME Data Processor (GDP) algorithm Version 4.8. The second (herewith called S5P_TO3_GODFIT) is based on the direct-fitting algorithm developed at BIRA-IASB and used for generating the total ozone ECV as part of the Ozone CCI activities. S5P_TO3_DOAS is used for generation of near-real-time products, while S5P_TO3_GODFIT generates the off-line and reprocessed products. Both algorithms are based on ozone absorption in the UV Huggins bands (325-335 nm).

The main product parameters (total ozone in [DU], ozone temperature in [K]) are the same in both cases. Both algorithms will also deliver a profile correlation matrix based on a priori column-classified ozone profile climatology. However, the full Level 2 TO3 data products differ in some aspects - this reflects intrinsic differences in the two retrieval algorithms. For example, the S5P_TO3_DOAS L2 product will contain intermediate output (slant columns, air mass factors) resulting from the two-step DOAS style of retrieval. Both L2 TO3 products output additional diagnostics (ancillary retrieval state vector elements, cloud information), again depending on the style of retrieval.

The present document is the ATBD (Algorithm Theoretical Basis Document) for these two algorithms. The purpose of the ATBD is to provide detailed mathematical and physical descriptions of the two algorithms, along with discussions of algorithm inputs and outputs, data products, algorithm validation and error analysis. The ATBD contains a summary description of the TROPOMI instrument.

1.2 Document overview

Following reference sections on applicable documentation (Chapter 2) and terms of reference (Chapter 3), the instrument is summarized in Chapter 4. The main algorithm descriptions are found in Chapters 5 and 6, while Chapters 7, 8 and 9 contain notes on Feasibility, Error Analyses and Validation respectively.

1.3 Acknowledgements

The authors would like to thank the following people working on algorithm development, validation, and performing the internal review of this document:

- Jeroen van Gent (BIRA), Pieter Valks (DLR), Adrian Doicu (DLR), Nan Hao (former DLR, now at EUMETSAT), Walter Zimmer (DLR), Fabian Romahn (DLR).
- Jean-Christopher Lambert (BIRA), Jose Granville (BIRA), Dimitris Balis (AUTH), MariLiza Koukouli (AUTH), Katerina Garane (AUTH)
- J.F. de Haan (KNMI), J.P. Veefkind (KNMI), Tijl Verhoelst (BIRA), Mark Weber (IUP-B).

2 Applicable and reference documents

2.1 Applicable documents

[AD1] 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

[AD2] Sentinel-5P Level 2 Processor Development – Statement of Work -; **source**: ESA; **ref**: S5P-SW-ESA-GS-053; **issue**: 1; **date**: 2012-03-02

[AD3] Sentinel-5 Precursor L2 UPAS Processor – Software Development Plan; **source**: DLR; **ref**: S5P-L2-DLR-SDP-1007; **issue**: 1.0; **date**: 2012-09-21

[AD4] GMES Sentinels 4 and 5 Mission Requirements Traceability Document; **source**: ESA/ESTEC; **ref**: EOP-SM/2413/BV-bv; **issue**: 1.9; **date**: 2012-09-20

2.2 Standard documents

There are no standard documents

2.3 Reference documents

[RD1] 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

[RD2] Terms, and symbols in the TROPOMI Algorithm Team; **source**: KNMI; **ref**: SN-TROPOMI-KNMI-049; **date**: 2011-09-28

[RD3] S5P/TROPOMI Clouds ATBD; **source**: DLR; **ref**: S5P-L2-DLR-ATBD-400I; **issue**: 2.2; **date**: 2020-06-15

[RD4] Algorithm theoretical basis document for TROPOMI L01b data processor; **source**: KNMI; **ref**: S5P-KNMI-L01B-0009-SD; **issue**: 8.0.0; **date**: 2017-06-01

[RD5] Input output data specification for TROPOMI L01b data processor; **source**: KNMI; **ref**: S5P-KNMI-L01B-0012-SD; **issue**: 8.0.0; **date**: 2017-06-01

[RD6] S5P/TROPOMI Static input for Level 2 processors; **source**: KNMI/SRON/BIRA/DLR; **ref**: S5P-KNMI-L2CO-0004-SD; **issue**: 4.0.0; **date**: 2013-06-21

[RD7] TROPOMI Instrument and Performance Overview; **source**: KNMI; **ref**: S5P-KNMI-L2-0010-RP; **issue**: 0.10.0; **date**: 2014-03-15

[RD8] Sentinel-5 precursor/TROPOMI Level 2 Product User Manual O_3 Total Column **source:** DLR; **ref:** S5P-L2-DLR-PUM-400A; **issue:** 2.2 **date:** 2020-06-15

[RD9] Quarterly Validation Report on 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

There are no electronic references.

3 Terms, definitions and abbreviated terms

Terms, definitions and abbreviated terms that are used in development program for the TROPOMI/S5P L0-1b data processor are described in [RD1]. Terms, definitions and abbreviated terms that are used in development program for the TROPOMI/S5P L2 data processors are described in [RD2]. Terms, definitions and abbreviated terms that are specific for this document can be found below.

3.1 Terms and definitions

Term	Definition

3.2 Acronyms and abbreviations

AK	Averaging Kernel
AMF	Air Mass Factor
ATBD	Algorithm Theoretical Basis Document
AUTH	Aristotle University of Thessaloniki
BDM	Brion-Daumont-Malicet
BIRA-IASB	Belgian Institute for Space Aeronomy
BRDF	Bi-directional Reflectance Distribution Function
BOA	Bottom of Atmosphere
CAL	Clouds As scattering Layers
CBH	Cloud base height
CBP	Cloud base pressure
CFR	Cloud Fraction
CCI	Climate Change Initiative
COT	Cloud optical thickness
CRB	Clouds as Reflecting Boundaries
CTA	Cloud top albedo
CTH	Cloud top height
CTP	Cloud top pressure
DFS	Degrees of Freedom for Signal
DLR	German Aerospace Center (Deutsches zentrum für Luft- und Raumfahrt)
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit
ECV	Essential Climate Variable
ECMWF	European Centre for Medium-range Weather Forecasting
ENVISAT	Environmental Satellite
EO	Earth Observation

ERS-2	European Remote Sensing Satellite-2
EOS-AURA	(NASA's) Earth Observing System Aura
ESA	European Space Agency
ESRIN	European Space Research Institute
ETOP	Earth Topography
FD	Finite Difference
FM	Flight Model
GAW	Global Atmospheric Watch
GDP	GOME Data Processor
GMES	Global Monitoring for Environment and Security
GODFIT	GOme Direct FITting
GOME	Global Ozone Monitoring Experiment
GTOPO	Global TOPOgraphic (Data set)
hPa	Hectopascals
IOP	Inherent Optical Property
IPA	Independent Pixel Approximation
IUP	Institut für Umwelt-Physik
K	Kelvin
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LER	Lambertian Equivalent Reflectivity
LIDORT	LInearized Discrete Ordinate Radiative Transfer
LM	Levenberg-Marquardt
LLM	Labow-Logan-McPeters
LOS	Line-Of-Sight
LUT	Look Up Table
MetOp	Meteorological Operational
ML	McPeters-Labow
MLS	Microwave Limb Sounder
MSG	Meteosat Second Generation
MRTD	Mission Requirements Traceability Document
NASA	National Aeronautics and Space Administration
NetCDF-CF	Network Common Data Format – Climate and Forecast (CF) convention
NDACC	Network for the Detection of Atmospheric Composition Change
NISE	Near-real-time global Ice and Snow Extent
NDSC	Network for the Detection of Stratospheric Change
NOAA	National Oceanic and Atmospheric Administration

NRTI	Near-real-time
OCRA	Optical Cloud Recognition Algorithm
OE	Optimal Estimation
OFFL	Off-line
OMI	Ozone Monitoring Instrument
P-S	Pseudo-Spherical
PTH	Pressure-Temperature-Height
QOS	Quadrennial Ozone Symposium
RRS	Rotational Raman Scattering
RT	Radiative Transfer
ROCINN	Retrieval of Cloud Information using Neural Networks
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY
SBUV	Solar Backscatter UltraViolet
S5P	Sentinel 5 Precursor
SGP	SCIAMACHY Ground Processor
SRD	Systems Requirement Document
SZA	Solar Zenith Angle
TO3	Total Ozone
TOA	Top of Atmosphere
TOMS	Total Ozone Mapping Spectrometer
TROPOMI	TROPOspheric Monitoring Instrument
UPAS	Universal Processor for UV/VIS Atmospheric Spectrometers
UV	Ultra Violet
V8	Version 8
VCD	Vertical Column Density
VLIDORT	Vector LInearized Discrete Ordinate Radiative Transfer
WMO	World Meteorological Office
WOUDC	World Ozone and Ultraviolet Data Centre

4 Introduction to the S5P Data Products

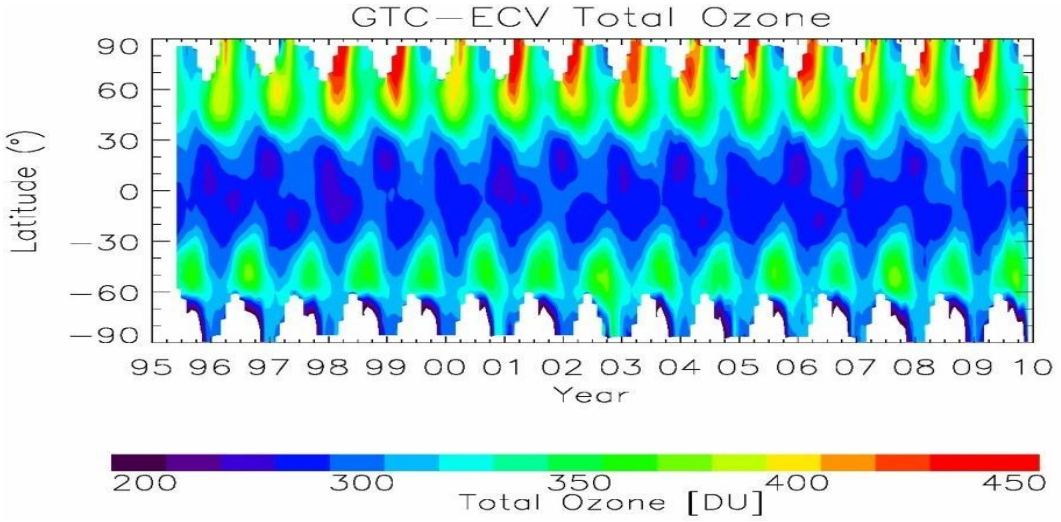

Figure 4.4.1: Example of a merged total ozone data set based on GOME, SCIAMACHY and GOME-2 observations. Zonally-averaged columns are plotted as a function of time and latitude.

4.1 Heritage

In the USA, total ozone has been measured since 1978 by NASA and NOAA using a series of TOMS and SBUV instruments [*Bhartia et al.*, 2003; *Miller et al*., 2002]. In Europe, the GOME instrument [*Burrows et al.*, 1999b] launched on the ERS-2 platform in April 1995 was the first in a new generation of hyperspectral sensors. Following GOME, SCIAMACHY [*Bovensmann et al.*, 1999] aboard ENVISAT has provided total ozone measurements from 2002 until April 2012 with an increased spatial resolution. The OMI instrument [*Levelt et al.*, 2006] is a joint Dutch-Finnish contribution to the American EOS-AURA platform launched in 2004. The GOME-2 [*Munro et al.*, 2006] instruments on the successive MetOp platforms monitor the Earth atmosphere with almost daily global coverage. The European instrument TROPOMI on S5P is the direct successor of OMI.

The baseline retrieval approach for TROPOMI/S5P is designed to take into consideration not only the requirements to provide fast near-real-time (NRTI) data products fitting COPERNICUS needs, but also requirements for off-line (OFFL) data sets compatible with long-term historical data records. For total ozone, two different approaches have been used to retrieve total ozone columns from the European sensors: DOAS and Direct-fitting.

The DOAS technique for total ozone retrieval was used from the start of the GOME/ERS-2 mission in 1995 (GDP Versions 1-4, see for example [*Spurr et al*., 2005] and [*Van Roozendael et al*., 2006]), DOAS retrieval is currently implemented in the SCIAMACHY and GOME-2 operational processors, respectively SGP 5.0 [*Lerot et al.*, 2009], GDP 4.4 [Loyola *et al.*, 2011] and GDP 4.7 [Hao *et al.*, 2014]. The DOAS technique is fast and provides ozone columns at the 1% level of accuracy in most situations [*Van Roozendael et al*., 2006].

The Direct-fitting approach provides more accurate total ozone columns, at the cost of slower computational performance. Heritage can be traced back to the analysis of ozone column measurements obtained from the continuous scan Nimbus-7 data [*Joiner and Bhartia*, 1997]. The upgraded GOME operational processor GDP 5 is based on Direct-fitting; the complete 16-year GOME data record has been recently reprocessed [*Van Roozendael et al*., 2012]. For GDP 5, the algorithm theoretical basis document (ATBD) [*Spurr et al*., 2012], validation report and disclaimer [*Lambert et al*., 2012b] and product user manual [*Loyola et al*., 2012] can be downloaded from the website at https://atmos.eoc.dlr.de/app/documents. The Direct-fitting algorithm also generates the total ozone essential climate variable for the ESA Ozone_CCI project with an homogenous data record from GOME, SCIAMACHY, GOME-2 and OMI [*Lerot et al.*, 2014] that agree at the −0.32 to 0.76% level with the recently published monthly zonal mean total ozone SBUV data [Chiou *et al.*, 2013] and has a 1 ± 1% bias with respect to the ground-based observations [*Koukouli et al.*, 2015; Garane et al., 2018]. The Ozone_CCI ATBD describes the latest version of the GODFIT algorithm and is available at http://www.esa-ozone-cci.org/.

It should be noted that there are a number of other total ozone algorithms both in previous and existing missions – these are not strictly speaking heritage algorithms for TROPOMI/S5P. These algorithms include the TOMS total ozone algorithm [*Bhartia et al*., 2003] (in use for various TOMS missions and also for OMI total ozone), the OMI-DOAS algorithm [*Veefkind et al*., 2006], and the GOME “weighting function” DOAS algorithm [*Coldewey-Egbers et al*., 2005].

4.2 Product requirements

The uncertainty requirements for the total ozone products are 3%, see Sentinel 4/5 MRTD mission requirements document [AD4] Tables A1, A2, A3, B1, B2, B3, C1, and C2. The retrieval algorithms selected for S5P can fulfil these requirements, as has been demonstrated with the previous missions GOME, SCIAMACHY and GOME-2.

4.3 Overview of the retrieval method

4.3.1 Near-Real-Time (NRTI) algorithm

The first major step in the DOAS-based NRTI algorithm (S5P_TO3_DOAS) is the least-squares fitting of the effective total ozone slant column, based on Beer-Lambert extinction; ancillary parameters in this inversion include effective temperature, a semi-empirical correction for Ring interference effects, and closure coefficients. The second step is a conversion to the vertical total ozone density, achieved through an iterative radiative transfer (RT) computation (at one wavelength) of appropriate air mass factors (AMFs) for clear sky and cloudy scenes in the IPA (independent pixel approximation); ozone profiles in RT simulations are derived from the column-classified Labow et al. climatology [2015] and OMI/MLS tropospheric climatologies.

4.3.2 Off-Line (OFFL) algorithm

The one-step OFFL algorithm (S5P_TO3_GODFIT) comprises a non-linear least-squares inversion based on the direct comparison of simulated and measured backscattered radiances. Simulated radiances are computed (again for a multiple scattering atmosphere) at all nominal wavelengths in the UV fitting window, along with corresponding analytically-derived weighting functions for total ozone, albedo and effective temperature. The algorithm also contains closure fitting coefficients and a semi-empirical correction for Ring interference. Although more accurate, the S5P_TO3_GODFIT algorithm involves many more RT simulations than S5P_TO3_DOAS, and is therefore slower by an order of magnitude approximately. This is currently the main driver for selection of S5P_TO3_DOAS for the NRTI product.

4.3.3 Advantages of Direct-fitting

The main motivation for direct-fitting is that, in contrast to DOAS, this approach encompasses a physically sound treatment of the ozone absorption in a wavelength region providing excellent sensitivity to total ozone. Direct-fitting allows for consistent treatment of the atmospheric temperature variability and its impact on ozone (while the effective temperature approach used with DOAS is subject to biases related to poor ozone absorption fitting). A full radiance closure (including polarization corrections) is possible with direct-fitting, thus enabling simultaneous retrieval of surface albedo or effective fractional cloud cover. Direct-fitting will retrieve accurate ozone columns for all regimes, even at large solar zenith angle (SZA) and for high ozone content when the atmospheric optical depth is such that the optically thin approximation intrinsic to DOAS is no longer valid.

4.4 Foreseen update approach

Originally, the proposed DOAS operational algorithm assumed clouds to be Lambertian reflectors (CRB), which implied the use of a semi-empirical intra-cloud correction algorithms and a ghost column correction. The cloud product for S5P is based on a more precise treatment of clouds as scattering layers (CAL), and the algorithm S5P_TO3_DOAS has been upgraded accordingly to allow for ingestion of the cloud parameters derived from the CAL algorithm. The direct-fitting algorithm assumes one single effective scene of which an effective albedo is internally fitted. The latter has as advantages to reduce the dependence on cloud parameters, and also to implicitly account for presence of aerosols.

Further developments for both algorithms have been included for better polarization correction and Ring effect treatment, improved treatment of surface effects (aerosols and albedo), and specific adjustments necessary to deal with the instrument characteristics of TROPOMI (e.g. a destriping correction has been implemented in the NRTI algorithm).

4.5 General design considerations

The science behind both prototype algorithms is well established; RT simulations in both cases are based on the widely used LIDORT models. Both heritage algorithms (GDP 4.x and GODFIT) have been written up in the literature [*Van Roozendael et al.*, 2006; *Van Roozendael et al., 2012*] for GOME; these algorithms have been applied to other satellite platforms measuring ozone [*Lerot et al.*, 2010; *Loyola et al.*, 2012].

The NRTI DOAS heritage algorithm has already been employed as an input to merge ozone data sets from GOME, SCIAMACHY and GOME-2, and it has been extensively used in recent trend studies of the ozone record, as evidenced at the WMO Ozone Assessment 2014 and the recent Quadriennal Ozone Symposium 2016.

Level 1b input requirements for both algorithms are the same (calibrated geolocated radiance and irradiance measurements in the UV window 325-335 nm). For more details on these requirements, see section 6.2.

Both heritage algorithms are robust, stable and operationally qualified. Both heritage algorithms (more so for GDP 4.x) have been validated against ground-based ozone networks and other satellite instruments [*Balis et al.*, 2007; *Koukouli et al.*, 2012 and 2015]. Both heritage algorithms retrieve ozone at the "1%" accuracy level for large parts of the global distribution; the GODFIT algorithm is overall more accurate in extreme geophysical conditions and therefore has smaller seasonal and other dependencies.

Current operational products from GDP 4.x (GOME, SCIAMACHY) and GDP 5 (GOME) are generated at DLR and are freely available from ESA. The complete reprocessed GOME data record is available to the public at no cost via ftp server using a simple registration procedure through the ESA ESRIN - EO Help Desk at eohelp@esa.int. Further information can be found at https://atmos.eoc.dlr.de/ghttome. GDP 4.x products for GOME-2 are freely available from EUMETSAT AC-SAF, see https://atmos.eoc.dlr.de/gome2.

At BIRA-IASB, a scientific prototype implementing both DOAS and direct-fitting approach has been used as a development test-bed for all sensors. The scientific direct-fitting prototype has been extensively used for ECV product development in the ESA Ozone_CCI project. L2 data sets from GOME, SCIAMACHY, GOME-2A/B and OMI produced in this context are available at http://www.esa-ozone-cci.org/. This prototype, has been extended and modified to build the development TROPOMI/S5P total ozone algorithms described in this ATBD. Based on this development, the operational algorithms are part of the UPAS system at DLR.

5 This ATBD Algorithm descriptions

5.1 Preamble

In this chapter, we describe the two main algorithms to be used for S5P total ozone retrieval.

S5P_TO3_DOAS is the default algorithm for the S5P Near-Real-Time (NRTI) total ozone product and is described in section 5.2. S5P_TO3_GODFIT is the default algorithm for the S5P Off-Line (OFFL) total ozone product and is described in section 5.3.

There is some overlap between the algorithms - both use the LIDORT radiative transfer model (though in different ways), and the auxiliary algorithms providing cloud information and Ring effect treatments are similar in the two cases. Algorithm components common to the two algorithms are dealt with in section 5.4.

5.2 S5P_TO3_DOAS, NRTI total ozone (GDP 4.x Heritage)

5.2.1 Overview with diagram

S5P_TO3_DOAS is based on the latest DOAS GDP 4.x algorithm [*van Roozendael et al.*, 2006, *Lerot et al.,* 2009, *Loyola et al.,* 2011*, and Hao et al.,* 2014] and uses a two-step approach consisting of a slant-column fitting followed by the AMF conversion to vertical column density. The slant column fitting is based on the Beer-Lambert extinction law for trace gas absorption, whereas the AMF computations require use of a multiple scatter RT model (in this case the LIDORT forward model [*Spurr*, 2008]) for simulation of backscattered radiances. There are several ancillary parameters in the DOAS fitting including polynomial closure coefficients, amplitudes for Ring and undersampling corrections, and a wavelength registration shift.

Figure 5.1 is a general flow diagram for the S5P_TO3_DOAS algorithm, assuming a granularity of one orbit (this is the GDP 4.x baseline for GOME). Following the orbital ingestion of Level 1 radiance/irradiance data, and the input of auxiliary data (topography fields, cloud information if pre-calculated), the algorithm then enters the pixel loop. For each pixel, the cloud ancillary information is derived (fractional cloud cover, cloud-top height (pressure) and optical thickness (albedo) derived from dedicated cloud retrieval algorithms such as OCRA and ROCINN). Next, the DOAS fitting step delivers the effective slant column of total ozone, plus a number of auxiliary fitted parameters and error diagnostics. The next steps are concerned with the generation of Air Mass Factors *M* (AMFs), which are needed to convert the DOAS-fitted slant column density N_s to the final vertical column density N_v (VCD) of total ozone, The AMF/VCD computation proceeds iteratively - at each iteration, AMFs (clear and cloudy scenes) are computed for the current VCD guess, the slant column is adjusted for the molecular Ring effect, and the VCD guess is updated. Once this iteration has converged, pixel processing is completed with a destriping algorithm and the generation of Level 2 output (total columns, errors/diagnostics, and auxiliary fitted parameters).

5.2.2 DOAS inverse model and state vector

The DOAS (Differential Optical Absorption Spectroscopy) spectral fitting technique is widely used in ground-based and satellite trace gas retrievals. DOAS focuses on narrow band-absorption signatures in the UV visible in order to separate and measure concentrations of trace species. For a review of the method, see *Platt* [1994]. The DOAS fitting is a least-squares regression between measured and simulated optical densities $\tau_{meas}(\lambda_i)$ and $\tau_{sim}(\lambda_i)$ respectively, for a set of measurement wavelengths $\{\lambda_i\}$ in a given fitting window.

Here, $\tau(\lambda) = \ln[I(\lambda)/E_0(\lambda)]$, the logarithm of the ratio of the backscattered radiance and extra-terrestrial solar irradiance on the same wavelength grid.

The DOAS fitting minimizes the least-squares functional $\mathfrak{F} \equiv \sum_i \left[\frac{\tau_{meas}(\lambda_i)-\tau_{sim}(\lambda_i)}{\epsilon_\tau(\lambda_i)}\right]^2$, where the error weighting is given by $\epsilon_\tau(\lambda_i) = \sqrt{[\epsilon_I(\lambda_i)]^2+[\epsilon_{E0}(\lambda_i)]^2}$, assuming normally distributed uncorrelated relative measurement errors $\epsilon_I(\lambda_i)$ and $\epsilon_{E0}(\lambda_i)$ on respective measurements $I_{meas}(\lambda_i)$ and $E_{0,meas}(\lambda_i)$.

In DOAS, the atmospheric model is the Beer-Lambert extinction law for trace gas absorbers. This model is acceptable for optically thin trace gas absorption. The simulated optical density is written:

$$\tau_{sim}(\lambda) = -\sum_g N_{s,g}(\Theta)\sigma_g(\lambda) - \sum_{m=0}^{3} \alpha_m \left(1-\frac{\lambda}{\lambda^*}\right)^m. \qquad 5.1$$

For wavelength λ and geometrical path Θ, $N_{s,g}(\Theta)$ is the effective slant column density of gas g and $\sigma_g(\lambda)$ the associated trace gas absorption cross-section. Note that cross-sections may depend on temperature. Broadband radiative effects (aerosol scattering/extinction, surface reflection and molecular scattering) are filtered out using a low-order polynomial in wavelength. The polynomial coefficients are α_m (a cubic has been assumed in Eq. (5.1), with λ^* a reference wavelength. The DOAS model in Eq. (5.1) is linear for the fitting variables $N_{s,g}$ and α_m; the fitting is then a multi-linear regression.

The polynomial filtering is an *external closure* term in Equation 5.1 that deals with all other radiative transfer effects like surface albedo - there is no attempt to model any of these effects from first principles. The resulting fitted parameters α_m are used to retrieve the surface reflectance based on a neural network [Loyola et al.,2020] In the AMF radiative transfer computation (see section 5.2.4), aerosols may be reintroduced in an *ad hoc* manner; however, there is no real justification for this, as these aerosols then become a hard-to-quantify source of error.

Wavelength grids for backscattered radiance and solar irradiances are different. Using the solar data as a baseline, this mismatch is treated by fitting shift and squeeze parameters to the backscattered radiance wavelength grid and re-sampling the measurements. The DOAS fitting has non-linear dependence on these shift and squeeze parameters, which must be determined iteratively. In practice, the multi-linear regression based on Eq. (5.1) must be repeated at each iteration step until a suitable convergence is reached for the shift and squeeze parameters is reached.

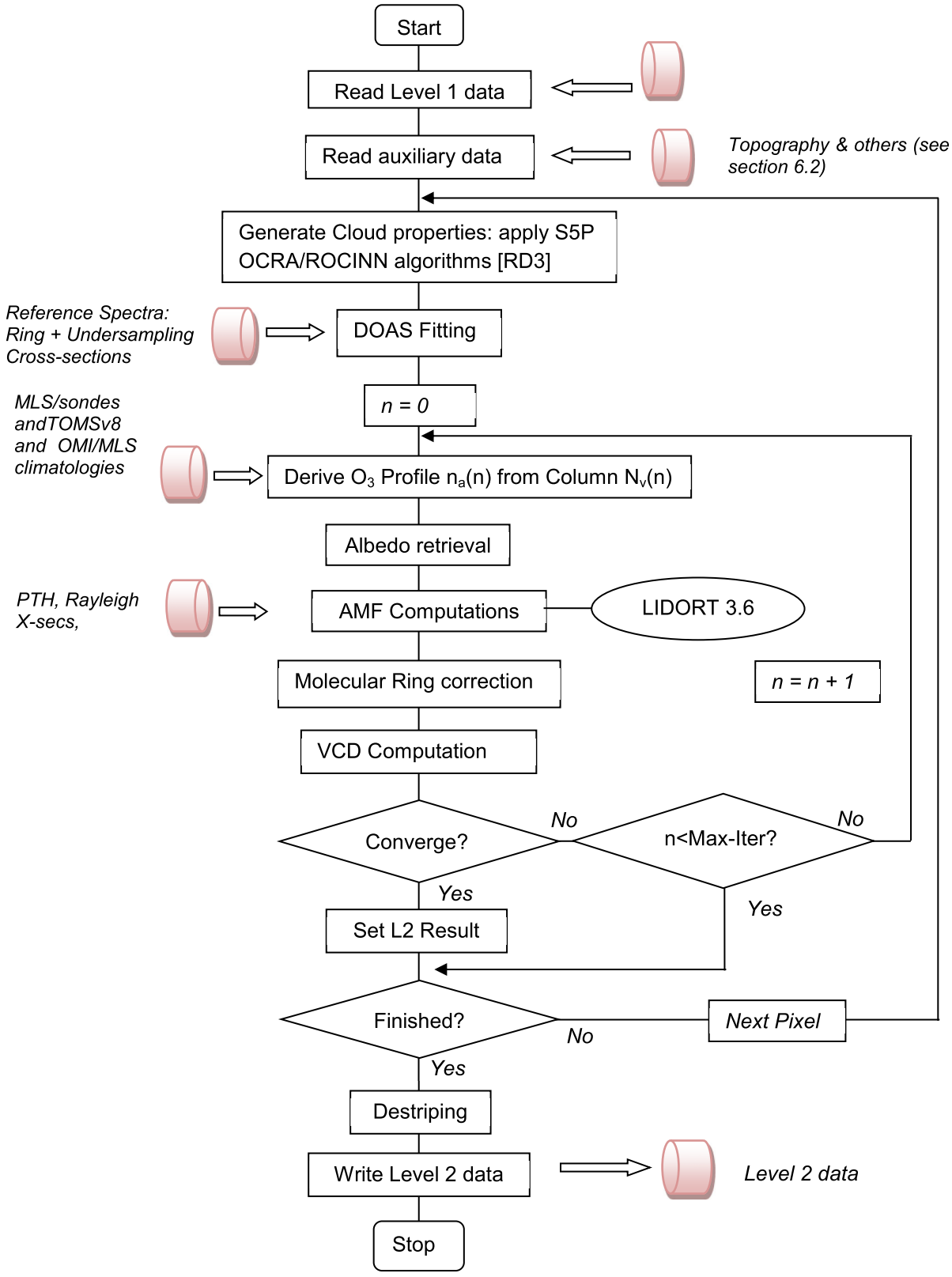

Figure 5.1: Flow Diagram for the S5P_TO3_DOAS retrieval algorithm.

Ozone cross-sections in the Huggins bands are highly temperature dependent, and are often parameterized according to a quadratic function of temperature. In earlier versions of the GOME DOAS algorithm, temperature dependence was introduced through an external choice of "effective" temperature T_{eff}, which is equal to the temperature at the altitude of the maximum number density.

A marked improvement in DOAS total ozone fitting was established with the use of two ozone cross-sections at different temperatures [*Richter and Burrows*, 2002]. Now, T_{eff} is retrieved along with the slant column, allowing the slant column to adjust to the actual O_3 profile weighted-mean atmospheric temperature [*van Roozendael et al.*, 2002]. In practice, the fitting retrieves an effective slant column N_s corresponding to ozone cross-section $\sigma(T_1)$ at temperature T_1, and a difference parameter *D* corresponding to an ozone *difference* cross-section $\Delta\sigma_{12} = \sigma(T_1) - \sigma(T_2)$. Taking cross-section temperature dependence to be linear (a good assumption in the stratosphere where temperature ranges are limited), the effective temperature may then be written $T_{eff} = T_1 + D(T_1 - T_2)/Ns$. Effective slant columns are thus independent of assumed temperature climatology. Values of T_{eff} can be compared not only with climatological data but also with analysis fields from numerical weather prediction forecast models such as that of the ECMWF (European Centre for Medium-range Weather Forecasts). Note that the slant column fitting algorithm is extracting temperature information directly from the measurements - we are not using the dynamical ECMWF data except as an aid to the algorithm verification.

The Ring effect (the filling-in of Fraunhofer solar spectral features and telluric absorption-band signatures due to rotational Raman scattering) induces a small but significant amplitude distortion in the backscattered radiance spectrum [*Grainger and Ring*, 1962]. Ring spectra may be defined as (logarithms) of the ratios of radiances with and without rotational Raman scattering (RRS). In DOAS, Ring interference is typically treated as “pseudo absorption”, based on a scaled additive Ring spectrum for which the Ring scaling factor is an ancillary DOAS fitting parameter. The simplest Ring spectrum is obtained by convolution of Raman cross-sections at a fixed temperature with a high-resolution solar reference spectrum [*Chance and Spurr*, 1997]. This Ring spectrum is often sufficiently accurate for DOAS fitting of optically thin trace gas absorbers; it was used in earlier versions of GDP.

In the Huggins bands, the telluric component is significant, and the Ring effect distortion is large enough to seriously compromise the fitting accuracy [*Sioris and Evans*, 1999]. The lack of a *molecular* Ring correction for GOME total ozone retrieval was recognized as an important source of error (systematic underestimation of ozone total columns by as much as 10%) [*Van Roozendael et al.*, 2002]. Spectral dependence in molecular Ring effect correlates quite strongly with the behaviour of the ozone absorption. A more sophisticated Ring treatment was employed in GDP 4 [*Van Roozendael et al.*, 2006], and this is summarized in section 5.2.6 below. An improved semi-empirical Ring correction was also developed for total ozone direct-fitting (section 5.3.6).

In GDP 4.x DOAS, a *molecular* Ring correction C_{Ring} is now applied to effective slant column Ns, and in the rest of the GDP 4.x algorithm, we use a corrected slant column $N_{s,corr} = Ns/C_{Ring}$ in the VCD computation. As we will see in the next section on the AMFs, C_{Ring} depends on the AMF, and hence C_{Ring} and $N_{s,corr}$ will be updated for each new AMF computation.

5.2.3 Albedo retrieval

Machine learning can be used for solving inverse problems, see for example [Loyola et al., 2016]. During the last years we developed an approach called full-physics inverse learning machine (FP_ILM) technique that was successfully applied for retrieving profile shapes from

GOME-2 (Xu et al., 2017) and retrieving SO_2 layer height from GOME-2 (Efremenko et al., 2017) and TROPOMI (Hedelt et al., 2019).

In the first step a measurement is simulated, the basic features are extracted from the simulated data and finally a Neural Network is trained to retrieve some of the input parameters form the extracted features.

The forward model has two components: first a radiative transfer model (RTM) that computes the spectral intensity as a function of the viewing geometry, atmospheric components and surface properties; and second a sensor model that transforms the RTM spectra to simulated spectra using sensor information like the instrument spectral response function and the instrument signal to noise ratio.

The forward model *F* can be used to compute simulated spectra radiances R_{sim} for a given wavelength λ as

$$R_{sim}(\lambda) \pm \varepsilon_R = F(\lambda, \Theta, \Omega, A_e, P_e) \quad 5.2$$

where ε_R denotes the expected instrument error, Θ is the light path geometry (solar and satellite zenith and azimuth angles), Ω are the atmospheric composition components, and the surface properties A_e for the geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) and P_e for the effective surface pressure.

Retrieval of trace gas, cloud and aerosol concentrations from UVN sensors requires spectrometers with sufficient spectral resolution to resolve features in the electromagnetic spectrum; therefore the fitting-window used for the retrieval of a trace gas usually contains radiances at a high-dimensional space (tens to hundreds of wavelengths). Machine learning techniques perform best with low-dimensional datasets by avoiding the effects of the curse of dimensionality.

Feature extraction is a mapping function that transforms a dataset from a high- to a low-dimensional space removing redundant information and noise. In previous FP_ILM applications [Loyola et al., 2006; Xu et al., 2017] we used principal component analysis for the feature extraction, however for the GE_LER retrieval we take advantage of the DOAS fitting results (5.1)

The feature extraction step consists in applying the DOAS fit to the simulated radiances. Combining (5.2) and (5.1) for a given fitting window Λ we obtain the following approximation with simulated datasets that representing the forward problem

$$\{N_{s,g}(\Theta), p(\Lambda)\} \cong \{F(\Theta, A_e(\Lambda), P_e)\} \quad 5.3$$

Machine learning approximates a function represented by input/output datasets using either linear or non-linear regression algorithms. Here we use artificial neural networks (NN) to learn the non-linear inverse problem by reorganizing the datasets from (5.3) to represent the inverse problem

$$\{A_e(\Lambda)\} \cong \{F_{NN}^{-1}(p(\Lambda), N_{s,g}, \Theta, \mathrm{P}_e)\} \quad 5.4$$

In other words, a neural network solves the inverse problem and retrieves the GE_LER as function of the DOAS closure polynomial p, the DOAS fitted effective slant column density N, the viewing geometry Θ and the effective surface pressure P_e. The inverse operator are the weights and biases of the neural network approximating F_{NN}^{-1}.

Obtaining the inverse operator is very time consuming mainly due to the relative large amount of RTM simulations needed to properly represent the forward problem. Finding a NN topology that learns the inverse function with a small error is also computational intensive. But all these steps are done offline and only once for a given sensor and trace gas fitting window. There is no extra computational needed for the feature extraction part as we are using the results from the DOAS fitting and the application of the NN to retrieve GE_LER is extremely fast as it only involves simple matrix multiplications.

The conversion of the DOAS effective slant column to a geometry independent total column requires the calculation of air mas factors (AMF) using either the effective scene approximation [*Coldewey-Egbers et al.*, 2005] or the independent pixel approximation (see section 5.2.4). The GE_LER can be used directly for the computation of AMFs using the effective scene approximation, whereas a LER is needed for the computation of AMFs using the independent pixel approximation.

A global gapless geometry-dependent LER (G3_LER) daily map can be easily created from GE_LER retrieved under clear-sky conditions. The SUOMI-NPP satellite mission is flying in loose formation with Sentinel 5 p with a time shift of ~5 min. The VIIRS (Visible Infrared Imaging Radiometer Suite) instrument on SUOMI-NPP has a spatial resolution of 375 m in the cloud imaging channel. The sum of all cloudy VIIRS pixel within a TROPOMI pixel divided by the number of all VIIRS data results in the respective cloud fraction. If the VIIRS cloud fraction is smaller than 0.1 the data are assumed to be cloud free. The VIIRS cloud fraction is included in the OFFL TROPOMI Cloud data. The G3_LER map for a given day is created by merging the clear-sky LER data from the same day with the G3_LER map based on the LER data from previous days. For the cloudy scenes an initial fallback map is generated as follows: The actual albedo is retrieved for each TROPOMI observation, and the viewing geometry correction is applied. For some rare cases negative outliers in the albedo are found, these data are filtered out and for each grid cell (0.1° resolution) the second smallest data is used, to avoid contamination by individual outliers. However, for some persistent clouds the local albedo might still be affected by clouds, using a longer time period this issues will hopefully be solved soon.

If the albedo of a pixel is updated the orbit number is also stored. The updated albedo will be used until it is either updated again or if the pixel is not updated for more than 30 days (450 orbits) the fallback data will be used again.

It is important to note that the GE_LER takes into account the bidirectional reflectance distribution function (BRDF) effects as it is based on radiative transfer model simulations using the actual viewing geometry. But when combining GE_LER data their BRDF dependencies $\rho(\Lambda, \theta, \psi)$ as function of the wavelength in the fitting window Λ, the viewing zenith angle θ, and the surface types ψ must be considered. The function can be easily obtained separately for the used fitting windows Λ, different surface types ψ (land, water, snow/ice) and time periods (e.g. monthly) by fitting a polynomial of clear-sky LERs averaged as function of θ.

The G3_LER daily map contains normalized LER, i.e. GE_LER retrieved under clear-sky conditions divided by the fitted BRDF dependency, as well as the multiplicative factors $\rho(\theta)$ to compute the geometry-dependent LER as a function of the actual satellite viewing zenith angle θ.

It is necessary to aggregate normalized LER retrievals over several days (between one to four weeks depending on cloudiness) in order to obtain a global gapless map. In contrast to LER climatologies, the G3_LER map represents the actual surface properties as it is updated on a daily basis.

5.2.4 AMF computation and VCD corrections

The AMF definition is

$$\mathrm{M} = \frac{\ln(\mathrm{I_{nog}}/\mathrm{I_g})}{\tau_\mathrm{v}}, \qquad 5.5$$

a quantity requiring two radiative transfer calculations of backscatter intensity: one (I_g) for an atmosphere including ozone, the other (I_{nog}) for an atmosphere excluding ozone absorption; τ_v is the vertical optical depth of ozone for the whole atmosphere. Other equivalent definitions have been suggested for optically thin trace gases such as HCHO and NO_2; see for example [Palmer et. al., 2001].

Simulated intensities in Eq. (5.5) are computed using the radiative transfer models LIDORT (no polarization) or VLIDORT (including polarization). The scalar model LIDORT has been the default for GDP, with the (for the most part, smoothly varying) polarization signature included in the closure polynomial. For fuller accuracy, it is ultimately better to include the effects of polarization in the UV (where Rayleigh scattering predominates) using VLIDORT. For the S5P DOAS, we continue to use LIDORT for the main calculation, but include a polarization correction in the form of a look-up table based on off-line VLIDORT computations. For more details on this discussion, see section 5.4.1.

For a clear sky scenario, the VCD is $\mathrm{N_v} = \frac{\mathrm{N_s}}{\mathrm{M_{clear}}}$, where $\mathrm{N_s}$ is effective slant column, and $\mathrm{M_{clear}}$ the clear-sky AMF. For partially cloudy scenes, AMF computations rely on the independent pixel approximation (IPA), and the VCD is then given by:

$$N_v = \frac{N_s}{(1-\Phi)M_{clear}+\Phi M_{cloud}} \qquad 5.6$$

where M_{cloud} is the AMF for the cloudy atmosphere down to the surface, computed from the input choice of ozone profile. Here, $\Phi = f_c\, I_{cloud}/I_{total}$ is the effective "intensity-weighted" cloud fraction, in which the total intensity is $I_{total} = (1-f_c)I_{clear} + f_c I_{cloud}$ for effective cloud fraction f_c.

For RT calculations of M_{cloud}, we require cloud-top height and cloud optical thickness. Cloud properties are retrieved with dedicated cloud retrievals such as generated by the OCRA/ROCINN algorithms (section 5.4.5); the application to the S5P_TO3_DOAS algorithm is discussed below in section 5.2.5.

In traditional DOAS ozone retrievals, slant column fitting and AMF calculation steps are decoupled; AMF simulations are then based on seasonally and geographically classified climatological profile inputs. However, the shape and total content of the selected ozone profile may bear little resemblance to the true profile, and (particularly for scenarios with high ozone content and/or high solar zenith angles) the AMF may then be incorrect due to a poor choice of input profile.

The *iterative* AMF approach [*Spurr et al*., 2005] circumvents this issue by using the fitted slant column result N_s to make an adjustment to the AMF (and hence the VCD) that reflects the trace gas content as expressed in the value of N_s. This adjustment depends on the use of an ozone profile climatology that is *column-classified*: for a given time and location, the choice of profile is then uniquely determined by the VCD. We use a linear one-to-one column-profile map based the climatological database developed for the TOMS Version 8 total ozone retrieval [*Bhartia et al.,* 2003].A (clear-sky) AMF M_0 is computed (with the RT model) using profile $n_{a,0}$ corresponding to the initial choice $N_{v,0}$ of vertical column density for ozone. The first guess $N_{v,0}$ may be taken from climatology (e.g. the TOMS zonal-mean column climatology [*McPeters et al.*, 2007]), or from a previously-retrieved result from an

adjacent footprint (this latter option is usually the default). Then, the value M_0 is used to compute a molecular Ring correction $C_{ring,0}$. Given the slant column N_s from the DOAS fitting, we compute the correction $N_{s,corr,0} = N_s/C_{ring,0}$. An updated vertical column $N_{v,1}$ is calculated through the relation $N_{v,1} = N_{s,corr,0}/M_0$; V_1 selects a new profile $n_{a,1}$, which then results in a new AMF result M_1. A new Ring factor $C_{ring,1}$ and a new corrected slant column $\mathrm{N_{s,corr,1}} = \mathrm{E_s/C_{ring,1}}$ are then calculated. The next guess $\mathrm{N_{v,2}}$ for the vertical column follows from $\mathrm{N_{v,2}} = \mathrm{N_{s,corr,1}/M_1}$, and so on. This process is repeated until convergence is attained (absolute relative difference between iterations of $\mathrm{N_v}$ is less than some small number 10^{-3}, for example). For the partially cloudy footprint, the iteration proceeds via iterative use of Eq. (5.6), and we note that the ghost column is also updated at each iteration. For the great majority of cases, convergence for $\mathrm{N_v}$ is rapid (3 or 4 iterations for a relative change of 0.1% in the vertical column in [DU]).

The number of RT simulations for this DOAS-style algorithm is small. GDP 4 uses "on-the-fly" LIDORT calculations to generate AMFs, obviating the need for large multidimensional look-up tables (LUTs) which are time-consuming to create and cumbersome to manipulate. For GDP versions up to and including 3.0, the ozone AMF was always calculated at 325.0 nm, at the lower end of the DOAS fitting window. In GDP 4.0+, we use 325.5 nm as the baseline – it has been shown that computing the AMF at this higher wavelength minimizes the error caused by the DOAS approach itself for solar zenith angles larger than 80° [Van Roozendael et al., 2002]. For the Sentinel 5P analysis the AMF is calculated at 328.125 nm thereby reducing the error, even more.

For clear-sky RT simulations, the static surface albedo climatology in GDP 3.0 was replaced in GDP 4 with a dynamic albedo data set derived from accumulated satellite reflectance data. The GOME Lambertian equivalent reflectivity (LER) data set of albedos prepared from 5.5 years of reflectivity data [*Koelemeijer et al.,* 2003] was used in GDP 4 in conjunction with Nimbus-7 TOMS LER values (N7-TOMS) prepared from 14.5 years of data from 1978 [*Herman and Celarier,* 1997], and valid for 340 and 380 nm. The GOME LER data has monthly and yearly entries on a 1°x1° latitude/longitude grid, at 12 different wavelengths spanning the GOME range; the N7-TOMS data is also monthly. We use GOME LER data at 335 and 380 nm, and N7-TOMS LER data at 380 nm; the desired albedo is $\mathrm{A_s}(\lambda) = \mathrm{s}(\lambda)\mathrm{A_{s,TOMS}}(380)$, with scaling $\mathrm{s}(\lambda) = \mathrm{A_{s,GOME}}(\lambda)/\mathrm{A_{s,GOME}}(380)$ and $\lambda = 335$ nm for total ozone fitting [*Boersma et al.*, 2004]. In S5P we first used the OMI based monthly climatology with a resolution of 0.5x0.5° [*Kleipool et al.*, 2008] for version 1. With version 2 this has been replaced by the albedo retrieval described in section 5.2.3.

We note that the AMF calculations in this algorithm are based on the selection of a temperature profile (by latitude and season) from static temperature climatology such as the one provided for the TOMS Version 8 climatology.

In the DOAS slant column fitting, it is implicitly assumed that the broad-band signatures such as polarization and scattering are subsumed by the closure polynomial. Nevertheless it has been shown [Lerot et al., 2014]) that polarization introduces structures correlating with the ozone signature. This is taken into account by applying polarization correction factors to the simulated spectra required for the AMF computation (see section 5.4).

5.2.5 Cloud treatments

As noted already, we use the IPA to treat clouds as a first-order correction in this algorithm. Cloud optical properties come from the OCRA/ROCINN algorithms [Loyola et al., 2018, Schuessler et al., 2014]. OCRA provides the cloud fraction using the broad-band polarization

measurements, and ROCINN provides cloud-top height and cloud optical thickness from measurements in and adjacent to the oxygen *A*-band around 760 nm.

The S5P baseline is to treat clouds as scattering layers (CAL), the AMF simulations are based on ROCINN-CAL output. In this case it is not necessary to use a ghost column in the VCD formulation, because the cloudy-scene RT calculation is based on the entire ozone profile down to the surface - intra-cloud ozone absorption is treated implicitly. For the snow/ice mode, the cloud fraction is again assumed to be 1.0, and the retrieved cloud-top height and optical depth represent the effective values for the underlying scene.

5.2.6 Molecular Ring effect auxiliary algorithm for DOAS

The molecular Ring correction is based on a simplified forward model of the intensity $\mathrm{I}(\lambda)$ at the satellite; this includes an explicit contribution due to RRS, as described by:

$$I(\lambda) = E_0(\lambda)\exp\left[-N_{s,O3}\sigma_{O3}(\lambda) - P(\lambda)\right] + N_{s,Ring}I_0^{RRS}(\lambda)\ \exp\left[-N_{s,O3}^{RRS}\sigma_{O3}(\lambda)\right] \qquad 5.7$$

The first term in 5.7 on the right-hand side is the Lambert-Beer extinction, with $E_0(\lambda)$ the solar intensity, $P(\lambda)$ the closure polynomial, and $\sigma_{O3}(\lambda)$ and $N_{s,O3}$the ozone absorption cross-section and effective slant column respectively. The second term contains an approximation of the Ring effect. First, the Raman light is assumed to be produced close to the surface, with a spectral shape given by a source spectrum for Raman scattering $I_0^{RRS}(\lambda)$. This source spectrum only treats the spectral smoothing effect of RRS on the solar intensity, and is calculated by convolution of a GOME irradiance spectrum using rotational Raman cross sections appropriate to inelastic scattering into the wavelength of interest.

The fractional intensity of Raman light (the $N_{s,Ring}$ parameter) is freely adjustable; this will depend on parameters such as cloud coverage, cloud altitude and surface albedo. Ozone absorption as expressed by the term $N_{s,O3}^{RRS}\sigma_{O3}(\lambda)$ is then treated consistently, assuming that Raman photons produced at the surface and/or above clouds travel upward to the satellite. Ozone absorption taking place in the incoming light is assumed to be fully smeared out in the inelastic process.

In the DOAS context, we expand the Ring correction term to first-order, and write:

$$\ln\left[\frac{I(\lambda)}{E_0(\lambda)}\right] = -N_{s,O3}\sigma_{O3}(\lambda) - P(\lambda) - N_{s,Ring}\sigma_{Ring}(\lambda); \quad \sigma_{Ring}(\lambda) = -\frac{I_0^{RRS}(\lambda)}{F_0(\lambda)} \qquad 5.8$$

Equation (5.8) has the same form as the familiar DOAS forward-model law, from which $N_{s,O3,}$ $N_{s,Ring}$ and polynomial coefficients can be determined in the usual manner. The major new feature comes with the definition of the modified effective slant column $N_{s,O3}$, which is related to the effective slant column for *elastic* scattering ($N_{s,corr,O3}$) by the following formula:

$$N_{s,O3} \cong N_{s,corr,O3}\left[1 + N_{s,Ring}\bar{\sigma}_{Ring}\left(1 - \frac{\sec\theta_0}{M_{total}}\right)\right] = N_{s,corr,O3}C_{Ring}, \qquad 5.9$$

where M_{total} is the ozone AMF, θ_0 the solar zenith angle, and $\bar{\sigma}_{Ring}$ an average Ring cross-section calculated over the spectral fitting interval. In this formulation, the DOAS fitting is essentially unchanged; it gives fitted parameters $N_{s,O3}$ and $N_{s,Ring}$. The effective slant column for ozone is then adjusted ***after*** the fit through the relation $N_{s,corr,O3} = N_{s,O3}\left[C_{Ring}\right]^{-1}$. Note that the molecular Ring term C_{Ring} can also be used to quantify the error due to an incorrect estimation of the Ring effect in previous GDP versions.

Eq. (5.9) defines the semi-empirical Ring correction for the DOAS algorithm; the definition requires knowledge of AMFs. Studies have shown that for moderate solar zenith angle (SZA), the geometrical AMF is sufficiently accurate to approximate $\mathrm{M_{total}}$. For high SZA with

a long path through absorbing ozone layers, a more precise calculation is needed. Here, we use the total AMF already computed at each AMF/VCD iteration step – thus the molecular Ring correction must also be applied at each iteration to the DOAS slant column result before a new estimate is made for the VCD.

5.2.7 Destriping algorithm

The previous versions of the GDP 4.x algorithm were applied to the satellite instruments GOME, SCIAMACHY and GOME-2. These instruments use a scanning optical system, the earth radiance from different ground pixels is reflected by a scanning mirror to the same entrance slit and detected by the same 1-dimensional detector. TROPOMI on S5P uses an imaging approach. The incoming radiance from different ground pixel (perpendicular to the flight direction) is imaged on different parts of the spectrometer and different lines of a 2-dimensional detector. One dimension (column)is used for the wavelength and the other one (row) for the ground pixel. Therefore, the instrumental parameters such as slit function and wavelength calibration depend on the ground pixel. Although the slit function has been measured and optimized separately for each row, intrinsic calibration uncertainties remain, which may impact the retrievals and lead to a small systematic row dependency of the total ozone.

During the first month of operation, such an effect was observed for TROPOMI. Therefore, after the last iteration of the AMF and VCD algorithm (Sec. 5.2.4), the vertical columns were multiplied by a fixed correction array. To avoid the influence of natural stratospheric variability on the correction factor we averaged the tropical total ozone column for each detector row between 15° south and 15° north from January 1^{s}, 2018 to April 2nd, 2018. The respective mean columns were normalised to the total mean of the tropical columns from all ground pixels. A decrease of roughly 1% was observed between rows 415 and 450. The final correction array (Figure 5.2) is obviously the inverse of the row dependent mean values. The destriping correction was implement for version 1 based on the preoperational data. The correction factor did not change until version 2.2. Initial comparisons of S5P_TO3_DOAS with other observations including S5P_TO3_GODFIT and ground-based measurements show a slight overestimation, therefore it was decided to also apply a general offset correction of -1.5%. This destriping correction is only applied to the NRTI product.

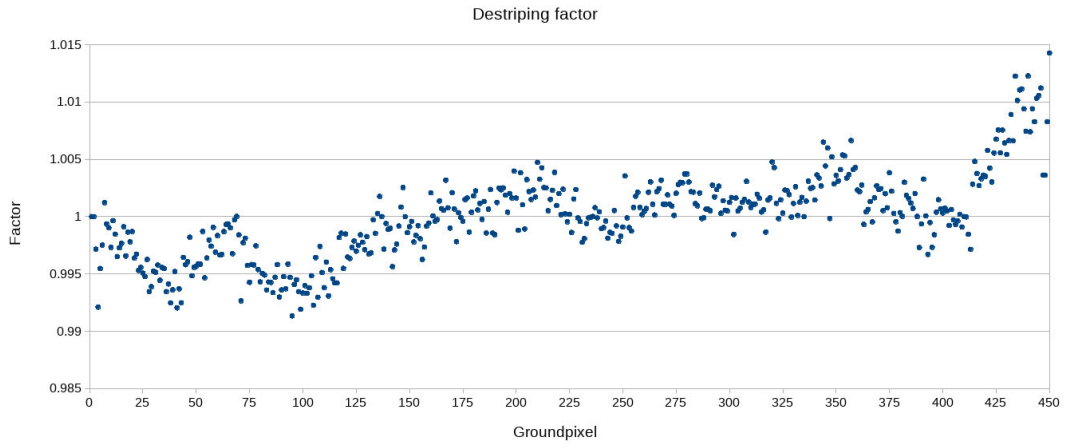

Figure 5.2: Destriping correction factor for all ground lines

5.3 S5P_TO3_GODFIT, OFFL total ozone (GDP 5 Heritage)

5.3.1 Overview

S5P_TO3_GODFIT is based on the GODFIT direct fitting algorithm [*Lerot et al.,* 2010 and 2014; *van Roozendael et al.*, 2012]. The algorithm uses an iterative least-squares cost function minimization method, based on the differences between satellite-measured and model-simulated radiances. Heritage goes back to the analysis of ozone column measurements from 1979-1986 Nimbus-7 data [*Joiner and Bhartia*, 1997].

Least-squares minimization algorithms also require forward model simulations of radiance sensitivity functions (Jacobians, or weighting functions) which are the partial derivatives of the radiances with respect to state vector parameters that are to be retrieved. For the forward model component, GODFIT uses the LIDORT radiative transfer model [*Spurr*, 2008] for multiple-scatter RT simulation of satellite radiances and analytically calculated Jacobians with respect to the total O_3 column, the temperature shift and a number of albedo closure coefficients. In addition to these parameters, retrieval state vector also has a parameter for Ring-effect correction, and for a wavelength registration shift. Unlike S5P_TO3_DOAS, S5P_TO3_GODFIT is a one-step algorithm; there is no separation of slant-column fitting and AMF conversion. Although the baseline is to use on-the-fly RT simulations, an alternative to have faster retrievals is to extract radiances and Jacobians from pre-computed tables.

For each satellite orbit data, the L1 wavelength grid is first recalibrated to have the required very high accuracy for the alignment between the earthshine and reference irradiance spectra on one hand and the cross-sections on the other. This recalibration consists in aligning the solar Fraunhofer lines in the L1 sun spectrum to the position provided by a reference solar atlas [*Chance and Kurucz,* 2010] degraded to the resolution of the instrument. Once this is done, total ozone can be inverted for each measured radiance.

The flowchart in Figure 5.3gives an overview of the inversion procedure. Following the initial reading of satellite radiance and irradiance data, and the input of auxiliary data (topography fields, optional temperature profiles, fractional cloud cover and cloud-top-height), the iteration counter is set (n=0), and an initial guess is made for the state vector (total ozone amount, temperature shift, closure coefficients, etc.). A unique ozone profile *P(n)* is then constructed from the total column estimate *C(n)*, using a 1-1 column-profile map based on column-classified ozone profile climatology. For this, we use the climatological database by *Labow et al.* [2015] and constructed using MLS and sondes observations, combined with the tropospheric climatology based on OMI and MLS measurements published by *Ziemke et al.* [2011]. Next, pressure, temperature and height profiles are constructed; this is where the current value of the temperature shift *S(n)* is applied.

The algorithm then enters the forward model step, in which optical properties are created and the LIDORT model is called to deliver top-of-atmosphere (TOA) radiances $I(n)$, and a vector $\mathbf{K}(n)$ containing the associated ozone column, albedo, T-shift and other weighting functions. These simulated quantities are then corrected for the molecular Ring effect. Next, the inversion module (variable-regularization Levenberg-Marquardt least squares, with line-searching) yields a new guess for the ozone column and ancillary state vector parameters. The iteration stops when suitable convergence criteria have been satisfied, or when the maximum number of iterations has been reached. The total ozone column and other diagnostic errors are computed directly from the inverse variance-covariance matrix.

One option if faster retrievals are needed is to extract simulated spectra from a lookup table (LUT) of precomputed radiances instead of performing online LIDORT simulations. When doing this, the optical properties do not have to be computed and the calls to the RT model LIDORT are replaced by interpolation procedures through the LUT using directly the state

vector variables as input in addition to the geolocation parameters. The radiance LUT was computed using the same forward model as the online scheme in order to have full consistency between the two approaches. However, the available computational infrastructure allows retrievals based on online simulations, which guarantees the highest accuracy level. Therefore, the baseline in version 2 remains the online calculation.

If the maximum number of iterations is reached, there is no established convergence and the final product is not meaningful. In this case the corresponding entry in the Level 2 product will be flagged as invalid.

5.3.2 Forward model: Temperature shifting

Temperature profiles are required for an accurate simulation of the radiative transfer and of the ozone optical depths. In S5P_TO3_GODFIT, temperature profiles are taken from an external source. One such source is the *static* 10° latitude-zone 12-monthly temperature climatology supplied with the TOMS Version 8 ozone profiles [*Bhartia*, 2003, *McPeters et al.,* 2007]. Although temperature profiles can be extracted from meteorological analysis fields such as those provided by ECMWF, it is more desirable to adjust the temperature profile to reflect dependence of the ozone absorption signature on temperature.

In this way, temperature information is obtained directly from the measurements, with the static TOMS climatology used to generate the initial temperature profile before adjustment. Thus, no *dynamic* temperature fields (e.g. ECMWF) are required in the operational S5P GODFIT algorithm. The ECMWF fields are regarded as one of the best source of temperature information, and these fields have been used extensively in the development and verification of both the DOAS and GODFIT types of algorithm. This will become apparent in the following paragraphs.

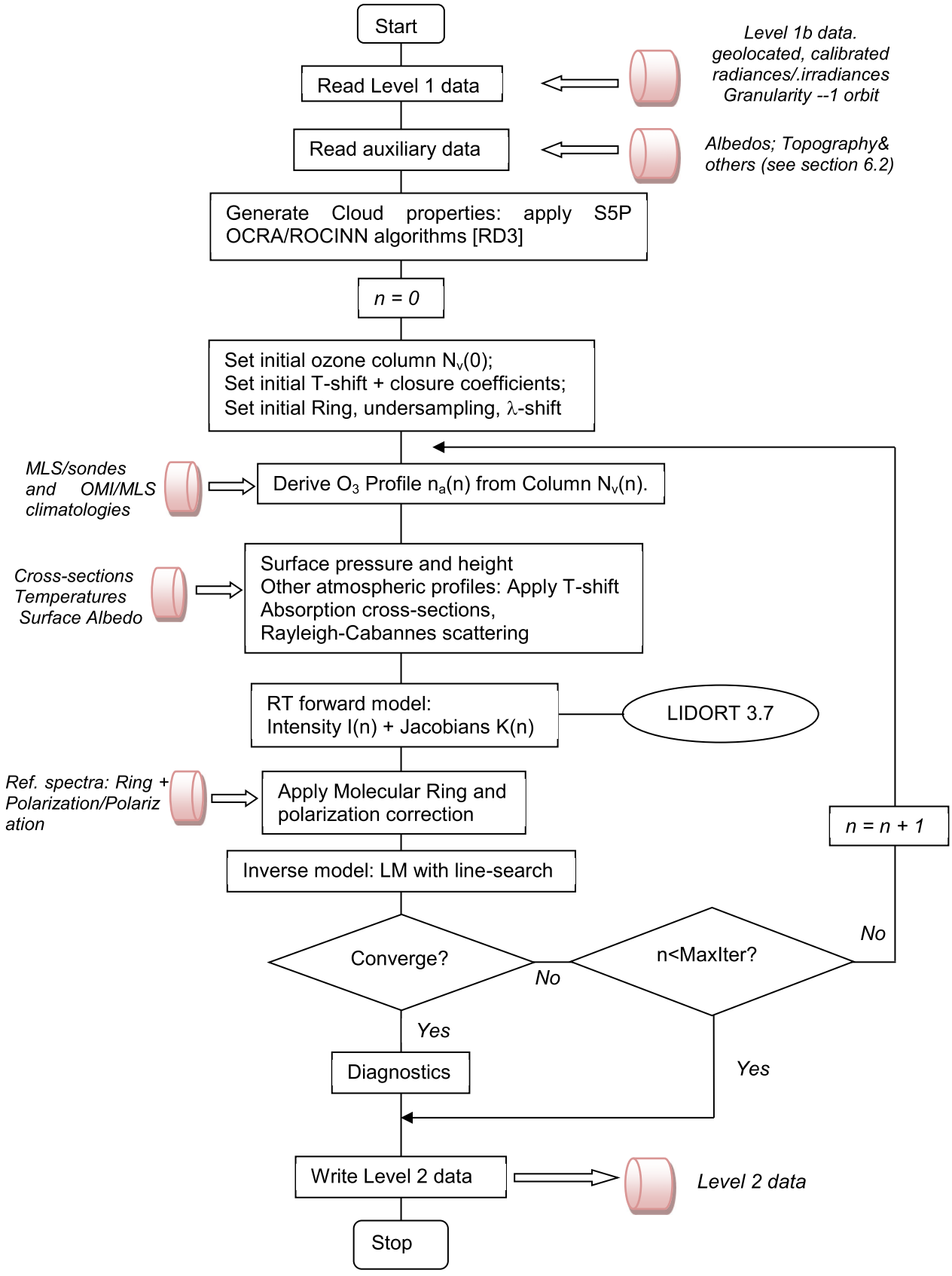

Figure 5.3: Flow Diagram of the S5P_TO3_GODFIT direct fitting retrieval algorithm.

:

The “effective temperature” determined from GDP 4.x DOAS fitting has been described in section 5.2.2. In S5P GODFIT, we define the effective temperature through $T_{eff} \sum_n n_{a,n} = \sum_n T_n n_{a,n}$, that is, from the input temperature profile $\{T_n\}$ weighted with the layer ozone column amounts $\{n_{a,n}\}$. Comparing effective temperatures calculated using different Temperature profiles (e.g. TOMS and ECMWF), we find a clear almost-linear correlation between fitted ozone columns and differences between the T_{eff} differences; a difference of 10 K in effective temperature results in a 3% change in total ozone. This originates from the significant temperature-dependence of the ozone absorption in the Huggins bands and points to the importance of the temperature shift adjustment to improve total ozone accuracy.

The ozone profile $\{n_{a,n}\}, n = 1, \cdots N_L$ (the number of layers) depends on input total column N_v. A set of temperatures $\{t_n\}$ are specified at pressure levels $\{p_n\}, n = 0, \cdots N_L$. We correct the temperature profile with a shift amplitude S by writing $\theta_n = t_n + S\phi_n$, where S is treated as a free parameter to be retrieved, and ϕ_n is a pre-specified *temperature shape function*. We compute column air density $a_n(S)$ (in [mol/cm^2]) as a two-point integration of the profile air density (in [mol/cm^3]) over the layer height thickness d_n, and thence obtain the partial derivative of $a_n(S)$ with respect to the T-shift:

$$a_n = \frac{Rd_n}{2}\left[\frac{p_{n-1}}{\theta_{n-1}} + \frac{p_n}{\theta_n}\right]; \qquad \frac{\partial a_n}{\partial S} = -\frac{Rd_n}{2}\left[\frac{\phi_{n-1}p_{n-1}}{\theta_{n-1}^2} + \frac{\phi_n p_n}{\theta_n^2}\right]. \tag{5.10}$$

Here, R is a constant proportional to Loschmidt’s number.

Temperature-dependent ozone cross-sections $\sigma_n(S)$ are thus functions of the T-shift S. Indeed, for the well-known parameterized quadratic temperature dependence [*Bass and Paur*, 1984] based on three coefficients $\{\alpha_0, \alpha_1, \alpha_2\}$, we have ($\theta^* = 271.15$ K):

$$\sigma_n = \alpha_0 + \alpha_1(\theta_n - \theta^*) + \alpha_2(\theta_n - \theta^*)^2; \quad \frac{\partial \sigma_n}{\partial S} = \phi_n[\alpha_1 + 2\alpha_2(\theta_n - \theta^*)]. \tag{5.11}$$

Since we are retrieving S, the T-shift derivatives in Eqs. (5.10) and (5.11) will be required in the generation of linearized optical property inputs for LIDORT to output Jacobians with respect to *S*; we deal with this below. The default temperature shape function is a constant throughout the whole atmosphere; that is, $\phi_n = 1$ in all layers. However, other choices such as box or triangle functions within specified altitude ranges in the atmosphere are possible. (Results were found to be weakly dependent on the choice of shape function).

Studies have shown that the T-shift fitting is robust; the retrieved effective temperature is virtually independent of any prescribed *a priori* value. In general, effective temperatures retrieved with T-shift adjustments are close to the values derived with ECMWF temperature profiles. Also, the T-shift procedure applied to TOMS temperature profiles generally leads to small differences with respect to the ozone columns derived using ECMWF temperatures. Thus, ozone columns are consistent and unbiased, no matter the choice of the *a priori* temperature profile. This is an important remark, as it means that the ECMWF profiles are not required as a *sine-qua-non* for accurate total columns. Consequently, S5P GODFIT uses the climatological TOMS temperature profiles as *a priori* in the total ozone retrievals.

Finally, we note that introduction of the additional T-shift parameter to the fitting could imply a minor loss in precision for the fitted ozone product. For the 325-335 nm fitting window, this loss is small (compared with the DOAS precision), and is offset by the greater accuracy of the ozone product. This issue will be revisited.

5.3.3 Forward model, LIDORT Jacobians

In this algorithm, simulation of backscattered radiance and retrieval-parameter Jacobians is done using LIDORT, which generates analytic Jacobians (weighting functions) with respect to atmospheric and surface properties. The LIDORT summary is in section 5.4.1 below, and this should be referred to in order to follow the rest of this sub-section.

LIDORT requires as input a set of *inherent optical properties* (IOPs); these are (for each layer *n*) the total extinction optical thickness Δ_n, the total layer single scatter albedo ω_n and the set of Legendre expansion coefficients β_{ln} for the total phase function. Inputs $\{\Delta_n, \omega_n, \beta_{ln}\}$ are constructed from atmospheric profiles of temperature, pressure, air density and trace gas distributions, Rayleigh (molecular) scattering parameters and trace gas absorption cross-sections, and (if present) aerosol extinction and scattering properties.

In S5P_TO3_GODFIT, we require atmospheric-property Jacobians with respect to the total ozone column N_v and the T-shift S, and a surface-property Jacobian with respect to the Lambertian albedo. For the N_v and S Jacobian output, we must specify a set of *linearized IOP inputs* for LIDORT to generate these Jacobians, that is, we require $\left\{\frac{\partial\Delta_n}{\partial N_v}, \frac{\partial\Delta_n}{\partial S}\right\}$ and $\left\{\frac{\partial\Delta_n}{\partial N_v}, \frac{\partial\Delta_n}{\partial S}\right\}$. In the Rayleigh-only case (no clouds/aerosols), there are no derivatives of β_{ln} (this is not the case with clouds and/or aerosols present). Linearized IOP inputs may be defined for other model parameters (such as the aerosol profile) that are sources of uncertainty in the fit; for examples appropriate to ozone profile modelling in the UV, see [*van Oss and Spurr*, 2002].

The derivatives with respect to N_v come through the dependence $\frac{\partial n_{a,n}(N_v)}{\partial N_v}$ which is a reflection of the ozone profile-column mapping; details of this dependence are given in section 5.4.2 below. Derivatives with respect to S will emerge through application of the cross-section and air-density dependences as noted above in Equations 5.10 and 5.11. Derivations of IOPs and their linearizations are given in section 5.4.1.

Surface property Jacobians are discussed next.

5.3.4 Forward model: Albedo internal/external closure; Cloud treatment

In S5P GODFIT, aerosol scattering and absorption and surface reflectivity can be brought together in an *internal albedo closure term* that is fitted internally in the sense that coupling between the surface and tropospheric scatterers is treated properly in a full multiple scattering context. In this case, direct fitting makes a simultaneous determination of an *effective* wavelength-dependent albedo in a *molecular* atmosphere, thus avoiding introduction of a lot of extraneous (and highly uncertain) aerosol information into the algorithm.

Assuming the surface albedo $A_s(\lambda) = \gamma_0 + \gamma_1(1-\lambda/\lambda_0) + \gamma_2(1-\lambda/\lambda_0)^2$ is a 3-parameter quadratic function of wavelength, then we can define Jacobians with respect to fitted parameters γ_0, γ_1 and γ_2 as follows:

$$K_0 = K_R(\lambda); \quad K_1 = (1-\lambda/\lambda_0)K_R(\lambda); \quad K_2 = (1-\lambda/\lambda_0)^2 K_R(\lambda); \qquad 5.12$$

where $K_R(\lambda)$ is the albedo weighting function obtained from LIDORT. We assume first guess values $\gamma_1 = \gamma_2 = 0$, and an initial value for γ_0 is taken from the albedo database.

The direct fitting also has an option for *external albedo closure*; this is preferred for retrievals of Level 1 data having significant radiometric calibration uncertainty or even unknown radiometric calibration. The external closure is a multiplicative polynomial that is applied to the LIDORT radiance simulation; the closure is thus an adjustment to the effect of broad-scale effects due to surface reflection and molecular scattering.

[In DOAS, the external polynomial closure represents the sum of all broad-scale radiative effects, which are decoupled from the Beer-Lambert absorption].

The external closure operates in the following manner. A LIDORT simulation is performed with a fixed albedo $A_{s,0}$; there is no Jacobian with respect to this variable. Then the LIDORT output is scaled with the molecular Ring correction – call this result the "corrected" LIDORT radiance $I_{LC}(\lambda, A_0)$. The complete simulated TOA radiance is then $I_{sim}(\lambda) = I_{LC}(\lambda, A_{s,0})R_{ext}(\lambda)$, where the external closure polynomial $R_{ext}(\lambda)$ can again be taken as quadratic with fitted parameters γ_0, γ_1 and γ_2. The weighting functions for these parameters are then given by $K_J = (1 - \lambda/\lambda_0)^{J-1} I_{LC}(\lambda, A_0)$ for $J = 0, 1, 2$. Note that this external closure is applied after the molecular Ring correction, but before the addition of an (optional) undersampling reference spectrum.

When dealing with clouds in this algorithm, we use the effective scene approach. The observed scene can be treated as a single effective surface, located at an altitude resulting from the cloud fraction weighted mean of the ground and cloud altitudes (Coldewey-Egbers, et al. 2005). The effective surface albedo is retrieved simultaneously to the total ozone column using the internal closure mode of GODFIT. We found that this approach minimizes the impact of cloud contamination on the retrieved ozone columns, especially for high clouds and it has been consequently adopted in the current version of the algorithm. The cloud fraction and cloud altitudes used to compute the effective scene altitude are provided by the OCRA/ROCINN CRB parameters. In addition, the use of an effective scene strategy has the advantage to only require one forward simulation instead of two as is the case with the IPA.

5.3.5 Lookup tables of LIDORT sun-normalized radiances

The goal of the lookup table (optional if faster retrievals are needed) approach is to replace the online radiative transfer calculation by an interpolation of precalculated radiances. Therefore, we construct a multi-dimensional lookup table of radiances as a function of all varying parameters that enter the LIDORT simulation: the fitted parameters (total ozone column and the ancillary fitting parameters scene albedo and temperature shift), angles describing the observation geometry, surface pressure, as well as latitude, by which we select the appropriate temperature and ozone profile shapes from the TOMSv8 and MLS/sonde climatologies (see section 5.4.2.). The tabulated radiances are then calculated for a fixed wavelength grid spanning the 325nm-335nm range at 3 times the instrument sampling rate, using cross sections convolved with the instrument's slit function. Note that, to maintain consistency with the online simulation mode of the processor, LIDORT is used in scalar mode to construct the radiance look-up table. Polarization correction factors are applied afterwards as explained in section 5.4.1.

The forward model calculation for a set of parameter values now becomes an interpolation of the radiances at surrounding grid points. For the total ozone column and solar zenith angle, we use quadratic interpolation through 3 surrounding grid points. For the other dimensions of the table, linear interpolation is sufficient. This results in an interpolated radiance as a function of the lookup table's wavelength grid, which is then resampled onto the wavelength grid of the observed spectrum using cubic spline interpolation. The derivative of this interpolation procedure produces the needed Jacobians.

In order to keep the interpolation procedure simple and efficient, the LUT uses a wavelength-independent scene albedo. Within the inversion procedure, only a wavelength-independent albedo is fitted, and the possible wavelength dependence of the spectrally-smooth variation of the measured radiance is taken into account via the fit of a polynomial of which the constant term is neglected.

In order for the lookup table approach to be faster than the online algorithm, frequent hard disk access must be avoided. Because all forward model parameters, except for the time of year, vary rapidly within a single orbit file, this restriction translates into the requirement that the radiances for the full range of those parameters fit in memory. This puts a limit on the density of the table's parameter grid, and some experimentation is necessary to obtain a grid which fits in memory and produces accurate interpolation results over the whole parameter space. To save space, the parameter grid does not include a longitudinal dimension. The precalculated radiances are therefore based solely on the MLS/sonde profile database, which has no longitudinal dependence, and do not take into account the tropospheric climatology OMI/MLS, which would be used in the online approach. After the retrieval, we use the averaging kernels (see section 7.1.2) and the difference between the profile used in the lookup table and a more accurate profile which matches the tropospheric column prescribed by the OMI/MLS climatology, to apply a correction to the retrieved total column. Using these techniques, we have managed to construct a lookup table which reproduces the retrieved columns of the online algorithm with an accuracy better than 1%, and a tenfold performance improvement. Note that the computational resources of the S5p operational environment allowed finally to use online simulations of the radiances. Therefore this look-up approach is only used in case of need of very fast processing.

5.3.6 Ring effect auxiliary algorithm

The semi-empirical treatment of molecular Ring effects for the direct fitting case has some similarity with that for the DOAS algorithm, but there are sufficient differences to justify a separate discussion here (Lerot et., 2014). The elastic-scattering sun-normalized radiance $\mathrm{I_{LIDORT}}(\lambda)$ simulated by LIDORT can be corrected for inelastic scattering using the expression:

$$I'(\lambda) = I_{LIDORT}(\lambda)\{1 + \mathrm{FF}(\lambda)\} \quad 5.13$$

Here, $\mathrm{FF}(\lambda)$ is the so-called Ring Filling-Factor, which can be expressed as follows [Lerot et al., 2014]:

$$\mathrm{FF}(\lambda) = \mathrm{C_{ring}}\left[\sigma_{\mathrm{ring}}(\lambda)\exp\left(-\tau^{\mathrm{v}}_{\mathrm{O_3}}(\lambda)\mathrm{M_{out}}(\lambda) - \tau^{\mathrm{RRS,\,v}}_{\mathrm{O_3}}(\lambda)\mathrm{M_{in}}(\lambda) + \tau^{\mathrm{v}}_{\mathrm{O_3}}(\lambda)\mathrm{M_{tot}}\right) - 1\right]. \quad 5.14$$

In this expression, the terms M_{tot}, $M_{\mathrm{in}}(\lambda)$ and $M_{\mathrm{out}}(\lambda)$ represent effective air mass factors respectively for the total, incoming and outgoing light paths. The ozone optical thicknesses $\tau^{\mathrm{v}}_{\mathrm{O_3}}$ and $\tau^{\mathrm{RRS,\,v}}_{\mathrm{O_3}}$ are defined as follows:

$$\tau^{\mathrm{v}}_{\mathrm{O_3}} = \sum_{i=1}^{n} N_{v,i}\sigma_{\mathrm{O_3}}(T_i)\,; \quad 5.15$$

$$\tau^{\mathrm{RRS,\,v}}_{\mathrm{O_3}} = \sum_{i=1}^{n} N_{v,i}\sigma^{\mathrm{RRS}}_{\mathrm{O_3}}(T_i), \quad 5.16$$

where $N_{v,i}$ are the ozone partial columns in different layers of the atmosphere and the ozone absorption cross-sections are interpolated at the appropriate temperature in each layer. Also, σ^{RRS}_{o3} is an ozone absorption cross-section for Raman scattering, representative of ozone absorption after smoothing by the Rotational Raman Scattering (RRS) effect.

The effective air mass factors $M_{\text{in}}(\lambda)$ and $M_{\text{out}}(\lambda)$ depend linearly on the wavelength and have been pre-computed so that the formulation in Eq. () closely reproduces filling-in factors calculated with the LIDORT-RRS radiative transport code. These terms are tabulated and depend on the total ozone column itself, the observation and solar geometry, and on the surface albedo and altitude. M_{tot} is a classical ozone air mass factor calculated at one wavelength as described in section 5.2.4. The quantity C_{ring} represents the probability for a photon to undergo an inelastic scattering process [*Wagner et al.,* 2009]. This quantity is the only unspecified part of the Ring correction, and is regarded as an ancillary element in the retrieval state vector, to be fitted simultaneously to the ozone column and other parameters. Since the filling factor $\text{FF}(\lambda)$ is linear in C_{ring}, the Jacobian (derivative of $I(\lambda)$ with respect to C_{ring}) is easy to derive from Eq. (5.13,).

5.3.7 Inverse model and state vector

Inverse problems for satellite retrievals are inherently ill-posed, and it is necessary to make additional assumptions in order to obtain (numerical) solutions - this is the regularization process. However, the two total ozone algorithms described in this ATBD are well-posed in one sense, since stable solutions can be obtained without the need for additional regularization. This is not the case for ozone profile retrieval, which requires *a priori* profile specifications when using the optimal estimation (OE) method.

In the original direct-fitting GODFIT algorithm for GOME, the OE inverse method [*Rodgers*, 2001] was used, with loose *a priori* constraints on the state vector elements - this regularization was imposed only to ensure numerical stability. However, a faster alternative was used in GODFIT for GDP 5: this is a variant of the Levenberg-Marquardt (LM) least squares algorithm with line searching [*Doicu et al*, 2004; *Doicu et al*., 2007]. Specifically, the following functional at the k^{th} iteration step is minimized:

$$\mathfrak{F}(\mathbf{x}) = \|\mathbf{G}(\mathbf{x}) - \mathbf{y}\|^2 + \alpha_k \|\mathbf{L}_k(\mathbf{x} - \mathbf{x}_k) - \mathbf{y}\|^2. \qquad 5.17$$

Here, the symbol $\|..\|^2$ indicates a Euclidean norm, $\mathbf{L}_k$ is a constant invertible square matrix acting as a constraint, and the factor α_k is a regularization parameter which determines the strength of this constraint contribution to the functional. The positive-valued regularization parameter α_k changes with each iteration step, and is always less than the initial value α_0. Note that the regularization is mainly for numerical stability, something that is not always easy to guarantee with an un-regularized standard Levenberg-Marquardt fitting in the present inverse problem.

Forward model $\mathbf{G}(\mathbf{x})$ is linearized about its value $\mathbf{G}(\mathbf{x}_k)$ according to the result $\mathbf{G}(\mathbf{x}) = \mathbf{G}(\mathbf{x}_k) + \mathbf{K}_k(\mathbf{x} - \mathbf{x}_k)$, where $\mathbf{K}_k = \mathbf{K}(\mathbf{x}_k)$ is the weighting function matrix evaluated at $\mathbf{x}_k$. Setting the first derivative of $\mathfrak{F}(\mathbf{x})$ around $\mathbf{x}_k$ to zero yields the next guess:

$$\mathbf{x}_{k+1} \equiv \mathbf{x}_k + \tau_k \mathbf{p}_k = \mathbf{x}_k + \left[\mathbf{K}_k^{\mathrm{T}} \mathbf{S}_y^{-1} \mathbf{K}_k + \alpha_k \mathbf{L}_k^{\mathrm{T}} \mathbf{L}_k\right]^{-1} \cdot \mathbf{K}_k^{\mathrm{T}} \mathbf{S}_y^{-1} \left[\mathbf{y} - \mathbf{G}(\mathbf{x}_k)\right] \qquad 5.18$$

Here, superscript T denotes matrix transpose, and $\mathrm{S_y}$ is the measurement error covariance matrix. Eq. 5.18 allows us to find the search direction $\mathrm{p_k}$ (a vector with unit modulus), and the step length τ_{k}. Given that the functional is minimized for a certain neighbourhood around $\mathrm{x_k}$ in state vector space defined by the length τ_{k}, we define a series of minimum values $\mathrm{x_i} = \mathrm{x_k} + \tau_{\mathrm{i}} \mathrm{p_k}$ according to step lengths $\tau_{\mathrm{i}} \in (0, \tau_{\mathrm{k}})$. We then search this series for a special $\mathrm{x_j}$ such that $\mathfrak{F}(\mathrm{x_j}) \leq \mathfrak{F}(\mathrm{x_i})$ for all $\mathrm{i} \neq \mathrm{j}$. This is the “line-search” procedure.

Although the line-search algorithm can in principle produce faster convergence through a smaller number of iterations, it can also lead to additional computational effort, since the line-search routine (and by default, the forward model) is usually called for all wavelengths in a

given fitting window. However, performance gains can be achieved by application of "line-search" at a reduced number of wavelengths, and for GDP 5, 3 “line-search” wavelengths grouped closely around the 328 nm Huggins bands maximum gave the most satisfactory results in terms of accuracy and performance trade-off (10-15% faster than OE). Additional speed-ups are obtained with the use of the total ozone previous pixel value as first guess, and the use of a slightly smaller fitting window [*Van Roozendael et al*., 2012].

In addition to the total ozone column, the temperature shift and the closure coefficients, the state vector also includes a fitting parameter C_{Ring} dealing with the Ring effect interference, and an backscattered radiance wavelength shift (with 3 closure coefficients, this makes 8 parameters in total). Undersampling and wavelength shift parameters function in the same manner as those for the DOAS fitting (see section 5.2.2). The Ring amplitude is dealt with in the next section.

5.4 Common components to S5P_TO3_DOAS and S5P_TO3_GODFIT

5.4.1 LIDORT and VLIDORT radiative transfer models

In both algorithms, forward model simulations are done "on-the-fly" using the multi-layer multiple scattering radiative transfer LIDORT model ([*Spurr*, 2008] and references therein). LIDORT is a discrete ordinate code with a full capacity for the simultaneous generation of atmospheric and surface Jacobians (radiance sensitivities). LIDORT is a scalar code, in which polarization is neglected. Polarization effects can be modelled using the vector LIDORT code (VLIDORT) [*Spurr*, 2008] which will generate the full Stokes 4-vector with components [I, Q, U, V]. VLIDORT also has a full Jacobian facility and is completely compatible with LIDORT in terms of input and output. VLIDORT is normally an order of magnitude or more slower in terms of performance than LIDORT.

Before entering into a summary description of the LIDORT model, we first discuss the issue of polarization in relation to the two algorithms in this ATBD. Until recently, the default with both the direct-fitting and DOAS algorithms has been to use the LIDORT scalar code for RT simulations, and to regard the polarization signatures as being subsumed in the closure treatments in both algorithms. This assumption is in general accurate enough for small fitting intervals such as used here (325-335 nm). In addition, RT with polarization is much slower and this is another reason for using the faster scalar model. In contrast, for ozone profile retrieval, the fitting window and dynamic range are much larger; the polarization correction applied to Level 1 data is an appreciable source of error and the inclusion of polarization in the RT simulations is necessary (see for example [*Liu et al.*, 2005]).

For performance reasons, VLIDORT is not used directly in GODFIT. However, a polarization correction has now been added to the Direct-fitting algorithm in the form of a look-up table (LUT) of intensity differences (that is, the relative difference in radiant intensity between calculations with and without polarization). This LUT is classified according to surface albedo, total column amount, surface height and solar and viewing geometries.

These differences are determined from two calculations with VLIDORT (one with polarization included, the other without polarization). The resulting corrections are compatible with the LIDORT simulations. We have found that total ozone accuracy has not been compromised by the use of this LUT as opposed to the deployment of full VLIDORT calculations. This LUT of polarization corrections is also used for DOAS AMF simulations, as noted in section 5.2.4.

LIDORT solves the RT equation in each layer using the discrete-ordinate method [*Chandrasekhar*, 1960; *Stamnes et al*., 1988]; application of boundary conditions (surface reflectance, level continuity, direct sunlight at top-of-atmosphere) generates the whole-atmosphere field at discrete ordinates; the source function integration method is then used to generate the radiance at any desired viewing geometry and output level. The entire discrete

ordinate RT solution is analytically differentiable with respect to any atmospheric and/or surface parameter [*Spurr*, 2002], and this allows Jacobians to be determined analytically without the use of computationally expensive (and dubiously accurate) finite-differencing estimation.

The use of the pseudo-spherical (P-S) approximation (solar beam attenuation treated for a curved atmosphere) is standard in LIDORT; for a discussion see [*Spurr*, 2008]. LIDORT also has an *outgoing sphericity correction*, in which both solar and viewing angles are allowed to vary along the line-of-sight (LOS) path treated for a spherical-shell atmosphere. In general, for line-of-sight view angles ~30° or less, the regular P-S approximation gives sufficient accuracy, but for the wider off-nadir viewing scenarios that apply as one moves towards the swath ends, the outgoing sphericity correction is necessary [*Caudill et al.*, 1997; *Spurr* 2003; *Rozanov et al*., 2000]. Thus for regular cross-track GOME viewing (maximum swath 960 km), the P-S approximation is accurate enough, but for the GOME and SCIAMACHY polar view mode [*ESA*, 1995] and for other instruments such as GOME 2 (swath 1920 km) [*Munro et al*., 2006] and OMI ([*Levelt et al*., 2006], swath width 2600 km), the outgoing correction is necessary. The outgoing correction is required for S5P.

Aside from the geolocation quantities (solar and viewing zenith angles, relative azimuth angles, earth radius) and discretization control (number of atmospheric layers and atmospheric height grid, number of discrete ordinates), the main inputs for LIDORT are the surface and atmosphere optical properties, and these are discussed below. LIDORT will generate as output the top-of-atmosphere (TOA) upwelling radiance field for both the DOAS and GODFIT algorithm. For the GODFIT algorithm, LIDORT will generate in addition a set of Jacobians (partial derivatives of the TOA radiance) with respect to the total ozone, the temperature shift and the surface albedo.

Generation of optical property input for LIDORT

LIDORT is a pure scattering RT model that requires specification of a suitable set of atmospheric optical properties $\{\Delta_n, \omega_n, \beta_{ln}\}$. These properties are the layer extinction optical thickness Δ_n, the layer single scatter albedo ω_n and phase function expansion coefficients β_{ln}. These quantities depend on atmospheric pressure and temperature profiles, distributions of absorbing trace species and aerosols (if present), plus scattering properties of molecules and other particulates. Here, we outline the determination for an atmosphere with Rayleigh scattering and trace gas absorption.

In hydrostatic equilibrium, air column density $a_n(T_n)$ in layer *n* depends on the layer temperature T_n (and the temperature shift S), as does the ozone cross section $\sigma_n^{O3}(T_n)$. These dependencies were given explicitly for the direct fitting algorithm in Eqns.(5.10 and 5.11). Ozone partial columns $n_{a,n}(N_v)$ depend only on the total column N_v, as outlined in section 5.4.2 below. The Rayleigh scattering coefficient σ_{Ray} is determined from a standard formulation [*Bodhaine et al*., 1999]. For a Rayleigh-only atmosphere:

$$\Delta_n(N_v, T_n) = \sigma_{Ray} a_n(T_n) + \sigma_n^{O3}(T_n) n_{a,n}(N_v);\ \ \omega_n(N_v, T_n) = \frac{\sigma_{Ray} a_n(T_n)}{\Delta_n(N_v, T_n)}. \qquad 5.19$$

In Rayleigh scattering, the phase function has $\cos^2\Theta$ dependence on scattering angle Θ, and expansion coefficients $\beta_0 = 1, \beta_1 = 0, \beta_2 = (1-\delta)/(2+\delta)$, where the depolarization ratio δ is taken from [*Chance and Spurr*, 1997]. These coefficients are the same in all layers.

Jacobians are required for the direct fitting GODFIT algorithm. In order for LIDORT to generate the total ozone and T-shift Jacobians, it is necessary to specify a set of *linearized* optical property inputs. These are the derivatives of optical properties $\{\Delta_n, \omega_n, \beta_{ln}\}$ with

respect to the total column N_v and the T-shift S, and they may be obtained by straightforward partial differentiation of the quantities in Eq. (5.19). We find:

$$\frac{\partial \Delta_n}{\partial N_v} = \sigma_n^{O3} \frac{\partial n_{a,n}}{\partial N_v}; \qquad \frac{\partial \Delta_n}{\partial S} = \frac{\partial \sigma_n^{O3}}{\partial S}\, n_{a,n} + \sigma_{Ray} \frac{\partial a_n}{\partial S}. \tag{5.20}$$

$$\frac{\partial \omega_n}{\partial N_v} = -\frac{\omega_n}{\Delta_n}\frac{\partial \Delta_n}{\partial N_v}; \qquad \frac{\partial \omega_n}{\partial S} = \frac{1}{\Delta_n}\left[\sigma_{Ray}\frac{\partial a_n}{\partial S} - \omega_n \frac{\partial \Delta_n}{\partial S}\right]. \tag{5.21}$$

Derivatives $\frac{\partial \sigma_n^{O3}}{\partial S}, \frac{\partial a_n}{\partial S}$ are taken from the T-shift analysis in section 5.3.2. No additional input is required for LIDORT to generate the surface albedo Jacobian.

5.4.2 A priori ozone profile climatology

The forward model needs as input a complete ozone profile. In both algorithms, the ozone profile is parameterized by a single quantity – the total column. The use of total column as a proxy for the stratospheric ozone profile was recognized a number of years ago by scientists at NASA, and column-classified ozone profile climatologies were created for the TOMS Version 7 [*Wellemeyer et al.*, 1997] and Version 8 (V8) retrieval algorithms [*Bhartia*, 2003]. A TOMS-based column-profile mapping was developed for GOME AMF computations in GDP 3.0 [*Spurr et al*., 2005], and GDP 4.0 [*Van Roozendael et al*., 2006]. Labow et al. (2015) have constructed a total ozone-classified profile climatology based on aggregation of MLS and sondes measurements. The latter database is used for direct-fitting retrieval. An additional ozone climatology derived from OMI and MLS measurements [Ziemke et al. (2011)] is included in both algorithms for constructing the tropospheric ozone.

Pressure levels defined by the TOMS V8 climatology determine the vertical layering for forward model simulations. Note that the Labow climatology has been reformatted to follow the same layering. Starting at a surface pressure of 1013 hPa, pressures are halved at each successive atmospheric boundary; the scale height is 5.0 to 5.7 km. There are 12 levels corresponding to 11 'Umkehr' layers, with TOA set at 0.03 hPa. Height levels are determined by hydrostatic equilibrium based on a suitable temperature profile. In hydrostatic balance, acceleration due to gravity varies with latitude and height according to the specification in [*Bodhaine et al*., 1999]. Layering is set up for clear sky and cloudy scenes, with the lowest layer adjusted to fit the boundary pressure (surface or cloud-top); height and temperature adjust linearly with the logarithm of the pressure. For each satellite footprint, we can extract surface height from the DEM database. Surface pressure is either taken from analysis fields, or interpolated against the surface height using a standard atmosphere.

In the TOMSv8 climatology, profiles are specified for columns at every 50 DU, for 18 latitude bands from pole to pole (10° intervals), and for each month of the year. Latitude and time variations are treated using a bilinear interpolation scheme. Columns range from 125 to 575 DU at higher latitudes down to 225 to 325 DU in the tropics. In the Labow climatology, a similar ozone column mapping is provided but only in 6 latitude bands and it is assumed that the seasonal variation is implicitly taken into account via the column dependence. The tropospheric content of the profiles provided by those two data bases is not highly realistic as, for example, longitudinal variations are not considered. Therefore, we combine these two data bases with the tropospheric ozone column climatology constructed by Ziemke et al. (2011). Depending on the month and pixel coordinates, we apply to the a priori profile extracted from the TOMSv8 or Labow database scale factors to ensure that the integrated tropospheric column matches the column provided in the Ziemke climatology. We also ensure to maintain the mapping between the integrated a priori profile and the associated column. Adjustments are also made to the lowest Umkehr profile layers to account for

surface, cloud-top or scene pressure. These adjustments assume that the partial column in any layer is proportional to the logarithm of the layer pressure drop (i.e., constant mixing ratio in such a layer).

In both algorithms, the profile-column mapping is defined as follows. For profiles $n_{a,n}^{(1)}$ and $n_{a,n}^{(2)}$ associated with total columns ${N_v}^{(1)}$ and ${N_v}^{(2)}$ such that ${N_v}^{(2)} - {N_v}^{(1)} = 50\ \mathrm{DU}$, we define an intermediate profile according to a linear map $n_{a,n}(N_v) = \frac{1}{50}\left[\left(N_v - {N_v}^{(1)}\right)n_{a,n}^{(2)} + \left({N_v}^{(2)} - N_v\right)n_{a,n}^{(1)}\right]$. In effect, we are defining an ensemble of ozone profiles based on the sample set represented by the climatology. If the vertical column lies outside the range of values classifying the climatology, we use the limiting profile value: this situation may occur in extreme ozone-hole scenarios (total column < 125 DU). Note that $\partial n_{a,n}(N_v)/\partial N_v = \frac{1}{50}\left[n_{a,n}^{(2)} - n_{a,n}^{(1)}\right]$; this derivative is essential for defining linearized optical property input for LIDORT in the Direct Fitting algorithm (see section 5.3.3).

Remark. It is possible in both algorithms to use ozone profiles from global transport models. However, model data can have biases and drifts, and using model profiles as a priori implies that these biases/drifts would be propagated to the satellite retrievals. Using a priori profiles from a model will probably not improve the retrieval results but will create unnecessary external dependencies. We prefer to use an ozone climatology, noting that it has been demonstrated in past studies that ozone profile shapes are strongly linked to ozone columns.

5.4.3 Ozone cross-sections

Cross-section uncertainty is a crucial source of retrieval error. For the total ozone retrieval from backscatter UV sensors, the choice of ozone cross-sections has either been the older Bass/Paur laboratory data from the 1980s [*Bass and Paur*, 1984] or flight model (FM) data (GOME: [*Burrows et al.*, 1999]; SCIAMACHY: [*Bogumil et al*., 2003]) derived from pre-flight calibration measurements. For more discussion on these data sets, see [*Orphal*, 2003]. In the Huggins bands range, the GOME GDP 4.0 baseline used re-sampled FM O_3 data with the solar I_0 correction [*Aliwell et al.*, 2002] included, and were pre-shifted (optimized pre-shift +0.016 nm). For more details on these specifications, see [*Van Roozendael et al.,* 2006]. However, the high spectral resolution laboratory data (the so-called “Brion-Daumont-Malicet [BDM]” cross sections [*Daumont et al.*, 1992; *Malicet et al*., 1995; *Brion et al.,* 1998]) produce lower uncertainties than those from GOME FM data [*Liu et al*., 2007].

These cross-section data sets have been intercompared [*Lerot et al.*, 2009] - there are large differences in terms of differential amplitudes and wavelength calibration. The high-resolution BDM data can be used for ozone retrievals from any space-borne UV spectrometer; BDM data produces high quality fits and has accurate wavelength calibration.

Also, its temperature dependence appears to be reliable since the retrieved effective temperatures are generally in good agreement with ECMWF temperatures. For these reasons, BDM was selected as the reference absorption cross-section data set for the recent operational products from the GOME and GOME-2 sensors. This is also the choice for the S5P NRTI DOAS product. More recently, a new laboratory data set of high-resolution ozone cross-sections measured at 10 different temperatures has been released [*Serdyuchenko et al.*, 2014]. This data set is also of very high quality with reliable wavelength registration and well characterized temperature dependence. In the fit window 325-335 nm, there is no clear indication that either BDM or Serdyuchenko data produce better retrievals, although differences in terms of ozone column range within 1-2% depending on this choice. This illustrates the current uncertainty associated to ozone cross-section measurements. Total ozone column ECVs data sets have been produced from GOME, SCIAMACHY, GOME-2

and OMI sensors using Serdyuchenko data. For consistency, the S5P OFFL direct-fitting processor also uses this data set.

Note that one should ideally convolve simulated radiances calculated at higher resolution. However, errors introduced by the use of pre-convolved cross-sections at nominal resolution are largely eliminated by making use of the solar-I0 correction in the convolution: this provides a first-order correction for the impact of solar lines on the shape of ozone absorption structures. To further improve this in the OFFL product, the solar-I0 correction applied to the ozone cross-section is applied iteratively depending on the fitted ozone optical depth at the previous iteration.

5.4.4 Other trace species

NO_2 absorption is present in the Huggins' bands fitting window, but this effect is small compared with that for ozone, and this trace species is only included in the fit for the DOAS product. The effect of neglecting NO_2 needs to be quantified, and we return to this issue in the discussion on error sources in section 7.1.3.

Note that SO_2 absorption is not included in either of the S5P total algorithms. This is in contrast to the ozone profile algorithms for GOME and successor instruments (including TROPOMI), which require use of shorter wavelengths in the UV for which SO_2 absorption is stronger and needs to be taken into consideration. In general, SO_2 absorption is very weak in the 325-335 nm fitting window. However, sensitivity tests have been carried out to assess the impact of SO_2 absorption on total ozone accuracy – the effect is generally negligible, except for extreme volcanic scenarios.

5.4.5 Cloud treatments

The treatment of cloud-contaminated pixels in ozone retrieval is crucial. Clouds are both spatially inhomogeneous and highly variable over short time scales, and it is very difficult if not impossible to model clouds accurately. 3-D radiative transfer modelling of clouds is too complex and time-consuming, and it is usual to approximate cloud effects using 1-D RT models and simple parameterizations. Traditionally with total ozone retrieval, clouds have been treated as perfectly reflecting boundaries, or as optically uniform layers of scattering particles. It is then only necessary to specify a small number of parameters characterizing the cloud - these are the cloud-top pressure, the cloud albedo or optical thickness, and the cloud geometrical thickness. In addition, the radiance for partially cloudy scenes is computed as a linear combination of separately calculated radiances for clear-sky and cloud-filled scenes - the combination being weighted by the effective cloud fraction of the scene in question. This is the well-known Independent Pixel Approximation (IPA). For total ozone, ancillary cloud parameters are obtained from dedicated retrieval algorithms and then used in forward model computations of UV backscatter radiances.

The treatment of clouds in the TROPOMI/S5P total ozone algorithm is done through a combination of the OCRA (for the cloud fraction) and ROCINN (for other cloud parameters) algorithms. Descriptions of these algorithms are given in a separate S5P ATBD [RD03].

The OCRA deployment in S5P retrieves effective cloud fraction through use of a cloud-free background and thresh-holding based on TROPOMI measurements. In version 1 the cloud-free background is taken from OMI data. With version 2 the cloudfree background maps are switched to TROPOMI based maps.

ROCINN is a neural network algorithm based on trained data sets of reflectances in and around the O_2 *A*-band. ROCINN Version 2.0, as used in the GDP 4.x and GDP 5 heritage algorithms, provided the cloud height (pressure) and the cloud albedo, based on the

assumption of a Lambertian reflecting boundary at cloud top. The optical thickness was then obtained from the albedo by means of a secondary neural-network algorithm linking the two quantities).

In S5P, ROCINN Version 3.0 is used to retrieve cloud-top height and cloud optical thickness, with reflectances now calculated for atmospheres containing clouds as layers of water droplets having optical properties determined from Mie scattering layers. This is clearly a more physically reasonable model that the Lambertian reflector assumption, the cloud-top albedo is not required in this setting.

The main consequence for the S5P_TO3 NRTI algorithms is that all appropriate RT simulations needed for AMF computations must be done with scattering-layer clouds characterized by these OCRA/ROCINN_CAL retrieved variables. In the same way, the RT corrections (polarization, Ring) are adapted Optical properties for the cloud types (as used in the cloud algorithm) are extrapolated to the UV.

For the S5P_TO3 OFFL product, the approach is simpler as we use an effective scene approach as described before, which has been demonstrated to work well as part of Ozone_CCI activities [Lerot et al., 2014]. The OCRA/ROCINN_CRB cloud fraction and cloud altitude are used to compute the altitude and the apriori effective albedo of the effective scene, of which the effective albedo is fitted simultaneously to the ozone column.

5.4.6 Processing Flags and QA Values

All processing errors and warnings are collected in the processing_quality_flags. A detailed description can be found in the Product User Manual (PUM) for the ozone total column [RD8].

For a retrieval not affected by any processing error or warning, the qa_value is assumed to be 1.0 (i.e. highest possible product quality). Each potential processing error or warning will reduce the product quality to result in qa_values below 1.0. The following processing results, warnings and errors are translated to the quality assurance value for the total ozone algorithms:

ozone_range_error, ozone_vertical_column, fitted_root_mean_square, solar_zenith_angle, effective_albedo (OFFL), qa_value_cloud, saturation_warning cloud_retrieval_warning, cloud_inhomogeneity_warning,

Each warning has a different impact on the product quality since some warnings are probably less critical then others. The respective reduction is shown in Table 5.1

For all processing errors the qa_value is set to 0.0, and in most cases the respective data will be replaced by a fillvalue. The ozone_range_errors labels all total ozone columns outside the range of 0 to 0.45 mol/m² (~1000 DU). The respective values are written in the dataset but the qa_value indicates that these data should not be used (Table 5.1).

Table 5.1: The qa_values are reduced according to the following table depending on the individual input parameters. For each parameter a valid range is given where the qa-value is not changed. If the data are within the reduction range the qa_value is reduced in case it is beyond the skipping threshold the qa_value is set to zero. The fitted_root_mean_square is lower in the OFFL algorithm, the thresholds for the OFFL are half of the values for the NRTI. The SZA θ_0 is weighted by $cosweight = \frac{\cos(\theta_0)-\cos(75°)}{\cos(89)-\cos(75°)}$.

	Reduction	valid range	reduction range	skip threshold
ozone_vertical_column	0.2	(0,0.35) mol m^{-2}	[0.35,0.45] mol m^{-2}	< 0 and > 0.45 mol m^{-2}

fitted_root_mean_square (NRTI)	0.3	(0,5e-3)	[5e3,1e-2]	> 1e-2
fitted_root_mean_square (OFFL)	0.3	(0,25e-3)	[2.5e3,5e-3]	> 5e-3
effective_albedo (OFFL)	0.2	[0,1.1]	[-0.5,0) and (1.1,1.5]	< -0.5 and > 1.5
solar_zenith_angle	0.5*cosweight	(0,75°]	(75°,89°]	> 89°
saturation_warning	0.3			
cloud_retrieval_warning	0.3*(1-qa_value cloud)			
cloud_inhomogenity_warning	0.3			
qa_value_cloud				=0

6 Input-Output file description

6.1 S5P ozone product description and size

The S5P total ozone product will be provided in netCDF-CF. The following information will be included for each ground pixel:

- measurement time and geolocation, taken from the L1b product
- total ozone, error, averaging kernel
- climatology and other relevant parameters used in the retrieval
- fit results (RMS, etc.)
- quality flag

Table 6.1 and Table 6.2 present detailed lists of the output fields that are required in the total ozone level-2 files for S5P_TO3_DOAS and S5P_TO3_GODFIT, respectively. Additional output parameters such as geolocation, quality control, input data etc are also included in the L2 product and specified in the corresponding PUM [RD8]

The estimated product size for near-real-time products is ~30 MB per 5 min granule and for one orbit off-line products is ~320 MB for the smaller ground pixel since since August 2019.

Table 6.1: **List of output fields required in the total O_3 level-2 product generated with the near-real time algorithm S5P_TO3_DOAS.**

Name/Data	Symbol	Unit	Description	Data type per pixel	Dimensions
Number of measurements	*N*	---	Number of measurements included in the file. N = nAlong x nAcross	Integer	1
Orbit number	n_0	---	Satellite orbit number	integer	1
Time	---	---	Date and time of measurement [YYMMDDHHMMSS.MS]	Character	1
Latitudes	*lat*	degree	Latitudes of the pixel center and corners	float	5
Longitudes	*lon*	degree	Longitudes of the pixel center and corners	float	5
SZA	θ_0	degree	Solar zenith angle at pixel center	float	1
LoSZA	θ	degree	Viewing zenith angle at pixel center	float	1
RAA	φ	degree	Relative azimuth angle at pixel center	float	1

SCD	N_s	mol m^{-2}	Slant column density	float	1
Ring-corrected SCD	$N_{s,\,corr}$	mol m^{-2}	Slant column density corrected for molecular Ring effect	float	1
VCD	N_v	mol m^{-2}	Vertical column density	float	1
Ring scale factor	C_{ring}	---	Molecular Ring correction factor	float	1
Effective temperature	T_{eff}	K	Effective ozone temperature from the DOAS fit	float	1
Cloud free AMF	M_{clear}	---	Cloud free air mass factor	float	1
Cloud AMF	M_{cloud}	---	Cloudy air mass factor	float	1
CFR	f_c	---	Cloud fraction	float	1
CTP/CTH	p_c / Z_c	Pa	Cloud top pressure	float	1
CTA/COT	A_c / τ_c	---	Cloud top albedo / optical thickness	float	1
Albedo	A_s	---	Climatological surface albedo	float	1
Surface pressure	p_s	Pa	Effective surface pressure of the satellite pixel	float	1
fitted_root_mean_square	*RMS*	---	residual of the DOAS fit	float	1
SCD random error	σN_{s_rand}	mol m^{-2}	N_s random error	float	1
Pressure grid	p	Pa	Atmospheric pressure grid	float	15
Averaging kernels	*AK*	---	Layer averaging kernels	float	14
O_3 profile	n_a	mol m^{-2}	a-priori partial ozone column profile	float	14

Table 6.2: List of output fields required in the total O_3 level-2 product generated with the offline time algorithm S5P_TO3_GODFIT

Name/Data	Symbol	Unit	Description	Data type per pixel	Dimensions
Number of measurements	*N*	---	Number of measurements included in the file. N = nAlong x nAcross	Integer	1

Orbit number	n_0	---	Satellite orbit number	integer	1
Time	---	---	Date and time of measurement [YYMMDDHHMMSS.MS]	character	1
Latitudes	*lat*	degree	Latitudes of the pixel center and corners	float	5
Longitudes	*lon*	degree	Longitudes of the pixel center and corners	float	5
SZA	θ_0	degree	Solar zenith angle at pixel center	float	1
LoSZA	θ	degree	Viewing zenith angle at pixel center	float	1
RAA	φ	degree	Relative azimuth angle at pixel center	float	1
VCD	N_v	mol m^{-2}	Vertical column density	float	1
Ring scale factor	C_{ring}	---	Ring correction scale factor	float	1
Effective temperature	T_{eff}	K	Effective ozone temperature resulting from shifted temperature profile	float	1
Effective Albedo	A_{eff}	---	Effective scene albedo	float	1
Scene pressure	p_{eff}	Pa	Effective scene pressure computed from surface and cloud pressures	float	1
Ghost column	G	mol m^{-2}	Ghost column: Partial vertical ozone column below the effective scene	float	1
CFR	f_c	---	Cloud fraction	float	1
CTP/CTH	p_c / Z_c	Pa	Cloud top pressure/height	float	1
COT	A_c / τ_c	---	Cloud top albedo / optical thickness	float	1
Albedo	A_s	---	Climatological surface albedo	float	1
Surface pressure	p_s	Pa	Effective surface pressure of the satellite pixel	float	1
Chi square	Chi^2	---	Chi square of the DOAS fit	float	1
VCD random error	σN_{v_rand}	mol m^{-2}	N_v random error	float	1
Pressure grid	p	Pa	Atmospheric pressure grid	float	15
Averaging kernels	AK	---	Layer averaging kernels	float	14

O_3 profile	n_a	mol m^{-2}	a-priori partial ozone column profile	float	14

6.2 Auxiliary information needs

All dynamic and static auxiliary data are described in Table 6.3 and Table 6.4. The needs are similar for both algorithms.

Table 6.3: Dynamic auxiliary information needs in the total ozone retrieval algorithms

Name/Data	Symbol	Unit	Source	Pre-process needs	Backup if not available
S5P level 1B Earth radiance	I	W $m^{-2}nm^{-1}sr^{-1}$	S5P L1b product		No retrieval
S5P level 1B sun irradiance	E_0	W $m^{-2}nm^{-1}$	S5P L1b product	Wavelength recalibrated using a high-resolution reference solar spectrum	Use previous measurement
Cloud fraction	f_c	---	S5P L2 operational cloud product	---	No retrieval
Cloud top height	p_c	Pa	S5P L2 operational cloud product	---	No retrieval
Cloud top albedo/optical thickness	A_c / τ_c	---	S5P L2 operational cloud product	---	No retrieval
Snow/ice flag	---	---	Near real-time global Ice and Snow Extent (NISE) data from NASA.	---	Use snow/ice climatology

Table 6.4: Static auxiliary information needs in the total ozone retrieval algorithms

Name/Data	Symbol	Unit	Source	Pre-process needs	Comments
Instrument slit function	*SF*	---	Slit function provided by wavelength/detector	---	---
High-resolution reference solar spectrum	E_s	W $m^{-2}nm^{-1}$	Chance and Kurucz [2010]	---	---
Absorption O_3 cross-section at 5 temperatures	σ_{O3}	$cm^2molec.^{-1}$	Brion et al. [1993], Daumont et al. [1992], Malicet et al. [1995] Serdyuchenko et al. [2014]	Convolution at the instrumental spectral resolution using the provided slit function	---
Ring cross-section	σ_{ring}	---	Generated internally.	A high-resolution reference solar spectrum and the instrument slit function are needed to generate the data set.	---
Ring-smoothed O_3 cross-sections at 5 temperatures	$\sigma_{O3\text{-}RRS}$	$cm^2molec.^{-1}$	Generated internally using the original O_3 cross-section.	A high-resolution reference solar spectrum and the instrument slit function are needed to generate the data set.	---
A-priori O_3 vertical profile shapes	n_a	DU	TOMSv8 database MLS/sondes data base (Labow et al. [2015], McPeters et al.[2012)] combined with OMI/MLS tropospheric O_3 climatology (Ziemke et al. [2011])	---	---
Digital elevation map	Z_s	m	GMTED2010 (Danielson et al., 2011) same DEM for all L2 products	---	---

6.3 Level 1 information needs

In the heritage algorithms for GOME, the GOME backscattered radiance and solar irradiance were generated by the GDP Level 0-to-1b extractor [*Slijkhuis et al.*, 2004]. In addition, Level-1 wavelength calibration was improved selectively through application of window-dependent pre-shifts to parts of the solar spectrum [*Van Roozendael et al*., 2006].

As with the heritage GDP4 and GODFIT algorithms, the main Level 1 measurement data sets for both S5P total ozone algorithms are the Level 1b geolocated and calibrated backscatter UV Earthshine radiances and the Level 1b calibrated solar irradiances for wavelengths in the range 325-335 nm. These measurements are provided in the TROPOMI UVN band 3 (310-405 nm) radiance and UVN irradiance products. Unless otherwise specified, Level 1b radiances are in units of [Photon/nm/sr/cm^2]. Radiances and irradiances are accompanied by error quantities in the same units. Wavelengths are calibrated values in [nm]. To ensure an accurate wavelength registration, a recalibration procedure, based on a cross-correlation with a high-resolution solar spectrum (e.g. [*Chance and Kurucz*, 2010]), is applied to both the radiance and irradiance measurement spectra [*Van Roozendael et al*., 2006].

The radiance data is assumed corrected for polarization (TROPOMI uses a scrambler to remove polarization at the instrument level). The Level 1b product also comes with full geolocation information - solar and viewing zenith and azimuth angles at the bottom of the atmosphere (BOA) in addition to those at TOA and at the spacecraft.

The cloud products from OCRA and ROCINN require additional Level 1b data. OCRA needs broad-band integrated measurement data (radiances/irradiances) from the TROPOMI, UVIS detector bands 3 and 4 (310-405 nm and 405-500 nm) and in the future NIR detector band 6 (725-775 nm) will be included. The ROCINN algorithm needs measurements in and around the oxygen *A*-band (TROPOMI NIR detector band 6). For more details on the cloud algorithm Level 1b requirements, refer to the accompanying ATBD [RD3].

7 Error analyses

7.1 General formulation

7.1.1 Error classifications

In a classical inversion involving least-squares minimization it is possible to characterize and quantify errors. In the optimal estimation inverse method [*Rodgers*, 2001], the optimal estimate X is given by the following:

$$\mathrm{X} \approx \mathrm{X}_{true} + (\mathrm{A} - \mathrm{I})(\mathrm{X}_{true} - \mathrm{X}_a) + \mathrm{D}_y\epsilon, \quad 7.1$$

where $A = D_y K,$ and $D_y \equiv S_x \cdot K^T S_y^{-1} = \left[K^T S_y^{-1} K + S_a^{-1}\right]^{-1} \cdot K^T S_y^{-1};$ this defines the solution error covariance matrix S_x (cf. Eq. (5.18)). Also, K is the weighting function matrix, subscript *T* denotes matrix transpose, S_y is the measurement error covariance matrix, X_a is the *a priori* state vector with random normally distributed error covariance matrix S_a. Measurement errors are usually assumed to be normally distributed and uncorrelated (S_y is diagonal). Matrix A contains the averaging kernels and is an indicator of the sensitivity of the retrieval to the true state. The second term in Eq. (7.1) is the smoothing error. The remaining error ϵ can be divided into three components as follows.

Measurement errors. This error is $\epsilon_{me} = D_y \epsilon_y$ which may be systematic or random; in the latter case, the contribution to the total error covariance is just the standard error propagation of uncertainty $S_{noise} = D_y S_y D_y^T$. Systematic errors in this category include stray light, slit function and radiometric calibration uncertainties. In practice, only the random error (or precision) is provided for each pixel; systematic measurement errors are estimated offline.

Model parameter errors. The retrieval error due to this source of uncertainty is $\epsilon_{mpe} = D_y K_b \Delta b$ where K_b is the sensitivity (Jacobian) of the forward model to a vector b of model parameters, and Δb is the error on b. If the error is random (invariably assumed Gaussian) with covariance S_b, then the associated solution covariance contribution is $S_{param} = D_y K_b S_b K_b^T D_y^T$. Estimation of these errors is greatly helped when the forward model is able to deliver Jacobians K_b in an efficient and accurate manner (LIDORT has this capability). Model parameters can be any atmospheric variables that are not fitted (cloud fraction and cloud top pressure, cross-section amplitudes, temperature profile entries, etc.). All model parameter errors are estimated through off-line analysis

Forward model errors. Here the retrieval error due to this source of uncertainty is given by $\epsilon_{fwd} = D_y \Delta F$, where ΔF is the forward model error due either to incorrect physical assumptions in the RT model (e.g. neglect of polarization, omission of rotational Raman scattering) or to a certain level of mathematical approximation (number of stratifications, number of discrete ordinates, plane-parallel scattering, etc.). These are systematic errors which require offline estimation.

With no regularization, the gain D_y is the identity matrix. This is the case for the DOAS fitting, which is a traditional least-squares regression. For the operational line-search Levenberg-Marquardt inversion in S5P_TO3_GODFIT, the constraint contribution in D_y is provided by Tikhonov regularization. However, the constraint is deliberately chosen to be very loose and the main purpose of the regularization is just to provide numerical stability.

Note that many of these errors are difficult to quantify. In the S5P_TO3_DOAS algorithm (GDP 4.x DOAS heritage) the DOAS least squares fitting will yield the solution covariance S_x, but this is difficult to interpret for the closure coefficients, which have no physical

interpretation. Also in this algorithm, there are many sources of forward model error, starting with the very first physical assumption of Beer-Lambert extinction for the DOAS fitting.

In the direct fitting algorithm S5P_TO3_GODFIT (GDP 5 heritage), the situation with error quantification is simpler and more clearly defined, and the above error classification is generally valid.

The error analysis results are given in section 7.1.3 below.

7.1.2 Profile averaging kernel for total ozone retrieval

The averaging kernel (AK) matrix AK was defined quite generally in the previous section. In optimal estimation, the trace (diagonal sum) of AK is the "degrees-of-freedom-for-signal" (DFS) for the retrieval. *A priori* regularization is essential for an ill-posed problem such as ozone profile retrieval from GOME measurements; the DFS for ozone in this inverse problem is typically 4.5-5.5 for a state vector with 11-layer ozone profile [*Liu et al.*, 2005].

Ozone amount (slant column in GDP 4.x, vertical column in GODFIT) is represented in the state vector as a single element, and the corresponding average kernel element is just the weighting function with respect to total ozone (for a well-posed inversion). However, we can still define a *profile* averaging kernel as a *vector* that indicates the sensitivity of the retrieved total vertical column to changes in ozone concentration in different layers. [Recall that in both algorithms, the total column amount defines a unique profile with the column-classified climatology section 5.4.2].

One can establish this profile averaging kernel vector using a DOAS-based approach for an optically thin absorber [*Eskes and Boersma*, 2003]. Here, AK vector values are approximated as the ratios of layer AMFs to the total AMF. For the determination of layer AMF values, it is necessary to make one final call to the RT model (LIDORT) to generate a vector of profile layer Jacobians for the ozone profile corresponding to the final total column. Layer AMFs are calculated at one wavelength (328.125) for the optically-thin absorber assumption.

For the OFFL product, a more accurate approach is followed. At each wavelength, LIDORT is called to derive the ozone profile layer Jacobians K* using the ozone a priori profile corresponding to the final retrieved total column. The contribution function D_y is obtained making use of the column weighting function K_i calculated as part of the retrieval process. The averaging kernel is then given by $AK = D_y K^*$.

When using the LUT approach, calculating the averaging kernels would require that all Jacobians at all wavelengths are stored in a table as well, which would multiply the size of the table, again making it impossible to keep all the required data in memory. Therefore we chose to store precomputed averaging kernels for each grid point, fixing the fitted forward model parameters which are not part of the lookup table grid (closure, Ring amplitude and wavelength shift) at their initial values. We found that this approximation does not have any significant impact.

7.1.3 Error estimates

Random error due to instrumental signal-to-noise:

An estimation of the random error can be derived by examining the propagation of the level-1 radiance and irradiance statistical errors through the inversion algorithm. The random error caused by the radiance and irradiance noise is about 1% and according to the Verification report [RD9] it does not exceed 2.5%.

Error due to radiometric uncertainties:

This error is very difficult to assess and depends strongly on the nature of the calibration limitations. In general, the sensitivity to multiplicative errors on the level-1 spectra is relatively

limited. These errors have an impact only when an effective (internal-closure) albedo is retrieved simultaneously with the ozone column. Additive errors due to this source are much more significant in general; the magnitude of these errors will be established once the instrument is calibrated.

Error due to the O_3 profile shape

Once we have the averaging kernel vector AK, the error $S_p$due to the profile shape may be estimated as $S_p = AK^T S_a AK$, where S_a is the covariance matrix associated with the a priori profile climatology used in the inversion. What is really required here is the covariance associated with the particular retrieved total column for a specific latitude band and season. This being unavailable, we have used another total ozone classified climatology (the dynamical ozone climatology developed at IUP/Bremen [Lamsal et al., 2009) as a proxy to construct S_a.

As illustrated in Figure 7.1, the mean total ozone error due to the profile shape is less than 0.5 % at low SZAs and is as large as 4% at extreme SZA for clear sky pixels. In the case of cloud contamination, the error increases, especially at low SZA, where it may reach 1%.

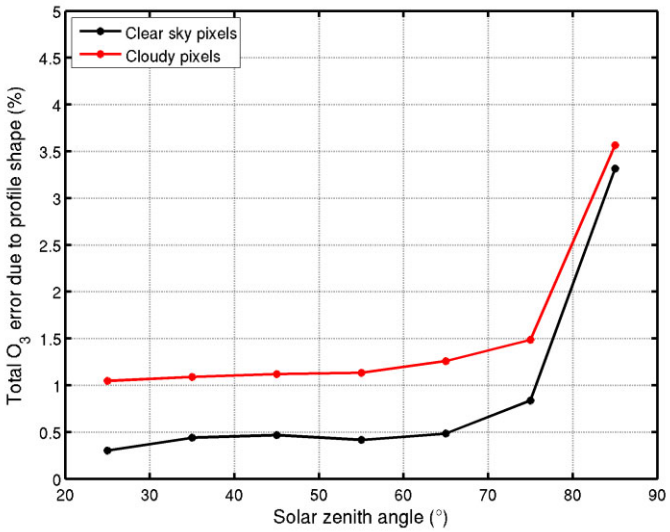

Figure 7.1: Estimation of total ozone error due to the O_3 profile shape as a function of the SZA for clear sky and cloudy pixels.

Error due to other trace gases and the presence of aerosols.

Some closed-loop tests have been performed to simulate the impact of neglecting other trace gases besides O_3 and of the inclusion of aerosols in the forward model. Synthetic radiances are generated using the GODFIT forward model, including the presence of these sources of error (e.g NO_2 or aerosols). Then, a total ozone retrieval is performed based on these synthetic spectra using the retrieval settings baseline (i.e. neglecting other trace gases or aerosols in the forward model). The difference with respect to the “true” state gives the error estimate.

To simulate the impact of stratospheric NO_2, a typical stratospheric profile, as depicted in Figure 7.2, has been used to generate the synthetic radiances. The total ozone columns retrieved from the resulting synthetic spectra show errors of less than 0.5% for all SZAs and all surface albedos. When considering a profile with a large amount of NO_2 in the lowermost layer (e.g. representative of a heavily polluted scenario), the total ozone errors increase slightly but remain less than 0.5% for low surface albedo (0.05). The errors are slightly larger

than 1% when the surface albedo is very high (0.8), but the likelihood of such a high NO_2 concentration above a bright surface is unlikely. The errors due to the neglect of BrO and SO_2 are negligible, except for the latter species for a major volcanic eruption scenario. In this case, the errors may reach a few per cent.

The same approach has been followed to estimate the error due to the neglect of aerosols in the forward model. Different scenarios have been considered for the total error estimate, including a background aerosol case, a heavily polluted scenario with a large amount of absorbing aerosol in the lowermost layer and a dust storm scenario with a large amount of scattering aerosol in the lowermost layer. The optical property profiles for these scenarios are plotted in Figure 7.3. The associated total ozone errors, plotted as a function of SZA in Figure 7.3, are generally within 1%. This small impact is mainly due to the simultaneous fit of the effective surface albedo. As illustrated in Figure 7.3 for the pollution scenario, the total ozone errors are much larger (up to 4%) if the surface albedo is fixed to a climatological value. For a scenario with strong injection of stratospheric aerosols due to a major volcanic eruption such as Pinatubo, the total errors may reach 10% (not shown here).

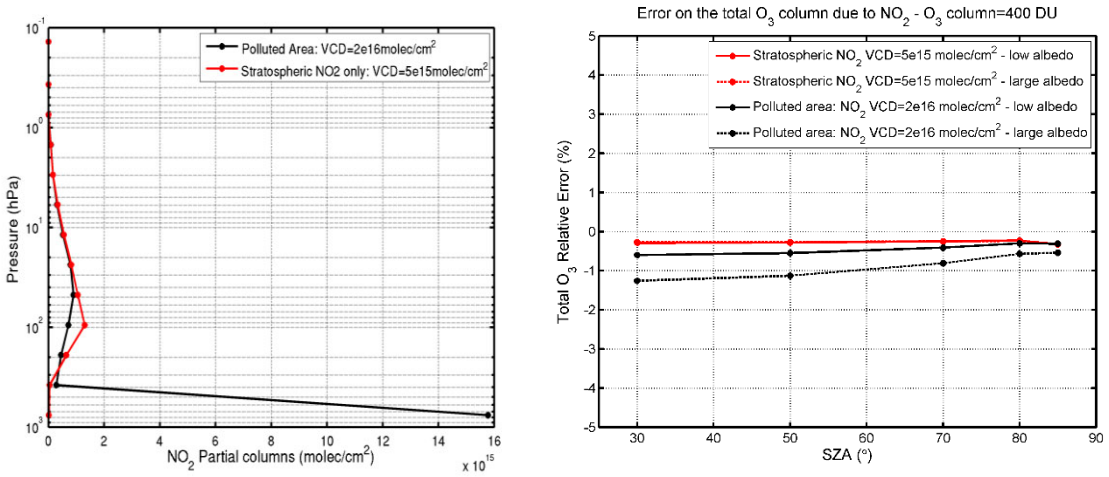

Figure 7.2: Left panel: NO_2 vertical profiles used for generating synthetic radiances. Right panel: Total ozone error (%) due to the neglect of the NO_2 in the retrieval scheme as a function of SZA. For the two profiles presented in the left panel, the errors are estimated for both low and large surface albedo (0.05 and 0.8 respectively).

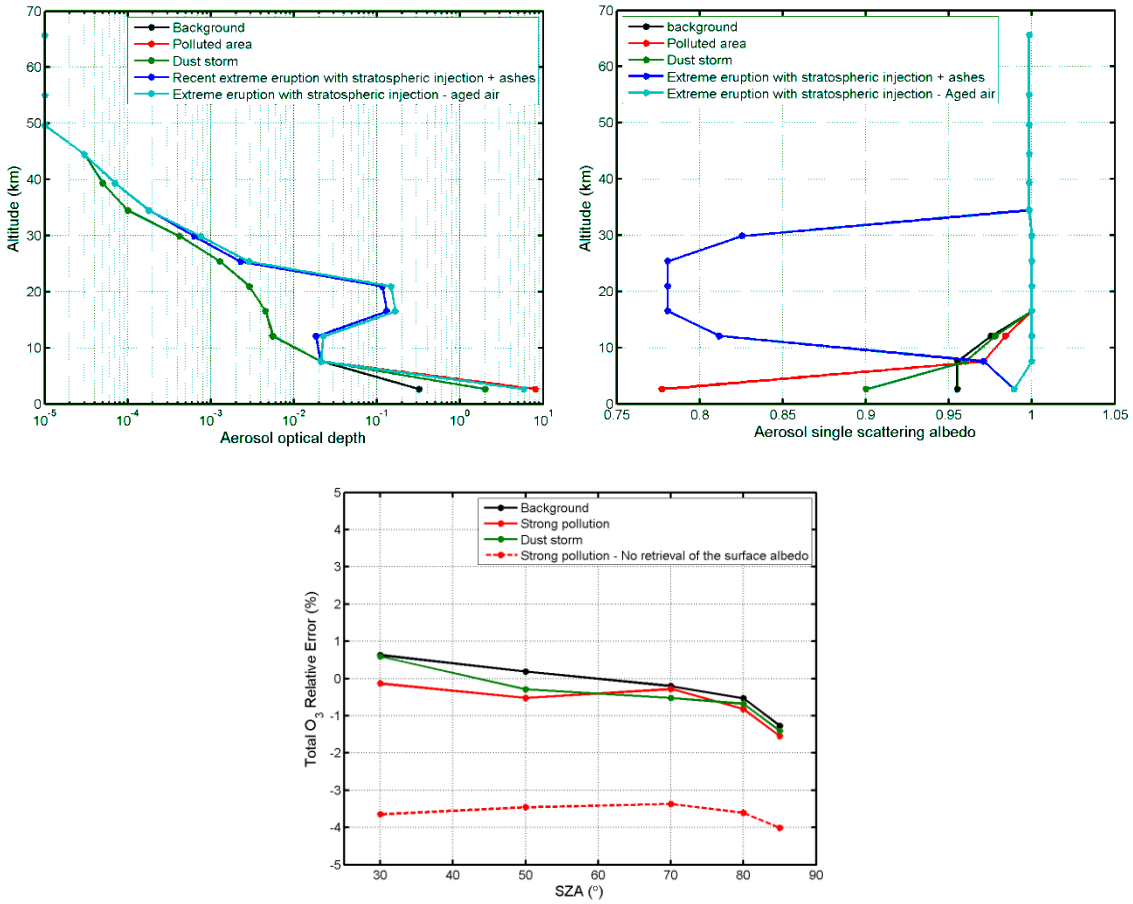

Figure 7.3: Upper panels: Profiles of aerosol optical properties used for generating synthetic radiances. Lower panel: Total ozone error (%) due to the neglect of aerosols in the retrieval scheme, plotted as a function of SZA. Note that the fit of an effective surface albedo implicitly accounts for aerosols to some degree. The total ozone errors are much larger without this albedo adjustment procedure (red dashed line).

Error due to cloud parameters:

Using typical values of uncertainties on the cloud fraction and cloud top height values generated by a cloud algorithm, sensitivity tests have been performed to estimate the impact of these uncertainties on total ozone retrievals. We compare the ozone columns retrieved from one GOME orbit with modified cloud parameters to those retrieved with the original parameters. The global effective scene approach has been used for these sensitivity tests. Figure 7.4. shows ozone column relative differences due to an error in the cloud fraction on the one hand, and in the cloud top height on the other hand. The cloud fraction errors generally lead to small total ozone errors (< 0.5%). The cloud top height error may lead to total ozone errors of up to 1.5%.

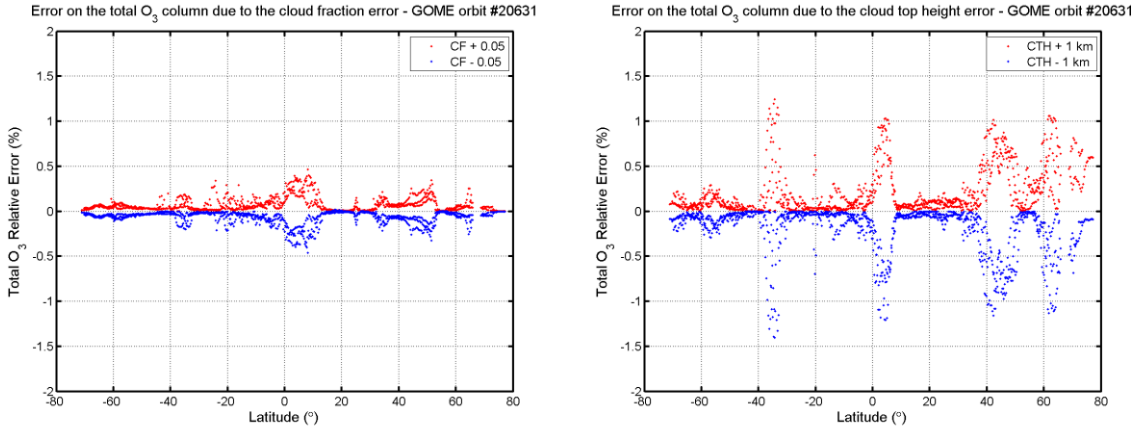

Figure 7.4: Left panel: Total ozone errors due to an error on the cloud fraction of 0.05. Right panel: Total ozone errors due to an error on the cloud top height of 1 km. These have been estimated for one GOME orbit.

Error due to the ozone cross-section:

As mentioned before, the ozone cross-sections provided by Brion, Daumont and Malicet are the baseline for ozone retrievals. An alternative cross-section data set has been released by the University of Bremen (Serdyuchenko et al., 2014). Preliminary tests show that a switch to this data set would lead to an increase of the ozone columns by 1 to 2.5%, depending on the effective temperature (see Figure 7.5). This represents a good estimate of the intrinsic error in the cross-section data.

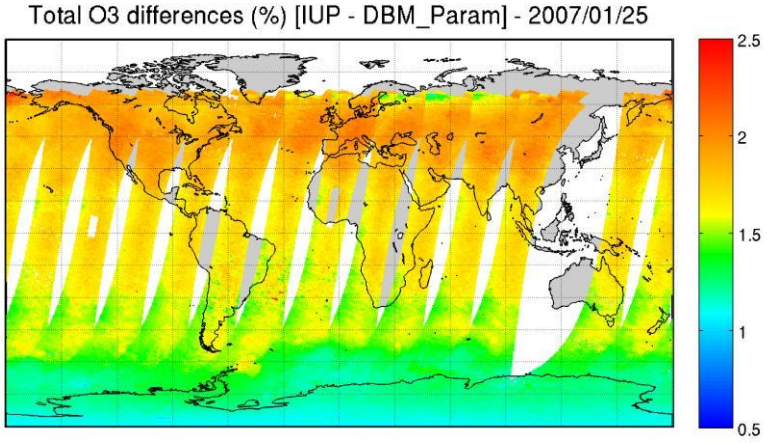

Figure 7.5: Relative ozone differences due to the switch from BDM to Serdyuchenko et al. 2014 ozone cross-section. This is illustrated for one day of GOME-2/MetOp-A measurements.

An estimate of the total error budget:

Table 7.1 contains a total error budget for the column ozone retrieval algorithm. Most of the figures are taken from the analyses discussed here. The level 1 calibration error source has been neglected since the instrument performance is difficult to foresee.

However we can expect that this contribution will be small if level-1 data is within the specifications. We have omitted other minor sources of error (Solar I_0 effect, the Ring effect, spectral stability, forward model errors such as the atmospheric layering scheme), as these make virtually no difference to the total budgets. A forward model error should however be added for S5P_TO3_DOAS for large SZAs (~3-4% for SZA > 80°).

Table 7.1: Total error budget for column ozone retrieval. [Blue indicates random error, red systematic error, green random or systematic (depending on time scale)]

Error source	Per cent error	
	SZA < 80°	SZA > 80°
Instrument signal-to-noise	<0.5	<2
O_3 absorption cross-sections and its atmospheric temperature	<2.5	<2.5
Interferences with other species (except in case of volcanic eruption)	<2	<1
Aerosols (except in case of volcanic eruption)	<1	<1.5
O_3 profile shape	<1	<4
Cloud fraction	<0.5	<0.5
Cloud top height	<1.5	<1.5
Total random error	**<1.6**	**<2.5**
Total systematic error	**<3.5**	**<5.0**

8 Validation

Please refer to the Mission Performance Centre (MPC) validation reports [RD9] that are regularly updated on http: mpc-vdaf.tropomi.eu.

9 Conclusions

The prototype algorithms for the generation of the operational S5P total column products are presented in this ATBD. Two different retrieval approaches are selected taking into consideration the requirements to provide accurate and fast near-real-time data products fitting Copernicus needs and the requirements for off-line and reprocessed products compatible with long-term historical data records needed for the creation of climate data sets.

S5P_TO3_DOAS and S5P_TO3_GODFIT are the default algorithms for the S5P total ozone near-real-time and off-line/reprocessed products respectively. Both algorithms use the LIDORT radiative transfer model together with the TROPOMI cloud information provided by the operational S5P cloud products. The DOAS technique for total ozone retrieval was used from the start of the GOME/ERS-2 mission; DOAS retrieval is currently implemented in the SCIAMACHY and GOME-2 operational processors. The DOAS technique is fast and provides ozone columns at the 1% level of accuracy in most situations The GODFIT approach provides even more accurate total ozone columns, at the cost of slower computational performance.

The GODFIT algorithm is the baseline for the total ozone reprocessing component of the ESA Ozone CCI project and has been used to generate an integrated record of total ozone from the GOME, GOME-2A/B, SCIAMACHY and OMI instruments [*Lerot et al.*, 2014]. The agreement between the GODFIT and the SBUV v8.6 data records is excellent with differences in the monthly zonal mean total ozone of only −0.32% to 0.76% [*Chiou et al.*, 2013].

The algorithms for the S5P total column products meet the accuracy requirements formulated for this mission. The near-real-time products are available 3 hours after sensing. The products are provided in NetCDF-CF and contain total ozone, ozone temperature, and error information including averaging kernels.

10 References

Aliwell, S. R., M. van Roozendael, P. V. Johnston, A. Richter, T. Wagner, D. W. Arlander, J. P. Burrows, D. J. Fish, R. L. Jones, K. K. Tørnkvist, J.-C. Lambert, K. Pfeilsticker, and I. Pundt, Analysis for BrO in zenith-sky spectra: An intercomparison exercise for analysis improvement, *J. Geophys. Res.,* **107**(D14), doi: 10.1029/2001JD000329, 2002.

Balis, D., M. Kroon, M. E. Koukouli, et al., Validation of Ozone Monitoring Instrument total ozone column measurements using Brewer and Dobson spectrophotometer ground-based observations, *J. Geophys. Res.*, **112**, D24S46, doi:10.1029/2007JD008796, 2007.

Bass, A. M., and R. J. Paur, The Ultraviolet Cross Sections of Ozone: I. The Measurements, in *Atmospheric Ozone, Proceedings of the Quadrennial Ozone Symposium*, edited by C. S. Zerefos and A. Ghazi, D. Reidel Publishing Company, 1984.

Bhartia, P. K., Algorithm Theoretical Baseline Document, TOMS v8 Total ozone algorithm, (http://toms.gsfc.nasa.gov/version8/version8_update.html), 2003.

Bhartia P.K, and C. Wellemeyer, TOMS version 8 Algorithm Theoretical Basis Document, http://toms.gsfc.nasa.gov, 2004-11-24, 2004.

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

Bowker, D. E., R. E. Davies, D. L. Myrick, K. Stacy, and W. T. Jones, Spectral Reflectances of Natural Targets for Use in Remote Sensing Studies, NASA Reference Publication 1139, 1985.

Bodhaine, B., N. Wood, E. Dutton, and J. Slusser, On Rayleigh optical depth calculations, *J. Atmos. Ocean. Tech.*, **16**, 1854-1861, 1999.

Bogumil, K., J. Orphal, T. Homann, S. Voigt, P. Spietz, O. Fleischmann, A. Vogel, M. Hartmann, H. Bovensmann, J. Frerick, and J. Burrows: Measurements of molecular absorption spectra with the SCIAMACHY Pre-Flight Model: instrument characterization and reference data for atmospheric remote-sensing in the 230--2380 nm region, *Journal of Photochemistry and Photobiology A,* **157**, 167-184, 2003.

Bovensmann, H., J. Burrows, M. Buchwitz, J. Frerick, S. Noel, V. Rozanov, K. Chance, and A. Goede, SCIAMACHY: Mission Objectives and Measurement Modes, *J. Atmos. Sci.,* **56**, 127–150, 1999.

Brion, J., Chakir, A., Charbonnier, J., Daumont, D., Parisse, C. and J. Malicet.Absorption spectra measurements for the ozone molecule in the 350-830 nm region, *J. Atmos. Chem.*, **30**, 291-299, 1998.

Burrows, J., A. Dehn, B. Deters, S. Himmelmann, A. Richter, S. Voigt, and J. Orphal, Atmospheric Remote-Sensing Reference Data from GOME: Part 1. Temperature-dependent absorption cross-sections of NO_2 in the 231-794 nm range, *J. Quant. Spectrosc. Radiat. Trans.*, **60**, 1025-1031, 1998.

Burrows, J., A. Richter, A. Dehn. B. Deters, S. Himmelmann, S. Voigt, and J. Orphal, Atmospheric remote sensing reference data from GOME: Part 2. Temperature-dependent absorption cross-sections of O_3 in the 231-794 nm range, *J. Quant. Spectrosc. Radiat. Transfer*, **61**, 509-517, 1999a.

Burrows, J., M. Weber, M. Buchwitz, V. Rozanov, A. Ladstaetter-Weissenmeyer, A. Richter, R. de Beek, R. Hoogen, K. Bramstadt, K.-U. Eichmann, M. Eisinger, and D. Perner, The Global Ozone Monitoring Experiment (GOME): mission concept and first scientific results, *J. Atmos. Sci*, **56**, 151-175, 1999b.

Casadio, S., D. Loyola, and C. Zehner, GOME-MERIS cloud products inter-comparison on global scale, Atmospheric Science Conference, Frascati, 2006.

Caudill T.R., D.E. Flittner, B.M. Herman, O. Torres, R.D. McPeters, Evaluation of the pseudo-spherical approximation for backscattered ultraviolet radiances and ozone retrieval. *J. Geophys. Res.*, **102**, 3881-3890, 1997.

Chance, K., and R. 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., Analysis of BrO measurements from the Global Ozone Monitoring Experiment, *Geophys. Res. Lett*., **25**, 3335-3338, 1998.

Chance, K, Kurosu T., and C. Sioris, Undersampling correction for array-detector based satellite spectrometers, Applied Optics **44**(7), 1296-1304, 2005.

Chandrasekhar, S., Radiative Transfer, Dover Publications Inc., New York, 1960.

Chiou, E. W., Bhartia, P. K., McPeters, R. D., Loyola, D. G., Coldewey-Egbers, M., Fioletov, V. E., Van Roozendael, M., Lerot, C., Spurr, R., and Frith, S. M.: Comparison of profile total ozone from SBUV(v8.6) with GOME-type and ground-based total ozone for 16-yr period (1996 to 2011), Atmos. Meas. Tech. Discuss., **6**, 10081-10115, 2013.

Coldewey-Egbers, M., M. Weber, L. Lamsal, R. de Beek, M. Buchwitz, and J. Burrows, Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach, *Atmos. Chem. Phys.,* **5**, 1015–1025, 2005.

Coldewey-Egbers, M., S. Slijkhuis, B. Aberle, and D. Loyola, Long-term analysis of GOME in-flight calibration parameters and instrument degradation, *Appied Optics*, **47**, 4749-4761, 2008.

Dahlback, A., and K. Stamnes, A new spherical model for computing the radiation field available for photolysis and heating at twilight, *Planet. Space Sci.*, **39**, 671, 1991.

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, D., J. Brion, J. Charbonnier, and J. Malicet, Ozone UV spectroscopy. I. Absorption cross-sections at room temperature, *J. Atmos. Chem*., **15**, 145-155, 1992.

De Backer, H., and D. De Muer, Intercomparison of Total Ozone Data Measured With Dobson and Brewer Ozone Spectrophotometers at Uccle (Belgium) From January 1984 to March 1991, Including Zenith Sky Observations, *J. Geophys. Res.*, Vol. 96, pp. 20,711-20,719, 1991.

Doicu, A., F. Schreier, and M. Hess, Iterative regularization methods for atmospheric remote sensing, *J. Quant. Spectrosc. Radiat. Transfer*, **83**, 47-61, 2004.

Doicu, A., F. Schreier, S. Hilgers, A. von Bargen, S. Slijkhuis, M. Hess, and B. Aberle, An efficient inversion algorithm for atmospheric remote sensing with application to UV limb observations, *J. Quant. Spectrosc. Radiat. Transfer*, **103**, 193-208, 2007.

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

Eskes, H., R. van der A, E. Brinksma, J. Veefkind, J. de Haan, and P. Valks, Retrieval and validation of ozone columns derived from measurements of SCIAMACHY on Envisat, *Atmos. Chem. Phys. Discuss.,* **5**, 4429–4475, 2005.

ESA, GOME Global Ozone Monitoring Experiment Users Manual, ed. F. Bednarz, ESA SP-1182, 1995.

ESA, The ESA Climate Change Initiatiave, Decription, EOP-SEP/TN/0030-09/SP, 2009.

ETOP05, Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado, 1988.

Fayt, C., and M. van Roozendael, WINDOAS Users Manual, http://www.oma.be/BIRA-IASB/Molecules/BrO, BIRA-IASB, Brussels, 2001.

Fioletov, V., J. Kerr, E. Hare, G. Labow, and R. McPeters, An assessment of the world ground-based total ozone network performance from the comparison with satellite data, *Journal of Geophysical Research*., 1999.

Fioletov, V. E., J. B. Kerr, C. T. McElroy, et al., The Brewer reference triad, *Geophys. Res. Lett.*, **32**, L20805, doi:10.1029/2005GL024244, 2005.

Fioletov, V. E., et al., Performance of the ground-based total ozone network assessed using satellite data, J. Geophys. Res., 113, D14313, doi:10.1029/2008JD009809, 2008.

Fortuin, J. P. F., and H. Kelder, An ozone climatology based on ozonesonde and satellite measurements, *J. Geophys. Res.*, **103**, 31709-31734, 1998.

Garane, K., Lerot, C., Coldewey-Egbers, M., Verhoelst, T., Koukouli, M. E., Zyrichidou, I., Balis, D., Danckaert, T., Goutail, F., Granville, J., Hubert, D., Keppens, ., Lambert, J.-C., Loyola, D., Pommereau, J.-P., Van Roozendael, M., Zehner, C., Quality assessment of the Ozone_cci Climate Research Data Package (release 2017) – Part 1: Ground-based validation of total ozone column data products, 115194, 1385–1402. https://doi.org/10.5194/amt-11-1385-2018, 2018

GDOAS Delta-validation report, ESA contract AO/1-4235/02/I-LG, 28 January 2004.

GOME Global Ozone Monitoring Experiment Users Manual, ed. F. Bednarz, ESA SP-1182, 1995.

Grainger, J., and J. Ring, Anomalous Fraunhofer Line Profiles, *Nature*, **193**, 762, 1962.

Greenblatt, G. D., J. J. Orlando, J. B. Burkholder, and A. R. Ravishankara, Absorption measurements of oxygen between 330 and 1140 nm, *J. Geophys. Res.*, *95*, 18577-18582, 1990.

Hassler, B., G. Bodeker, and M. Dameris, Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements, *Atmos. Chem. Phys*., **8**, 5403-5421, 2008.

Hasekamp, O., J. Landgraf, and R. van Oss, The need of polarization monitoring for ozone profile retrieval from backscattered sunlight, *J. Geophys. Res*., **107**, 4692, 2002.

Hao N., Koukouli M. E., Inness A., Valks P., Loyola D. G., Zimmer W., Balis D. S., Zyrichidou I., Van Roozendael M., Lerot C., Spurr R. J. D., GOME-2 total ozone columns from MetOp-A/MetOp-B and assimilation in the MACC system, Atmospheric Measurement Techniques, **7**, 2937-2951, 2014.

Hedelt, P., Efremenko D.S., Loyola D.G., Spurr, R., and Clarisse L.: SO2 Layer Height retrieval from Sentinel-5 Precursor/TROPOMI using FP_ILM, Atmos. Meas. Tech.12, 5503-5517 doi: 10.5194/amt-12-5503-2019, 2019.

Heinen T., S. Kiemle, B. Buckl, E. Mikusch and D. Loyola, The Geospatial Service Infrastructure for DLR's National Remote Sensing Data Library, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, **2**, 260-269, 2009.

Herman, J., and E. Celarier, Earth surface reflectivity climatology at 340 nm to 380 nm from TOMS data, *J. Geophys. Res*., **102**, 28003-28011, 1997.

Hollman, R., et al., The ESA Climate Change Initiative: satellite data records for essential climate variables, *Bull. Am. Met. Soc*, doi" 10.1175/BAMS-D-11-00254.1, 2013.Joiner, J., P.K. Bhartia, R.P. Cebula, E. Hilsenrath, R. McPeters, and H.W. Park, Rotational Raman Scattering (Ring Effect) in Satellite Backscatter Ultraviolet Measurements, Appl. Opt., 34, 4513-4525, 1995.

Joiner, J., and P. Bhartia, Accurate determination of total ozone using SBUV continuous spectral scan measurements, *J. Geophys. Res.*, **102**, 12957-12969, 1997.

Kleipool, Q. L., M. R. Dobber, J. F. de Haan, and P. F. Levelt, Earth surface reflectance climatology from 3 years of OMI data *J. Geophys. Res.*, **113**, D18308, doi:10.1029/2008JD010290, 2008.

Kneizys, F. X., E. P. Shettle, L. W. Abreu, J. H. Chetwynd, G. P. Anderson, W. O. Gallery, J. E. A. Selby, and S. A. Clough, Users Guide to LOWTRAN 7, Air Force Geophysics Laboratory, Environmental Research Papers, No. 1010, AFGL-TR-88-0177, 1988.

Koelemeijer, R., and P. Stammes, A fast method for retrieval of cloud parameters using oxygen *A* band measurements from the Global Ozone Monitoring Experiment, *J. Geophys. Res.*, **106**, 3475-3490, 2001.

Koelemeijer, R., J. de Haan, J. Hovenier, and P. Stammes, A database of spectral surface reflectivity in the range 335-772 nm derived from 5.5 years of GOME observations, *J. Geophys. Res.*, **108,** D4070, doi:10.1029/2002JD0024, 2003.

Komhyr, W. D., C. L. Mateer, and R. D. Hudson, Effective Bass-Paur 1985 Ozone Absorption Coefficients for Use With Dobson Ozone Spectrophotometers, *J. Geophys. Res.*, **98**, 20,451-20,465, 1993.

Koukouli, M. E., D. S. Balis, D. G. Loyola, P. Valks, W. Zimmer, N. Hao, J-C. Lambert, M. Van Roozendael, C. Lerot and R. J. D. Spurr, Geophysical validation and long-term consistency between GOME-2/MetOp-A total ozone column and measurements from the sensors GOME/ERS-2, SCIAMACHY/ENVISAT and OMI/Aura, *Atmos. Meas. Tech.,* **5**, 2169-2181, 2012.

Koukouli, M. E., C. Lerot, J. Granville, F. Goutail, J.-C. Lambert, J.-P. Pommereau, D. Balis, I. Zyrichidou, M. Van Roozendael, M. Coldewey-Egbers, D. Loyola, G. Labow, S. Frith, R. Spurr and C. Zehner, Evaluating a new homogeneous total ozone climate data record from GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, *J. Geophys. Res. Atmos,* **120**, 2015.

Kuze, A., and K. V. Chance, Analysis of Cloud-Top Height and Cloud Coverage from Satellites Using the O_2 *A* and *B* Bands, *J. Geophys. Res.*, **99**, 14481-14491, 1994.

Labow, G. J., R. D. McPeters, and P. K. Bhartia, A Comparison of TOMS, SBUV and SBUV/2 Version 8 Total Column Ozone Data with Data from Ground Stations, Proceedings of QOS2004, Kos, 1-8 June, Kos, Greece, 2004.

Labow, G. J., Ziemke, J. R., McPeters, R. D., Haffner, D. P. and Bhartia, P. K.: A Total Ozone Dependent Ozone Profile Climatology based on Ozone-Sondes and Aura MLS Data, J. Geophys. Res. Atmos., n/a-n/a, doi:10.1002/2014JD022634, 2015.

Lacis, A., J. Chowdhary, M. Mishchenko, and B. Cairns, Modeling errors in diffuse sky radiance: vector vs. scalar treatment, *Geophys. Res. Lett.*, **25**, 135-138, 1994.

Lambert, J.-C., M. Koukouli, D. Balis, J. Granville, C. Lerot, D. Pieroux, M. Van Roozendael, GOME/ERS-2 – GDP5.0 Upgrade of the GOME Data Processor for Improved Total Ozone Columns – Validation Report, 2012.

Lamsal, L. N., M. Weber, S. Tellmann, and J. P. Burrows, Ozone column classified climatology of ozone and temperature profiles based on ozonesonde and satellite data,

J. Geophys. Res., 109, D20304, doi:10.1029/2004JD004680, 2004.Lerot, C., M. Van Roozendael, J. van Geffen, J. van Gent, C. Fayt, R. Spurr, G. Lichtenberg, and A. von Bargen, Six years of total ozone column measurements from SCIAMACHY nadir observations, *Atmos. Meas. Tech.,* **2**, 87–98, 2009.

Lerot, C., M. Van Roozendael, J. van Geffen, J. van Gent, C. Fayt, R. Spurr, G. Lichtenberg, and A. von Bargen, Six years of total ozone column measurements from SCIAMACHY nadir observations, Atmos. Meas. Tech., 2, 87–98, doi:10.5194/amt-2-87-2009, 2009.

Lerot, C., M. van Roozendael, J. van Gent, D. Loyola, and R. Spurr, The GODFIT algorithm: a direct fitting approach to improve the accuracy of total ozone measurements from GOME, *Int. J. Remote Sensing*, **31**(2), 543-550, doi: 10-1080/01431160902893576, 2010.

Lerot, C., M. van Roozendael, R. Spurr, D. Loyola, M. Coldewey-Egbers, S. Kochenova, J. van Gent, M. Koukouli, D. Balis, J.-C. Lambert, J. Granville, and C. Zehner, Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat and GOME-2/MetOp-A, *J. Geophys. Res.,*, 119, doi: 10.1002/2013JD020831, 2014.

Levelt, P., G. van den Oord, M. Dobber, A. Malkki, H. Visser, J. de Vries, P. Stammes, J. Lundell, and H. Saari, The Ozone Monitoring Instrument, IEEE *Trans. Geosc. Rem. Sens*., **44**(5), 1093-1101, 2006.

Liu, X., K. Chance, C. Sioris, R. Spurr, T. Kurosu, R. Martin, and M. Newchurch, Ozone Profile and Tropospheric Ozone Retrievals from Global Ozone Monitoring Experiment: Algorithm Description and Validation, *J. Geophys. Res.*, **110**, D20307, doi:10.1029/2005JD006240, 2005.

Liu, X., Chance, K., Sioris, C. E., and Kurosu, T. P.: Impact of using different ozone cross sections on ozone profile retrievals from Global Ozone Monitoring Experiment (GOME) ultraviolet measurements, Atmos. Chem. Phys., 7, 3571–3578, https://doi.org/10.5194/acp-7-3571-2007, 2007.

Loyola D.: Applications of Neural Network Methods to the Processing of Earth Observation Satellite Data, Neural Networks, 19/2, 168-177, https://doi.org/10.1016/j.neunet.2006.01.010, 2006.

Loyola D., Coldewey-Egbers M., Dameris M., Garny H., Stenke A., Van Roozendael M., Lerot C., Balis D. and Koukouli M., Global long-term monitoring of the ozone layer - a prerequisite for predictions, International Journal of Remote Sensing, vol. 30, no. 15, pp. 4295-4318, 2009.

Loyola, D., M. Koukouli, P. Valks, D. Balis, N. Hao, M. Van Roozendael, R. Spurr, W. Zimmer, S. Kiemle, C. Lerot, and J-C. Lambert, The GOME-2 Total Column Ozone Product: Retrieval Algorithm and Ground-Based Validation, J. Geophys. Res., 116, D07302, doi:10.1029/2010JD014675, 2011.

Loyola, D., W. Zimmer, S. Kiemle, P. Valks, Product User Manual for GOME Total Columns of Ozone, NO2, tropospheric NO2, BrO, SO2, H2O, HCHO, OClO, and Cloud Properties, 2012.

Loyola, D., Pedergnana, M., and Gimeno García, S., Smart sampling and incremental function learning for very large high dimensional data, Neural Networks, 78, 75–87, https://doi.org/10.1016/j.neunet.2015.09.001, 2016.

Loyola, D., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R., Pedergnana, M., Doicu, A., Molina García, V., and Schüssler, O., The operational cloud retrieval

algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11, 409–427, 2018.

Loyola, D. G., Xu, J., Heue, K.-P., Zimmer, W.: Applying FP_ILM to the retrieval of geometry-dependent effective Lambertian equivalent reflectivity (GGE_LER) to account for BRDF effects on satellite UV/VIS trace gas measurements, accepted for Atmos. Meas. Tech., 2020

Malicet, J., D. Daumont, J. Charbonnier, C. Parisse, A. Chakir, and J. Brion, Ozone UV Spectroscopy. II. Absorption Cross-Sections and Temperature Dependence, *J. Atmos. Chem*., **21**, 263-273, 1995.

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

McPeters, R.D. and G.J. Labow: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, *J. Geophys. Res.*, **117**, D10303, doi:10.1029/2011JD017006, 2012.

Miller, A. J., Nagatani, R. M., Flynn, L. E., Kondragunta, S., Beach, E., Stolarski, R., McPeters, R. D., Bhartia, P. K., De-Land, M. T., Jackman, C. H., Wuebbles, D. J., Patten, K. O., and Cebula, R. P.: A cohesive total ozone data set from the SBUV(/2) satellite system, J. Geophys. Res., **107**(D23), 4701, doi:10.1029/2001JD000853, 2002.

Munro, R., M. Eisinger, C. Anderson, J. Callies, E. Corpaccioli, R. Lang, A. Lefebvre, Y. Livschitz, and A. Perez Albinana, GOME-2 on MetOp: From in-orbit verification to routine operations, in: Proceedings of EUMETSAT Meteorological Satellite Conference, Helsinki, Finland, 12-16 June 2006.

Orphal, J., A critical review of the absorption cross-sections of O_3 and NO_2 in the 240-790 nm region, *Journal of Photochemistry and Photobiology A,* **157**, 185-209, 2003.

Platt, U., Differential optical absorption spectroscopy (DOAS), in Air monitoring by spectroscopic techniques, ed. M. Sigrist, *Chem. Anal. Ser.*, *127*, 27-84, 1994.

Richter, A., and J. Burrows, Tropospheric NO_2 from GOME measurements, *Adv. Space Res.*, **29**, 1673-1683, 2002.

Rodgers, C., Inverse Methods for Atmospheres: Theory and Practice, *World Scientific Press*, 2001.

Rozanov A.V., V.V. Rozanov, and J.P. Burrows, Combined differential-integral approach for the radiation field computation in a spherical shell atmosphere: Non-limb geometry. *J. Geophys. Res.*, **105**, 22937-22942, 2000.

Schuessler, O., D. Loyola, A. Doicu, and R. Spurr. Information Content in the Oxygen A-band for the Retrieval of Macrophysical Cloud Parameters, IEEE Transactions on Geoscience and Remote Sensing, 52, 3246-3255, 2014.

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, https://doi.org/10.5194/amt-7-625-2014, 2014.

Sioris, C. and W. Evans, Filling-in of Fraunhofer and gas-absorption lines as caused by rotational Raman scattering, *Applied Optics*, **38**, 2706-2713, 1999.

Slijkhuis, S., B. Aberle, and D. Loyola, GOME Data Processor Extraction Software User's Manual, ER-SUM-DLR-GO-0045, 2004.

Spurr, R., Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, *J. Quant. Spectrosc. Radiat. Transfer*, **75**, 129-175, 2002. +

Spurr, R., LIDORT V2PLUS, a comprehensive radiative transfer package for nadir viewing spectrometers, Proceedings SPIE conference, Barcelona, 2003.+

Spurr, R., D. Loyola, W. Thomas, W. Balzer, E. Mikusch, B. Aberle, S. Slijkhuis, T. Ruppert, M. Van Roozendael, J.-C. Lambert, and T. V. Soebijanta, GOME Level 1-to-2 Data Processor Version 3.0: A Major Upgrade of the GOME/ERS-2 Total Ozone Retrieval Algorithm, *Applied Optics*, **44**, 7196-7209, 2005.

Spurr, R., LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. *Light Scattering Reviews*, Volume 3, ed. A. Kokhanovsky, Springer, 2008.

Spurr R., M. van Roozendael, D. Loyola, C. Lerot, J. van Geffen, J. van Gent, C. Fayt, J.-C. Lambert, W. Zimmer, A. Doicu, S. Otto, D. Balis, M. Koukouli, C. Zehner, GOME/ERS-2 – GDP5.0 Upgrade of the GOME Data Processor for Improved Total Ozone Columns – Algorithm Theoretical Basis Document, 2012.

Stamnes, K., S.-C.Tsay, W. Wiscombe, K. Jayaweera, Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media, *Applied Optics*, **27**, 2502-2509, 1988.

Van Oss, R., and R. Spurr, Fast and accurate 4 and 6 stream linearized discrete ordinate radiative transfer models for ozone profile retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, **75**, 177-220, 2002.

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, Ed. J-C. Lambert, ERSE-DTEX-EOAD-TN-02-0006, ESA/ESRIN, Frascati, Italy, Chap.6, pp.97-129, 2002.

Van Roozendael, M., D. Loyola, R. Spurr, D. Balis, J-C. Lambert, Y. Livschitz, P. Valks, T. Ruppert, P. Kenter, C. Fayt, and C. Zehner, Ten years of GOME/ERS2 total ozone data: the new GOME Data Processor (GDP) Version 4: I. Algorithm Description, *J. Geophys Res.*, doi: 10.1029/2005JD006375, 2006.

Van Roozendael, M., R. Spurr, D. Loyola, C. Lerot, D. Balis, J-C. Lambert, W. Zimmer, J. van Gent, J. van Geffen, M. Koukouli, J. Granville, A. Doicu, C. Fayt, C. Zehner, Sixteen years of GOME/ERS2 total ozone data: the new direct-fitting GOME Data Processor (GDP) Version 5: I. Algorithm Description, *J. Geophys. Res.*, **117**, D03305, doi: 10.1029/2011JD016471, 2012.

Veefkind, J. P., J. F. de Haan, E. J. Brinksma, M. Kroon, and P. F. Levell, Total Ozone from the Ozone Monitoring Instrument (OMI) using the DOAS technique, IEEE Trans. Geosci. Remote Sens., **44**, 1239–1244, 2006.

Wagner, T., Beirle, S., and Deutschmann, T.: Three-dimensional simulation of the Ring effect in observations of scattered sun light using Monte Carlo radiative transfer models, Atmos. Meas. Tech., 2, 113-124, doi:10.5194/amt-2-113-2009, 2009.

Wellemeyer, C., S. Taylor, C. Seftor, R. McPeters and P. Bhartia, A correction for total ozone mapping spectrometer profile shape errors at high latitude, *J. Geophys. Res.,* **102**, 9029-9038, 1997.

Xu, J., Schüssler, O., Loyola R, D., Romahn, F. and Doicu, A.: A novel ozone profile shape retrieval using Full-Physics Inverse Learning Machine (FP_ILM). IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 10(12), pp. 5442–5457, doi: 10.1109/JSTARS.2017.2740168, 2017.

Ziemke, J.R., S. Chandra, G. J. Labow, P. K. Bhartia, L. Froidevaux, and J. C. Witte: A global climatology of tropospheric and stratospheric ozone derived from Aura OMI and MLS measurements, Atmos. Chem. Phys., **11**, 9237–9251, 2011.