

Royal Netherlands Meteorological Institute Ministry of Infrastructure and Water Management

TROPOMI ATBD Ozone Profile





document number	:	S5P-KNMI-L2-0004-RP
author	:	Pepijn Veefkind, Arno Keppens, Johan de Haan
CI identification	:	CI-7340-ATBD
issue	:	1.0.0
date	:	2021-10-22
status	:	Released

Document approval record

	digital signature
prepared:	
checked:	
approved PM:	
approved PI:	
approved client:	

Document change record

issue	date	item	comments
0.2.0	2012-09-27		First incomplete draft
0.3.0	2013-04-10		Comments by ESA on issue 0.2.0 worked in
0.4.0	2013-06-03		Added error analysis
0.5.0	2013-06-20		Comments internal review worked in
0.9.0	2013-12-10		Comments external review worked in
0.10.0	2014-04-15		Updated; removed instrument description; added results using OMI data; changed numbering of figures, tables, and equations
0.11.0	2014-10-06		RIDs (a.o. comments by Joerg Langen) worked in, updated tables, and added Appendix C.
0.13.0	2015-09-15		Minor changes dealing with efficiency
0.15.0	2021-10-12	All	Major update
1.0.0	2021-10-22	Section 1.3 Section 4.2 Section 5.1.5 Section 6.1 Section 8.1 Section 8.3 Section 9	Textual comments from ESA on version 0.15.0 worked in. First released version.

Contents

Docume	ent approval record	.2
Docume	ent change record	.3
Content	ے۔ ایجا ہے۔	.4
1	Introduction	
∎ 1.1	Identification	
1.1		
	Purpose and objective	
1.3	Document overview	
2	Applicable and reference documents	
2.1	Applicable documents	
2.2	Reference documents	
2.3	Electronic references	.7
3	Terms, definitions and abbreviated terms	.8
3.1	Terms and definitions	
3.2	Acronyms and abbreviations	
4	Introduction to ozone profile retrieval	
4 4.1		
	Product description	
4.2	Heritage	
4.3	Product requirements	
4.4	Overview of the retrieval method	
4.5	Foreseen update approach	
4.5.1	Improved quality flags and qa_value	
4.5.2	Radiometric soft calibration	
4.6	Design considerations	
5	Algorithm description	14
5.1	Pre-processing	
5.1.1	Spectral Calibration	
5.1.2	Spatial Regridding	
5.1.3	Spectral Regridding	
5.1.4	Polarisation and Raman Correction	
5.1.5	Radiometric Correction	
5.1.6	Signal-to-Noise Ratio	
5.2	Forward model	
5.2.1	Output forward model	
5.2.2 5.2.2.1	Model atmosphere	
5.2.2.1	Molecular scattering Ozone absorption	
5.2.2.2	SO ₂ absorption	
5.2.2.4	Lambertian surface and cloud model	
5.2.3	Processing Steps.	
5.3	Retrieval algorithm	
5.3.1	Optimal Estimation	
5.3.1.1	Cost function	
5.3.1.2	State vector elements and prior information	22
5.3.1.3	Updating the state vector.	
5.3.1.4	Convergence	26
5.4	Error characterization	
5.4.1	Measurement errors	26
5.4.2	Fit residue	
5.4.3	A posteriori errors	
5.4.4	Averaging kernel and DFS	
5.4.5	Gain matrix	
5.4.6	Values and errors for sub columns	
6	The Ozone Profile Product	28

6.1	High level data product description	
6.2	Applying the Averaging Kernel	
6.3	Quality Indicator	
7	Error analyses	
7.1	Standard Case	
7.2	Model errors	
7.2.1	Spectroscopic data	
7.2.2	A priori ozone profile	
7.2.3 7.2.3.1	Radiative transfer	
7.2.3.1	ISRF Convolution Number of streams	
7.2.3.2	Polarization and Raman Scattering	
7.2.3.3	Lambertian cloud model and fitting of cloud properties	
7.2.5	Sun glint	
7.2.6	Corrections for a spherical atmosphere	
7.2.7	Absorbing aerosol.	
7.3	Instrumental errors	
7.3.1	Radiometric errors	
7.3.2	ISRF errors	
7.3.3	Noise	
7.4	Summary of Error Analysis	
8	Validation	44
8.1	Reference datasets	
8.1.1	Ozonesondes	
8.1.2	DIAL ozone profile reference measurements	
8.1.3	Campaign needed with reference to specific algorithm features	
8.1.4	Satellite inter-comparison	
8.2	Validation approaches	45
8.2.1	Automated validation server	45
8.2.2	Daily global maps	
8.2.3	Parameter correlation checks	
8.3	Initial validation results	45
9	Conclusion	50
Referen	ces	51
Append	ix A Description of prototype software	53
A.1	DISAMAR	53
A.1.1	Some strong points of DISAMAR	53
A.1.2	Limitations of DISAMAR	
A.2	Calculating properties of finite sub columns	55
A.3	Special interpolation to improve the efficiency	
Append	ix B Description of Test Cases	58
B.1	Description of Geophysical Test Cases	
Append	ix C Input Data	63
C.1	Dynamic input	
C.2	(Semi) static input data	
		-

1 Introduction

1.1 Identification

This document is the Algorithm Theoretical Basis Document (ATBD) for the S5P/TROPOMI ozone profile product. The ATBD together with the Product User Manual (PUM) [RD10] and the Product Readme File (PRF) [RD11] contain the main documentation for the user of the product.

1.2 **Purpose and objective**

The purpose of the document is to give a high-level description of the algorithm that will be used to retrieve ozone profiles from TROPOMI measurements. The algorithm computes ozone concentrations at about 30 levels in the atmosphere.

1.3 Document overview

Section 4 gives an introduction to ozone profile retrieval. Section 5 provides a description of the retrieval algorithm. In Section 6 the content of the data product is described. In Section 7 the effects of error analyses are presented. In Section 8 validation and tools for validation are discussed. In Section 9, conclusions are drawn.

2 Applicable and reference documents

2.1 Applicable documents

- [AD01] Sentinel-5P Level 2 Processor Development Statement of Work; source: ESA/ESTEC; ref: S5P-SW-ESA-GS-053; issue: 1; date: 2012-03-2.
- [AD02] GMES Sentinel-5 Precursor S5p System Requirement Document (SRD); source: ESA/ESTEC; ref: S5p-RS-ESA-SY-0002; issue: 4.1.
- [AD03] GMES Sentinels 4 and 5 Mission Requirements Document; source: ESA/ESTEC; ref: EOP-SMA/1507/JL-dr; issue: 3; date: 2011-09-21.

2.2 Reference documents

- [RD01] Terms, definitions and abbreviations for TROPOMI L01b data processor; source: KNMI; ref: S5P-KNMI-L01B-0004-LI; issue: 1.0.0; date: 2011-09-15
- [RD02] Terms and symbols in the TROPOMI Algorithm Team; source: KNMI; ref: SN-TROPOMI-KNMI-049; date: 2011-09-28.
- [RD03] Science Requirements Document for TROPOMI. Volume 1; source: KNMI & SRON; ref: RS-TROPOMI-KNMI-017; issue: 2.0; date: 2008-10-30.
- [RD04] CAPACITY: Operational Atmospheric Chemistry Monitoring Missions Final report; source: KNMI; ref: CAPACITY; date: Oct. 2005.
- [RD05] CAMELOT: Observation Techniques and Mission Concepts for Atmospheric Chemistry; **source:** KNMI; **ref:** RP-CAM-KNMI-050; **date:** Nov. 2009.
- [RD06] TROPOMI ATBD of the total and tropospheric NO2 data products; **source**: KNMI; **ref**: S5P-KNMI-L2-0005-RP; **issue**: 2.2.0; **date**: 16-06-2021
- [RD07] TROPOMI ATBD of the UV Aerosol Index; source: KNMI; ref: S5P-KNMI-L2-0008-RP; issue: 2.0; date: 05-07-2021
- [RD08] S5L2PP: Sentinel 5 L2 Prototype Processors ATBD for Cloud; **source**: KNMI; **ref**: KNMI-ESA-S5L2PP-ATBD-005; **issue**: 4.1; **date**: 04-10-2021
- [RD09] Sentinel 5-precursor/TROPOMI KNMI and SRON level 2 Input Output Data Definition source: KNMI; ref: S5P-KNMI-L2-0009-SD; issue: 15.0.0; date: 05-08-2021
- [RD10] Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Ozone profiles; **source:** KNMI; **ref:** S5P-KNMI-L2-0020-MA ; **issue:** 1.0.0; **date**: 25-06-2021
- [RD11] S5P Mission Performance Centre Ozone Profile [L2_O3_PR] Readme; **source:** KNMI; **ref:** S5P-MPC-KNMI-PRF-O3_PR ; **issue:** 1.0; **date**: TBD

2.3 Electronic references

- [ER01] http://www.unidata.ucar.edu/software/netcdf/docs/
- [ER02] <u>https://mpc-vdaf.tropomi.eu</u>
- [ER03] http://evdc.nilu.no
- [ER04] http://ndacc.org
- [ER05] http://woudc.org
- [ER06] <u>https://www-air.larc.nasa.gov/missions/TOLNet/</u>
- [ER07] <u>https://tropo.gsfc.nasa.gov/shadoz/</u>
- [ER08] https://atmosphere.copernicus.eu

3 Terms, definitions and abbreviated terms

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

3.1 Terms and definitions

Effective cloud fraction	Cloud fraction obtained when the model cloud is a Lambertian surface with a fixed albedo of 0.80. The effective cloud fraction is in general wavelength dependent.
Effective cloud pressure	Cloud pressure obtained when the model cloud is a Lambertian surface with a fixed albedo of 0.80. The effective cloud pressure is in general wavelength dependent, because the illumination of the cloud is with wavelength dependent.

3.2 Acronyms and abbreviations

AMF	Air Mass Factor
CAMELOT	Chemistry of the Atmosphere Mission concEpts and sentineL Observations Techniques
CAPACITY	Composition of the Atmosphere: Progress to Applications in the user CommunITY
CTM	Chemical Transport Model
DFS	Degrees of Freedom of the Signal
DISAMAR	Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval - The DISAMAR tool is KNMI prototype code for simulations of radiance and retrievals
DOAS	Differential Optical Absorption Spectroscopy
EVDC	ESA Validation Data Centre
FRESCO	Fast REtrieval Scheme for Clouds from the Oxygen A band
	This is also the name for the L2 KNMI cloud support product
ISRF	Instrument Spectral Response Function
MLS	Microwave Limb Sounder
MPC	Mission Performance Center
NRT	Near-Real Time (i.e. processing within 3 hours of measurement)
OE	Optimal Estimation
OMPS	Ozone Mapping and Profiler Suite
PBL	Planetary Boundary Layer
RRS	Rotational Raman Scattering
SAA	Solar Azimuth Angle
SNR	Signal to Noise Ratio
SZA	Solar Zenith Angle
TOA	Top Of Atmosphere (scattering and absorption above TOA is ignored)
UVAI	UV Absorbing Aerosol Index
VAA	Viewing Azimuth Angle
VMR	Volume Mixing Ratio
VZA	Viewing Zenith Angle

4 Introduction to ozone profile retrieval

Retrieved ozone profiles are used to monitor the evolution of stratospheric and tropospheric ozone. Such monitoring is important as the ozone layer protects life on Earth, and climate change might affect the ozone layer. In contrast, tropospheric ozone is toxic and it plays an important role in tropospheric chemistry. The aim of the TROPOMI ozone profile product is to continue the record of stratospheric ozone, monitor changes, and improve the accuracy of the retrieved stratospheric profiles. In addition, we focus on the retrieval of the tropospheric ozone. This is not only important for studies of atmospheric pollution in the troposphere, but also important for climate studies as tropospheric ozone is a greenhouse gas.

4.1 Product description

The TROPOMI ozone profile product provides vertical information of ozone in two ways:

- As a number density profile at 33 pressure levels;
- As 6 sub-columns with a vertical sampling of 6 km up to an altitude of 24 km and lower sampling above (24-32 km, and 32 km to TOA).

The full profile information is intended for advanced users that also consider the averaging kernel. The sub-columns are intended for more simple usage, where to effect of vertical smoothing can be neglected, i.e., assuming roughly one retrieval degree of freedom per layer). It is noted that the UV spectrum allows retrieving approximately 6 independent pieces of information on the ozone profile. Furthermore, the product requirements (see the beginning of Sec. 4.3) are specified for sub-columns. Examples of single retrieved profile and of curtain plot a shown in Figure 4-1 and Figure 4-2.

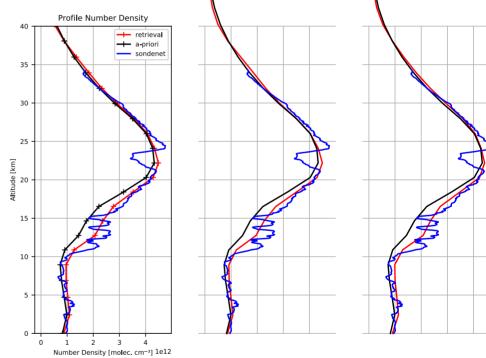


Figure 4-1. Example of a retrieved ozone profile from TROPOMI (red line) over De Bilt, The Netherlands on 8 August 2019. Also shown are the a priori profile (black) and the co-located balloon sounding (blue). Note that the vertical resolution of the retrieved profile is much lower compared to the balloon sounding.

The spatial sampling of the ozone profile and sub-columns is approximately 28x28 km² in nadir. This is a lower sampling than the TROPOMI Level 1B data and is achieved by co-adding ground pixels. This reduction of the spatial sampling is performed to reduce the computational cost and to increase the signal to noise ratio of the measured radiances.

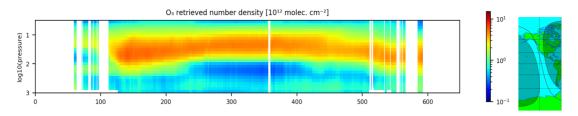


Figure 4-2. Curtain plot (time versus pressure) of all valid ozone profile retrievals for ground pixel number 30 for orbit 4958 measured on 27 September 2019. The right panel show the coverage of the orbit, which starts over Antarctica.

In addition to the a priori ozone profile, the retrieved ozone profiles and their errors, the following diagnostic information is provided: diagonal elements of the a priori error covariance matrix, a correlation length for the a priori errors, the full a posteriori error covariance matrix, and the averaging kernel (for the elements corresponding to the ozone profile).

Because a large amount of SO_2 (e.g. after a volcanic eruption) interferes with ozone profile retrieval, an SO_2 column amount is also retrieved. However, this column is derived using an assumed fixed profile for SO_2 and will not be the actual SO_2 column when the assumed profile differs significantly from the true profile. The main purpose is to identify pixels with less reliable ozone profiles due to the interference with SO_2 .

4.2 Heritage

The retrieval algorithm for the ozone profile uses Optimal Estimation (OE) (see Rodgers, 2000). In the retrieval the measured radiance is simulated using radiative transfer calculations. The radiative transfer calculations are initially based on the adding-doubling method (De Haan et al, 1987, Hovenier et al, 2004). The adding part of this algorithm is replaced by the layer based orders of scattering method to increase the efficiency of the calculations. The Jacobians needed in OE are calculated based on reciprocity, which is similar to using the adjoint version of the equation of radiative transfer. In many ways, the retrieval algorithm is similar to the algorithm used for ozone profile retrievals from the OMI measurements (Kroon et al., 2011). Differences are that we retrieve ozone at pressure levels in the atmosphere and not for atmospheric layers. Extensions are that we also retrieve ozone for sub-columns. In addition, many smaller changes have been implemented and the operational software is a completely new implementation. This implementation wraps the prototype algorithm called DISAMAR.

4.3 **Product requirements**

Accuracy requirements for sub columns of ozone are listed in Table 4.3-1 (taken from [RD03]). There are no requirements on the profile itself.

Sub column	PBL ¹⁾	0 - 6 km	6-12 km	12-18 km	18-50 km
Required accuracy	≤ 60 %	≤ 20 %	≤ 12 %	≤ 5 %	≤ 3 %

Table 4.3-1. Accuracy requirements for sub columns of the ozone profile.

¹⁾ In practice the pressure at the top of the planetary boundary layer is not known and results for this sub column cannot be provided. Simulation studies can be used to estimate the accuracy for the PBL.

Requirements on ozone were formulated in the CAPACITY study [RD04], but these requirements did not consider a specific instrument to perform the measurements and formulated generic requirements. The required vertical resolution ranges from 0.5 to 3 km, depending on the user category. Such a vertical resolution can only be met for limb sounding instruments and not for nadir viewing instruments like TROPOMI.

In the GMES Sentinels 4 and 5 Mission Requirements Document [AD03] the requirements are taken from the CAPACITY study.

Ozone in the atmosphere changes strongly with latitude and with longitude due to dynamical and (photo) chemical processes. In order to characterize the ozone profile, it should be determined at 20 or more levels in the atmosphere which explains the vertical resolution stated in the CAPACITY study. For nadir looking instruments like TROPOMI the vertical resolution is about 5 - 6 km (or lower outside the stratosphere) and retrieval of the ozone profile at more levels is only possible if a priori information on the ozone profile and a priori information on the variability of ozone is used as a constraint. Often, an ozone climatology is used for the a priori information, and we propose to use such a climatology also for the algorithm described here. Therefore, large smoothing errors can be expected, often much larger than the errors specified in Table 4.3-1. This means that we cannot meet the requirements listed in Table 4.3-1, unless we have accurate a priori information. Accurate a priori information, at least more accurate than the information obtained from a climatology, could be obtained from a chemistry transport model (CTM). In that case the retrieved ozone profiles will be accurate if the retrieved profile agrees with the a priori information. If, however, the retrieved ozone profiles differ significantly from the a priori information, the smoothing error becomes important and the retrieved profiles are inaccurate.

Instead of trying to obtain accurate a priori information on the ozone profiles, we prefer to focus on the use of measured ozone profiles in the context of data assimilation. When data assimilation is used to combine models and measurements most of the smoothing error can be removed by applying the averaging kernel to the model ozone profiles. In Section 7, Error Analysis, we will therefore consider errors in the retrieved ozone profile with and without application of the averaging kernel. Furthermore, this makes it possible to use relatively simple a priori information such as a climatology, so that patterns in the retrieved ozone profiles are due to the measurements themselves and not due to the complicated a priori information obtained from a CTM.

4.4 **Overview of the retrieval method**

The Optimal Estimation (OE) method is used for the retrieval of the ozone profile. In order to simulate the radiance measured above the atmosphere a realistic ozone profile is needed which has to be specified at approximately 30 levels in the atmosphere. However, the reflected Earth radiance contains independent information on not more than 5 - 7 levels, which means that the problem is ill posed. Hence, additional information, often based on an ozone climatology, is used to constrain the solution. The mathematical framework for Optimal Estimation is fully described in Rodgers (2000). One of the challenges of ozone profile retrieval is to use proper a priori information on ozone, not only with regard to the climatology used, but also on the variability of ozone as represented by the a priori error covariance matrix. Use of different a priori error covariance matrices leads to significantly different solutions.

Before the OE algorithm can be applied several pre-processing steps are performed on the data, including:

- Spectral calibration;
- Spatial co-adding and spectral regridding;
- Radiometric correction;
- Polarisation and Raman scattering correction.

Figure 4-3 shows a schematic diagram of the retrieval method that is based on OE. Critical elements are the radiative transfer calculations, systematic errors in the measured reflectance (e.g. by stray light), and the a priori information on the ozone profile and its priori error covariance matrix (not shown in Figure 4-3).

The update of the state vector during an iteration step is shown in Figure 4-4. One of the inputs is the measured reflectance, calculated from the measured radiance and irradiance. For the current state of the atmosphere the reflectance derivatives are calculated. Other input values (on the left) are the a priori state vector, the current state vector, the previous estimate of the state vector, and the a priori error covariance matrix. Further, constraints can be set on the change in the state vector.

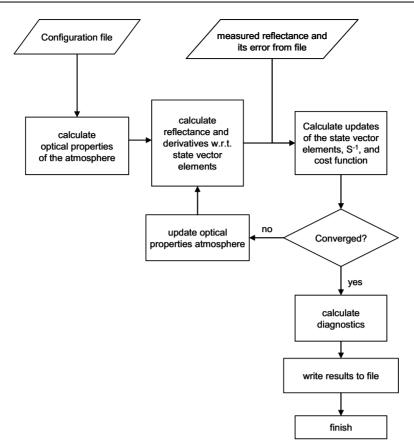


Figure 4-3. Simplified schematic diagram for the retrieval.

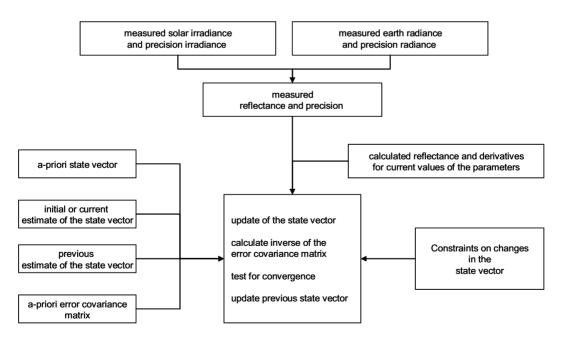


Figure 4-4. Schematic diagram for the update of the state vector.

4.5 Foreseen update approach

4.5.1 Improved quality flags and qa_value

Currently we have used the operational processor only to process a limited set of test data. Once the processor is running operationally, more data will become available. Using these data and validation results, the quality flags and the qa_value have to be fine-tuned.

4.5.2 Radiometric soft calibration

To improve on the radiometric calibration of the TROPOMI Level 1B data, soft calibration parameters are applied (see section 5.1.5). Because of the optical degradation of TROPOMI, these parameters change over time and therefore the soft calibration has to be performed regularly to keep the parameters up-to-date.

4.6 Design considerations

Our experience with the ozone profile retrieval algorithm for OMI has shown that the retrieval is complicated as there are many issues involved, such as using proper a priori information, efficient radiative transfer, corrections for polarized light and accounting for rotational Raman scattering (RRS). Decisions concerning fitting surface albedo, cloud properties, and/or stray light have to be made. The prototype software (DISAMAR) has been designed to be very flexible. The operational processor uses DISAMAR for the retrievals, thus the flexibility is also maintained for the operational processor and can be controlled by changing the configuration files.

5 Algorithm description

In this section the algorithm is described which is used to retrieve the ozone. Section 5.1 describes the pre-processing steps performed on the TROPOMI measured spectra, Section 5.2 describes the forward model and Section 5.3 describes the retrieval algorithm.

5.1 Pre-processing

Before the OE algorithm is a number of pre-processing steps are applied. The output of these steps are recalibrated radiance and irradiance spectra corrected for polarisation and Raman scattering, covering the spectral range from 270 to 330 nm.

5.1.1 Spectral Calibration

In the TROPOMI Level 1B product a fixed wavelength assignment is provided. The true barycentres of the spectral pixels may differ from this assignment due to thermal changes in the instrument, and in addition for the radiance due to inhomogeneous filling of the slit. The spectral calibration procedure is based on precise knowledge of the Fraunhofer lines in the solar spectrum [RD06, Appendix A]. The spectral calibration of the radiance spectra is more complex compared to the irradiance spectra. This is caused by the significant variation of ozone absorption and Rayliegh scattering over the spectral range between 270 to 330 nm. For shorter wavelength ozone itself has to be taken into account for th spectral calibration. In addition, variability of the spectral calibration derived at the longer wavelength is not applicable for the entire spectral range, because the variability of the signal in the flight direction (causing inhomogeneous filling of the slit) is significantly smaller for the shorter wavelengths. Therefore, the spectral calibration is performed only on the solar irradiance. Shift and squeeze parameters are used to recalculate the wavelength grid of the solar irradiance and the Earth radiance, assuming that the instrument is sufficiently stable of the course of one orbit.

5.1.2 Spatial Regridding

The algorithm uses data from the TROPOMI Bands 1 and 2, which are both derived from the same UV detector of the instrument (Kleipool et al 2018; Ludewig et al. 2020). The difference between the Bands is in the read-out of the detector. Because of the lower radiance signals in Band 1, more detector pixels are binned during the read-out of the detector to reduce the read noise, resulting in a lower spatial sampling in the across track direction. The nadir across track pixel size of Band 1 pixels is 28 km and for Band 2 this is 3.5 km. We reconstruct the full radiance and irradiance spectra by averaging the Band 2 pixels that match the Band 1 pixels. This averaging is also applied to the ISRF. The signal to noise ratio (snr) for the combined pixel is computed as $\overline{snr} = \sum_{i=1}^{n} \frac{snr_i}{\sqrt{n}}$, which assumes shot noise and accounts for the increased SNR for the averaged spectrum.

We also apply an averaging for the radiances of 5 scanlines in the flight direction, reducing the spatial sampling from 5.5 km to 28 km. For orbits measured before 6 August 2019 the along track sampling was 7 km and averaging 5 pixels leads to a sampling of 35 km. For the averaging of the radiances and SNR the same procedures as for the across track averaging of Band 2 are used.

After applying the averaging of Band 2 to match the Band 1 ground pixels, and the along track averaging of 5 scanlines, we end up with continuous radiance spectra for the fit window from 270 – 330 nm, for 77 across track pixels with a spatial resolution of 28x28 km in nadir after August 6, 2019, and 28x35 km before.

In addition to averaging the Level 1B, also the FRESCO cloud data [RD08] derived from the TROPOMI Band 6 are collocated to the Band 1 ground pixels.

5.1.3 Spectral Regridding

The spectra of TROPOMI Band 1 and 2 have a very large spectral oversampling (ratio of spectral resolution and spectral sampling) of more than 6.9. The algorithm performs line-by-line radiative transfer computations on the spectral grid of the measured spectra. To reduce the number of lineby-line calculations, 3 spectral pixels are averaged, resulting in an oversampling ratio of >2.3. Similar to the averaging in the spatial direction, the SNR is computed assuming shot-noise and the ISRF is averaged to take the effect of the spectral averaging on the spectral resolution into account.

5.1.4 Polarisation and Raman Correction

The radiative transfer model of DISAMAR can run in vector mode, including polarisation of the light, or in scalar mode without polarisation. For both the vector and the scalar mode, DISAMAR can be configured to include rotational Raman scattering (RRS). Including polarisation and/or RRS has a significant impact on the computational effort. To increase the computational speed of the algorithm, we run the forward model in scalar mode and without RRS. Therefore, we first apply a correction for polarization and RRS. After applying this correction, the radiance spectra can be compared to radiative transfer model results in scalar mode and without RRS.

In DISAMAR, RRS is taken into account following Landgraf et al. 2004. First, the internal radiance field is calculated at the interfaces between the atmospheric layers, assuming molecular scattering is described by normal Rayleigh scattering. This also provides the radiance at the sensor when RRS is ignored and is called the original radiance at the sensor. From the internal radiation field, we know the distribution of light incident at an optically (very) thin layer located at a particular interface. It is then straightforward to calculate the light reflected by this thin layer assuming RRS takes place in the thin layer. Repeating this for each interface and performing integration over the altitude while accounting for the attenuation of light when it travels from the interface to the sensor, provides the amount of RRS that arrives at the sensor. The same is done assuming Rayleigh scattering in the thin layer, which provides the contribution when wavelength shifts by RRS are ignored, called here the Rayleigh contribution. Adding the RRS contribution to and subtracting the Rayleigh contribution from the original radiance at the sensor provides the radiance when single RRS is considered. The properties for RRS are taken from Chance and Spurr (1997) and Stam et al. (2002).

Using DISAMAR, we compute the radiance in vector mode including RRS and in scalar mode without RRS. These computations are performed for a large number of wavelengths, ozone profiles, Sun-satellite geometries and surface albedos. In total 1.2 10⁹ radiances are collected over the spectral window from 268 to 332 nm. The DISAMAR calculations are done monochromatically, i.e. without applying the ISRF. This dataset is used to train a neural network that predicts the polarization and RRS correction factor. The predicting variables for the neural network are:

- Sun-satellite geometry: solar zenith and azimuth angle, viewing zenith and azimuth angle;
- surface pressure;
- surface albedo;
- total ozone column;
- wavelength;
- single scattering degree of linear polarisation for Rayleigh scattering, which is computed from the Sun-satellite geometry;
- the direct ozone transmittance from the Sun to the satellite, which depends on the total ozone column, the ozone absorption cross section, the Sun-satellite geometry and the wavelength, and is computed for an effective ozone temperature of 220K;
- a high-pass filtered solar spectrum, which depends on the wavelength.

The neural network used for the training consists of four hidden layers with 64 nodes each. 60% of the data is used for training the neural network, whereas 20% is used for validation and another 20% for evaluation. An example of the correction factor computed with the neural network is provided in Figure 5-1. Before applying this correction factor, it first has to be convolved with the ISRF.

We apply the polarization and RRS correction as a pre-processing step. For the surface albedo the scene albedo fitted at 330 nm is used and the total ozone column from the CAMS forecast is used.

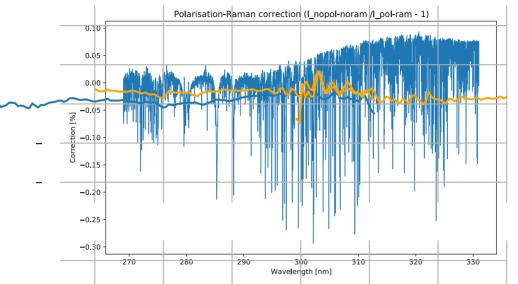


Figure 5-1. Example of the polarization and RRS correction factor computed with the neural network. This example is for a solar zenith angle of 45°, solar azimuth angle of 0°, viewing zenith angle o

5.1.5 Radiometric Correction

TROPOMI Bands 1 and 2 shows stematic radiometric deviations, as compared to the NPP-OMPS and as compared to forward models. This is illustrated in Figure 5-2, showing results for a pixel with low reflectance (low cloud fraction) over the tropical Pacific Ocean. As can be seen, significant deviations between TROPOMI and the other datasets occur for a wide spectral range between 280 and 315 nm.

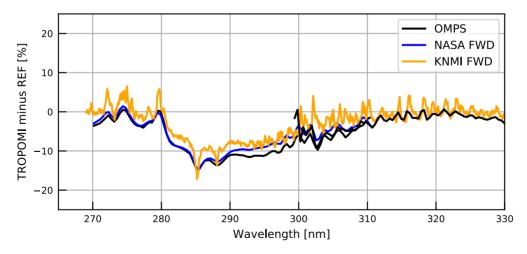


Figure 5-2. Difference between TROPOMI Band 1-2 and OMPS (black) NASA forward modelling (Blue) and KNMI forward modelling. This comparison is performed for OMPS NP Pixel 31 orbit 40301, 2019-08-07 and S5P Orbit 9414. All TROPOMI co-located with the OMPS NM ground pixel are averaged. The OMPS NM pixel is selected based on the latitude and on the low reflectance. NASA forward modelling is performed on lower spectra resolution. TROPOMI L1b version 2 are used.

To correct the systematic radiometric errors in Bands 1 and 2 of TROPOMI, the radiances are compared to a forward model. Differences between the measured radiances are collected for a large number of observations, covering different conditions and seasons. From this dataset correction parameters are derived. Because the instrument optical sensitivity as well as detector characteristics vary over the swath, the correction parameters are a function of the across flight ground pixel number. Other instrument artefacts that are not well corrected for, e.g. straylight, vary probably with the measured signal. Therefore, the correction parameters are also a function of the radiance itself. Note that both the measured radiances as well as the forward model will have errors. Model errors can be caused by differences between the model input and the real atmosphere, as well as by systematic errors, for example because of errors in the radiative transfer. In the residuals between measured and modelled radiance spectra, we cannot distinguish between instrument and model errors, although from comparisons with other satellite instrument we know that most of the residual is from the TROPOMI radiances (see Figure 5-2). When we use the same forward model setup as in the retrieval, the soft calibration may thus also reduce systematic model errors.

For the forward model it is important to discuss the model parameters itself as well the inputs that are used.

The DISAMAR orders of scattering radiative transfer model (RTM) is used. The RTM applies a correction to account for the sphericity of the atmosphere and uses 8 streams. Before computing the radiance residuals, we correct the measured spectra for polarization and Raman scattering, using the same procedure as used in the ozone profile retrieval. Thus, the RTM is run is scalar mode, ignoring polarization, and with Raman scattering switched off. Pressure, temperature and ozone profiles are obtained from the Copernicus Atmosphere Monitoring Service (CAMS) global analyses [ER08]. The CAMS ozone profiles are scaled to match the total column ozone derived from daily OMPS L3 data (Jaross, 2017). The atmosphere is modelled without clouds and aerosols. Effects of these are compensated for by adjusting the surface albedo. Therefore, we first fit the surface albedo in a small spectral window between 328 and 330 nm. For this fit, the RTM does account for polarization, and the measured radiance without the polarization and Raman correction is used. This is done, because the polarization and Raman correction is a function of, amongst others, the surface albedo. The fitted surface albedo for the 328 to 330 nm window is applied to the entire window from 270 to 330 nm.

We use TROPOMI Level 1B version 2 test data for 8 orbits. All ground pixels with solar zenith angle <85° are used. In total 402,565 residual spectra are computed. Because Band 1 contains 77 ground pixels in the across flight direction, the data set contains more than 5000 radiance residual spectra per across flight ground pixel number. The radiance residual data are collected into a single HDF5 file, which in addition to the residuals also stores for each ground pixel the wavelength, radiance, irradiance, orbit number, ground pixel number, scan line number, solar zenith angle, surface albedo, latitude and longitude.

We correct the radiance spectra using a piecewise linear correction which depends on the radiance, the across track index and the wavelength index. To determine the soft calibration parameters for the piecewise linear (PWL) correction, we first select all the radiances and residuals per ground pixel number and per wavelength index. Next, the data is binned into 20 equally large groups, based on the radiance. For each of these bins the mean radiance and the mean and standard deviation of the residual is computed. To force a smooth behaviour of the correction as a function of the radiance, a 3rd order polynomial is fitted through these points. The fit function is evaluated for mean radiance for the 20 radiance groups. The soft calibration correction is determined by PWL interpolation between these points. Note that extrapolation is done by using the value for the minimum or maximum bin. Figure 5-3 shows examples of the PWL radiance correction for different wavelengths.

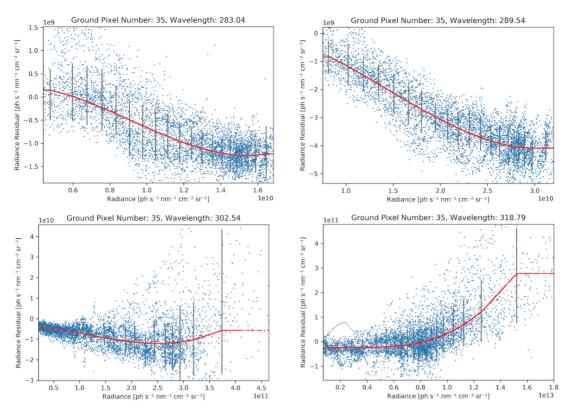


Figure 5-3. Examples of the PWL method. The blue plots show the absolute radiance residual as a function of the radiance for ground pixel number 35 and 4 wavelengths. The grey symbols show the mean radiance and radiance residuals and their standard deviation for 20 bins. The red dots show the correction based on the PWL method.

Figure 5-4 shows the median relative radiance residuals before and after correction, as a function of wavelength. Before correction, the residuals show the typical shape with the strong negative residuals between approximately 280 and 310 nm. Also note that the residual shows a lot a spectral fine structure. This may be because the figure shows the relative radiance residual, which is the radiance residual divided by the radiance, where the radiance itself contains fine structure due to the Fraunhofer lines.

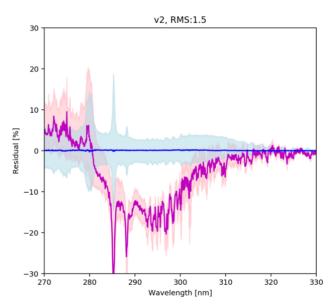


Figure 5-4. Relative radiance residuals for ground pixel number 35 without correction (magenta) and after PWL version 2 correction. The lines represent all the median of all spectra, and the coloured are the range from the 16th to the 84th percentile.

5.1.6 Signal-to-Noise Ratio

As a final step of the pre-processing, the SNR of the radiance is clipped to 150, thus implementing a noise floor. The noise floor will have a positive impact on the convergence, which will be reached in fewer steps, thus reducing the computational effort. On the other hand, it decreases the amount of information that is derived from the measurement data.

5.2 Forward model

5.2.1 Output forward model

The output of the forward model is a simulated measured reflectance that can be compared with the measured reflectance (measured radiance divided by measured irradiance). Derivatives with respect to the parameters that are to be fitted are also calculated by the forward model. In the configuration used for the ozone profile retrievals, the main parameters that will be fitted, and for which derivatives are calculated, are:

- Ozone profile at 33 pressure levels
- SO₂ column
- Surface albedo at 3 wavelengths
- Cloud albedo at 3 wavelengths

5.2.2 Model atmosphere

The model atmosphere contains dry air, ozone, SO₂, and a Lambertian cloud, while the atmosphere is bounded by a Lambertian surface. The model atmosphere is described by a pressure-temperature profile, the ozone profile, the SO₂ profile, the surface albedo, and the cloud albedo, fraction and pressure. The cloud pressure is obtained from the TROPOMI level 2 KNMI cloud support product (FRESCO algorithm) and the effective cloud fraction is fitted at 330 nm; all other inputs are part of the state vector. For SO₂ we are fitting the SO₂ column by scaling an a priori profile. Furthermore, it is assumed that the atmosphere is in hydrostatic equilibrium, which means that an altitude grid can be calculated from the pressure grid when the temperature profile is known.

The temperature, ozone and SO₂ profiles are provided on pressure levels, which may differ among each other. Interpolation to other pressure levels required by the forward model is performed using spline interpolation.

In the forward model scattering and absorption by aerosol particles is ignored. The UV Absorbing Aerosol Index (UVAI) [RD07] is used to identify pixels that contain absorbing aerosol. The retrieved ozone profile for these pixels is expected to be inaccurate.

The cloud fraction and pressure are derived from the O_2 A-band around 760 nm where the radiation field is vastly different from the radiation field around 310 nm. Therefore, we fit the cloud albedo during the retrieval of the ozone profile in order to remove errors introduced by the differences in radiation fields.

5.2.2.1 Molecular scattering

As there are different expressions for Rayleigh scattering in the literature, we specify here what we use. The volume scattering coefficient for Rayleigh scattering at altitude z, $k_{sca}^{Ray}(z, \lambda)$ in cm⁻¹, follows from

$$k_{sca}^{Ray}(z,\lambda) = n_{air}(z)\sigma^{Ray}(\lambda) = \frac{p(z)}{k_B T(z)} \frac{24\pi^3}{\lambda^4 N^2} \frac{(n^2(\lambda)-1)^2}{(n^2(\lambda)+1)^2} F_{air}(\lambda)$$
(eq. 5.1)

where λ is the wavelength [nm], $n_{air}(z)$ is the number density of air [molecules cm⁻³], $\sigma^{Ray}(\lambda)$ is the Rayleigh cross section [cm²], $k_B = 1.380658 \times 10^{-19}$ is Boltzmann's constant, p(z) is the pressure [hPa], T(z) is the temperature [K], $N = 2.5468993 \times 10^{-19}$ molecules cm⁻³ [288.15 K

and 1013.25 hPa]. $n(\lambda)$ is the refractive index of air which follows from the expression given by Peck and Reeder (1972):

$$n(\lambda) = 1 + 10^{-8} \left(8060.51 + \frac{2480990.0}{132.274 - \frac{1}{\lambda^2}} + \frac{17455.7}{39.32957 - \frac{1}{\lambda^2}} \right)$$
(eq. 5.2)

which holds for dry air, a pressure of 1013.25 hPa and a temperature of 15 degrees Celsius (288.15 K). Here λ is given in micrometer. Finally, $F_{air}(\lambda)$ is the King factor. It can be calculated from (see Bates 1984):

$$F_{\rm N_2}(\lambda) = 1.034 + 0.000317 / \lambda^2$$
 (eq. 5.3)

$$F_{O_2}(\lambda) = 1.096 + 0.001385 / \lambda^2 + 0.0001448 / \lambda^4$$
 (eq. 5.4)

$$F_{\rm Ar}(\lambda) = 1.000$$
 (eq. 5.5)

where λ is given in micrometer.

Here we ignore water vapour and CO_2 here as they contribute very little. See Bodhaine et al. (1999) for a detailed discussion. The King factor for air follows from

$$F_{air}(\lambda) = \frac{78.084 \times F_{N_2}(\lambda) + 20.946 \times F_{O_2}(\lambda) + 0.934F_{Ar}(\lambda)}{78.084 + 20.946 + 0.934}$$
(eq. 5.6)

The scattering phase function (phase matrix when polarization is included) is taken from Stam et al. (2002).

5.2.2.2 Ozone absorption

The absorption cross section for ozone is plotted in Figure 5-5 (Malicet et al., 1995). It shows that the absorption cross section increases very strongly over the wavelength range considered which makes it possible to derive ozone profile information. At 270 nm the ozone absorption is so strong that only light scattered in top part of the stratosphere reaches the TROPOMI instrument. At 320 nm the absorption is so weak that light reflected by the surface or cloud and light scattered in the entire atmosphere can reach TROPOMI. Further, the figure shows that at longer wavelengths the absorption cross sections differ at different pressure levels due to the different temperatures at those levels.

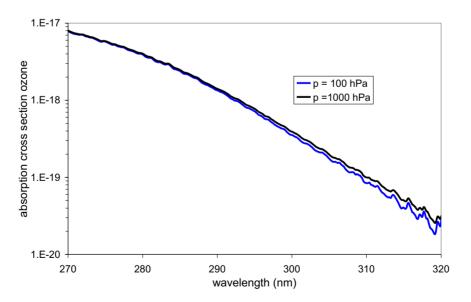


Figure 5-5. Ozone absorption cross section (cm²/molecule) plotted as function of wavelength for two pressure levels which correspond to the temperatures 216.7 K (100 hPa) and 288.2 K (1000 hPa).

5.2.2.3 SO₂ absorption

The SO₂ reference spectrum is a composite based on Bogumil *et al.* (2000) and Hermans *et al.* (2009).

5.2.2.4 Lambertian surface and cloud model

The ground surface and clouds are represented as Lambertian surfaces with a wavelength dependence given by a low-degree polynomial. The surface albedo is specified at 3 wavelengths and the albedo values at those wavelengths can be fitted.

5.2.3 Processing Steps

Below we list the main processing steps to arrive at the simulated reflectance.

- 1. Derive altitude grid (z_{OE}) from pressure and temperature grid assuming hydrostatic equilibrium. At this grid the ozone profile is retrieved and the derivatives $\partial R/\partial \ln(v_i)$ are used in the retrieval procedure, where *R* is the reflectance for a spectral pixel and v_i is the volume mixing ratio at the altitude i where the ozone profile is retrieved.
- 2. Determine a second altitude grid used for the radiative transfer calculations. This altitude grid consists of two sets of Gaussian division points for the intervals (z_s , z_c) and (z_c , z_{TOA}) where z_s is the surface altitude, z_c is the cloud altitude, and z_{TOA} is the altitude of the top of the atmosphere (typically 60 km). z_s , z_c , z_{TOA} are added to this grid with a Gaussian weight equal to zero. The atmospheric layers between the grid points are homogeneous layers. At this grid the derivatives $\partial R/\partial \ln(v_i)$ are calculated. The derivatives at the altitude grid defined in step 1 are obtained by using cubic spline interpolation on the second altitude grid. It is important that the second grid has enough points to make the interpolated values sufficiently accurate. Typically, the second grid has about 40 80 altitudes whereas the OE grid has about 33 altitudes.
- 3. Calculate the optical thickness and single scattering albedo for the homogeneous layers defined in step 2, using numerical integration. Here we use that the temperature profile, the number density of air, the absorption cross section of ozone, the volume mixing ratios of O₃ and SO₂, and the Rayleigh scattering coefficient are known (from cubic spline interpolation) at all altitudes of the sub layer grid shown in Figure 5.1. Typically, 2–4 points are used to perform the integration providing the average values for the layers.
- 4. Use the doubling method to calculate the reflection and transmission properties of the homogeneous layers defined in step 2. Here the calculations are started with single scattering for an optically thin layer (i.e. of the order 2⁻⁷). The doubling method then gives a recipe to calculate the optical properties of a layer that has twice the optical thickness of the original layer (see e.g. De Haan et al. 1987). By repeating this procedure 4 7 times the optical properties (reflectance and transmittance) of a layer with a substantial optical thickness is calculated.
- 5. Combine the optical properties of the homogeneous layers using the Layer Based Orders of Scattering method (LABOS) method. Here the internal radiation field at the interfaces between the layers is also calculated.
- 6. Use integration over the source function (see e.g. Chandraksekar, 1950) over the altitude to calculate the reflectance. This makes it possible to reduce the number of homogeneous layers that are needed to calculate the reflectance accurately.
- 7. Use the internal radiation field calculated in step 5 to calculate the basic derivatives $\frac{\partial^2 R}{\partial \ln(v) \partial z}$

at the grid defined in step 2. These derivatives are used in subsequent steps.

- 8. Calculate the derivatives $\partial R/\partial \ln(v_i)$ on the altitude grid defined in step 1 taking into account that the profile is defined through cubic spline interpolation where *i* is the index for the altitude. Note that these derivatives and the ones calculated in step 9, after convolution with the instrument slit function, are input for the optimal estimation method.
- 9. Calculate the other derivatives (surface albedo, cloud albedo, SO₂ column) as needed.

- 10. Repeat steps 3 9 for all wavelengths on a high-resolution wavelength grid. This grid consists of sets of repeated Gaussian division points and corresponding weights so that integration over the instrument slit function is straightforward.
- 11. Multiply the reflectance and derivatives with a high resolution solar irradiance spectrum that has been interpolated to the high resolution wavelength grid defined in step 10. This provides the simulated radiance and associated derivatives on a high resolution wavelength grid.
- 12. Convolute the irradiance, radiance and derivatives with the instrument slit function for each spectral pixel involved using numerical integration. This provides the irradiance, radiance and derivatives on the instrument wavelength grid.
- 13. Divide the simulated radiance and the derivatives by the solar irradiance to obtain the modelled reflectance and derivatives that are input for the OE retrieval algorithm.

The forward model is summarized in Figure 5-6 which shows a schematic diagram of the operations involved.

5.3 Retrieval algorithm

5.3.1 Optimal Estimation

OE is a method developed to retrieve parameter values and their errors assuming that a priori information, including an a priori error covariance matrix, is available. OE assumes Gaussian distributions for the errors. It is widely used for retrievals of atmospheric properties from remotely sensed data and is described in Rogers (2000). Here a brief summary is given.

5.3.1.1 Cost function

The cost function or χ^2 that is minimized in OE is defined as follows

$$\chi^{2} = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^{T} \mathbf{S}_{\varepsilon}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + (\mathbf{x} - \mathbf{x}_{a})^{T} \mathbf{S}_{a}^{-1} (\mathbf{x} - \mathbf{x}_{a})$$
(eq. 5.8)

where **y** is the vector of measured reflectance containing values for the different wavelengths; $\mathbf{F}(\mathbf{x})$ is the vector of calculated reflectances, also called the forward model; **x** is the state vector containing the parameters that are to be fitted; \mathbf{S}_{ε} is the error covariance matrix of the measurement, which is diagonal when the measurement errors are assumed to be uncorrelated; \mathbf{x}_{a} is the a priori state vector; and \mathbf{S}_{a} is the a priori error covariance matrix.

The cost function differs from the usual least-square cost function due to the additional term containing the a priori information. If we remove this second term the retrieval may not be stable. The purpose of the second term is to constrain possible solutions, i.e. perform regularization. A side effect of this approach is that the retrieved state vector is based on the measurements as well as on the a priori information. State vectors that differ strongly from the a priori but are consistent with the measured reflectance tend to be rejected.

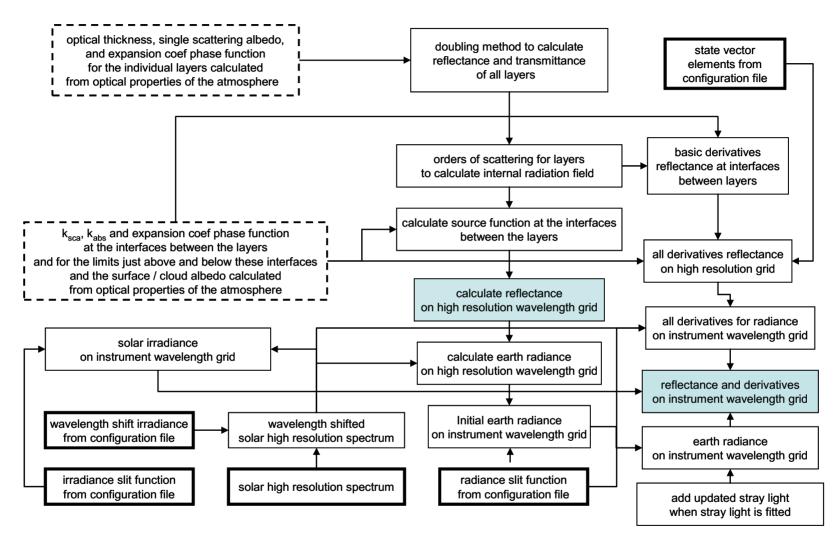
5.3.1.2 State vector elements and prior information

The state vector elements used for ozone profile retrieval and the corresponding a priori values are listed in Table 5-1.

The non-diagonal elements for the part of S_a corresponding with the ozone profile are filled using a correlation length of 6 km. The correlation length is used to minimize oscillations in the retrieved profile. The non-diagonal elements of S_a pertaining to the profile are calculated from

$$\mathbf{S}_{a}(i,j) = \exp\left(-\frac{\left|z_{i}-z_{j}\right|}{l}\right) \sqrt{\mathbf{S}_{a}(i,i)} \sqrt{\mathbf{S}_{a}(j,j)}$$
(eq. 5.9)

where $|z_i - z_j|$ is the absolute difference in altitude and l is the correlation length. If *i* or *j* pertain to another fit parameter than the ozone profile, e.g. the cloud fraction or surface albedo, then $S_a(i, j) = 0$ for $i \neq j$.



1

Figure 5-6. Schematic diagram of the forward model. Blocks with thick lines are input. Blocks with dashed lines are pre-processed input. The block 'Calculate reflectance on high resolution grid' has two modes, one for improved efficiency and one for accurate line-by-line calculations. The blue/gray block at the right gives the desired result.

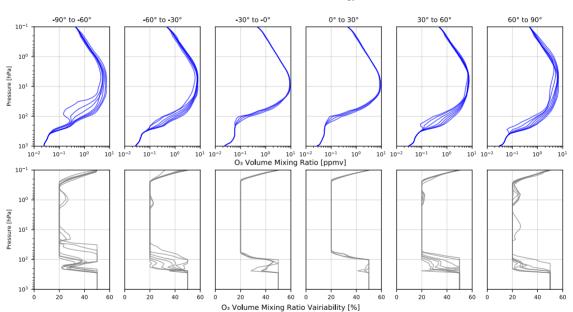
State vector elements	A priori value	A priori variance (1 sigma)
The logarithm of the ozone volume mixing ratio at different pressure levels	From climatology.	From climatology.
SO ₂ column	0.01 DU	0.1 DU
Surface albedo at 3 wavelengths	From LER climatology	1.0 (cloud fraction < 0.2) 1e-8 (cloud fraction >= 0.2)
Cloud albedo at 3 wavelengths	From FRESCO	1e-8 (cloud fraction < 0.2) 0.01 (cloud fraction $>= 0.2$)

Table 5-1. State vector elements.

The ozone profile climatology is based on Labow et al. (2015). This climatology provides the ozone profile and standard deviation as a function of latitude and total ozone column. Compared to the original climatology, the following adjustments have been made:

- In the troposphere and upper atmosphere (pressure <0.1 hPa) the original ozone values are replaced by the median of the values along the total ozone axis. This prevents that in these altitude ranges the a priori varies with the total ozone column, which is dominated by the stratosphere.
- The a priori error is limited to the range 20 50%. For pressure levels larger than 250 hPa the error is set to 50%.

The a priori profile is derived from the climatology using the CAMS forecasted total ozone and is linearly interpolated to the latitude of the ground pixel.



A Priori Ozone Profile Climatology

Figure 5-7. Ozone profile climatology used as a priori. Top row shows the volume mixing ratio as a function of pressure, bottom row the variability. The columns show the data for the different latitude bins. The climatology is based on a modified version of Labow (2015) (see text).

5.3.1.3 Updating the state vector

Let \mathbf{x}_a be the a priori state, \mathbf{x}_i (i = 0, 1, 2...) subsequent iterations of the state vector. It is possible to use $\mathbf{x}_0 = \mathbf{x}_a$, but this is not required. The Gauss-Newton method gives the following update for the state vector: (Eq. (5.9) in Rodgers, 2000)

$$\mathbf{x}_{i+1} = \mathbf{x}_a + \left[\mathbf{K}_i^T \mathbf{S}_{\varepsilon}^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1} \right]^{-1} \mathbf{K}_i^T \mathbf{S}_{\varepsilon}^{-1} \left[\mathbf{y} - \mathbf{F}(\mathbf{x}_i) + \mathbf{K}_i (\mathbf{x}_i - \mathbf{x}_a) \right]$$
(eq. 5.10)

Here $\mathbf{K}_i = \partial \mathbf{F}(\mathbf{x}_i) / \partial \mathbf{x}_i$ is the Jacobian or weighting function matrix for iteration number *i*. The a posteriori covariance matrix is given by

$$\mathbf{S}_{i} = \left[\mathbf{K}_{i}^{T} \mathbf{S}_{\varepsilon}^{-1} \mathbf{K}_{i} + \mathbf{S}_{a}^{-1}\right]^{-1}$$
(eq. 5.11)

where i is the iteration index. S_i is calculated after the solution has converged.

5.3.1.4 Convergence

To test the convergence, we have to check whether the state vector elements do not change significantly for subsequent iterations, compared to the expected accuracy. A proper measure (see Rodgers, 2000, Sec. 5.6.3) is

$$d_i^2 = (\mathbf{x}_{i+1} - \mathbf{x}_i)^T \mathbf{S}_i^{-1} (\mathbf{x}_{i+1} - \mathbf{x}_i) \ll n$$
 (eq. 5.12)

where *n* is the number of state vector elements. And S_i is the a posteriori error covariance matrix after *i* iterations. The specific value for the convergence criterion is set in the configuration file. This value is be based on a compromise between accuracy and calculation time. The maximum number of iterations is limited to 12.

5.4 Error characterization

Following Rodgers (2000) different types of errors can be distinguished, namely ozone profile errors due to

- measurements errors (noise and calibration errors)
- forward model errors
- a priori errors
- smoothing errors

5.4.1 Measurement errors

The error covariance matrix for the measurement errors S_{ε} is diagonal if there are no calibration errors. The diagonal then contains the errors in the measured reflectance and it is assumed that these errors are dominated by shot noise and are uncorrelated. Let S_{ε} be the sum of S_{ε}^{noise} and

 $\mathbf{S}_{\varepsilon}^{calib}$. Here $\mathbf{S}_{\varepsilon}^{noise}$ is diagonal and its elements are calculated from the SNR specified for

TROPOMI. The diagonal elements for S_{ε}^{calib} represent the calibration errors which typically are a factor of 2 to 10 larger than the noise error. It is assumed that these calibration errors are correlated, and the degree of correlation is specified by the wavelength distance for which the correlation reduces with a factor of *e*. Typically, the correlation length used is 10-100 nm. The correlation length is used to ensure that errors, such as errors in gain settings, usually affect different spectral pixels in the same manner and are therefore highly correlated.

In the current version of the algorithm $\mathbf{S}_{\varepsilon}^{calib}$ is set to zero. For computing the measurement errors

 S_{ε} the SNR of the radiance is maximised to 150 and that of the irradiance to 5000, thus acting as a noise floor.

5.4.2 Fit residue

In view of the size of the product (see Sec. 6.3), the residue of the fit will be written only to an intermediate output file used for testing. Averaged values such as the mean square root of the fit will be written to the output.

5.4.3 A posteriori errors

After the retrieval has converged to a certain ozone profile, or the maximum number of iterations has been reached, the posterior error covariance matrix is calculated from Eq. 5.11. The errors in the ozone profile are then obtained by taking the square root of the diagonal elements of this matrix (only the part of the matrix that corresponds to the ozone profile). The entire error covariance matrix will also be part of the output product. It is possible to separately calculate the error covariance matrix due to noise. However, in order to reduce the size of the output product this quantity will not be provided. Moreover, it can be calculated from the averaging kernel and the error covariance matrix, as we have (Rodgers, 2000, Eq. 3.30) $S_{a}^{noise} = AS$.

5.4.4 Averaging kernel and DFS

The degrees of freedom of the measurement signal (DFS) (both total and for the ozone profile only) will also be written to output. This provides information on the number of independent pieces of information that is available in the measurement, given the a priori information.

The averaging kernel, A, is calculated from

$$\mathbf{A} = \mathbf{S}^{-1} \mathbf{K}_i^T \mathbf{S}_{\varepsilon}^{-1} \mathbf{K}_i$$

(eq. 5.13)

and the trace of **A** provides the DFS.

In addition, the averaging kernel is provided in the output as it can be used to account for smoothing difference errors in the comparison of retrieved profiles with profiles obtained from other measurements. Moreover, the averaging kernel provides information on the retrieval sensitivity, retrieval offset, smoothing error, the vertical resolution of the measurement and the noise error, all of which can be used in data assimilation.

5.4.5 Gain matrix

The gain matrix maps measurement errors into errors in the retrieved state vector. In order to reduce the size of the output product the gain matrix is not part of the output product.

5.4.6 Values and errors for sub columns

The ozone profile is provided in the output as number density profile. The number density profile is integrated from the surface to the tropopause to determine tropospheric ozone column. The tropopause pressure level is computed using the WMO lapse rate method . If there is a double tropopause the one with the lowest pressure will be used. In addition, errors for these columns will be calculated using a procedure described in Appendix A2. Using a similar procedure, results will be provided for the sub columns 0 - 6 km, 6 - 12 km, 12 - 18 km, 18 - 24 km, 24 - 32 km, and 32 km - TOA.

6 The Ozone Profile Product

6.1 High level data product description

In this section we describe the main parameters and their usage. A complete description of the file format is provided in [RD10].

The ozone profile product provides the data for ground pixels of $28 \times 28 \text{ km}^2$ at nadir. This spatial grid is constructed by regridding the data on the Level 1B Band 1 grid and averaging 5 pixels in the flight direction. This results in 77 ground pixels per scan line and of the order 835 scan lines per orbit.

The main parameters in the file are the retrieved ozone profile at 33 levels and the retrieved subcolumns of ozone in 6 layers. In addition, the total ozone column and tropospheric ozone columns are provided. An overview of the main data fields in the product is provided in Table 6-1.

For the ozone profile, the precision and smoothing errors, the a-priori profile and the averaging kernel are also provided. Users that intent to use the ozone profile data for data assimilation should make use of the averaging kernel data. Users comparing the data to other data sets that have a different vertical resolution and smoothing, should also use the averaging kernel, or consider the smoothing difference error.

The ozone profile is provided as number densities on 33 levels in units of mole per m³. To convert to molecules per cm³, the values have to be multiplied by 6.0221406 10¹⁷. To convert to volume mixing ratios vmr in ppmv (parts per million, 10⁻⁶), the values have to be multiplied by $8.31 \cdot T(z) \cdot p(z)^{-1}$, with pressure in Pa. When the ozone profile needs to be interpolated to a different vertical grid, it is recommend using altitude or the logarithm of the pressure as the vertical coordinate.

Ozone columns and sub-columns are provided in units of mole per m². To convert to molecules per cm², the values have to multiplied by 6.0221406 10¹⁹. To convert to Dobson units (DU), the values have to be multiplied by 2241.15.

origin of data set	description of data set	symbols
Level 1B data	time	t
	ground pixel centre and corner coordinates	$ heta_{geo}, \delta_{geo}$
	solar and viewing geometry	θ ₀ , θ, φ ₀ , φ
Static auxiliary data	surface altitude	Zs
	a priori surface albedo	A _{s, ap}
	a priori ozone profile	п _{ОЗ,ар} (z)
Dynamic auxiliary data	Pressure – temperature profile	p(z), T(z)
	a priori ozone column	Noз
	tropopause pressure using WMO definition	p trop
Cloud data	cloud fraction and cloud pressure FRESCO	f _{eff,} p _{eff}
	a priori cloud albedo FRESCO	Ac
	UV aerosol index	-
	cloud fraction derived at 330 nm	f eff,330
Optimal Estimation	ozone profile	п ₀₃ (z)
	ozone profile precision (noise error)	dn _{O3,prec} (z)
	smoothing error	dn _{O3, smooth} (z)
	averaging kernel for the ozone profile	Α
	ozone in 6 sub-columns and precisions	ΔN _{O3} , dΔN _{O3}
	total ozone column and precision	N оз, dN оз
	tropospheric ozone column and precision	No3,trop , dNo3,trop
	surface albedo and precision	As, dAs
	cloud albedo and precision	A _c , dA _c
	sulphur dioxide column and precision	Nso2, dNso2
Diagnostics	degrees of freedom for signal	ds
	degrees of freedom for signal for ozone profile	d _{s,O3}
	root-mean-square error of fit	R _{rms}
	cost function	X ²
	error covariance matrix for ozone	Sε
Flags	quality assurance value (qa_value)	f _{QA}
	processing quality flags	-
	snow/ice flag and land/water classification	-

Table 6-1. Overview of data sets for each ground pixel in the ozone profile data product.

6.2 Applying the Averaging Kernel

The TROPOMI ozone profile is provided at 33 levels, which is a higher sampling compared to the vertical resolution. To compare the TROPOMI ozone profile with datasets that have a significantly higher vertical resolution, for example from ozonesondes or from chemistry-transport models, the averaging kernel has to be applied on the dataset:

$$\hat{n}_{o3} = n_{o3,ap} + \mathbf{A} \left[n_{o3} - n_{o3,ap} \right]$$
(6.1)

where \hat{n}_{03} is the external data set with the TROPOMI averaging **A** kernel applied, n_{o3} is the ozone profile from the external dataset and $n_{o3,ap}$ is the a priori used in the TROPOMI retrieval.

It is noted that before applying equation 6.1, the TROPOMI and external datasets have to be on the same vertical grid. Preferably, this common vertical grid covers the complete TROPOMI pressure range from the surface to 0.01. When the external data set doesn't cover the complete atmosphere, it may be extended using the TROPOMI a priori profile.

6.3 Quality Indicator

The qa_value provided in the data product is a convenient way to filter the data. Ground pixels for which the qa_value<=0.5 should not be used, and ground pixels for which the qa_value <=0.8 should be used with care. The qa_value is initiated with a value of 1 and multiplied by a value smaller than one if a certain criterion is met. The definition of the qa_value multipliers is provided in Table 6-2. For a single ground pixel, different criterions may be met, leading to a further reduction of the qa_value. If for example an input spectrum warning is encountered and for measurements in the Southern Atlantic anomaly, the qa_value becomes 1x0.8x0.8=0.64.

More detailed information on the quality of the retrieval for a ground pixel is provided by the processing_quality_flags field, which provides bit-level flags. A full description of the processing_quality_flags field is given in [RD10].

In addition to the qa_value, also additional filtering may be recommended. The latest information on such recommendations is provided in the Product Readme File [RD11].

parameter	criterion	multiplier
Processing error	-	0.0
Spacecraft manoeuvre	-	0.1
Solar eclipse	-	0.2
Cost function	χ ² > 600	0.5
Ozone profile	$rms(n_{O3} - n_{O3,ap}) > 1.10^{-12}$	0.5
Ozone column	100% (N _{O3} / N _{O3,ap} -1) > 8%	0.5
Input spectrum warning	-	0.8
Wavelength calibration warning	-	0.8
Extrapolation warning	-	0.8
Sun glint warning	-	0.8
South Atlantic anomaly warning	-	0.8
Sun glint correction	-	0.8
Snow ice warning	-	0.8
Cloud warning	-	0.8
UV Aerosol Index warning	-	0.8
Pixel level input data missing	-	0.8
Data range warning	-	0.8
Low cloud fraction warning	-	0.8
Altitude consistency warning	-	0.8
Signal to noise ratio warning	-	0.8
Deconvolution warning	-	0.8

Table 6-2. Definition of qa_value. The last column is the maximum value of the qa_value when the criterion is met.

SO ₂ volcanic origin likely warning	-	0.8
SO ₂ volcanic origin certain warning	-	0.8
Interpolation warning	-	0.8

7 Error analyses

Ozone profile retrieval is complex and a full error characterization is beyond the scope of this ATBD. Instead, we discuss typical errors for a number of different common error sources. We distinguish two groups of error sources: model errors and instrumental errors. The errors are assessed using the DISAMAR prototype software. The standard settings are the same as for the operational configuration, except for the a-priori profile, which is set to the true profile for most of the simulations. Also, the radiometric correction is not performed. This is done to clearly distinguish the impact of the different error sources. Unless stated otherwise, the results shown in this section have been calculated assuming the European background ozone profile from CAMELOT [RD05] (see Appendix B1).

7.1 Standard Case

Before discussing the various error sources, we first examine the ideal standard case. Figure 7-1 shows the retrieval simulation used as the baseline case in this section. In this simulation no calibration errors or model errors are included, except for the convolution model error described below. As can be seen in the left and middle panels, the bias between the retrieval and the true profile is not zero. This is caused by a simplification in the radiative transfer, where for the retrieval the forward model uses cross sections that are convoluted with the ISRF, instead of convoluting the radiances that are computed on a high spectral resolution grid. In the lower troposphere, the impact of this is a few percent, and is small compared the noise and smoothing errors.

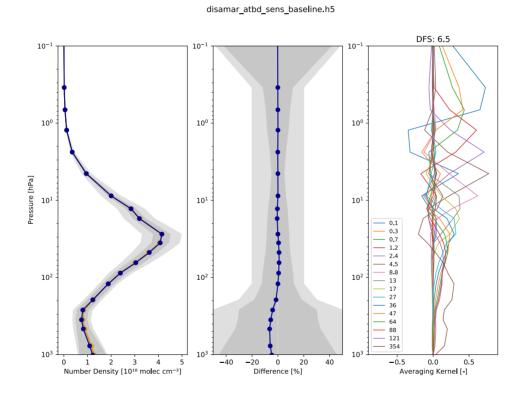


Figure 7-1. Retrieval simulation results for the European background profile. The surface albedo is 0.05 and the cloud fraction is 0.0. The solar zenith angle is 60° and the viewing zenith angle is 0°, Left panel: retrieved (blue) and true/a-priori ozone number density profile. The dark grey area shows the noise error and the light grey area the smoothing error. Middle panel: the bias (retrieved minus true) and error estimates. Right panel: averaging kernel for selected pressure levels).

TROPOMI ATBD ozone profile	S5P-KNMI-L2-0004-RP
issue 1.0.0, 2021-10-22 - Released	Page 33 of 64

Figure 7-2 shows the noise and smoothing errors for retrieval simulations for solar zenith angle 30°, 45° and 60°. This figure shows that the requirements are expected to be met for the noise error, however the smoothing error exceeds the requirement for sub columns below 25 km. It is noticed that the smoothing error strongly depends on the assumed a priori error. We have chosen to use quite large values for the a priori errors, with a minimum of 20%, which results in large smoothing errors.

Figure 7-3 shows the distribution of the total error (noise + smoothing error) for retrievals for orbit 7520. For some of the sub columns multi model distribution are observed, for example for the 12-18 km sub column. An important driver for this is the a priori error, which varies as a function of latitude due to the variation in the tropopause height. Other differences between Figure 7-2 and Figure 7-3 are the noise, which for the real retrievals is taken from the Level 1B data and for the simulations assumed on a shot noise model.

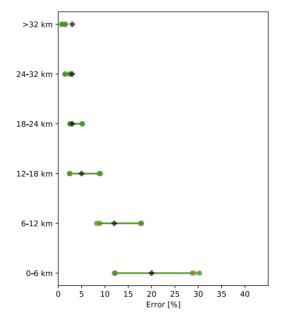


Figure 7-2. Retrieval simulation results for the European background profile. Horizontal colored lines show for each sub column the range of the noise and the total error (including smoothing error). Three different colors are used for SZA 30°, 45° and 60°, although the difference is small. The black plus signs indicate the requirements as discussed in section 4.3.

a posteriori precision for sub-columns

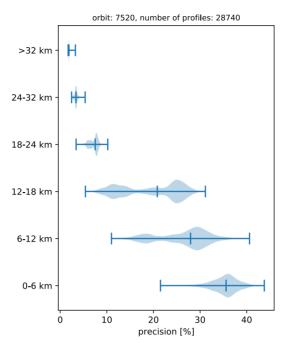


Figure 7-3. Distribution of the a posteriori precision (combined noise and smoothing error) for the sub-columns in the retrieval for orbit 7520. The 6-12 km and 12-18 km sub-columns show strong multi modal behaviour, which is caused by variability of the a priori errors with latitude. The bars show the extrema and the median value for each sub-column.

7.2 Model errors

In this section the impact of model errors on the retrieval is tested. To test these errors, we introduce differences between the forward model and the model used in the retrieval. Unless mentioned otherwise, we harmonize the way that the ISRF is treated by setting them to use preconvolved cross sections.

It is noted that the operational retrieval algorithm includes a soft calibration, based on comparisons between measured and modelled radiances (see section 5.1.5). While the primary reason for introducing the soft calibration is to correct the radiometric calibration of the TROPOMI Level 1B radiances, it may also mitigate model errors. Because a soft calibration is not applicable for the simulations, this mitigation is not included in the analyses performed in this section.

7.2.1 Spectroscopic data

To test the impact of errors in the spectroscopic data, we simulate the radiance spectra using the ozone absorption cross sections as measured by Birk and Wagner (2021), while using the Malicet et al (1995) cross section in the retrieval. As can be seen in Figure 7-4, this introduces an oscillating bias in the retrieval, with a maximum amplitude of the order 5%. This amplitude is relatively small compared to the a posteriori error in the troposphere but is comparable to these errors in the stratosphere.

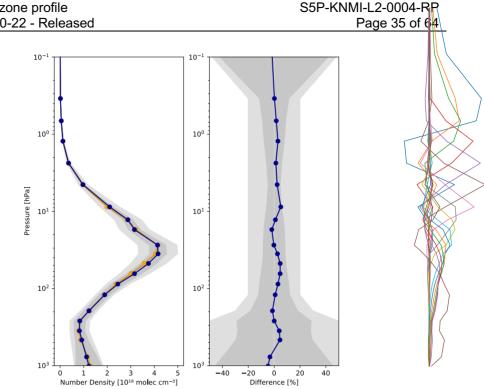
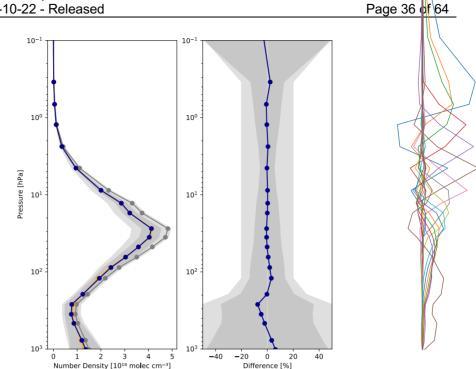


Figure 7-4. Retrieval simulation where different ozone absorption cross sections are used in the forward simulation and in the retrieval. Left panel retrieved ozone profile (blue), true ozone profile (orange), a priori error (light grey area) and a posteriori error (dark grey area). Right panel: retrieved minus true bias (blue), a priori error (light grey area) and a posteriori error (dark grey area).

7.2.2 A priori ozone profile

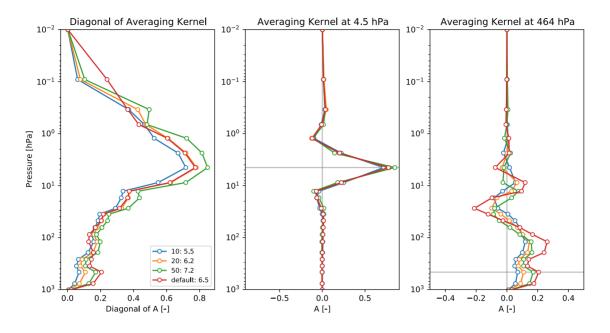
The optimal estimation algorithm combines the information derived from the measured spectra with an a priori ozone profile. Not only the a priori ozone values are imported, but also the assumptions on the a priori error. First, we test the impact of the a priori ozone profile, by scaling the profile from 300 to 350 DU. As can be seen in Figure 7-5, the bias in the retrieved profile is small in the stratosphere, and less than 10% for pressure levels in the troposphere. This shows that the algorithm can provide good results, even when the a priori is relatively far away from the true profile.

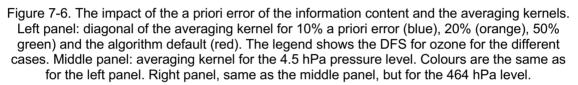


S5P-KNMI-L2-0004-RP

Figure 7-5. Retrieval simulation result when the a priori ozone profile is increased with 17% compared to the true profile. Left panel: retrieved profile (blue), true profile (orange) and a priori profile (grey). Right panel: bias of the retrieved profile (retrieved minus true). Light grey areas indicate the a priori error, dark grey areas the a posteriori error.

In addition to the a priori ozone profile, the assumptions on the a priori errors are also very important. A possible assumption would be the variability in the ozone profile climatology. However, this climatology is to a large extent based on ozonesondes, which have a very limited spatial coverage, especially in the tropics and in the southern hemisphere. The variability therefore may be underestimated by this climatology. In addition, model based variabilities tend to underestimate the variability. Other choices are more driven by the information content in the measurements: e.g. to make the a priori error sufficiently large to optimize the contribution of the measurements to the retrieval result. As described in section 5.3.1.2, we combine these approaches, by using the (adjusted) variability of the climatology but limiting it to the range 20-50%. To test the impact of this assumption, we compare retrieval results with a priori errors that are the same at all pressure levels, with results that have 50% in the troposphere and upper stratosphere and 20% at other levels. Figure 7-6 shows the diagonal of the averaging kernel and the averaging kernel for two pressure levels, for different a priori error assumptions. The diagonal of the averaging kernel provides insight on which pressure levels the retrieval gains the most information. The trace of the averaging kernel is the Degrees of Freedom for Signal (DFS). The peak in information content is between 1 and 10 hPa. As expected, the DFS is higher when the a priori error is large, i.e. when the retrieval gives more weight to the measurements. The DFS increase from 5.5 to 7.2 when the a priori error is increased from 10 to 50%. For the case which resembles the default approach, when the a priori error varies with pressure, the DFS is 6.5. It is noted that in the lower troposphere, the diagonal of the averaging kernel is the highest for the varying a priori error and is even larger than the results for constant 50% a priori error. To further analyse the information, Figure 7-6 also shows the averaging kernel at two levels: at the maximum information content (4.5 hPa) and in the troposphere (464 hPa). At 4.5 hPa, the averaging kernels are very similar, indicating that the effect of the different a priori error assumptions is relatively small. The averaging kernels have the same shape and the maxima vary approximately 15% between the smallest and the largest. For the 464 hPa level this is very different. The results vary by more than a factor of 2 and the averaging kernel contains much more contribution from other pressure levels. Whereas the varying a priori error has the largest contribution in the troposphere, it also has the largest positive and negative contributions at other levels, which are even larger than for the 50% a priori error case.





7.2.3 Radiative transfer

7.2.3.1 ISRF Convolution

As already discussed in section 7.1, in the retrieval we use pre-convolved ozone absorption cross sections, to reduce the number of wavelengths at which the forward calculations have to be performed. The impact of this assumption can be seen in Figure 7-1, where in the simulation the convolution is applied on the radiances. This assumption introduces a bias of the order 5% in the lower troposphere and smaller higher up. Everywhere in the profile these errors are small compared to the a posteriori error.

7.2.3.2 Number of streams

When the number of streams (number of Gaussian division points used for integration over polar angles) can be reduced, the radiative transfer calculations are less time consuming. However, when the number of streams is reduced too much, inaccurate integration leads to errors in the multiple scattering calculations. The default algorithm setting use 8 streams. To test the impact of this choice, we simulate the spectrum using 16 streams, and use 8 in the retrieval. It is found that the impact is smaller than 0.2% at all levels of the retrieval, which is negligible compared to other retrieval errors.

7.2.3.3 Polarization and Raman Scattering

As described in section 5.1.4, the retrieval algorithm corrects for polarisation and Raman scattering in a pre-processing step. By applying this correction, we obtain a spectrum as it would have been measured if there were no polarisation and Raman scattering in the atmosphere. To test this correction, we simulate a spectrum including polarisation and Raman scattering. As can be seen in Figure 7-7, the bias in the troposphere exceeds 40% if we do not apply the polarisation and Raman scattering correction. When the correction is enabled, the errors are reduced to approximately 10%, however also larger biases of the order 10% are introduced in the stratosphere. The polarisation will vary over the orbit and will also be modulated by cloud and surface reflectance. Therefore. It has to be investigated if there are systematic errors related to this correction that vary with latitude, surface albedo and cloud parameters.

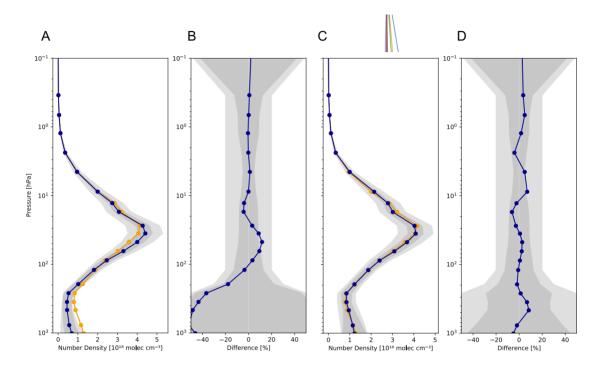


Figure 7-7. Retrieval simulation of polarization and Raman corrections. From left to right panels are identified as A-D. Panel A and B show the retrieval and bias when no polarization and Raman correction is performed; panels C and D show results when the corrections are applied. The Sun-satellite geometry is taken randomly from an existing orbit: solar zenith angle: 47.7°, solar azimuth angle: -178.7°, viewing zenith angle 7.1° and viewing azimuth angle 56.1°.

7.2.4 Lambertian cloud model and fitting of cloud properties

Clouds are represented in the algorithm as Lambertian reflecting surfaces, which cover part of the ground pixel and are placed at the cloud pressure. The cloud pressure is obtained from the FRESCO algorithm that uses the oxygen A-band of TROPOMI. The cloud fraction is fitted at 330 nm. To test the impact of this cloud model, we simulate spectra including spherical cloud droplets (Mie scattering). The cloud layer extends over a 50 hPa layer and we vary the optical thickness and top pressure. Figure 7-8 shows results for two cloud configurations. The impact of the clouds on the retrieved ozone profile is largest in the troposphere, with biases up to 20%. The impact shows an oscillating behaviour and extends to pressure levels smaller than the cloud top pressure. Note that in both cases shown in Figure 7-8 the bias is smaller than the a posteriori error. However, regarding tropospheric ozone information it is probably best to select ground pixels with low cloud fractions.

We also tested other cloud configurations. For high clouds with a large optical thickness, we found cases where the retrieval failed to converge. Also, the results are sensitive to the input of the cloud pressure, for which we simply used the middle of the cloud layer. For real data, convergence problems for thick high clouds may therefore be expected.

The algorithm has the option for fitting the cloud pressure or the surface pressure. Potentially, these could mitigate errors due to clouds, however in the current figurations these settings are not enabled.

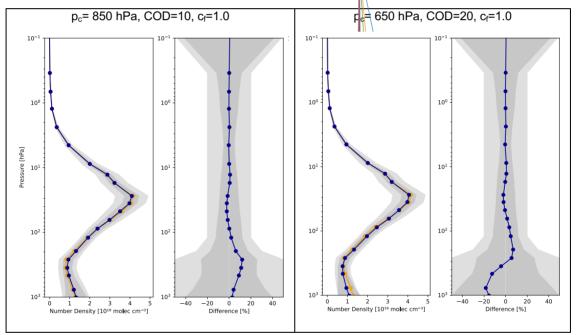


Figure 7-8. Impact of clouds on the ozone profile retrieval. Left panel: simulation for a Mie scattering cloud with an optical depth of 10 placed between 900 and 850 hPa. Right panel: same as left panel, but for an optical depth of 20 and between 700 and 650 hPa.

7.2.5 Sun glint

As the Rayleigh scattering optical thickness is rather large for wavelengths below 320 nm (> 0.92), the probability that photons from the sun travel directly towards the surface and then back to the sensor is small. It is assumed that sun glint can be modelled by increasing the albedo of the Lambertian surface below the atmosphere.

7.2.6 Corrections for a spherical atmosphere

Corrections for a spherical atmosphere have been implemented in the software for direct sunlight (see e.g. Caudill et al., 1997). Simulation studies (Spurr, 2002) indicate that more advanced corrections might be needed for viewing angles > 45 degrees. However, that would imply an increase in calculation time with a factor of 2 - 4. Because comparisons between OMI and MLS results (Kroon et al., 2011) did not show systematically larger differences for oblique viewing directions we will currently not implement a more advance correction for a spherical atmosphere.

7.2.7 Absorbing aerosol

The UV absorbing index product will be used to flag pixels with large amounts of UV absorbing aerosol, e.g. due to volcanic activity, biomass burning or desert dust. Absorption by aerosols can lead to significant overestimates of tropospheric ozone. It is noted that absorbing aerosol that is located low in the atmosphere (i.e. 0 - 2 km) hardly affects the UV aerosol index. However, the wavelength dependent surface albedo will mitigate effects of absorbing aerosol. We tested for aerosol layers with a single scattering albedo of 0.8 and optical depth larger than 4 in the UV. The impact of these layers was found to be less than a few percent throughout the atmosphere, both for aerosol layers near the surface as well as when they were situated 3 kilometres above the surface, which demonstrates the mitigation of fitting a spectrally varying surface albedo.

7.3 Instrumental errors

7.3.1 Radiometric errors

To test the impact of multiplicative radiometric errors, we increase the radiance spectrum with 5% over the entire range from 270 to 330 nm. As can be seen in panels A and B of Figure 7-9, the impact of such an error on the ozone column is limited to less than 5% in the troposphere. The error is mitigated by fitting the surface albedo. The bias in the total column ozone is less than 2 DU.

Compared to the impact of multiplicative errors, the impact of additive errors is much larger. Panels C and D of Figure 7-9 show simulation results for an additive error in the radiance of 5% at 270 nm. Larger biases are found in the stratosphere for pressure levels below approximately 20 hPa, whereas in the troposphere the bias is small. Because the total column is derived from longer wavelengths, where this additive error is relatively small, the bias in the total column is less than 0.5 DU. For the sub columns the bias is within 1.5 DU, except for the 6-12 km sub column which is 2.3 DU.

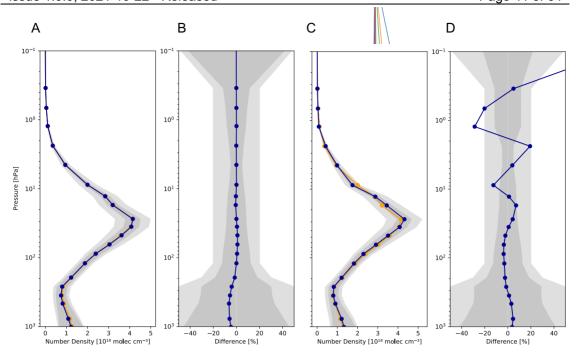
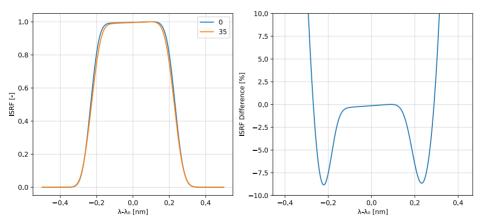


Figure 7-9. Retrieval simulation of impact of radiometric errors. From left to right panels are identified as A-D. Panels A and B show the retrieval and bias for a multiplicative error in the radiance of 5%; panels C and D show the impact of an additive error of 5% of the radiance at 270 nm (corresponding to 2.5 10⁸ ph/s/cm²/nm/sr).

7.3.2 ISRF errors

To test the impact of errors in the ISRF, we simulate the spectrum with the TROPOMI ISRF of ground pixel number 35 and perform the retrieval with the ISRF of number 35. The FWHM of ISRF of ground pixel 0 is 2.5% larger compared to ground pixel 35, resulting in ISRF difference of more than 7.5% for part of the curve where the ISRF exceeds 0.5. The impact of assuming this error in the ISRF on the ozone profile retrieval is less than 1% at all pressure levels.

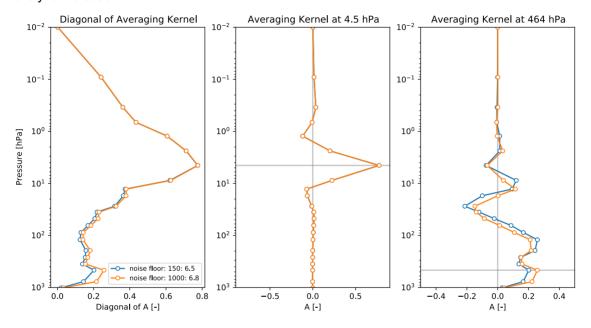


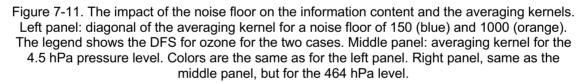
Difference in ISRF for ground pixel 0 and 35 at approximately 300 nm

Figure 7-10. Left panel: TROPOMI ISRF for Band 1 ground pixel 0 and 35 at 300 nm. Right panel: relative difference between the ISRF for ground pixel 0 and 35.

7.3.3 Noise

If the signal to noise ratio (SNR) increases, we get more precise information on the ozone profile. The gain in information if we go from a SNR from 100 to 1000 is often small as the error is dominated by the smoothing error. There is a second effect that may occur when the SNR increases, namely the effect of model errors on the retrieved ozone profile. A high SNR may lead to non-convergence because the system cannot find a solution that is consistent with such a high precision of the measurements. Therefore, a so-called noise floor of 150 has been introduced to improve the convergence. The impact of this noise floor is illustrated in Figure 7-11 by comparing simulation results of a noise floor of 150 and 1000. The noise floor of 150 decreases the DFS by 0.3 and the decrease in information is predominantly in the lower troposphere. In theory, increasing the noise floor thus increases the information on tropospheric ozone, however a prerequisite for this is that the model errors should also be comparable to or the noise floor, which rarely is the case.





7.4 Summary of Error Analysis

In this section the impact of various model and instrumental errors on the ozone profile retrieval were tested. The aim of the error analysis is to distinguish the leading errors and to derive typical numbers for these. Typical biases are reported in Table 7-1. For both the tropospheric and stratospheric pressure levels, incomplete correction for polarization and Raman scattering are an important source of error. For the troposphere, the effect of clouds is estimated to be the largest contribution. It is noted that this can be (partly) avoided by regarding ground pixels with small effective cloud fractions. For the stratosphere, the additive errors are important, especially at low pressure levels the errors can become large. It is noted that mitigations due to radiometric soft calibration are not considered for Table 7-1. The soft calibration is designed to account for both additive and multiplicative errors. However, additive errors of the order 5% at the shortest wavelength may be hard to avoid.

Table 7-1. Typical bias of various error sources on ozone profile retrieval. Errors are split up in tropospheric and stratospheric levels. Colours (green, yellow, red) indicate the severity of the error. Biases do not include mitigation due to radiometric soft calibrations.

Error Source	Troposphere Bias [%]	Stratosphere Bias [%]	Remarks
Ozone spectroscopy	5	5	
ISRF convolution	5	2	
Number of streams	0.2	0.2	
Pol. and Raman scattering	10	10	
Earth sphericity	2	2	Untested expert judgement
Cloud model	20	3	
Absorbing aerosol	2	2	
Sun glint	1	1	
Multiplicative radiometric	5	2	
Additive radiometric	2	20	Below 10 hPa errors are very large
ISRF knowledge	1	1	
RMS	7	7	

8 Validation

For the routine validation of the S5P/TROPOMI ozone profiles, the automated validation server (AVS) deployed by the MPC VDAF [ER02] facility is used. This facility collects reference data and collocated ozone profile retrievals. On a quarterly basis, a report (ROCVR) provides detailed results on the status of the operational satellite data, based on the comparisons with the reference data. In addition to the routine validation, there are also other opportunities for validation, including routine checks by the product validation leads and activities within the S5P validation team (S5PVT). An overview of the most important validation approaches and reference data is provided in this section.

8.1 Reference datasets

Ground-based validation data (FRM) are obtained through ESA's Validation Data Centre (EVDC). Ozonesonde soundings and both stratospheric and tropospheric lidar measurements are considered for ozone profile validation within the S5P MPC VDAF. Intercomparison of the TROPOMI ozone profile data with other satellite measurements is not part of VDAF, but may be considered in different validation exercises.

8.1.1 Ozonesondes

As part of WMO's GAW, the ozonesonde network has been providing accurate, high vertical resolution measurements of the ozone profile on a pseudo-global basis for several decades. Launched on board of small meteorological balloons, electrochemical ozonesondes measure the vertical distribution of atmospheric ozone from the ground up to burst point, typically 30 km. The vertical resolution of ozonesondes (typically 100m) is at a minimum thirty times greater than what is possible from satellite retrieval. Caveats for using ozonesonde datasets include the dependence on launch time, variable balloon burst height, errors depending on instrument set up (buffer solution, pump efficiency correction...), and changes in ozonesonde type and measurement parameters with time. Therefore, ad hoc selections based on the experience must be done before considering an ozonesonde data record as eligible FRM dataset.

In the framework of S5P MPC, it is planned to make use of homogenized ozonesonde FRM datasets as they become available through EVDC. These data originate from the NDACC Data Host facility, the SHADOZ archive, and WOUDC.

8.1.2 DIAL ozone profile reference measurements

As a contribution to WMO's GAW, ground-based differential absorption lidars (DIAL) measuring either the tropospheric ozone profile or the stratospheric ozone profile performs network operation in the framework of NDACC and the US TOLnet (Tropospheric Ozone Lidar Network). Details on the measurement techniques, on network operation and on data policy, as well as primary access to the original data, are available on the NDACC website [ER_NDACC], and the TOLnet website [ER_TOLnet].

In the framework of S5P MPC, it is planned to make use of homogenized DIAL ozone profile data available through EVDC, and originating from the NDACC Data Host facility, and the TOLnet data archive.

8.1.3 Campaign needed with reference to specific algorithm features

It is not foreseen that specific campaigns will be needed to support validation for specific algorithm features. However, results of campaigns that measure ozone profiles are certainly interesting for validation purposes.

8.1.4 Satellite inter-comparison

Comparison to other satellite data may be useful as a large number of intercomparisons can be performed. For the stratospheric ozone profile comparisons to limb sounders (MLS, OMPS) are appropriate. For nadir profiles, comparisons can be made to OMI and OMPS, where the OMPS measurements will have the best spatial and temporal co-registration with TROPOMI.

8.2 Validation approaches

8.2.1 Automated validation server

The MPC VDAF automated validation server provides curtain plots (ozone concentration as a function of altitude and time) of the satellite data at a selection of ground-based reference measurement sites, together with curtain plots showing the difference with the sites' FRMs. Summarizing comparison statistics in terms of bias and dispersion are added.

Comparison of retrieved ozone profiles is complicated if the observations have different horizontal and vertical resolutions. Averaging kernels have to be taken into account to correct for vertical smoothing differences when different grids are used (see e.g. Rodgers and Conner, 2003, Calisesi et al. 2005, Keppens et al., 2019). Further, one must be careful to minimize interpolation errors when profiles are given on different vertical grids.

8.2.2 Daily global maps

The MPC VDAF facility creates daily global maps of the six partial columns provided in the ozone profile product, together with the integrated total column. The latter is visually compared with the daily global map of the TROPOMI total column retrieval to check for product consistency. Daily global maps easily allow identifying data gaps, retrieval artefacts, striping behaviour, and other large-scale features that are not typically observed through FRM comparisons.

8.2.3 Parameter correlation checks

Using the in-house PyCAMA software, correlation checks are performed by KNMI on a broad selection of satellite data parameters within the orbit files. These checks provide a view on single-orbit features, correlations between retrievals of subsequent pixels, the appropriateness of the data flagging, etc.

8.3 Initial validation results

Fourteen preselected satellite orbits have been processed by the PDGS using the NL-L2 O3_PR processor v2.3.0 and based on the recalibrated L1b v2.0.0. These orbits, taken between 2018/05/09 and 2020/07/15, maximize the number of coincidences with ground-based reference measurements within the same period, i.e., more than one hundred in total from 21 ozonesonde stations and 2 lidar stations.

Daily global maps (here limited to one orbit due to the preselection) for the last day (July 15, 2020) are shown in Figure 8-1. Slight along-orbit striping can be observed, especially in the middle stratosphere (24-32 km). On the other hand, the integrated ozone profile map and the total column map (bottom row) look very consistent (same colour scale).

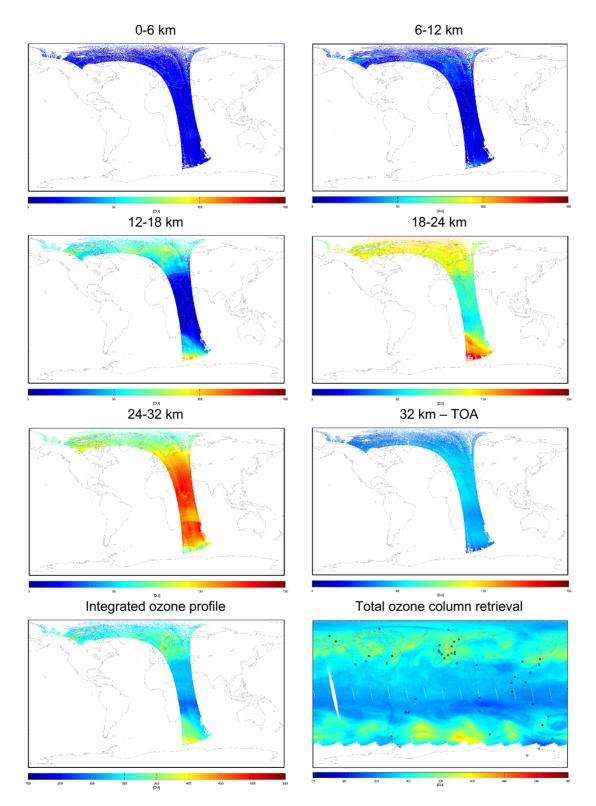
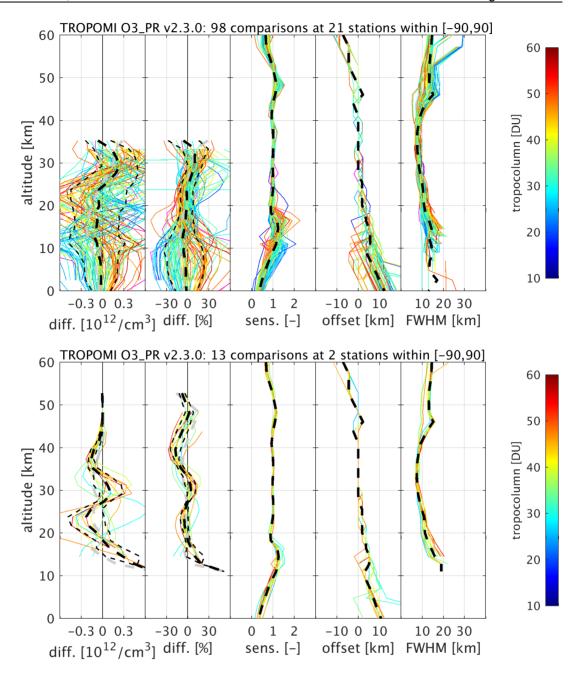


Figure 8-1: Daily global maps (limited to a single orbit on July 15, 2020 here) for the six partial columns in the TROPOMI ozone profile product. The bottom row shows the integrated ozone profile (left) and the total ozone column global map for the same day (right) to check for their consistency. Only pixels with QA_value > 0.5 are included.



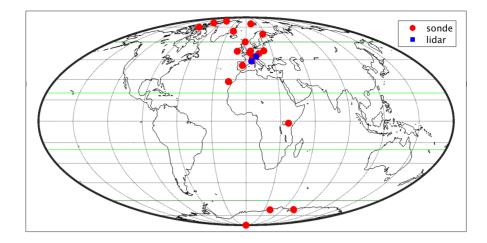


Figure 8-2: Comparison between TROPOMI nadir ozone profile data and all collocated groundbased reference measurements globally, including ozonesondes (top) and stratospheric lidars (middle). Individual graphs show (from left to right) the absolute difference, the relative difference, the vertical sensitivity, the retrieval offset, and the averaging kernel FWHM as a function of the tropospheric ozone column (color scale). Black dashed lines show median values, while grey dashed lines indicate median differences between the prior profiles and the FRM. The bottom panel shows the geographical distribution of the FRM sites.

First comparison results between ground-based reference measurements and 111 coincident satellite pixels (closet pixel with QA > 0.5 on the same day) from the fourteen preselected TROPOMI orbits were obtained through the versatile Multi-TASTE validation system at BIRA-IASB, as part of both S5P MPC and S5PVT validation activities. In Figure 8-2, differences with respect to ozonesonde and stratospheric lidar measurements are shown as a function of tropospheric ozone column, together with the vertical sensitivity, retrieval offset, and averaging kernel width (measured as full width at half maximum, FWHM). The geographical distribution of the FRM sites is added as well, indicating that the lidar comparisons provide a local view only.

The first comparisons show a bias below 10 % in the troposphere to UTLS, with order of 30 % dispersion. The stratosphere shows vertically oscillating biases of 10-20 % positive to negative, but with a smaller dispersion of about 5-10 %. As such, the full profile is within the combined mission requirements (30 % bias and 10 % dispersion), but at this moment does not seem to fulfil the dispersion requirement in the troposphere. The stratospheric bias oscillations may be due to larger a-priori error in the mid and high stratosphere as compared to other retrievals (usually above 20 %).

In the above plots, a bias dependence on the tropospheric ozone column is observed, but this needs to be confirmed. A dependence on other influence quantities (cloud fraction, DFS, latitude, QA_value, surface albedo, SZA, scan angle) has not been detected, but a thorough assessment requires more comparisons. Orbit curtain plots have revealed that for some pixels, particularly in the beginning of the orbit, the retrieved ozone profiles deviate strongly and non-physically from the a-priori. This issue needs further examination and might require an update of the data flagging (in terms of QA_value definition) in the future.

The TROPOMI ozone profile retrieval information content looks as expected for the fourteen orbits of the test dataset. The DFS is typically and on average close to six, with the vertical sensitivity nearly equal to unity from about 10 (UTLS) to 50 km, and decreasing above and below (with small bumps above one in the transition region, as can be expected for nadir profile retrievals). The barycentre of the retrieved information is typically at the retrieval altitude, with positive and negative offsets of up to 10 km below and above the 10-50 km altitude range, respectively (as again expected for nadir profile retrievals). The averaging kernel width (FWHM) as a measure of the effective vertical resolution of the profile retrieval equals 10-15 km on average, reaching a minimum below 10 km in the middle stratosphere.

9 Conclusion

This document provides a high-level description of the ozone profile retrieval. The algorithm is based on Optimal Estimation. Main characteristics of the algorithm are:

- A pre-processing spatial regridding that combines the TROPOMI bands 1 and 2 into larger ground pixels of 28x28 km² at nadir.
- A pre-processing spectral regridding that averages three spectral pixels, thus reducing the spectral oversampling to >2.3, increasing the signal-to-noise and reducing the computational effort.
- A pre-processing correction for polarisation and Raman scattering, to reduce the computational effort.
- A pre-processing radiometric correction, based on soft calibration procedure required to overcome radiometric calibration errors.
- A priori ozone profiles based on a climatology that varies with latitude and ozone column. A priori errors vary with altitude between 20 and 50%.
- Online radiative transfer using the scalar approach and eight streams.
- A state vector that contains the ozone concentration at 33 levels in the atmosphere, the surface albedo at 3 wavelengths, the cloud albedo at 3 wavelengths and the SO₂ column.
- The output product contains the number density ozone concentrations at 33 levels, ozone concentrations for six sub-columns as well as an extensive set of diagnostic information.

The degrees-of-freedom for signal for ozone is expected to be in the range 5-7, depending on the Sun-satellite geometry, and confirmed to be close to six on average for a dedicated test data set. In the troposphere and the lower stratosphere, the total error will be dominated by the smoothing error, resulting from the limited amount of information on the ozone profile in the measurements. Due to this limitation, and the choice for rather large a priori errors, the requirements are expected to be met for the noise error, but not for the total error, that includes the dominating smoothing term. It is noted that for applications that make use of the averaging kernel, the effects of the vertical smoothing differences are taken into account.

From sensitivity analysis, the largest errors in the troposphere are expected from the incomplete correction for polarisation and Raman scattering and from the cloud model. For the stratosphere, the largest errors maybe also be caused by the incomplete correction for polarisation and Raman scattering, as well as from additive radiometric errors.

References

Bates, D. R., Rayleigh scattering by air, Planet. Space Sci., 32, 785-790, 1984.

Birk, M. and Wagner G, ESA SEOM-IAS – Measurement and ACS database O3 UV region, <u>https://zenodo.org/record/4423918#.YVbYcC8Rr0p</u>, 2021

Bodhaine, B.A., N.B. Wood, E.G. Dutton, and J.R. Slusser, On Rayleigh Optical Depth Calculations, *J. Atmospheric and Oceanic Technology*, vol. 16, 1854-1861, 1999.

Bogumil, K., Orphal, J. and Burrows, J. P.: Temperature dependent absorption cross sections of O_3 , NO_2 , and other atmospheric trace gases measured with the SCIAMACHY spectrometer, in Looking down to Earth in the New Millennium, vol. SP-461, Gothenburg., 2000.

Chance, K. and R.J.D. Spurr, Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering and the Fraunhofer spectrum, *Applied Optics*, 36, 5224-5230, 1997.

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

De Haan, J. F., P.B. Bosma, J.W. Hovenier, The adding method for multiple scattering calculations of polarized light, *Astronomy and Astrophysics*, 183, 371-391, 1987.

Calisesi, Y., V. T. Soebijanta, and R. van Oss, Regridding of remote soundings: Formulation and application to ozone profile comparison, *J. Geophys. Res.*, 110, D23306, doi:10.1029/2005JD006122, 2005

Chandrasekhar, S., Radiative transfer, Oxford Univ. Press, 1950.

Dobber, M., R. Voors, R. Dirksen, Q. Kleipool, and P. Levelt, The high-resolution solar reference spectrum between 250 and 550 nm and its application to measurements with the Ozone Monitoring Instrument, *Solar Physics* 249, 281-291, DOI 10.1007/s11207-008-9187-7, 2008

Hermans, C., Vandaele, A. C. and Fally, S.: Fourier transform measurements of SO₂ absorption cross sections: I. Temperature dependence in the 24,000–29,000 cm⁻¹ (345–420 nm) region, jqsrt, 110, 756–765, doi:<u>10.1016/j.jqsrt.2009.01.031</u>, 2009.

Hovenier, J.W., C.V.M. van der Mee, and H. Domke. 2004. Transfer of polarized light in *planetary atmospheres; basic concepts and practical methods*. Dordrecht: Kluwer.

Jaross, G. (2017), OMPS-NPP L3 NM Ozone (O3) Total Column 1.0 deg grid daily V2, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [01-10-2021], 10.5067/7Y7KSA1QNQP8

Keppens, A., Compernolle, S., Verhoelst, T., Hubert, D., and Lambert, J.-C.: Harmonization and comparison of vertically resolved atmospheric state observations: methods, effects, and uncertainty budget, Atmos. Meas. Tech., 12, 4379–4391, https://doi.org/10.5194/amt-12-4379-2019, 2019.

Kleipool, Q., Ludewig, A., Babić, L., Bartstra, R., Braak, R., Dierssen, W., Dewitte, P.-J., Kenter, P., Landzaat, R., Leloux, J., Loots, E., et al.: Pre-launch calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite, Atmospheric Measurement Techniques, 11(12), 6439–6479, doi:10.5194/amt-11-6439-2018, 2018.

Kroon, M., J. F. de Haan, J. P. Veefkind, L. Froidevaux, R. Wang, R. Kivi, and J. J. Hakkarainen, Validation of operational ozone profiles from the Ozone Monitoring Instrument, *J. Geophys. Res.*, 116, D18305, doi:10.1029/2010JD015100, 2011.

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 ozonesondes and Aura MLS data, Journal of Geophysical Research: Atmospheres, 120(6), 2537–2545, doi:https://doi.org/10.1002/2014JD022634, 2015.

Landgraf, J., O. P. Hasekamp, M. Box, and T. Trautmann, A linearized radiative transfer model using the analytical perturbation approach, *J. Geophys. Res.*, 106, 27,291–27,305, 2001.

Landgraf, J., O. P. Hasekamp, R. van Deelen, and I. Aben, Rotational Raman scattering of polarized light in the Earth's atmosphere: A vector radiative transfer model using the radiative transfer perturbation theory approach, *J. Quant. Spectrosc. Radiat. Transfer*, 87, 399–433, 2004.

Ludewig, A., Kleipool, Q., Bartstra, R., Landzaat, R., Leloux, J., Loots, E., Meijering, P., van der Plas, E., Rozemeijer, N., Vonk, F. and Veefkind, P.: In-flight calibration results of the TROPOMI payload on-board theSentineI-5 Precursor satellite, Atmospheric Measurement Techniques Discussions, 2020, 1–28, doi:10.5194/amt-2019-488, 2020.

Malicet, C., Daumont, D., Charbonnier, J., Parisse, C., Chakir, A. and Brion, J.: Ozone UV spectroscopy, II. Absorption cross-sections and temperature dependence, *J. Atmos. Chem.*, 21, 263-273, 1995.

Peck, E. R., and K. Reeder, Dispersion of air. J. Opt. Soc. Amer., 62, 958–962, 1972.

Rodgers, C.D., Inverse Methods for Atmospheric Sounding – Theory and Practice, *World Scientific*, Singapore, 2000.

Rodgers, C. D., and B. J. Connor, Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108(D3), 4116, doi:10.1029/2002JD002299, 2003

Stam, D. M., I. Aben, and F. Helderman, Skylight polarization spectra: Numerical simulation of the Ring effect, *J. Geophys. Res.*, 107(D20), 4419, doi:10.1029/2001JD000951, 2002.

Spurr, R.J.D. 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.

Tran, H, Boulet C, Hartmann J-M. Line mixing and collision-induced absorption by oxygen in the A band: laboratory measurements, model, and tools for atmospheric spectra computations. *J. Geophys.Res.* 111, D15210, 2006

Tran, H., J. Hartmann J-M., An improved O2 A band absorption model and its consequences for retrievals of photon paths and surface pressures. *J. Geophys. Res.* 113, D18104, doi:10.1029/2008JD-010011, 2008

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H. J., De Haan, J. F., Kleipool, Q., Van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F., TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Rem. Sens. Environment*, 120:70–83, 2012.

Appendix A Description of prototype software

A.1 DISAMAR

DISAMAR is a code package written in FORTRAN 90 that was developed for supporting remote sensing of the Earth from space in the UV-VIS-NIR-SWIR wavelengths range (270 – 2400 nm). Thermal emission is currently ignored. DISAMAR can simulate the spectrum of the sunlight reflected by the Earth and can apply retrieval algorithms to that simulated spectrum. It can be used to evaluate retrieval algorithms and to determine their sensitivity to measurement errors (noise and offsets). It can also be used to evaluate correction procedures to reduce measurements errors such as fitting stray light. Results of different retrieval algorithms with different assumptions and employing different wavelengths ranges can be compared in terms of bias and precision of the level 2 products. DISAMAR can be used for different remote sensing products such as the ozone profile, the columns of O₃, H₂O, BrO, NO₂, CH₄, CO, SO₂, CHOCHO, and HCHO, properties of clouds and aerosol, and surface reflectance properties. In addition, measured reflectance spectra can be read and level 2 products can be produced with DISAMAR. The retrieval algorithms implemented are

- OE using line-by-line calculations
- OE with a special interpolation procedure to improve the efficiency (not yet implemented in DISAMAR but used in the operational ozone profile algorithm for OMI; see Appendix A3)
- Two versions of DOAS, one is classical DOAS where a slant column is fitted and it is assumed that the air mass factor is wavelength independent, and a more advanced one where the wavelength dependence of the air mass factor is taken into account.

A.1.1 Some strong points of DISAMAR

DISAMAR is one package that can do a lot of different things. The main retrieval algorithm is Optimal Estimation. Therefore, DISAMAR provides full diagnostic information including error covariance matrices and averaging kernels. It can deal with combined retrievals using information from different wavelength bands, e.g. a combined retrieval of NO₂, cloud information from the O₂ A band, and evaluation of the impact of stray light in the O₂ A band on the bias in the retrieved NO₂ column.

The retrieved profiles are real profiles in the sense that the volume mixing ratio (vmr) is defined for each altitude (through spline interpolation or linear interpolation on the logarithm of the vmr given at a user defined pressure grid). This makes conversion from vmr to Dobson Units per layer and from vmr to number densities simple.

Radiative transfer can be performed with a high accuracy for the following reasons:

- The wavelength grid used for integration over the instrument slit function can be chosen by the user (for line absorbers the number of Gaussian division points between strong lines can be specified).
- The altitude grids where the pressure and temperature profiles and the volume mixing ratio profiles of trace gases are specified, are supplied by the user and can be specified with a high vertical resolution (e.g. NO₂ profile in the boundary layer). These grids can be different for different trace gases, which is useful during campaigns where measured profiles are available.
- A separate altitude grid is used for the radiative transfer calculations. This grid is independent of the grid used for specifying the atmospheric properties. This makes it possible to deal with strong vertical gradients in the radiation field, e.g. near the top of clouds.
- The number of streams, used for integration over polar angles when multiple scattering is involved, can be arbitrary large.
- The number of coefficients used for the expansion of the phase function in Legendre functions can be arbitrarily large.
- The user can choose between the adding/doubling method and the LABOS/doubling method for the radiative transfer calculations. Here LABOS denotes Layer Based Orders

of Scattering. Both methods are exact in the sense that they converge to the correct solution if more streams, more expansion coefficients of the phase function, and more Gaussian points for the numerical integrations over altitude and wavelength, are used.

- Vector radiative transfer is implemented to account for the polarized nature of light scattering. However, the retrieval only deals with intensities. Other Stokes parameters than the intensity can currently not be used for the retrieval.
- By using the newly developed DISMAS (DIfferential and SMooth Absorption Separated) approach, which combines elements of the DOAS retrieval method with OE, the number of wavelength where forward calculations are to be performed can be reduced significantly. For NO₂ retrievals the calculation times can be reduced by a factor of about 80. However, this approach does not work for line absorbers and seems to be unstable for ozone profile retrieval.
- For the oxygen A band collision induced absorption and line mixing is implemented, based on the work of Tran et al. (2006) and Tran and Hartmann (2008).
- Single rotational Raman scattering (RRS) is implemented using the perturbation approach (Landgraf et al., 2004).

DISAMAR provides not just the radiance spectrum of the backscattered sunlight, but also the derivatives with respect to the elements of the state vector (i.e. the parameters that are to be fitted). These derivatives are essential in OE and are used to find the solution in an iterative manner. They are also used to determine the error covariance matrix. In DISAMAR currently all derivatives are calculated in a semi-analytical manner. Generally the calculation of the derivatives takes no more than 10 - 30% of the total calculation time. The derivatives are calculated using reciprocity (equivalent to the ad joint method, see for example Landgraf et al. 2001). In contrast to most approaches the derivatives are initially calculated for specific altitude levels in the atmosphere, not for atmospheric layers. This provides a very flexible approach to calculate of secondary derivatives, such as the derivative with respect to the optical thickness of an aerosol layer.

Currently the following parameters can be fitted:

- Absorbing trace gas profiles defined through interpolation with nodes specified by the user. The number of nodes and the spacing of the nodes are read from the configuration file and can differ for each trace gas.
- The total column of one or more traces gases.
- The surface albedo, represented as a low degree polynomial in the wavelength. Correlations between albedo values at different wavelengths can be taken into account in the a priori error covariance matrix.
- Surface emission, in particular fluorescence the O₂ A band
- The surface pressure using the O₂ A-band.
- The top altitude of a cloud or aerosol layer when it is assumed that the pressure difference between the top and base of the layer is constant.
- The aerosol/cloud optical thickness at the reference wavelength 550 nm for one aerosol / cloud layer.
- The angstrom coefficient in case of Henyey-Greenstein scattering.
- The single scattering albedo of an aerosol layer in case of Henyey-Greenstein scattering. For Mie scattering the single scattering albedo is determined by the Mie model that is used.
- The altitude of a Lambertian cloud layer.
- The albedo of a Lambertian cloud, represented by a low-degree polynomial in the wavelength.
- The cloud fraction, which can differ between spectral bands and can be a polynomial in the wavelength in order to address intra-band co-alignment errors.
- Stray light represented by a low degree polynomial in the wavelength.
- A constant offset in the temperature which is the same for all altitudes and temperature profiles.

If profiles are fitted, e.g. the ozone profile, OE provides errors in the volume mixing ratio at the nodes where the profile is specified. However, level 2 user requirements are often specified for parts of the atmosphere, such as ozone in the boundary layer or ozone in the free troposphere. Therefore a module was added to calculate the value and precision of sub columns in the atmosphere (see Appendix A2). The boundaries of the sub columns can be specified freely by the user.

The simulated measured radiance can be modified by adding random Gaussian noise, by adding multiplicative and additive offsets, by adding stray light, by introducing wavelength shifts and by adding sinusoidal spectral features to the radiance. This makes it possible to calculate Level 1b requirements for new instruments, given a set of (user) requirements for the Level 2 products.

Retrieval can be performed for simulated spectra. In addition, it is possible to read measured reflectances (including noise) and perform retrievals for those measured spectra.

A.1.2 Limitations of DISAMAR

DISAMAR is still under development and has its limitations. Some of the main limitations are:

- Radiative transfer assumes a plane-parallel atmosphere with an optional correction for the curvature of the atmosphere for incident sunlight. This may become inaccurate for observation far from nadir (viewing nadir angle > 60 degree), in particular for stratospheric gases.
- Retrieval of aerosol profiles is not supported. It is assumed that one or more aerosol or cloud layers are specified, but retrieval of aerosol and cloud properties is currently restricted to one layer. For each layer it is assumed that the volume extinction coefficient, the single scattering albedo and the phase function of the aerosol/cloud particles is constant with altitude. The number density of absorbing trace gas and Rayleigh scattering by molecules generally varies within the aerosol/cloud layer.
- Raman scattering in water is currently ignored. It could perhaps be approximated using fluorescence, but that has not been investigated.
- Remote sensing from the ground and from aircraft is not supported.
- Limb observations are not supported.

A.2 Calculating properties of finite sub columns

The profile is defined through cubic spline interpolation on the logarithm of the volume mixing ratio and errors in the profile are given in terms of the errors in the volume mixing ratio at the retrieval nodes. However, requirements on the ozone profile are formulated in terms of errors for finite columns (see Sec. 4.3). It is therefore necessary to be able to calculate the errors for sub columns.

For that purpose we construct a new altitude grid using Gaussian division points for altitudes between the boundaries of the sub columns. For this altitude grid a new a priori error covariance matrix is constructed by interpolating on the diagonal elements of the original a priori error covariance matrix and using the correlation length to fill the complete error covariance matrix.

One can use the following formalism

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon} = \mathbf{F}(\mathbf{x}_a) + \mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \boldsymbol{\varepsilon}$$
(A2-1)

where we want to use integration to obtain the errors for the sub columns. This has some consequences: (1) the state vector elements have to be number densities, not volume mixing ratios, because integration over number densities gives the amount of trace gas in a sub column, and (2) $\mathbf{K}(\mathbf{x} - \mathbf{x}_a)$ should be a matrix – vector multiplication, which means that the Gaussian weights are included in \mathbf{K} . For each sub column a number of Gaussian division points, n_k are used so that

$$N_k = \sum_{n=1}^{n_k} w_{n,k} \ x_{n,k}$$

is the sub column k. Here $w_{n,k}$ are the weights for sub column k, and $x_{n,k}$ are the number densities of ozone for sub column k at the altitudes $z_{n,k}$

The values of **K** on the Gaussian altitude grid, $z_{n,k}$, are calculated using spline interpolation on the original RTM altitude grid, and are multiplied with the Gaussian weights. The a posteriori error covariance matrix for the number density can now be calculated on the Gaussian altitude grid, using

$$\mathbf{S} = \left[\mathbf{K}^T \mathbf{S}_{\varepsilon}^{-1} \mathbf{K} + \mathbf{S}_{a}^{-1}\right]^{-1}$$
(A2-2)

Finally, integration is used to calculate the error covariance matrix for the sub columns. The integration is performed using the **W** matrix approach described in Rodgers 2000. In this case **W** is a diagonal matrix containing Gaussian weights. The error covariance matrix for the sub columns, S_{subCol} is given by

$$\mathbf{S}_{\text{subCol}} = \mathbf{W}^T \mathbf{S} \mathbf{W}$$
(A2-3)

where the diagonal elements provide the error for each sub column. Similarly, averaging kernels for the sub columns are calculated.

The main differences with the calculation of the usual diagnostic information are therefore:

- A Gaussian altitude grid is used adapted to the sub columns used.
- The state vector elements are the number densities at different altitudes and can not be volume mixing ratios.
- No new radiative transfer calculations are performed, instead interpolation is used to get

 \mathbf{K} , and the diagonal element of \mathbf{S}_{ε} and \mathbf{S}_{a} on the proper Gaussian altitude grid. If a correlation length is applies for the non-diagonal elements of \mathbf{S}_{ε} and \mathbf{S}_{a} , the non-diagonal elements of these matrices are calculated with the proper correlation length. When a Lambertian cloud is present, \mathbf{K} is discontinuous at the altitude of the cloud

A.3 Special interpolation to improve the efficiency

which is taken into account during the interpolation.

In order to simulate the measure radiances integration over the instrument slit function is required. That is, we have to calculate the radiance for each spectral bin

$$L(\lambda_i) = \frac{\cos(\theta_0)}{\pi} \int S(\lambda_i, \lambda) R(\lambda) F(\lambda) d\lambda$$
(A3-1)

using numerical integration. Here θ_0 is the solar zenith angle, $S(\lambda_i, \lambda)$ is the instrument slit function, $R(\lambda)$ is the reflectance of the atmosphere-surface system, and $F(\lambda)$ is the solar irradiance. The solar irradiance varies strongly as a function of wavelength. We use a high resolution solar spectrum that is tabulated with a spacing of 0.01 nm and a spectral resolution of 0.025 nm. Noting that $S(\lambda_i, \lambda)$ has a width of 1 nm in band 1 and 0.5 nm in band 2, it means that we need to know the reflectance also on a spectral grid with a spacing of about 0.01 nm (see also Sec. 8.1.2.1). For the full ozone profile we use the interval 270 nm - 320 nm, so we need to know the reflectance at about 5000 spectral points. Performing radiative transfer calculations at all those spectral points is extremely time-consuming.

In this section we ignore complexities due to rotational Raman scattering and assume that spectral features in the reflectance are caused by spectral features in the absorption cross sections of the trace gases present in the atmosphere, such as O3 and SO2.

We define two spectral grids, a fine grid and a coarse grid. The fine grid is used for the integration over the slit function (See Eq. (A3-1)) and consists of repeated Gaussian integration points, typically 2000 - 4000 points for the interval 270 - 320 nm. The coarse grid is the grid where radiative transfer calculations are performed.

The values of the reflectance on the fine grid are obtained by a special interpolation approach, based on our knowledge of the spectral variation of the absorption cross section of the trace gases. Omitting the derivation we simply give the end result.

$$R^{f}(\lambda_{i}) = R^{c}(\lambda_{i}) + \sum_{n=1}^{N_{br}} \int_{0}^{TOA} \frac{\partial R^{c}(\lambda_{i}, z)}{\partial k_{abs}} n_{n}(z) \left[\sigma_{n}^{f}(\lambda_{i}, T(z)) - \sigma_{n}^{c}(\lambda_{i}, T(z)) \right] dz$$
(A3-2)

Here $R^{f}(\lambda_{i})$ is the reflectance on the fine grid, $R^{c}(\lambda_{i})$ is the reflectance calculated by cubic spline interpolation on the coarse grid, $\partial R^{c}(\lambda_{i},z)/\partial k_{abs}$ is the partial derivative with respect to the volume absorption coefficient at altitude z interpolated to the fine grid, $n_{n}(z)$ is the number density of trace gas n at altitude z, $\sigma_{n}^{f}(\lambda_{i},T(z))$ is the absorption cross section of trace gas n on the fine grid at temperature T(z), and $\sigma_{n}^{c}(\lambda_{i},T(z))$ is the absorption cross section at temperature T(z)of trace gas n cubic spline interpolated to the fine grid. The integral over the altitude is performed using Gaussian integration. The partial derivative $\partial R^{c}(\lambda_{i},z)/\partial k_{abs}$ is known from the radiative transfer calculations.

A variant of this interpolation procedure is used for the OMI operational ozone profile algorithm, separately for Cabannes and Rayleigh scattering. For the coarse grid 105 wavelengths are used on the interval 270 - 330 nm. This procedure is not yet implemented for the prototype algorithm, but we expect that we need about 100 radiative transfer calculations instead of the 2000 - 4000 mentioned above. When RRS is included a similar approach can be used, or one can fall back to the approach of using Rayleigh and Cabannes scattering as used in the operational OMI algorithm.

Appendix B Description of Test Cases

B.1 Description of Geophysical Test Cases

The geophysical test cases were developed for the CAMELOT project and were calculated using the chemical transport model TM4. Below the ozone profiles are specified for European polluted, European background, China polluted, and Tropical biomass burning cases.

European polluted			
Latitude	50 - 52		
Longitude	3 - 6		
Month	6		
Altitude (m)	Pressure (hPa)	Temperature (K)	O3 (ppbv)
70000.0	0.053	218.100	1.853E+02
65000.0	0.106	240.100	3.689E+02
60000.0	0.206	247.012	7.158E+02
55000.0	0.399	252.480	1.387E+03
50000.0	0.772	258.689	2.681E+03
47500.0	1.331	261.853	3.480E+03
45000.0	1.899	265.017	4.278E+03
42500.0	2.466	260.314	5.128E+03
40000.0	3.142	254.592	5.985E+03
37500.0	4.762	248.870	6.841E+03
35000.0	6.382	243.048	7.038E+03
32500.0	9.344	237.205	7.093E+03
30000.0	13.431	231.778	6.704E+03
27500.0	18.600	227.086	5.532E+03
25000.0	28.728		
24000.0	32.779	32.779 221.493 3.847E+03	
23000.0	36.830	36.830 220.474 3.339E+03	
22000.0	42.050	42.050 219.456 2.832E+03	
21000.0	49.699	218.438	2.324E+03
20000.0	57.347	217.714	1.942E+03
19000.0	67.421	216.996	1.563E+03
18000.0	78.890	216.659	1.225E+03
17000.0	91.496	216.818	9.410E+02
16000.0	107.833	217.111	6.779E+02
15000.0	124.751	218.070	5.180E+02
14000.0	146.953	218.886	3.690E+02
13000.0	170.513	218.850	2.848E+02
12000.0	199.909	219.526	2.064E+02
11000.0	233.133	221.575	1.390E+02
10000.0	270.302	226.757	1.018E+02
9000.0	314.561	233.437	8.032E+01
8000.0	363.078	241.119	7.223E+01
7000.0	416.582	248.509	6.740E+01
6000.0	476.430	255.522	6.376E+01
5000.0	543.953	262.075	6.078E+01
4000.0	618.594	268.110	5.812E+01
3000.0	701.221	273.654	5.582E+01
2000.0	793.521	278.682	5.462E+01
1000.0	895.617	283.300	5.567E+01
0.0	1008.900	289.304	4.308E+01

European background			
Latitude	44 - 46		
Longitude	0 - 3		
Month	6		
Altitude (m)	Pressure (hPa)	Temperature (K)	O3 (ppbv)
70000.0	0.046	218.100	1.802E+02
65000.0	0.091	240.100	3.586E+02
60000.0	0.177	245.957	6.967E+02
55000.0	0.345	250.249	1.356E+03
50000.0	0.671	256.502	2.639E+03
47500.0	1.232	259.990	3.500E+03
45000.0	1.801	263.478	4.362E+03
42500.0	2.371	260.144	5.242E+03
40000.0	2.941	254.088	6.129E+03
37500.0	4.508	248.033	7.016E+03
35000.0	6.139	242.335	7.271E+03
32500.0	8.758	236.814	7.215E+03
30000.0	12.863	231.462	6.916E+03
27500.0	17.236	226.706	5.766E+03
25000.0	27.412	221.950	4.617E+03
24000.0	31.483	220.472	4.049E+03
23000.0	35.553	219.092	3.456E+03
22000.0	39.697	217.712	2.863E+03
21000.0	47.417	216.332	2.270E+03
20000.0	55.155	215.096	1.808E+03
19000.0	64.290	213.927	1.404E+03
18000.0	75.955	213.033	1.033E+03
17000.0	87.660	213.430	8.120E+02
16000.0	104.169	213.829	5.910E+02
15000.0	120.739	215.062	4.653E+02
14000.0	142.548	216.343	3.453E+02
13000.0	165.124	216.582	2.649E+02
12000.0	194.665	217.339	1.903E+02
11000.0	226.860	220.586	1.326E+02
10000.0	263.809	226.217	9.734E+01
9000.0	306.683	233.520	7.890E+01
8000.0	353.748	241.396	7.266E+01
7000.0	405.708	249.084	6.832E+01
6000.0	464.238	256.462	6.492E+01
5000.0	529.615	263.398	6.212E+01
4000.0	601.810	269.839	5.983E+01
3000.0	681.566	275.872	5.795E+01
2000.0	770.370	281.145	5.675E+01
1000.0	868.598	285.652	5.662E+01
0.0	977.520	291.238	5.034E+01
L			1

China polluted			
Latitude	30 - 32		
Longitude	114 - 117		
Month	6		
Altitude (m)	Pressure (hPa)	Temperature (K)	O3 (ppbv)
70000.0	0.041	218.100	1.938E+02
65000.0	0.081	240.100	3.855E+02
60000.0	0.158	244.523	7.503E+02
55000.0	0.309	247.224	1.469E+03
50000.0	0.608	252.707	2.888E+03
47500.0	1.177	256.152	3.785E+03
45000.0	1.755	259.597	4.682E+03
42500.0	2.332	256.827	5.619E+03
40000.0	2.910	250.535	6.577E+03
37500.0	4.447	244.243	7.536E+03
35000.0	6.106	238.766	8.014E+03
32500.0	8.735	233.752	8.219E+03
30000.0	12.887	228.936	8.097E+03
27500.0	17.409	224.801	6.853E+03
25000.0	27.638	220.666	5.609E+03
24000.0	31.730	218.248	4.837E+03
23000.0	35.822	215.716	4.025E+03
22000.0	40.246	213.184	3.213E+03
21000.0	48.195	210.652	2.400E+03
20000.0	56.145	206.995	1.763E+03
19000.0	66.302	203.043	1.171E+03
18000.0	78.684	200.643	7.016E+02
17000.0	92.508	200.713	4.283E+02
16000.0	110.158	202.006	2.151E+02
15000.0	129.485	206.335	1.510E+02
14000.0	152.716	211.478	1.050E+02
13000.0	178.802	217.880	8.708E+01
12000.0	207.934	224.953	7.645E+01
11000.0	241.971	232.408	7.006E+01
10000.0	278.673	240.058	6.934E+01
9000.0	320.568	247.332	6.942E+01
8000.0	367.611	254.143	7.030E+01
7000.0	419.203	260.119	7.049E+01
6000.0	476.376	265.730	6.915E+01
5000.0	540.995	271.129	6.625E+01
4000.0	612.781	276.507	6.238E+01
3000.0	692.010	281.786	5.899E+01
2000.0	780.478	286.724	5.719E+01
1000.0	877.569	292.394	6.035E+01
0.0	984.600	298.254	5.818E+01

Tropical biomass burning			
Latitude	4 - 6		
Longitude	18 - 21		
Month	1		
Altitude (m)	Pressure (hPa)	Temperature (K)	O3 (ppbv)
70000.0	0.019	218.100	1.547E+02
65000.0	0.037	240.100	3.079E+02
60000.0	0.072	245.139	5.987E+02
55000.0	0.141	248.387	1.169E+03
50000.0	0.276	250.834	2.293E+03
47500.0	0.873	251.303	3.257E+03
45000.0	1.462	251.772	4.220E+03
42500.0	2.060	251.271	5.223E+03
40000.0	2.658	245.723	6.434E+03
37500.0	3.773	240.174	7.644E+03
35000.0	5.474	234.835	8.591E+03
32500.0	7.260	230.122	8.748E+03
30000.0	11.504	225.409	8.905E+03
27500.0	15.748	220.464	7.274E+03
25000.0	24.754	215.492	5.432E+03
24000.0		28.972 213.499 4.668E+03	
23000.0	33.191	211.489	3.788E+03
22000.0	37.409	209.478	2.909E+03
21000.0	43.534	207.468	2.029E+03
20000.0	51.649	204.979	1.256E+03
19000.0	59.767	200.813	8.561E+02
18000.0	72.222	196.647	4.562E+02
17000.0	84.890	194.510	2.762E+02
16000.0	102.047	192.684	1.298E+02
15000.0	120.462	197.442	7.940E+01
14000.0	143.715	203.513	4.840E+01
13000.0	168.324	210.981	5.085E+01
12000.0	197.896	218.583	5.511E+01
11000.0	230.476	226.462	6.289E+01
10000.0	266.144	234.400	6.521E+01
9000.0	307.981	242.297	6.496E+01
8000.0	353.635	249.998	6.302E+01
7000.0	403.816	257.311	6.159E+01
6000.0	459.804	264.215	6.150E+01
5000.0	522.716	270.298	6.321E+01
4000.0	592.272	275.189	6.855E+01
3000.0	669.370	280.068	8.142E+01
2000.0	755.161	286.886	9.578E+01
1000.0	848.733	294.887	1.057E+02
0.0	951.320	298.708	7.907E+01
			1

Ozone hole conditions			
Origin	Based on AFGL arctic winter model, but ozone between 12 and 20 km altitude is strongly reduced.		
Month	10		
Altitude (m)	Pressure (hPa)	Temperature (K)	O3 (ppmv)
	0.040	245.4	0.5000
	0.079	248.4	0.6500
	0.155	250.9	0.9500
	0.299	259.1	1.6000
	0.572	259.3	2.6000
	0.790	253.2	3.0000
	1.113	247.0	4.1000
	1.570	240.8	5.1000
	2.243	234.7	5.9000
	3.230	228.5	6.2500
	4.701	222.3	6.2000
	6.910	218.5	5.9000
	10.200	216.0	5.4000
	15.130	213.6	3.0000
	22.560	211.2	1.0000
	26.490	211.8	0.3000
	31.090	212.4	0.1000
	36.470	213.0	0.0300
	42.770	213.6	0.0100
	50.140	214.2	0.0100
	58.750	214.8	0.0100
	68.820	215.4	0.0100
	80.580	216.0	0.0100
	94.310	216.6	0.0100
	110.300	217.2	0.0100
	129.100	217.2	0.0100
	151.000	217.2	0.0100
	176.600	217.2	0.0100
	206.700	217.2	0.0300
	241.800	217.2	0.1000
	282.900	217.2	0.2100
	330.800	220.6	0.1040
	385.300	227.3	0.0725
	446.700	234.1	0.0445
	515.800	240.9	0.0380
	593.200	247.7	0.0325
	679.800	252.7	0.0277
	777.500	255.9	0.0234
	887.800	259.1	0.0207
	1013.000	257.2	0.0180

Appendix C Input Data

This appendix presents the dynamic and (semi-) static input data used by the algorithm. This information is copied from the IODD [RD09].

C.1 Dynamic input

Name/Data	Source	Comments
7 profiles	Forecast from ECMWF	3-hour interval forecast (T +3 to T +12)
Surface pressure	ECMWF	3 hours interval forecast (T +3 to T +12)
Ozone Column	ECMWF	3 hours interval forecast (T +3 to T +12)
KNMI FRESCO cloud	TROPOMI Level 2 KNMI cloud support product	Cloud fraction and cloud pressure.
Snow and Ice cover	ECMWF	3 hours interval forecast (T +3 to T +12)
UV Aerosol Index	TROPOMI Level 2 prouct	
Soft calibration spectra		

C.2 (Semi) static input data

Name/Data	Source	Comments
Algorithm configuration for the full O ₃ profile product (CFG_O3_PRF)	Ozone profile prototype algorithm	Specification of choices made for the algorithm
Processor configuration file for the full O₃ profile product (CFG_O3PR)	Ozone profile prototype algorithm	
Reference spectra for the O ₃ profile processor (REF_XS_O3P)	Absorption cross sections for O_3 and SO_2 taken from literature.	For O ₃ : Malicet et al. 1995. For SO ₂ : composite based on Bogumil et al. (2000) and Hermans et al. (2009).
High resolution solar reference spectrum (REF_SOLAR_)	Dobber et al. 2008	Sampling at 0.01 nm with a spectral resolution of 0.025 nm
High resolution digital elevation map, including land-sea mask (REF_DEM)		

Surface albedo database (REF_LER)			
Co-registration file, mapping pixels from one band onto another (LUT_COREG_)			
O ₃ profile climatology (AUX_O3M)		O3 profile used a	s a-priori
Instrument spectral response function (AUX_ISRF)	Calibration measurements		
Polarization correction NN model for O ₃ profile retrieval (LUT_POLCOR)			
O ₃ profile shape and temperature climatology, TOMS version 8 (AUX_O3M)	TOMS V8		O3 profile shape is used as a- priori
Instrument spectral response function (AUX_ISRF)	Calibration measurements		
Polarization correction lookup table for O ₃ profile retrieval (LUT_POLCOR)	KNMI (produced by DISAMAR)		