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

S5P/TROPOMI ATBD Cloud Products

TROPOMI

sentinel-5p

document number	:	S5P-DLR-L2-ATBD-400I
authors	:	Diego Loyola, Ronny Lutz, Athina Argyrouli, Rob Spurr
CI identification	:	CI-400I-ATBD
issue	:	1.5
date	:	2018-04-30
status	:	Released

Document approval record

	digital signature	
prepared:	Diego Loyola Digitally signed by Diego Loyola DN: cn=Diego Loyola, o=German Aerospace Center (DLR), ou, email=Diego.Loyola@dlr.de, c=DE Date: 2018.04.27 18:56:21 +02'00'	Athina Argyrouli Digitally signed by Athina Argyrouli DN: cn=Athina Argyrouli, o=DLR, ou=IMF, email=athina.argyrouli@dlr.de, c=DE Date: 2018.05.03 10:26:08 +02'00'
	Ronny Lutz Digitally signed by Ronny Lutz DN: c=DE, st=Bavaria, o=DLR, ou=IMF-ATP, cn=Ronny Lutz, email=Ronny.Lutz@dlr.de Date: 2018.05.03 10:44:26 +02'00'	
checked:	M. Sneep, H. Sihler, T. Wagner, L. Lelli, J.P. Veefkind	
approved PI:	Diego Loyola Digitally signed by Diego Loyola DN: cn=Diego Loyola, o=German Aerospace Center (DLR), ou, email=Diego.Loyola@dlr.de, c=DE Date: 2018.04.27 18:56:46 +02'00'	
approved PM:		

approved CM:	Pascal Hedelt Digitally signed by Pascal Hedelt DN: c=DE, st=Bavaria, o=DLR, ou=IMF-ATP, cn=Pascal Hedelt, email=pascal.hedelt@dlr.de Date: 2018.05.03 08:07:03 +02'00'

Document change record

issue	date	item	Comments
0.1.0	2012-10-31	All	Initial draft version
0.2.0	2012-11-23	All	Updated version following the internal L2-WG review
0.3.0	2013-06-04	All	Document revised according to SRR/PDR review of February 2013
0.5.0	2013-06-25	All	Updated version following the internal L2-WG review
0.9.0	2013-12-06	All	Updated version following the external review Added recently published paper on ROCINN
0.10.0	2014-04-15	All	Remove instrument description (former chapter 4) and complete the error analysis in chapter 7.
0.11.0	2014-09-30	All Section 6	Update following the L2-CDR actions. Harmonization of symbols across ATBDs. Update computational effort and product size following new information on the number of L1 spectra. Input/output tables added.
0.12.0	2014-12-15	Section 5.2	Updated following the L2-CDR disposition
0.13.0	2015-09		Minor changes for preliminary release to S5P validation teams
0.14.0	2015-12	6.1	Estimates updated with the new binning scheme
1.0	2016-02-01	All	Included comments from L2-WG reviewers First public release
1.1	2016-06-30	5.4	Included chapter on cloud co-registration inhomogeneity flag.
1.5	2018-04-30	5.1, 5.4, 6.2	Updates following Commisioning phase. Reference added and threshold for CCIF updated, information on qa_value added

Contents

1 Introduction

1.1 Purpose of the ATBD

Clouds are an important component of the global hydrological cycle and play a major role in the Earth's climate system through their strong impact on radiation processes. The interplay of sunlight with clouds imposes major challenges for satellite remote sensing, both in terms of the spatial complexity of real clouds and the dominance of multiple scattering in radiation transport. The retrieval of trace gas products from TROPOMI/S5P will be strongly affected by the presence of clouds. The physics behind the influence of cloud on trace gas retrieval is well understood, and in general, there are three different contributions [*Liu et al*., 2004; *Kokhanovsky and Rozanov*, 2008; *Stammes et al*., 2008; *Wagner et al*., 2008]: (1) the albedo effect associated with the enhancement of reflectivity for cloudy scenes compared to cloud-free sky scenes, (2) the so-called shielding effect, for which that part of the trace gas column below the cloud is hidden by the clouds themselves, and (3) the increase in absorption, related to multiple scattering inside clouds which leads to enhancements of the optical path length. The albedo and in-cloud absorption effects increase the visibility of trace gases at and above the cloud-top, while the shielding effect normally results in an underestimation of the trace gas column.

Using radiative transfer modelling, several papers have quantified the influence of cloud parameters on the retrieval of trace gas columns [*Liu et al*., 2004; *Ahmad et al*., 2004; *Boersma et al*., 2004; *Van Roozendael et al*., 2006; *Kokhanovsky et al*., 2007]. These studies have shown that cloud fraction, cloud optical thickness (albedo), and cloud-top pressure (height) are the most important quantities for cloud correction of satellite trace gas retrievals. Figure 1.1 is a global overview of these three cloud properties, as derived from GOME-2 measurements using cloud property retrieval algorithms that will be adapted for TROPOMI/S5P.

(a) (b) (c)

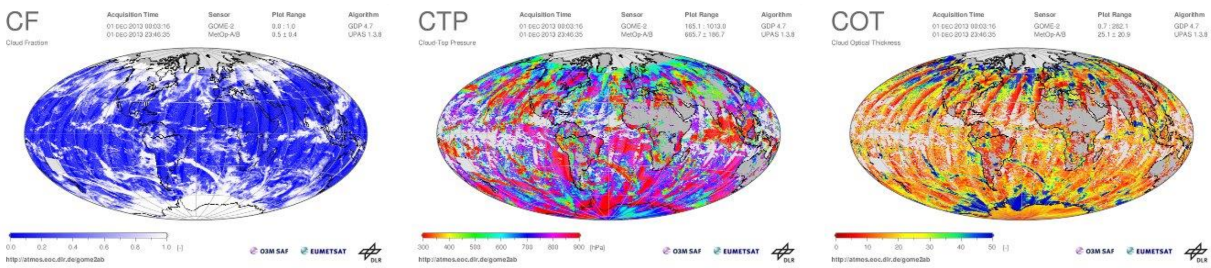


Figure 1.1: (a) Cloud fraction, (b) cloud-top pressure and (c) cloud optical thickness measured by GOME-2 on MetOp-A and MetOp-B in December 2012.

The TROPOMI (Tropospheric Monitoring Instrument) is the payload instrument for the Sentinel 5 Precursor (S5P) Mission. The S5P platform will be launched into a sun-synchronous low-earth orbit in 2016. TROPOMI is a nadir-viewing atmospheric chemistry instrument measuring at moderate spectral resolution from the UV to the near infrared [RD7].

In addition to global measurements of a number of trace species, including ozone, TROPOMI/S5P will also deliver several cloud properties as noted above, and the present ATBD (Algorithm Theoretical Basis Document) describes the two algorithms delivering these cloud properties. The purpose of the ATBD is to provide detailed mathematical and physical descriptions of the two algorithms, along with discussions of algorithm inputs and outputs, data products, algorithm validation and error analysis.

The first algorithm (herewith called S5P_CLOUD_OCRA) is based closely on OCRA - the Optical Cloud Recognition Algorithm [*Loyola and Ruppert*, 1998]. The main product parameter for S5P_CLOUD_OCRA is the fractional cloud cover (between 0 and 1), obtained through comparisons of broad-band reflectance measurements in the UV/VIS/NIR spectral regions with those from a global cloud-free composite data set made up of minimum reflectances.

The second algorithm (herewith called S5P_CLOUD_ROCINN) is based on the ROCINN algorithm [*Loyola et al.*, 2007, and references therein] (Retrieval Of Cloud Information through Neural Networks). The main product parameters for S5P_CLOUD_ROCINN are the cloud optical thickness and the cloud-top height. These are determined through a classical inversion method based on measurements in and around the O_2 *A*-band. However, unlike earlier versions of ROCINN which were based on the treatment of clouds as reflecting surfaces, the new algorithm version will treat clouds as scattering layers.

The OCRA/ROCINN combination has been used to provide auxiliary cloud information in the GOME Data Processor (GDP) algorithms Version 4.x and Version 5 for operational total ozone and other trace gas species. OCRA/ROCINN has also been applied to SCIAMACHY and GOME-2 measurements [*Lutz et al*., 2015]. We note that these two cloud property algorithms are applicable to UV/VIS/NIR atmospheric composition spectrometers; they are not based on thermal infrared and brightness temperatures (e.g. MODIS clouds [*Menzel et al*., 2008]).

It is important to note that cloud properties will be derived from TROPOMI/S5P using the same baseline of spectral measurements (available simultaneously, same spectral regions and samplings) as that for the retrieval of TROPOMI/S5P trace gas products.

Cloud parameters from TROPOMI/S5P will not only be used for enhancing the accuracy of trace gas retrievals, but they will also extend the satellite data record of cloud information derived from oxygen *A*-band measurements initiated with GOME [*Loyola et al*., 2010]. Use of the oxygen *A*-band generates complementary cloud information (especially for low clouds), as compared to traditional thermal infrared sensors (as used in most meteorological satellites) that are less sensitive to low clouds due to reduced thermal contrast.

1.2 Document overview

Following sections on applicable documentation (Chapter 2) and terms of reference (Chapter 3), the S5P cloud property products and retrieval algorithms are introduced in Chapter 4. The main algorithm descriptions are found in Chapter 5. Chapter 6 contains discussions on feasibility and external data sets, while Chapter 7 summarizes the error analysis. Chapter 8 provides an introduction to the validation of S5P cloud products. Concluding remarks are in Chapter 9, and references in Chapter 10.

1.3 Acknowledgements

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

- Sebastian Gimeno Garcia (formerly DLR, now EUMETSAT), Fabian Romahn (DLR), Mattia Pedergnana (DLR), Olena Schüssler (formerly DLR)
- Michel Van Roozendael (BIRA), Christophe Lerot (BIRA),
- Maarten Sneep (KNMI), Thomas Wagner (MPIC), Holger Sihler (MPIC), Luca Lelli (IUP).

2 Applicable and reference documents

2.1 Applicable documents

[AD1] GMES Sentinel-5 Precursor – S5P System Requirement Document (SRD);
source: ESA/ESTEC; **ref**: S5P-RS-ESA-SY-0002; **issue**: 4.1; **date**: 2011-04-29

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

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

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

2.2 Standard documents

There are no standard documents.

2.3 Reference documents

[RD1] Terms, definitions and abbreviations for TROPOMI L01b data processor;
source: KNMI; **ref**: S5P-KNMI-L01B-0004-LI; **issue**: 1.0.0; **date**: 2011-05-18

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

[RD3] S5P/TROPOMI Total Ozone ATBD;
source: DLR; **ref**: S5P-L2-DLR-ATBD-400A; **issue**: 1.0; **date**: 2016-02-01

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

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

[RD6] S5P/TROPOMI Static input for Level-2 processors;
source: KNMI/SRON/BIRA/DLR; **ref**: S5P-KNMI-L2CO-0004-SD; **issue**: 0.4.0; **date**: 2013-06-04

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

[RD8] Sentinel 5 precursor interband coregistration mapping tables;
source: KNMI; **ref** S5P-KNMI-L2-0129-TN; **issue**: 2.0.0; **date**: 2015-09-02

[RD9] Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Cloud Properties;
source: DLR; **ref** S5P-L2-DLR-PUM-400I; **issue**: 1.0; **date**: 2018-04-30

2.4 Electronic references

There are no electronic references.

3 Terms, definitions and abbreviated terms

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

3.1 Terms and definitions

Term	Definition

3.2 Acronyms and abbreviations

AK	Averaging Kernel
AMF	Air Mass Factor
ATBD	Algorithm Theoretical Basis Document
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BOA	Bottom of Atmosphere
CAL	Clouds As scattering Layers
CBH	Cloud base height
CBP	Cloud base pressure
CFR	Cloud Fraction
CCI	Climate Change Initiative
COT	Cloud optical thickness
CRB	Clouds as Reflecting Boundaries
CTA	Cloud top albedo
CTH	Cloud top height
CTP	Cloud top pressure
DLR	German Aerospace Center (Deutsches zentrum für Luft- und Raumfahrt)
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit
ENVISAT	Environmental Satellite
EO	Earth Observation
ERS-2	European Remote Sensing Satellite-2
EOS-AURA	(NASA's) Earth Observing System Aura
ESA	European Space Agency
FD	Finite Difference
FRESCO	Fast REtrieval Scheme for Clouds from the Oxygen *A*-band
GDP	GOME Data Processor

GODFIT	GOme Direct FITting
GOME	Global Ozone Monitoring Experiment
GTOPO	Global TOPOgraphic (Data set)
HICRU	HICRU Iterative Cloud Retrieval Utilities
HITRAN	HIgh-resolution TRANsmission
hPa	Hectopascals
ICFA	Initial Cloud Fitting Algorithm
IPA	Independent Pixel Approximation
ISCCP	International Satellite Cloud Climatology Project
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LIDORT	LInearized Discrete Ordinate Radiative Transfer
LBL	Line-by-Line
LUT	Look Up Table
MERIS	MEdium Resolution Imaging Spectrometer
MetOp	Meteorological Operational
MLER	Minimum Lambert equivalent Reflectivity
MSG	Meteosat Second Generation
MRTD	Mission Requirements Traceability Document
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NetCDF-CF	Network Common Data Format – Climate and Forecast (CF) convention
NIR	Near Infra-Red
NISE	Near-real-time global Ice and Snow Extent
NPP	NPOESS Preparatory Project (NASA platform)
NRT	Near-Real-Time
O3M-SAF	Ozone Monitoring Satellite Application Facility
OCRA	Optical Cloud Recognition Algorithm
OFL	Off-line
OMI	Ozone Monitoring Instrument
PMD	Polarization Measurement Device
RGB	Red-Green-Blue
RMS	Root-Mean-Square
RT	Radiative Transfer
ROCINN	Retrieval of Cloud Information using Neural Networks
SACURA	Semi-Analytical CloUd Retrieval Algorithm
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY

S5P	Sentinel 5 Precursor
SGP	SCIAMACHY Ground Processor
SRD	Systems Requirement Document
SZA	Solar Zenith Angle
TOA	Top of Atmosphere
TOMS	Total Ozone Mapping Spectrometer
TROPOMI	TROPOspheric Monitoring Instrument
UPAS	Universal Processor for UV/VIS Atmospheric Spectrometers
UV	Ultra Violet
VCD	Vertical Column Density
VIIRS	Visible Infrared Imager Radiometer Suite
VIS	Visible
VLIDORT	Vector LInearized Discrete Ordinate Radiative Transfer

4 Introduction to the S5P cloud products

4.1 S5P cloud retrieval heritage

Several cloud retrieval algorithms based on measurements in and around the O_2 *A*-band at 760 nm were developed for the GOME-family of sensors: these include the ICFA (Initial Cloud Fitting Algorithm), FRESCO (Fast REtrieval Scheme for Clouds from the Oxygen *A*-band), SACURA (Semi-Analytical CloUd Retrieval Algorithm), and ROCINN algorithms. These are all based on the Independent Pixel Approximation (IPA), which is the assumption that the "radiative properties of a single satellite "Pixel" are considered in isolation from neighbouring pixels" (definition of the American Meteorological Society). The IPA allows for the application of one-dimensional pseudo-spherical radiative transfer (RT) theory in the forward simulation of cloud-contaminated atmospheric scenarios.

The ICFA algorithm [*Kuze and Chance*, 1994] was used in the initial GOME data processing to derive the effective fractional cover. The FRESCO algorithm [*Koelemeijer et al*., 2001, *Wang et al*., 2008] was also developed for GOME. The FRESCO algorithm is based on the calculation of transmittances (later, single scattering radiances) around the O_2 *A*-band, and it retrieves effective cloud fraction and cloud-top pressure, assuming a fixed albedo of 0.8 for cloud-top. The SACURA algorithm [*Rozanov and Kokhanovsky,* 2004] was developed initially for the SCIAMACHY instrument and then modified to handle also GOME measurements [*Lelli et al*., 2012]. SACURA uses semi-empirical formulae from asymptotic radiative transfer theory to retrieve cloud optical thickness, cloud-top height, liquid water path and other parameters. The ROCINN algorithm [*Loyola et al.*, 2007] is also based on O_2 *A*-band measurements, and is currently being used in the operational GOME and GOME-2 products. ROCINN 2.0 retrieves as primary quantities the cloud-top height and cloud albedo.

The broad-band polarization measurements from GOME, SCIAMACHY and GOME-2 are used for computing cloud fraction, see for example OCRA [*Loyola et al.*, 1998] and HICRU [*Grzegorski et al.*, 2006]. Enhancements to these algorithms have been introduced continuously - see for example the detection of Sun glint effects by [*Loyola et al.*, 2011]. OCRA is also based on the IPA.

There are two competing cloud property algorithms for the OMI instrument. The first uses the cloud screening effect on Fraunhofer filling signatures (due to inelastic rotational Raman scattering) in the region 346-354 nm to derive an optical cloud centroid pressure [*Joiner and Vasilkov*, 2006; *Vasilkov et al*., 2008]. This algorithm uses the minimum Lambertian equivalent reflectivity (MLER) assumption. The second algorithm uses reflectances in and around the O_2-O_2 absorption band near 477 nm [*Acarreta et al*., 2004, *Sneep et al*., 2008]; DOAS-retrieved O_2-O_2 slant columns are compared with simulated look-up table entries to obtain effective fraction and cloud-top pressure. OMI has no O_2 *A*-band measurements.

4.2 S5P cloud product requirements

Unfortunately the Sentinel 4/5 MRTD [AD4] does not contain requirements for the cloud products to be retrieved from the atmospheric Sentinel missions. Based on sensitivity studies of the net effect on trace gas retrievals of errors induced by cloud property uncertainties (see for example [*Van Roozendael et al*., 2006]), we propose the following uncertainty requirements for the TROPOMI/S5P cloud products:

- Cloud Fraction: 20%
- Cloud-top Height (Pressure): 0.5 km (~100hPa)
- Cloud Optical Thickness (Albedo): 20%

The expected uncertainties for these cloud properties depend on the magnitudes of the cloud parameters themselves. Therefore, these values must be considered as average errors for a representative ensemble of observations covering a full range of geophysical conditions.

Note that these requirements are compatible with the threshold criteria defined for the algorithm verification work (S5P L2WG meeting held at DLR on April 15-16, 2013).

The timeliness requirements for the S5P cloud products are:

- Near-real-time (NRT): 3 hours after sensing
- Off-Line (OFL): two weeks after sensing

Note that the runtime budgets need to be apportioned with the trace gas processors that need the cloud products as input.

4.3 Overview of the retrieval algorithms

As noted, OCRA is the S5P_CLOUD_OCRA heritage. In OCRA, optical sensor measurements are divided into two components: a cloud-free background and a remainder expressing the influence of clouds. OCRA was first developed for GOME in the late 1990s, when enough data from the three sub-pixel broad-band PMDs (Polarization Measurement Devices) had accumulated to allow for the construction of the global cloud-free composite which is the key element in the algorithm. Over the course of the 16-year GOME record, the algorithm was refined and the cloud-free composite adjusted as more data became available.

OCRA has also been applied to SCIAMACHY and GOME-2. Initial cloud-free composites for these sensors were based on GOME data before dedicated measurements became available from SCIAMACHY and GOME-2. For S5P_CLOUD_OCRA, the initial cloud-free composite will be based on GOME-2 and OMI (see section 5.2).

ROCINN is the S5P_CLOUD_ROCINN heritage. ROCINN is based on the comparison of measured and simulated satellite sun-normalized radiances in and near the O_2 *A*-band, and it uses a neural network algorithm to retrieve cloud-top height and cloud-top albedo. ROCINN uses the cloud fraction input from OCRA as one starting point. Early versions of ROCINN used a transmittance model to compute simulated radiances, but the latest versions are based on the use of the VLIDORT radiative transfer scattering model.

For GOME and GOME-2, ROCINN Version 2.0 is the current operational algorithm in the GDP [GOME Data Processor]. This version is based on the assumption that clouds are simply Lambertian reflecting surfaces, so that the two main retrieval products are the cloud-top height and the cloud-top albedo itself. This is the “clouds-as-reflecting-boundaries” (CRB) model; see for example [*van Roozendael et al*., 2006] for GOME and [*Loyola et al*., 2011] for GOME 2.

Although ROCINN 2.0 is the heritage algorithm, there is an important point of departure for S5P. For TROPOMI/S5P we will use ROCINN Version 3.0, which is based on a more realistic treatment of clouds as optically uniform layers of light-scattering particles (water droplets). This is the “clouds-as-layers" (CAL) model – here, the two main retrieval products are the cloud-top height and the cloud optical thickness. Details of this algorithm prototype may be found in [*Schuessler et al*., 2014]. Although the CAL model will be the default for S5P, it has been requested that the CRB method should also be retained as an option.

CAL is the preferred method for the relatively small TROPOMI/S5P ground pixels (7x3.5 km^2). The CRB approach works best with large pixels such as those from GOME (footprint 320 x 40 km). Studies have shown that for the smaller GOME-2 pixels, CAL retrieval produces more reliable cloud information than that from CRB, not only with regard to the accuracy of the cloud parameters themselves, but also with regard to the effect of cloud parameter uncertainties on total ozone accuracy.

In OCRA, the intensity is regarded as a linear function of the radiometric cloud cover, and in ROCINN, TOA radiances for partially cloudy scenarios are computed using a linearly-weighted mean of the clear-sky and fully-cloudy calculations, the weighting factor being the cloud fraction. In the context of this IPA model, the two algorithms are consistent. With the notably smaller pixel size that comes with higher spatial resolution, 3-D cloud radiative effects will become an important consideration in error budgeting for the cloud algorithms.

4.4 Foreseen update approach

For OCRA, the main update will be the generation of a cloud-free composite data set. As noted above, the initial data set will be based on GOME-2 and OMI reflectances, and the cloud-free composite will be updated dynamically with S5P reflectance measurements as they become available during the mission. This cloud-composite update will be performed off-line, with no extra computational costs imposed on the retrievals.

For ROCINN, it will be necessary to create (again, off-line) sets of sun-normalized radiances for CAL and CRB – these tasks will be done once the TROPOMI slit function information becomes available. It should be noted that CAL and CRB template data sets for ROCINN have already been generated for existing satellite instruments, and the methodology for creation of corresponding data sets for TROPOMI/S5P will be employed. It is not expected that it will be necessary to update these templates once the mission is underway. Again, no extra computational costs will be imposed on the operational cloud property retrievals.

4.5 General design considerations

The science behind both cloud algorithms is well established, with RT simulations based on the widely used VLIDORT models. Both heritage algorithms have been described in published literature [*Loyola et al*., 2007; *Loyola et al*., 2010] and have been applied to GOME-type measurements.

Current operational products from GDP 4.x (GOME, SCIAMACHY) and GDP 5 (GOME) are generated at DLR and are freely available from ESA, while GDP 4.x products for GOME-2 are freely available from EUMETSAT O3M-SAF. The corresponding operational products contain cloud information retrieved with version 2.0 of the OCRA/ROCINN algorithms.

5 Algorithm descriptions

5.1 Preamble

In this chapter, we describe the two main algorithms to be used for S5P Cloud retrieval. S5P_CLOUD_OCRA is the default algorithm for obtaining the fractional cloud cover from S5P broad-band spectral measurements and is described in section 5.2. S5P_CLOUD_ROCINN is the default algorithm for the cloud optical thickness and cloud-top height products, and is described in section 5.3. Section 5.4 contains a discussion on the application of these algorithms to trace gas total column retrieval. An in-depth description of the operational cloud retrieval algorithms OCRA and ROCINN used for TROPOMI can be found in [*Loyola et al.*, 2018].

5.2 S5P_CLOUD_OCRA for fractional cover

The OCRA cloud fraction determination is based on the comparison of cloud-contaminated measurements with the corresponding measurements of the background (cloud-free) surface.

The core of the algorithm is a cloud-free composite of reflectances that is independent with respect to the atmosphere and to solar and viewing angles. The algorithm requires reflectances that are defined for ground-cover *projections* (i.e. the surface location) of the measurements. Thus, some pre-processing is required before multi-temporal (time series of measurements over the same location) data can be merged to develop the composite. In particular, a soft correction must be applied to the Level-1 data to handle viewing angle dependencies as well as latitudinal and seasonal dependencies as they occur in the data. Soft corrections will also be needed to deal with long-term degradation issues.

For a given location (x, y), we define the reflectance $\rho(x, y, \lambda_i)$ at wavelength range λ_i for the ground-cover projection of the measurement as

$$\rho(x, y, \lambda_i) = \frac{\pi \cdot I(\lambda_i)}{E_0(\lambda_i) \cdot cos\theta_0}. \qquad (5.1)$$

where $I(\lambda_i)$ and $E_0(\lambda_i)$ denote the measured earthshine backscattered radiance (in unit of [W m^{-2}nm^{-1}sr^{-1}]) and solar irradiance (in unit [W m^{-2}nm^{-1}]) spectra respectively, and θ_0 denotes the solar zenith angle. The reflectances used in this algorithm are derived from broad-band measurements of backscattered radiance and extra-terrestrial solar irradiance covering the spectral range of the Red-Green-Blue (RGB) colour system. For the GOME instrument, reflectances were generated from the three sub-pixel polarization measurement devices (PMDs) measuring at a relatively high spatial resolution (ground pixel size 20 x 40 km^2). GOME PMD measurements are directly mapped to the RGB colour space. The three PMDs span the UV (295-397 nm), the visible spectral range 397-580 nm and the range 580-745 nm extending into the NIR. SCIAMACHY has PMDs covering a spectral range similar to GOME, while for GOME-2, there are 15 PMDs (with varying band-widths) per polarization direction that are integrated to cover the required RGB spectral range (568-804nm, 400-556nm, and 322-384nm).

For the OMI instrument, the reflectances are mapped only to the GB colour space because the spectral coverage of this instrument ends at 500 nm. For S5P, the reflectances measured by the instrument will be mapped to the GB color space by integrating the data from the TROPOMI detector UVIS bands 3 and 4 (320-405 nm and 405-500 nm).

For the off-line creation of the cloud-free reflectance composites in the RGB case, the RGB reflectances are translated into normalized rg-colour space via the relations

$$r = \frac{\rho(x,y,\lambda_R)}{\sum_{i=RGB} \rho(x,y,\lambda_i)}, g = \frac{\rho(x,y,\lambda_G)}{\sum_{i=RGB} \rho(x,y,\lambda_i)}. \qquad (5.2)$$

The wavelength range covered by λ_R is 700–715 nm (band 5), λ_G is 410–495 nm (band 4), and λ_B is 356–390 nm (band 3). The wavelength range for red is selected based on the absence of strong absorption lines and the wavelength ranges for blue and green are adjusted to be similar to those used for OMI

If M is the set of n normalized multi-temporal measurements over the same location (x,y), then a cloud-free (or minimum cloudiness) pixel $rg_{CF} \in \mathrm{M}$ is selected using the brightness criterion $\|rg_{CF} - W\| \geq \|rg_k - W\|$, for $k = 1, \cdots n$, where $W = \left(\frac{1}{3}, \frac{1}{3}\right)$ is the *white point* in the rg-chromaticity diagram. Measurements under cloud conditions are projected to the *white point*, therefore the measurement that is most distant from *W* is considered to be cloud-free. A global cloud-free composite is constructed by merging cloud-free reflectances $\rho_{CF}(\lambda_i)$ (corresponding to rg_{CF}) at all locations.

The radiometric cloud fraction f_c is determined by examining separations between measured RGB reflectances and their corresponding cloud-free composite values:

$$f_c = min\left\{1, \sqrt{\sum\nolimits_{i=RGB} \alpha(\lambda_i)\max\{0, [\rho(\lambda_i) - \rho_{CF}(\lambda_i) - \beta(\lambda_i)]\}^2}\right\}. \quad (5.3)$$

This equation basically computes the distance between actual measurements and the corresponding cloud-free scene. Scaling factors $\alpha(\lambda_{i=RGB})$ define the upper limit for reflectances under fully cloudy conditions, while offsets $\beta(\lambda_{i=RGB})$ account for aerosol and other radiative effects in the atmosphere and as a lower limit basically define the cloud free conditions. The $max\{\}$ and $min\{\}$ functions ensure that the cloud fraction is mapped to the interval [0, 1].

The scaling and offset factors are determined off-line using representative global daily satellite measurements. The offsets are the histogram modes from the difference $\{\rho(\lambda_i) - \rho_{CF}(\lambda_i)\}$ and the scaling factors are inversely proportional to the 99th percentile of the cumulative histograms from the differences $\{\rho(\lambda_i) - \rho_{CF}(\lambda_i)\}^2$.

A detailed description of the OCRA algorithm and its application to satellite data is given in [*Loyola*, 2000]. A summary flow chart of the algorithm is given in Figure 5.1.

OCRA was validated by comparing cloud fractions to values derived from collocated measurements from the ATSR-2 instrument (like GOME, this sensor was on board ERS-2). Comparisons were also made with the FRESCO algorithm results. This ATSR-2 comparison confirms the excellent OCRA results reported in [*Tuinder et al*., 2004], where several algorithms for retrieving cloud fraction using GOME data were compared against synoptic surface observations. Similar results were obtained by OCRA applied to SCIAMACHY and compared with MERIS (both instruments on board ENVISAT) and OCRA applied to GOME-2 and compared with AVHRR (on the MetOp-A platform). See Chapter 8 for more details on the algorithm validation.

A cloud-free composite for every month as well as the offset and scaling factors based initially on reflectances from the OMI (colors B, G) instrument will be used during the S5P commissioning phase. Seasonally dependent RGB cloud-free composites will be created from existing TROPOMI/S5P measurements and dynamically updated as the mission proceeds.

5.3 S5P_CLOUD_ROCINN for cloud height, albedo and optical thickness

ROCINN is based on the comparison of measured and simulated radiances in and near the O_2 *A*-band. The sun-normalized radiance $R(\lambda)$ at wavelength λ is defined as

$$R(\lambda) = \frac{I(\lambda)}{E_0(\lambda)}. \tag{5.4}$$

where $I(\lambda)$ and $E_0(\lambda)$ denote the measured earthshine backscattered radiance (in unit of [W $m^{-2}nm^{-1}sr^{-1}$]) and solar irradiance (in unit [W $m^{-2}nm^{-1}$]) spectra respectively.

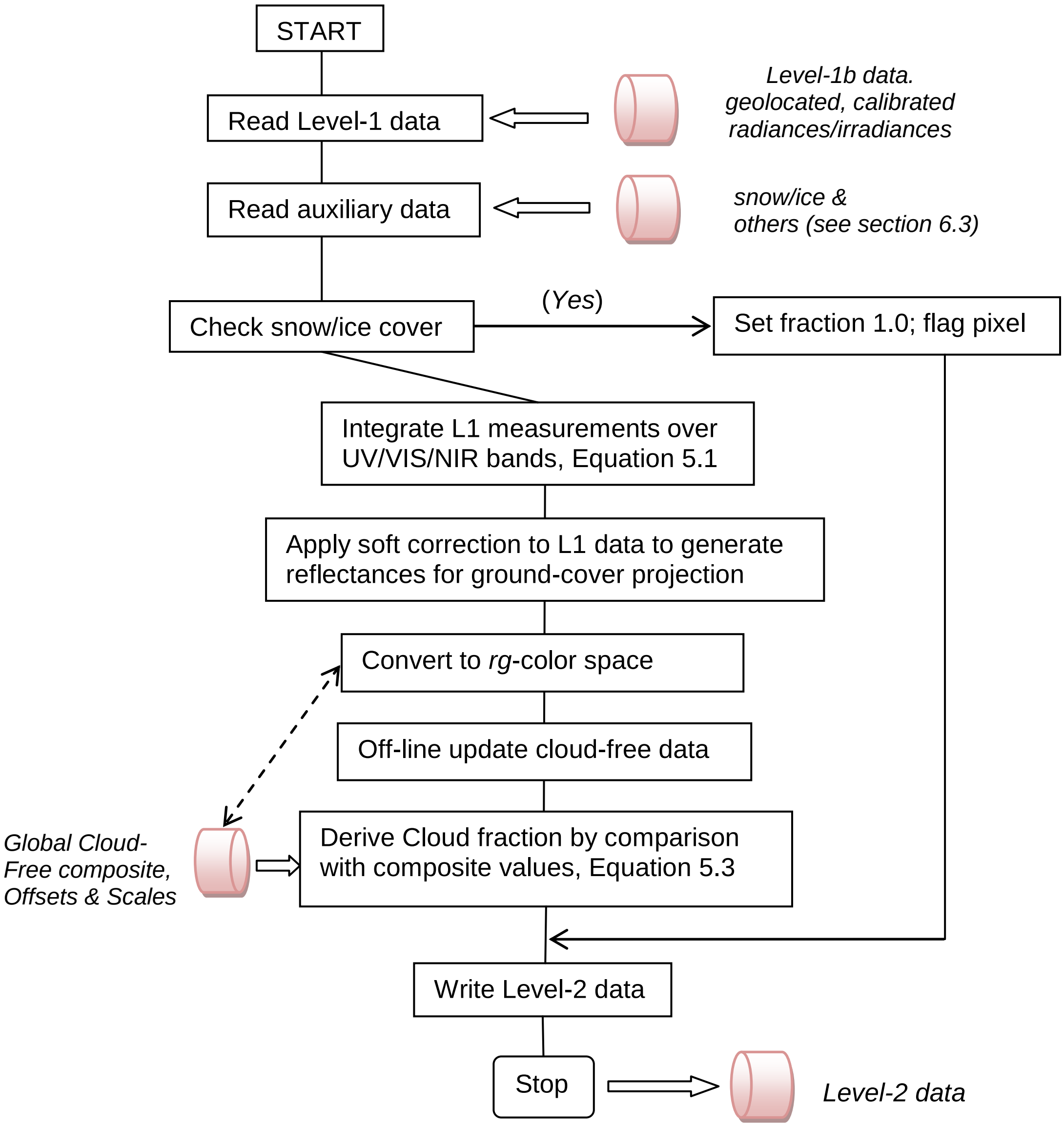


Figure 5.1: Flow Diagram for the S5P_CLOUD_OCRA (OCRA) retrieval algorithm.

5.3.1 ROCINN with CAL

For ROCINN with CAL (Clouds-As-Layers), the total sun-normalized radiance is taken to be a weighted sum of independent radiances from the surface R_s and cloud-top R_c^{CAL}, with the weighting expressed through the radiometric cloud fraction f_c. Both radiance contributions are calculated using standard one-dimensional radiative transfer models.

The sun-normalized radiance for a cloudy scene is calculated with the cloud treated as a set of contiguous scattering layers with geometrical extent characterized by cloud top height Z_{ct} and cloud-base height Z_{cb} (or alternatively the cloud geometrical thickness $H_c = Z_{ct} - Z_{cb}$).

The entire cloud is optically uniform with cloud optical thickness τ_c and its scattering properties are determined through Mie-scattering calculations for water droplet particles (microphysical properties are discussed below). In the IPA, we may write sun-normalized CAL simulated radiances R_{sim}^{CAL} as:

$$R_{sim}^{CAL}(\lambda) = f_c R_c^{CAL}(\lambda, \Theta, \tau_c, Z_{ct}, Z_{cb}, A_s, Z_s) + (1 - f_c) R_s(\lambda, \Theta, A_s, Z_s). \qquad (5.5)$$

Here, Θ denotes path geometry (solar and line-of-sight angles), with surface properties being the albedo A_s and lower boundary height Z_s.

Radiances for clear-sky and cloudy scenarios are calculated using the VLIDORT radiative transfer (RT) code [*Spurr*, 2006], at wavelengths in and adjacent to the O_2 *A*-band for sensors such as GOME/ERS-2, GOME-2/MetOp, SCIAMACHY/ENVISAT and now, TROPOMI/S5P. Details of the RT calculations are given in section 5.3.3 below.

A complete data set of simulated sun-normalized radiance templates is created off-line for an appropriate range of viewing/solar geometries and surface geophysical scenarios, and for various combinations of cloud properties.

The inverse problem uses least-squares fitting with a generalized form of Tikhonov regularization (details in section 5.3.4). Retrieval in the O_2 *A*-band with the 4-element state vector $\{\tau_c, Z_{ct}, Z_{cb}, f_c\}$ is an ill-posed problem that requires additional information in order to obtain an inverse solution, as there are only two degrees-of-freedom-of-signal [*Schuessler et al*., 2014]. For ROCINN-CAL, the retrieval state vector is just $\{\tau_c, Z_{ct}\}$ for cloud optical thickness τ_c and height Z_{ct}, and the cloud fraction f_c is taken from OCRA. The cloud base height Z_{cb} (or equivalently, the cloud geometrical thickness) is then taken from a suitable cloud-climatology or cloud-parameterization scheme.

Figure 5.2 presents the flow diagram for this algorithm. The granularity is one orbit. Following the ingestion of Level-1b and auxiliary data, and the OCRA cloud fraction results, the algorithm enters an iteration loop; at each step, the forward simulation (essentially a look-up table process) is followed by the inverse model, and the iteration ceases once a suitable convergence criterion has been satisfied.

5.3.2 ROCINN with CRB

ROCINN Version 2.0 uses the CRB treatment, with the cloud-top assumed to be a Lambertian reflector. The sun-normalized CRB simulated radiances R_{sim}^{CRB} are defined as:

$$R_{sim}^{CRB}(\lambda) = f_c R_c(\lambda, \Theta, A_c, Z_c, A_s, Z_s) + (1 - f_c) R_s(\lambda, \Theta, A_s, Z_s). \qquad (5.6)$$

ROCINN retrieves cloud-top albedo A_c and height Z_c in the IPA framework. The radiometric cloud fraction f_c is again from OCRA.

As noted already, the current operational GOME-2 cloud products are based on a combination of OCRA and ROCINN-CRB, and these cloud parameters are the ones that are used in the GOME-2 trace gas retrievals. ROCINN cloud parameters based on the CRB model will also be provided for S5P for trace gas retrieval algorithms that are based on the use of the CRB cloud model.

5.3.3 Forward model for sun-normalized radiance templates

ROCINN is based on a look-up table (LUT) of simulated sun-normalized radiances at wavelengths in and around the O_2 *A*-band. In earlier versions of the algorithm, these radiance templates were calculated with the CRB assumption. An initial LUT of sun-normalized radiance templates has been constructed for ROCINN with CAL, with cloudy-scene sun-normalized radiances R_c^{CAL}calculated for a full atmosphere including cloud-scattering in and below the cloud itself. Mie scattering was used to generate cloud optical properties, and a new classification scheme was developed for these templates (described below). Details may be found in [*Schuessler et al*., 2014].

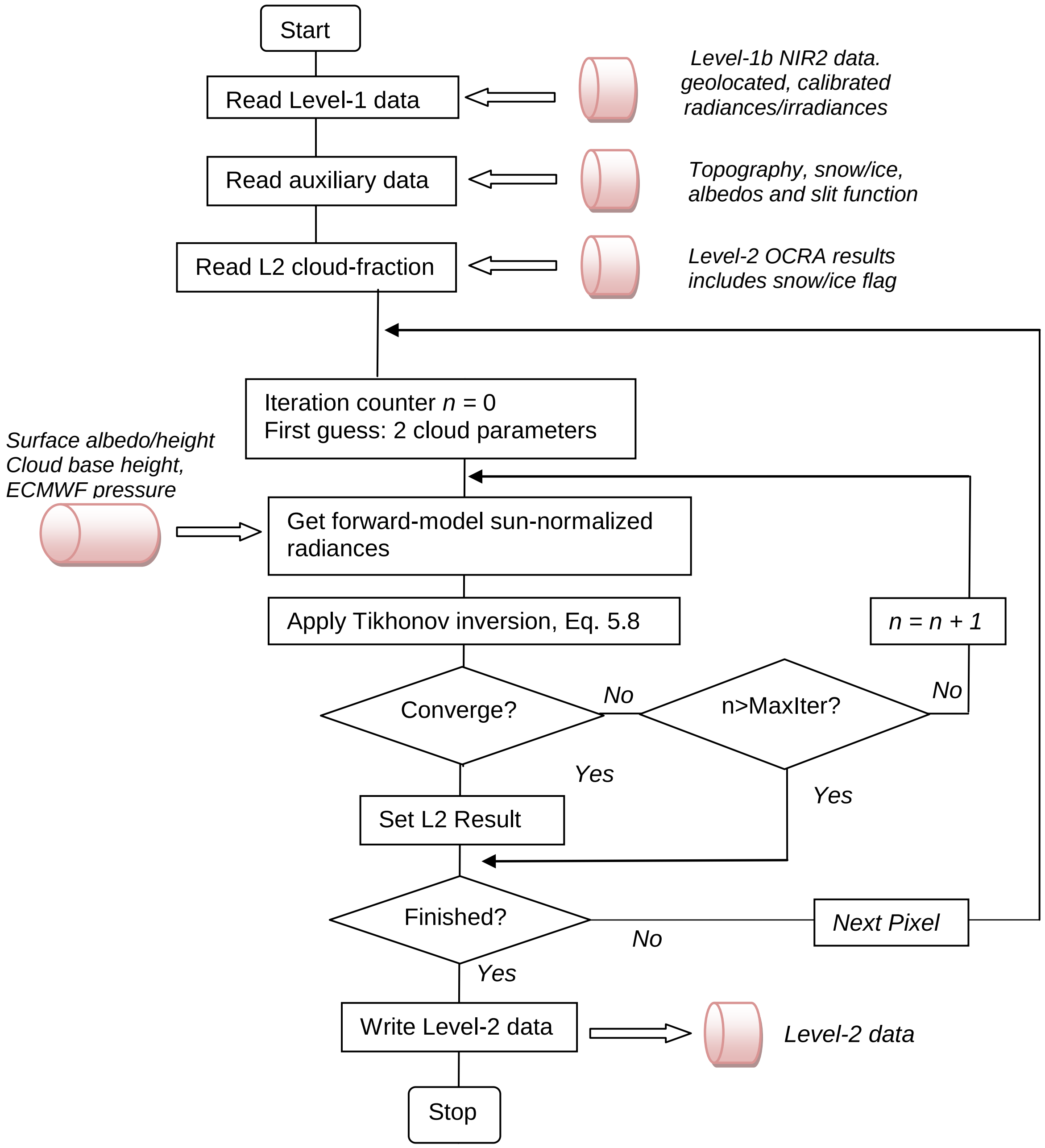


Figure 5.2: Flow Diagram for the S5P_CLOUD_ROCINN (CAL) retrieval algorithm.

Simulated sun-normalized radiances $R_{sim}(\lambda)$ are calculated using the vector (with polarization) VLIDORT multiple scattering multi-layer discrete ordinate RT model [*Spurr*, 2006]. This calculation is based on clear sky optical properties for line absorption by oxygen and Rayleigh scattering by air molecules. The desired total intensity *I* will incorporate the effects of polarization. VLIDORT uses a precise calculation of the single-scattering radiation field in a spherically-curved atmosphere. It should be noted here that VLIDORT is fully compatible with its scalar counterpart LIDORT (no polarization) which is used "on-the-fly" in the operational total ozone algorithms [RD3]. We discuss this issue in a little more detail in section 5.4.

For the line absorption, it is necessary to calculate line-by-line (LBL) radiances (typically at resolution 0.0025 wave number for the range 758-771 nm) using line-spectroscopic information for the O_2 *A*-band, before convolution with the sensor slit function.

In the first instance, spectroscopic data is taken from the HITRAN 2008 database [*Rothman et al.*, 2009], but subsequently, this information will come from the HITRAN 2012 database (released in June 2013). Absorption cross-sections are computed using LBL software from DLR [*Schreier and Schimpf*, 2001; *Schreier*, 2011], in which line absorption signatures are accurately modelled with the Voigt profile.

For Mie scattering calculations, we require microphysical optical properties but using water or ice properties have a relative small impact on the O_2 *A*-band spectral region. Note that consistency of models to be used in cloud and trace gas retrievals is far more critical than the optical properties selected for the RTM simulations.

The refractive index is 1.33 for water droplet clouds (no absorption). Droplets are assumed to be poly-dispersed according to the modified-Gamma size distribution function:

$$n(r) = Cr^{-\alpha}\exp\left[-\frac{\alpha}{\gamma}\left(\frac{r}{r_c}\right)^{\gamma}\right], \quad (5.7)$$

which is parameterized by the mode radius r_c in [μm] and constants α and γ describing the shape of the distribution (scheme follows that in [*Hess et al.*, 1998]). In Eq. (5.7), C is the normalization constant..

The cloud *macro-physical* properties (classifications of cloud-top height and cloud geometrical thickness) are based on the tables in [*Wang et al*., 2000]. Values of cloud-top height range from 0.5 km to as high as 14 km for some cloud types, with cloud optical thickness values ranging from 0 (clear sky) to 125 (typically 7 values in total), and cloud geometrical thickness values ranging from 0.5 km to 3.5 km at 0.5 km intervals. Over land, a small number of surface albedos (three values) and heights (four values) are used, while the ocean case is treated separately. Details of this algorithm prototype may be found in [*Schuessler et al*., 2014].

In addition to these surface and cloud parameter classifications, it is also necessary to generate LUT entries for a range of viewing and solar angles. Typically, we choose some 11 solar zenith angles in the range 15-88°, 15 viewing zenith angles from 0° to 70° at 5° intervals, and 5 relative azimuth angles from 0° to 180° at 45° intervals.

There are many millions of forward model calculations required in order to prepare the complete LUT of templates. This process is done off-line and normally takes many weeks to complete. In the S5P operational environment, the LUT extraction of forward-model sun-normalized radiances will be done by fast neural-network methods, and there will be no problems with data turn-over in ROCINN 3.0.

5.3.4 Details of the inverse model

If $\mathbf{x}$ is the state vector $\{\tau_c, Z_{ct}, Z_{cb}, f_c\}$ comprising possible cloud parameters for retrieval, and $\mathbf{b}$ denotes the vector of auxiliary forward-model parameters (surface properties, viewing geometry), we write the measurement vector as $\mathbf{y}^{\delta} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\delta}$, where $\mathbf{F}$ is the forward model and $\boldsymbol{\delta}$ is the data error vector. The inverse problem defined by this equation is nonlinear and ill-posed, and regularization is required in order to obtain a solution with physical meaning. The degree to which the problem is ill-posed is partly characterized by the condition number $c(\mathbf{K}) = \gamma_{max}/\gamma_{min}$ of the Jacobian matrix $\mathbf{K} = \mathrm{d}\mathbf{F}/\mathrm{d}\mathbf{x}$, where γ_{max} and γ_{min} are the largest and the smallest singular values of $\mathbf{K}$, respectively.

In the form of Tikhonov regularization that we use here, the regularized solution $\mathbf{x}_{\alpha}^{\delta}$ minimizes the objective functional

$$\mathfrak{F}_{\alpha}(\mathbf{x},\mathbf{b}) = \frac{1}{2}\{\|\mathbf{F}(\mathbf{x},\mathbf{b}) - \mathbf{y}\|^2 + \alpha\|\mathbf{L}(\mathbf{x}-\mathbf{x}_a)\|^{\mathbf{2}}\}. \quad (5.8)$$

Here, α denotes the regularization parameter, and $\mathbf{L}$ is the regularization matrix [*Doicu et al.*, 2010]. The function is assumed to be defined with the L_2 Euclidean norm. The minimizer for Eq. (5.8) can be computed with Gauss-Newton methods.

In statistical inversion theory, the Bayesian approach or the optimal estimation method can be regarded as a stochastic version of Tikhonov regularization. The maximum *a posteriori* solution coincides with the Tikhonov solution when the state vector $\mathbf{x}$ and the noise vector $\boldsymbol{\delta}$ are Gaussian random vectors with covariance matrices $\mathbf{C}_{\mathbf{x}} = \sigma_x^2\mathbf{I}_n$ and $\mathbf{C}_{\boldsymbol{\delta}} = \sigma^2\mathbf{I}_m$ respectively, where σ_x and σ are the corresponding standard deviations. In this case, the regularization parameter α is the ratio of these two variances, that is, $\alpha = \sigma^2/\sigma_x^2$.

As noted above, the operational ROCINN-CAL algorithm will be a 2-parameter retrieval of cloud optical thickness and cloud-top height, but the inverse framework is general enough to allow for other options.

Convergence is reached when the residual $\|\mathrm{F}(\mathrm{x},\mathrm{b}) - \mathrm{y}\|^2$ or the delta changes in the retrieved parameters Δ_{x} are smaller than pre-defined values (defaults 5E-3 and 5E-5 respectively), or the maximum number of iterations (default 50) is reached. The default value for the regularization parameter α is 1E-4.

5.4 Cloud co-registration inhomogeneity flag

The S5P cloud retrieval needs Level 1 data from different bands: OCRA cloud fractions (CF) uses broadband reflectances from bands 3 and 4 while the ROCINN algorithm uses the oxygen *A-band* information in band 6.

The combination of information from different spectral bands is not trivial since the spatial region covered by the ground pixels from different spectral bands do not match exactly. The problem is well explained in the dedicated technical note [RD8]. A possible method for combining information from different bands is by means of a co-registration table containing the fraction of overlapping area between the source and target pixels. In case of combinations between bands 3, 4, 5 and 6 a static co-registration table suffices. However, this method implies a smoothing of the source band product. For this reason, a cloud co-registration inhomogeneity flag (CCIF) has been created and the Level 2 products combining data from different bands are only processed when the CCIF is false.

The CCIF relies on a cloud co-registration inhomogeneity parameter (CCIP) which is defined as the weighted averaged gradient of cloud fractions

$$CCIP_j = \frac{\sum_i f_{ij}\,|CF_i - CF_j|}{\sum_i f_{ij}}. \quad (5.9)$$

where the weights f_{ij} correspond to the co-registration values between bands 4 (source) and 6 (target), i is the source index and j the target index.

The CCIF is defined as:

$$CCIF_j = CCIP_j > p\,. \quad (5.10)$$

where p is a fixed threshold which has been set to 0.35.

Figure 5.3 shows an example of the cloud co-registration inhomogeneity flag (CCIF) using the Suomi-NPP (VIIRS) cloud product resampled to the S5P-TROPOMI spatial grid. It can be seen that the CCIF is only true at the transition from cloud to cloud-free regions where the gradient between neighboring pixels is higher.

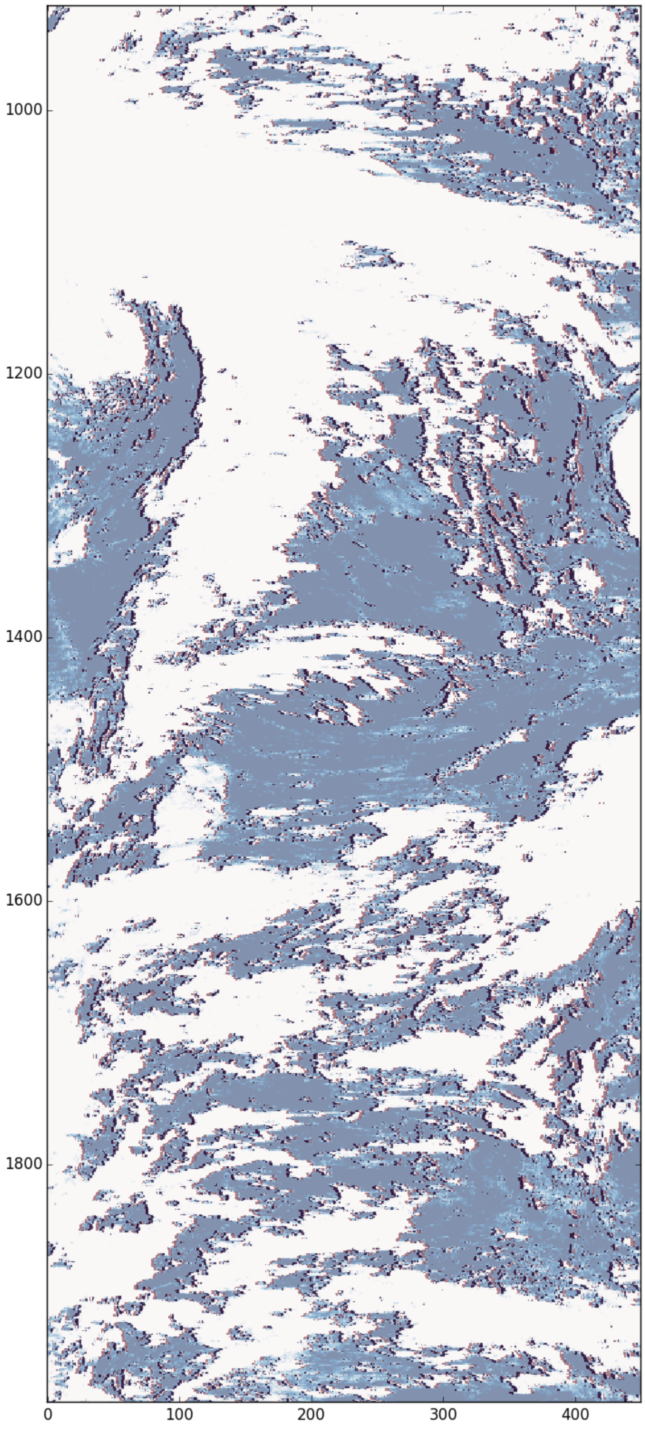


Figure 5.3: Example of the cloud co-registration inhomogeneity flag (CCIF) using the Suomi-NPP (VIIRS) cloud product regridded to the S5P-TROPOMI spatial sampling. CCIF have been overplotted (with a reddish colour map) to the cloud fraction (blueish colour map). It can be seen that the CCIF is only true at the edges of the cloud systems where the cloud fraction fields are more variable. VIIRS test data correspond to 2013.11.03 kindly provided by Richard Siddans, RAL.

5.5 Consistency of OCRA and ROCINN cloud parameters

The adequacy of the cloud parameters from OCRA/ROCINN for air mass factor calculations for trace gas retrievals had been demonstrated and published [Van Roozendael et al., 2006, Loyola et al., 2011]. OCRA/ROCINN are being used over 10 years in the operational L2 products from GOME and GOME- 2 under all kind of geophysical conditions (including low cloud fraction and/or thin clouds) and the geophysical validation shows that OCRA/ROCINN performs better than other GOME algorithms, see comparisons with synoptic data [Tuinder et al., 2004] and satellite data like MSG [Loyola et al., 2007], ATSR- 2 [Rozanov et al., 2007], and ISCCP [Loyola et al., 2010].

5.6 Application to trace gas retrievals

In this section, we discuss the usage of the S5P cloud algorithms, when they are called upon to generate cloud information required for reflectance simulations in trace gas retrievals. For the latter, we confine attention to the two S5P total ozone algorithms: S5P_TO3_DOAS, a DOAS-style NRT algorithm based on GDP4.x, and S5P_TO3_GODFIT, an off-line direct fitting algorithm based on GDP5. Please refer to the S5P Total Ozone ATBD for details [RD3].

Both total ozone algorithms use the Independent Pixel Approximation (IPA) in which, the total radiance I_{total} for a partially cloudy scene is modelled as a linearly-weighted combination of separate radiances for clear sky and full cloud: $I_{total} = (1 - f_c)I_{clear} + f_c I_{cloud}$. In both ozone algorithms, S5P_CLOUD_OCRA (OCRA heritage) will provide the radiometric cloud fraction f_c.

For ROCINN-CRB cloud parameters, RT simulations in the UV O_3 Huggins bands must assume clouds to be perfect reflectors situated at height Z_c with Lambertian albedo A_c. This has the effect of introducing an elevated surface boundary in the RT simulation. For RT simulations based on ROCINN-CAL inputs, additional layers must be inserted in the basic atmospheric pressure grid according to the specification of cloud geometrical extent, that is, the values of $[B_c, Z_c]$. These cloud boundaries have the effect of introducing two additional layers in the vertical grid, and the resulting atmosphere is used for both clear-sky and cloudy simulations. In addition, UV RT modelling requires the cloud optical thickness τ_c^{O3} (which is the ROCINN τ_c result translated to UV wavelengths). In the O_3 algorithms, cloud droplets are uniformly distributed, with optical properties pre-calculated from a Mie-scattering simulation that uses the same particle size distributions as those selected for the ROCINN-CAL template calculations.

Finally, we note that both TROPOMI/S5P total ozone algorithms use the scalar LIDORT discrete-ordinate RT model which is called "on-the-fly'" to generate the required simulated earthshine radiances and (for the GODFIT algorithm) the necessary analytic Jacobians.

Polarization signatures in the UV window (325-335 nm) are in general smoothly varying, and have traditionally been absorbed in the closure filters that are normally applied to forward model output in this UV window. However, the most recent GODFIT version now has a look-up table (LUT) of polarization corrections (relative differences in earthshine intensity with and without polarization included in VLIDORT), which are used to adjust the "on-the-fly" LIDORT scalar output. The use of a polarization correction LUT is also under investigation for AMF computations in the DOAS total ozone algorithm for TROPOMI/S5P.

VLIDORT is much slower than its counterpart LIDORT, and this is the main reason why VLIDORT is not used operationally in the ozone algorithms. With the ROCINN algorithm, it is perfectly possible to use VLIDORT from the start, since the sun-normalized radiance templates are calculated offline, and there is no data turn-over constraint. Polarization effects are generally small in the NIR, and sun-normalized radiance templates have in the first instance been calculated in scalar mode. With VLIDORT as the forward model, we are in a good position to assess the effect of forward model error on cloud property retrieval due to the neglect of polarization in the modelling. For more on the error analysis for ROCINN, see section 7.

5.6.1 Use of S5P_cloud information in the S5P_TO3_DOAS algorithm

S5P_TO3_DOAS has heritage from the GDP4 DOAS algorithm [RD3]. Determinations of the total ozone vertical column density (VCD) are based on the IPA combination of clear-sky and cloudy-scene Air Mass Factors (AMFs), with the “intensity-weighted cloud fraction” $\Phi = f_c\, I_{cloud}/I_{total}$ used in place of the OCRA-derived value f_c.

We consider first the CRB case. For a fully cloudy scene, that part of the ozone column lying below the effective cloud top height Z_c cannot be detected from a space-borne instrument if the optical thickness of the cloud is large enough. This is the “ghost column" G. For an opaque cloud, we have $G = V_{bc}$ (climatology ozone column below cloud-top). However, satellite measurements are sensitive to intra-cloud ozone, and neglecting this effect in the retrieval may give rise to significant error in the simulated backscatter signal, thereby inducing appreciable error in the ozone product. For ROCINN-CRB output, rather than ignoring intra-cloud ozone, we can use an empirical intra-cloud correction designed for the GDP4.x DOAS scheme [*Loyola et al.*, 2011]. The intra-cloud ozone column V_{ic} is characterized empirically using the simple linear formula $V_{ic} = V_{bc}(1 - A_c)\cos\theta_0$, where A_c is the cloud albedo and θ_0 the solar zenith angle (SZA). A more accurate value of the ghost column is then $G = V_{bc} - V_{ic}$. Over snow and ice conditions, ROCINN-CRB is configured to retrieve effective cloud-top height and albedo values for the underlying scene - the cloud fraction is assumed to be 1.0, and there is no ghost column.

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

For operational processing of GOME, GOME-2 and SCIAMACHY data, only the ROCINN-CRB method has been used to date. However, total ozone processing has already been done using ROCINN-CAL for GOME and GOME-2 in a prototype environment at DLR.

For GDP4 total ozone applied to GOME data, an initial attempt was made in 2008 to generate CAL templates using the GDP 4 classification system based on cloud-top albedos, with values of cloud optical thickness values obtained using a separate program linking albedos to cloud optical thickness [*Loyola et al*., 2010]. In summer 2009, GOME and GOME-2 total ozone products were computed using GDP4 with the CAL layering scheme, with cloud information coming from OCRA and the ROCINN prototype based on these CAL templates. Several years of GOME data were processed in this manner and given a first validation.

In general, CAL-derived ozone columns were found to be at least as good as CRB-derived values, with better validation (against Brewer and Dobson networks) for solar zenith angles in the range 60-85°.

5.6.2 Use of S5P_cloud information in the S5P_TO3_GODFIT algorithm

S5P_TO3_GODFIT is the GDP5-based direct fitting algorithm. For cloudy-scene intensity simulations based on ROCINN-CRB information, we still have the problem of total column overestimation. However, we cannot use the intra-cloud correction as described in section 5.6.1. Instead, we modify the OCRA/ROCINN-CRB output to transform optically thin clouds into equivalent optically thick clouds of reduced geometrical extent. If X and Y are retrieved cloud-top albedo and cloud fraction, then we define *effective* values as follows: $X^* = X,\ Y^* = Y$ for $X > 0.8$, and $X^* = 0.8,\ Y^* = XY/0.8$ for $X \leq 0.8$. In "snow/ice mode", the cloud fraction f_c is again set to 1.0, and ROCINN-CRB retrieves the effective scene albedo and reflecting surface height.

In the ROCINN-CRB case, the use of internal closure in the forward model becomes problematic for strongly cloud-contaminated scenes ($f_c > 0.85$). In this case the algorithm ignores any surface contributions, treating the scene as fully cloudy ($f_c = 1$) with the cloud-top treated as a reflecting surface characterized by the internal albedo closure. Consequently, the ROCINN albedo value A_c is ignored, and the actual cloud-top albedo is then retrieved as an ancillary state vector element in the total ozone inversion.

For S5P, the cloud algorithms will be OCRA and ROCINN-CAL. As noted already, RT simulations in the Huggins bands will be done for an otherwise-Rayleigh atmosphere with additional layers specified by geometrical parameters $[B_c, Z_c]$ encompassing an optically uniform cloud of optical thickness τ_c^{O3} (the ROCINN value translated to the UV), and scattering properties determined from off-line Mie calculations characterized by pre-set microphysical parameters. In this case, intra-cloud ozone is modelled properly and there is no need for an empirical correction. In addition, both clear-sky and cloudy-scene simulations include surface reflection effects, so there will be less of problem with the internal closure.

5.7 Processing Flags and QA Values

All processing errors and warnings are collected in the processing_quality_flags. A detailed description can be found in the Product User Manual (PUM) for the Cloud Properties [RD9].

For a retrieval not affected by any processing error or warning, the qa_value is assumed to be 1.0 (i.e. highest possible product quality). Each potential processing error or warning will reduce the product quality to result in qa_values below 1.0. Currently, the following processing warnings are translated to the quality assurance value for the CAL (qa_value) and CRB (qa_value_crb) cloud properties:
saturation_warning, input_spectrum_warning, high_sza_warning, cloud_inhomogeneity_warning, cloud_retrieval_warning, low_cloud_fraction_warning and sun_glint_warning.

Each of these warnings has a different impact on the product quality since some warnings are probably less critical while others are more critical. Currently, a very minor quality reduction of 0.05 is applied for the sun_glint_warning. A minor quality reduction of 0.10 is applied for the low_cloud_fraction_warning. A medium quality reduction of 0.50 is applied for the input_spectrum_warning, the cloud_retrieval_warning and the cloud_inhomogeneity_warning. Finally, a large quality reduction of 0.75 is applied for the saturation_warning.

A dynamic quality reduction between 0.0 and 1.0 is performed for the high_sza_warning depending on the actual SZA. The reduction starts with 0.0 at 75° and increases to 1.0 at 89°. For values above 89°, the warning is changed to an sza_range_error, the qa_value is set to 0.0 and the cloud retrieval is aborted.

A final setting and refinement for the contribution of the several processing errors and warnings to the qa_value will be done once a significantly large amount of TROPOMI data will be available in order to conduct an in-depth statistical analysis.

6 Feasibility

6.1 Estimated computational effort

The Sentinel 5P sensor TROPOMI samples the Earth's surface with an unprecedented spatial resolution of 7x3.5 km^2 at nadir. Although this sampling allows for the resolution of fine details in the observed products, it poses additional demands on the retrieval code regarding computational speed.

From the outset, OCRA and ROCINN have been fast-performance algorithms, with no data turnover issues - both algorithms are ready for NRT. It has always been straightforward to implement the two algorithms, since both are based on pre-computed quantities. For OCRA (S5P_CLOUD_OCRA), dynamic updating of the cloud-free composite is not time-consuming. For ROCINN-CAL (S5P_CLOUD_ROCINN), the retrieval is fast - simulations of sun-normalized radiance templates are done offline.

The S5P Level-1b data flow is expected to deliver spectral measurements with a size of 40 GB per orbit. In order to estimate the computational effort, extrapolations were made on the basis of 1,500,000 spectra for one full orbit of TROPOMI/S5P using the GOME-2 computational performance on one 2.7 GHz CPU.

- S5P_CLOUD_OCRA: OCRA for fractional cover
    - ~9 minutes/TROPOMI orbit
- S5P_CLOUD_ROCINN: ROCINN for cloud height and optical thickness
    - ~65 minutes/TROPOMI orbit

The performance figures were obtained using a Dell Precision T7500. The benchmark information for this hardware can found at:

http://www.spec.org/cpu2006/results/res2009q2/cpu2006-20090316-06712.html

6.2 S5P cloud product description and size

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

- measurement time and geolocation, taken from the Level-1b product;
- cloud fraction, cloud height/pressure, cloud optical-thickness/albedo and corresponding errors;
- climatology and other relevant parameters used in the retrieval;
- fit results (RMS, etc.);
- processing quality flags and quality assurance value (qa_value).

Table 6.1 lists the output fields that are required in the cloud level-2 files based on the S5P_CLOUD_OCRA and S5P_CLOUD_ROCINN algorithms.

The estimated product size for one orbit is 300 MB.

Table 6.1: List of output fields required in the cloud level-2 product generated with the S5P_CLOUD_OCRA and S5P_CLOUD_ROCINN algorithms.

Name/Data	Symbol	Unit	Description	Data type per pixel	Dimension
Number of measurements	N	---	Number of measurements included in the file. N = nAlong x nAcross	Integer	1
Orbit number	n_0	---	Satellite orbit number	integer	1
Time	---	---	Date and time of measurement [YYMMDDHHMMSS.MS]	character	1
Latitudes	*lat*	degree	Latitudes of the pixel center and corners	float	5
Longitudes	*lon*	degree	Longitudes of the pixel center and corners	float	5
SZA	θ_0	degree	Solar zenith angle at pixel center	float	1
LoSZA	θ	degree	Viewing zenith angle at pixel center	float	1
RAA	φ	degree	Relative azimuth angle at pixel center	float	1
CFR	f_c	---	Radiometric cloud fraction	float	1
CTH*	Z_{ct}	m	Cloud top height	float	1
CTP*	p_{ct}	Pa	Cloud top pressure	float	1
CBH	Z_{cb}	m	Cloud base height	float	1
CBP	p_{cb}	Pa	Cloud base pressure	float	1
CTA*	A_c	---	Cloud top albedo	float	1
COT*	τ_c	---	Cloud optical thickness	float	1
Surface albedo	A_s	---	Surface albedo	float	1
Surface height	Z_s	m	Surface height	float	1
OCRA Reflectances	$\rho(\lambda_{i=RGB})$	sr^{-1}	OCRA reflectance at wavelength ranges R, G, B	float	3
Snow/ice flag	---	---	Snow/ice flag	---	1

6.3 Auxiliary information needs

All auxiliary data in this section are static inputs. The TROPOMI slit function information will be taken from pre-flight calibration data.

* Two output parameters will be provided. One from ROCINN CAL and one from ROCINN CRB

OCRA is a threshold algorithm using broad-band (integrated) measurements from the UVIS and NIR bands. The major auxiliary data set is the global cloud-free composite of minimum reflectances - this is a dynamic data set which will be augmented with the measurements themselves as the mission proceeds. As noted in section 5.2, a previously-developed cloud-free composite data set from OMI will be used during the commissioning phase in order to start cloud fraction processing with OCRA. In this sense then, this OMI composite can be regarded as an initial static auxiliary data set.

For the ROCINN algorithm, we require some external data for surface properties (albedo, surface height/pressure). The snow/ice data set is used to flag snow-cover scenes. Terrain height and snow/ice data sets are the same as those used for the total ozone algorithms [RD3], though the surface albedo data set is different - the MERIS data at 760 nm (linearly interpolated between 754 and 775 nm) is more appropriate for the NIR. Harmonization on the format and usage of auxiliary data for the range of S5P L2 products (ATBDs, IODDs, etc.) is being determined within the framework of the L2WG (Level-2 Working Group).

Table 6.2: Dynamic input information needed in the cloud retrieval algorithms.

Name/Data	Symbol	Unit	Source	Pre-process needs	Backup if not available
S5P level 1B Earth radiance	*I*	W $m^{-2}nm^{-1}sr^{-1}$	S5P L1b product		No retrieval
S5P level 1B sun irradiance	E_0	W $m^{-2}nm^{-1}$	S5P L1b product	Wavelength recalibrated using a high-resolution reference solar spectrum	Use previous measurement
ECMWF pressure profiles	---	---	ECMWF forecast	Spatial grid: 1° ×1° or finer	Use fixed pressure profile
Snow/ice flag	---	---	Near real-time global Ice and Snow Extent (NISE) data from NASA.	---	Use TOMS snow/ice climatology (Tanskanen [2004])

Table 6.3: Static auxiliary information needed in the cloud retrieval algorithms.

Name/Data	Symbol	Unit	Source	Pre-process needs	Comments
Instrument slit function	*SF*	---	Slit function provided by wavelength/detector	---	---
High-resolution reference solar spectrum	E_s	W $m^{-2}nm^{-1}$	Chance and Kurucz [2010]	---	---
OCRA cloud-free composite	$\rho_{CF}(\lambda_i)$	sr^{-1}	OMI cloud-free composite reflectance	---	TROPOMI/S5P composite will be generated during the mission
OCRA scaling factors	$\alpha(\lambda_{i=RGB})$	sr^2	OMI measurements	---	---
OCRA offset factors	$\beta(\lambda_{i=RGB})$	sr^{-1}	OMI measurements	---	---
ROCINN regularization parameter	α	---	UPAS configuration file	---	Used in the Tikhonov inversion
ROCINN maximum iterations	---	---	UPAS configuration file	---	Maximum number of iterations
ROCINN convergence residual	---	---	UPAS configuration file	---	Convergence test
ROCINN convergence delta changes	Δ_x	---	UPAS configuration file	---	Convergence test
Surface albedo	A_s	---	MERIS black-sky albedo climatology (Popp et al. [2011])	Resolution: monthly on a 0.25° ×0.25° grid	Values at 760 nm will be used
Digital elevation map	Z_s	m	Same DEM for all L2 products	---	---

6.4 Level-1 requirements

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

As with the heritage algorithms, the main Level-1 measurement data sets for both S5P cloud algorithms are the Level-1b geolocated and calibrated backscatter UVIS and NIR Earthshine radiances and the Level-1b calibrated solar irradiances [RD4; RD5]. Radiances and irradiances are accompanied by error quantities in the same units. Wavelengths are calibrated values in [nm]. To ensure an accurate wavelength registration, a recalibration procedure, based on a cross-correlation with a high-resolution solar spectrum (e.g. [*Chance and Kurucz*, 2010]), is applied to both the radiance and irradiance measurement spectra [*Van Roozendael et al*., 2006].

Explicit calibration data are not required, other than the slit function parameters, which are regarded as static-input auxiliary data [RD6]. The radiance data is assumed corrected for polarization (TROPOMI uses a scrambler to remove polarization at the instrument level). The Level-1b product also comes with full geolocation information - solar and viewing zenith and azimuth angles at the bottom of the atmosphere (BOA) in addition to those at TOA and at the spacecraft; this set of angles was the specification for GOME and GOME-2 Level-1b geolocation, and this is expected to be the default for S5P Level-1 [RD5].

Level-1b requirements are different for the two cloud algorithms. OCRA needs broad-band integrated measurement data (both radiances and irradiances) from the TROPOMI/S5P UVIS detector, bands 3 and 4 (310-405 nm and 405-500 nm) and the two NIR bands 5 and 6 (675-725 nm and 725-775 nm). Thus, OCRA will digest virtually all the Level-1b spectral information from the visible and NIR TROPOMI/S5P detectors.

In contrast, the ROCINN algorithm needs radiance and irradiance measurements for a relatively narrow window in and around the oxygen *A*-band (TROPOMI/S5P NIR band 6), typically 758-772 nm.

Possible co-registration errors between the TROPOMI UVIS and NIR bands should be smaller than 10% in order to reach the requested accuracy of the S5P cloud products. Note that Level-1b geolocation mismatches between TROPOMI bands will have a direct impact on any retrieval algorithms using data from different bands.

7 Error analyses

7.1 General formulation and averaging kernels

7.1.1 Error classifications

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

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

where $\mathbf{A} = \mathbf{D}_y\mathbf{K}$, and $\mathbf{D}_y \equiv \mathbf{S}_x \cdot \mathbf{K}^{\mathrm{T}}\mathbf{S}_y^{-1} = [\mathbf{K}^{\mathrm{T}}\mathbf{S}_a^{-1}\mathbf{K} + \mathbf{S}_a^{-1}]^{-1} \cdot \mathbf{K}^{\mathrm{T}}\mathbf{S}_y^{-1}$. This result for the contribution (or gain) matrix $\mathbf{D}_y$ defines the solution error covariance matrix $\mathbf{S}_x$ (cf. Eq. (5.8)). Also, $\mathbf{K}$ is the weighting function matrix, superscript "T" denotes matrix transpose, $\mathbf{S}_y$ is the measurement error covariance matrix, with $\mathbf{X}_a$ the *a priori* state vector with random normally-distributed error covariance matrix $\mathbf{S}_a$. Matrix $\mathbf{A}$ contains the averaging kernels and is an indicator of the sensitivity of the retrieval to the true state. The second term in Eq. (7.1) is the smoothing error. The remaining error $\boldsymbol{\epsilon}$ can be divided into three components as follows.

<u>Measurement errors.</u> This error is $\boldsymbol{\epsilon}_{me} = \mathbf{D}_y\boldsymbol{\epsilon}_y$ which may be systematic or random. In the latter case the solution covariance contribution is $\mathbf{S}_{noise} = \mathbf{D}_y\mathbf{S}_y\mathbf{D}_y^{\mathrm{T}}$. Systematic errors in this category include stray light, slit function and radiometric calibration uncertainties.

<u>Model parameter errors.</u> The retrieval error due to this source of uncertainty is $\boldsymbol{\epsilon}_{me} = \mathbf{D}_y\mathbf{K}_b\boldsymbol{\Delta}\mathbf{b}$ where $\mathbf{K}_b$ is the sensitivity (Jacobian) of the forward model to a vector $\mathbf{b}$ of model parameters, and $\boldsymbol{\Delta}\mathbf{b}$ is the error on $\mathbf{b}$. If the error is random with covariance $\mathbf{S}_b$, then the associated solution covariance contribution is $\mathbf{S}_{param} = \mathbf{D}_y\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^{\mathrm{T}}\mathbf{D}_y^{\mathrm{T}}$. Estimation of these errors is greatly helped when the forward model is able to deliver Jacobians $\mathbf{K}_b$ in an efficient and accurate manner (LIDORT has this capability). Model parameters can be any atmospheric variables that are not fitted (cloud fraction and cloud top pressure, cross-section amplitudes, temperature profile entries, etc.).

<u>Forward model errors.</u> Here the retrieval error due to this source of uncertainty is given by $\boldsymbol{\epsilon}_{fwd} = \mathbf{D}_y\boldsymbol{\Delta}\mathbf{F}$, where $\boldsymbol{\Delta}\mathbf{F}$ is the forward model error due either to incorrect physical assumptions in the RT model (e.g. neglect of polarization, omission of rotational Raman scattering) or to a certain level of mathematical approximation (number of stratifications, number of discrete ordinates in the diffuse scattering quadrature approximation, plane-parallel scattering, etc.). These are systematic errors which require off-line estimation.

7.2 Error estimates

7.2.1 Random errors due to instrumental signal-to-noise

An estimation of random errors can be derived by examining the propagation of the Level-1 radiance and irradiance statistical errors through the inversion algorithm. For the GOME, GOME-2 and SCIAMACHY instruments, these errors are generally much less than 0.5% at moderate SZA and may reach 2% at extreme SZA (>80°). Given the anticipated TROPOMI signal-to-noise ratios, we can expect smaller random errors than these values.

7.2.2 Errors due to radiometric uncertainties

These errors are very difficult to assess and depend strongly on the nature of the calibration limitations. In general, the sensitivity to multiplicative errors on Level-1 spectra is relatively low. The magnitude of these errors will be established once the instrument is calibrated. Radiometric uncertainties have a direct impact on both OCRA and ROCINN algorithms.

7.2.3 Errors due to model parameter uncertainty

The most important sources of model parameter uncertainty in ROCINN are the values for geometrical cloud thickness and surface albedo. The corresponding cloud property retrieval errors are discussed in detail in [*Schuessler et al*., 2014]; here we present some results which summarize these findings. In Figure 7.1, we see that cloud-top height and cloud geometrical thickness can be retrieved with sufficient accuracy even when the cloud geometrical thickness is underestimated or overestimated by 30-40%. In Figure 7.2 (also taken from [*Schuessler et al*., 2014]), the retrievals are clearly more sensitive to uncertainty in the surface albedo, especially for the cloud optical thickness.

Less significant are ROCINN errors due to uncertainties in the choices of Mie-scattering particle size distribution parameters. An initial investigation (see section 7.2.3) has revealed that these error sources are small.

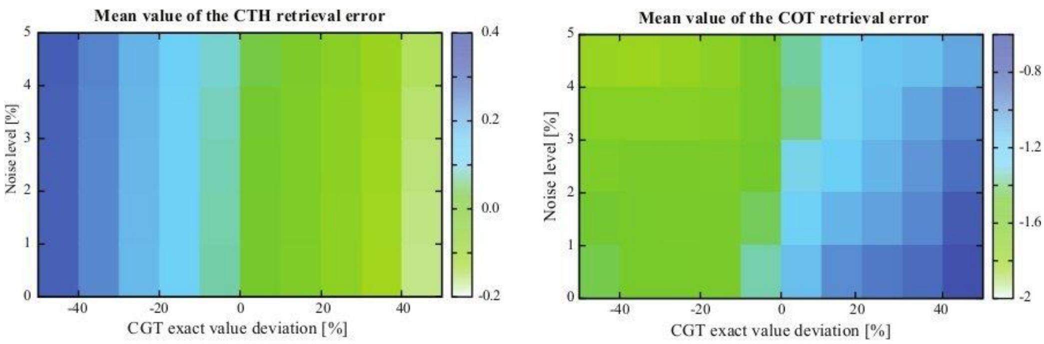


Figure 7.1: Mean values of the retrieval error for cloud-top height (left panel) and cloud optical thickness (right panel), plotted as functions of noise level (y-axis) and deviation (in %) of the cloud geometrical thickness from its true value (x-axis).

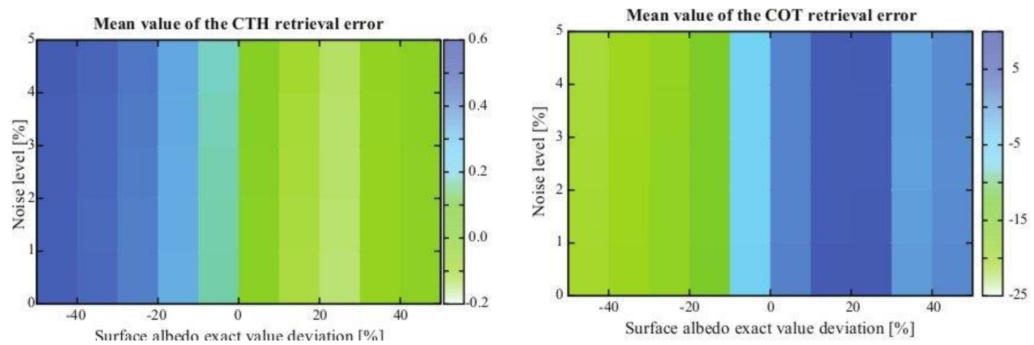


Figure 7.2: Mean values of the retrieval error for cloud-top height (left panel) and cloud optical thickness (right panel), plotted as function of noise level (y-axis) and deviation (in %) of the surface albedo from its true value (x-axis).

7.2.4 Errors due to forward-model uncertainty

This error category is the hardest to quantify, as it includes sources such as mathematical discretization choices and physical simplifications. The most basic assumption is of course the use of a simplified 1-D radiative transfer model as mandated by the IPA. With the relatively small TROPOMI/S5P spatial footprint, horizontal inhomogeneity in cloud fields will be an important consideration from both the geometrical and the radiation perspectives. 3-D RT modelling of atmospheres with clouds is notoriously difficult and time-consuming.

Some results have been reported using Monte-Carlo models for the 3-D baseline [*Marshak and Davis*, 2005], and in a more recent paper [*Doicu et al.*, 2014] focusing on solar-backscatter satellite retrievals, stochastic RT methods were used to quantify forward model error arising from the IPA.

7.2.5 An estimate of the total error budget

Table 7.1 contains an estimate of the total error budget for the cloud retrieval algorithms. Numbers are based on the above error estimation discussions and other sensitivity analyses as discussed in [*Schuessler et al.*, 2014] (see also below in section 7.3), with an assumed reflectance error of 1%, a cloud-free composite uncertainty of 5%, a surface albedo error of 5% and a cloud fraction error of 5%. No attempt has been made to include any sort of forward model error associated with the IPA simplification.

Table 7.1: Total error budget for cloud parameter retrieval using the OCRA and ROCINN algorithms.**Error Source**	**Cloud Fraction**		**Cloud-top Height (Pressure)**		**Cloud Optical Thickness (Albedo)**	
	CF > 0.2 Rel. Error	**CF ≤ 0.2 Abs. Error**	**CF > 0.2 Rel. Error**	**CF ≤ 0.2 Abs. Error**	**CF > 0.2 Rel. Error**	**CF ≤ 0.2 Abs. Error**
Measured reflectance or sun-normalized radiance	< 1.49%	< 0.01	< 0.7%	< 0.24	< 0.2%	< 0.01
Cloud-free Composite	< 4.06%	< 0.01	–	–	–	–
Surface Albedo	–	–	< 1.1%	< 1.12	< 1.4%	< 0.09
Cloud fraction	–	–	< 2.1%	< 0.12	< 1.6%	< 0.01
Total error	**< 4.33%**	**< 0.02**	**< 3.3%**	**< 1.18**	**< 3.2%**	**< 0.09**

7.3 Selected error and sensitivity studies

This section presents studies on the impacts of cloud parameter errors on total ozone retrievals; note however that the impact may be more critical for other trace gases.

7.3.1 Total ozone accuracy using CRB clouds

Here we report on some results obtained for the error budget pertaining to total ozone retrieval from GOME-type measurements in the UV Huggins bands. The analysis here is taken from [*Van Roozendael et al.*, 2006]. This study is a sensitivity analysis to investigate the effect of model parameter error (in this case, cloud fraction uncertainty) on the accuracy of the total ozone GDP4 product (which is generated via the heritage algorithm for S5P_TO3_DOAS), and also the accuracy of the ROCINN-retrieved cloud-top height and albedo products Figure 7.3 (top/bottom) shows normalized histograms for errors induced by a 10% overestimation/underestimation of cloud fraction. A 10% increase in the cloud fraction induces a ~5% decrease for the cloud top albedo and a ~5% increase for the cloud top height.

The OCRA algorithm retrieves a *radiometric* cloud fraction which is close to but not identical with the geometric cloud fraction. Deviations from the geometric cloud fraction could occur for thin clouds.

Interestingly, the ROCINN algorithm compensates a possible cloud fraction overestimation by underestimating the cloud top albedo and overestimating the cloud top height. Thus the net effect of combined OCRA/ROCINN uncertainties is to maintain the level of ozone total column error to the ±0.5% level.

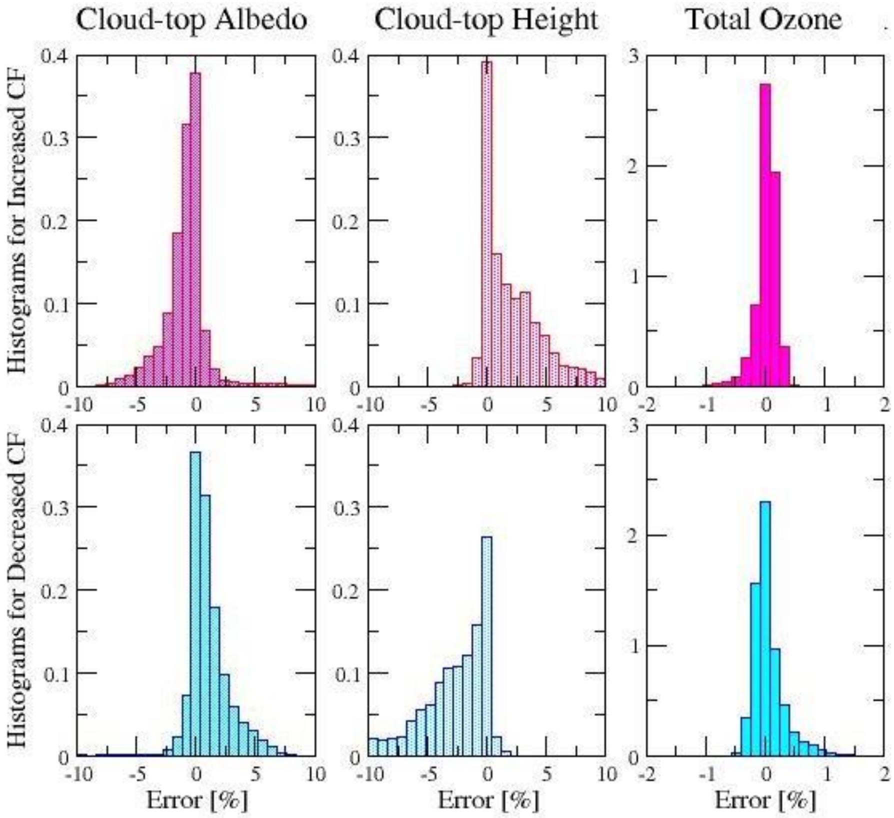


Figure 7.3: (top panels) Normalized histograms for errors induced by a 10% overestimate of OCRA's cloud fraction; and (bottom panels) a 10% underestimate of the cloud fraction. Relative errors are shown for cloud-top albedo, cloud-top height, and (right panels) total ozone. A 10% increase in cloud fraction induces a ~5% decrease in cloud-top albedo and a ~5% increase in the cloud-top height. The net error on total ozone stays at the 0.5% level.

7.3.2 Total ozone accuracy using CAL clouds; initial results

In this section, we summarize results obtained with another cloud-parameter sensitivity study for GOME-type total ozone retrieval, this time investigating the effect of model parameter errors (cloud property uncertainties) and selected forward model errors on the accuracy of the total ozone GDP5 product (heritage for S5P_TO3_GODFIT, see [RD3]).

A linear sensitivity analysis requires just one application of the least-squares fitting inverse model, since we are linearizing the atmosphere about a given state. We are not iterating to upgrade the state vector, and there is no requirement for a measurement vector. The error analysis is based on the solution error covariance matrix that emerges from the inversion, along with the contribution function. Since this is a theoretical study, we ignore instrument-related state vector parameters (undersampling amplitude, wavelength registration shift), and the Ring effect interference. The state vector $\mathbf{X} = [\Omega, S, \alpha_1, \alpha_2, \alpha_3]^T$ then has 5 elements, the total ozone Ω, the temperature shift S, and 3 albedo closure parameters α_1, α_2, α_3, such that the modelled Lambertian surface albedo is given by $\alpha = \alpha_1 + \alpha_2(1-\lambda/\lambda_0) + \alpha_3(1-\lambda/\lambda_0)^2$ as a function of wavelength λ, where $\lambda_0 = 330$ nm. For more on this retrieval, refer to [RD3].

We use a 30-layer atmosphere with pressure heights from an atmospheric chemistry model, and for given total ozone Ω, we use the TOMS Version 8 climatology to generate the associated ozone profile for inclusion in the optical property setups required for the radiative transfer simulations. We use pressures in [hPa] rather than heights in [km] for cloud boundaries. The (unshifted) temperature profiles also come from an ancillary TOMS temperature climatology; the temperature shift is assumed uniform throughout the atmosphere. There are no aerosols. For details of this set-up, refer to the S5P total ozone ATBD [RD3]. Cloud optical properties are generated using Mie scattering with a 3-parameter modified-gamma distribution (see Eq. (5.7) in section 5.3.3); only one set of parameters was used in this test. We also consider retrieval based on the CRB assumption; this will provide an estimate of forward-model error due to the CAL vs. CRB model choice.

VLIDORT calculations were performed for 10 solar zenith angles (SZAs) from 20° to 87.5°, 9 viewing angles from 0° to 72°, and 5 azimuth angles from 0° to 180°. VLIDORT generates the Stokes vector $\mathbf{I}$, the two column weighting functions $\partial\mathbf{K}/\partial\Omega$ and $\partial\mathbf{K}/\partial S$, and a surface albedo weighting function $\partial\mathbf{K}/\partial\alpha$ from which the closure-coefficient Jacobians can be derived. We are interested in the relative error (in %) on retrieved ozone, which is defined to be $100\delta X_1/X_1$, which is the first entry in the error vector $\delta\mathbf{X}$ divided by the first entry in state vector $\mathbf{X}$. Table 7.2 lists sources of error in this study. Errors 1-7 are model parameter errors. The final error #8 is a forward model uncertainty given by $\delta\mathbf{I} = \mathbf{I}' - \mathbf{I_0}$, where $\mathbf{I_0}$ is the simulated radiance obtained with the CAL set-up, and $\mathbf{I}'$ is the radiance simulated with the CRB assumption.

Table 7.2: Error settings and test setups used in this study.

Error #	Source of Error	Magnitude of error	Test 1 Settings	Test 2 Settings
1	Cloud-top pressure p_{ct}	50 hPa	750 hPa	550 hPa
2	Cloud-base pressure p_{cb}	50 hPa on p_{cb} - p_{ct}	950 hPa	900 hPa
3	Cloud optical thickness	15% (relative)	20	5
4	Mie parameter #1 (α)	33% (relative)	6	6
5	Mie parameter #2 (r_c)	33% (relative)	1.5	1.5
6	Mie parameter #3 (γ)	33% (relative)	1.0	1.0
7	Cloud Fraction f_c	0.1	20%,50%,80%	20%,50%,80%
8	CRB vs. CAL assumption	forward model error	CRB p_{ct} 750, albedo 0.8	CRB p_{ct} 750, albedo 0.8

The error budget was obtained for a range of total ozone amounts from 200 DU (or less) to 550 DU in mid- and high-latitudes, 200-350 DU in the tropics, and for temperature shifts from -10K to +10K. Errors are "scenario-averaged" RMS values, where for example, errors for one combination of (Ω,S) have been averaged over all solar/viewing geometry configurations. Two test scenarios were considered, with setups as indicated in the table.

Results for error tests 1 to 6 confirm that the total ozone algorithm is relatively insensitive to uncertainties in the cloud model parameters. In general, ozone errors are larger for increasing cloud fractions, and larger for low ozone values at lower SZAs (situations where photon penetration to the ground is highest). For cloud-top pressure, a 50 hPa error induces ozone errors up to 0.5%, with the largest values for relatively cloudy scenes and for lower ozone amounts at low SZA. Similar patterns are seen for the cloud geometrical thickness uncertainty, with values generally lower than those for the cloud-top pressure. Results were also relatively insensitive to cloud optical thickness, though errors this time are larger for the optically thin cloud case. Cloud fraction errors only appear to be significant for low optical thickness, thin but deep clouds, and for high cloud fraction scenarios. The errors due to poor knowledge of the Mie parameters were always below 0.2% for all cases - this is clearly not a significant source of error.

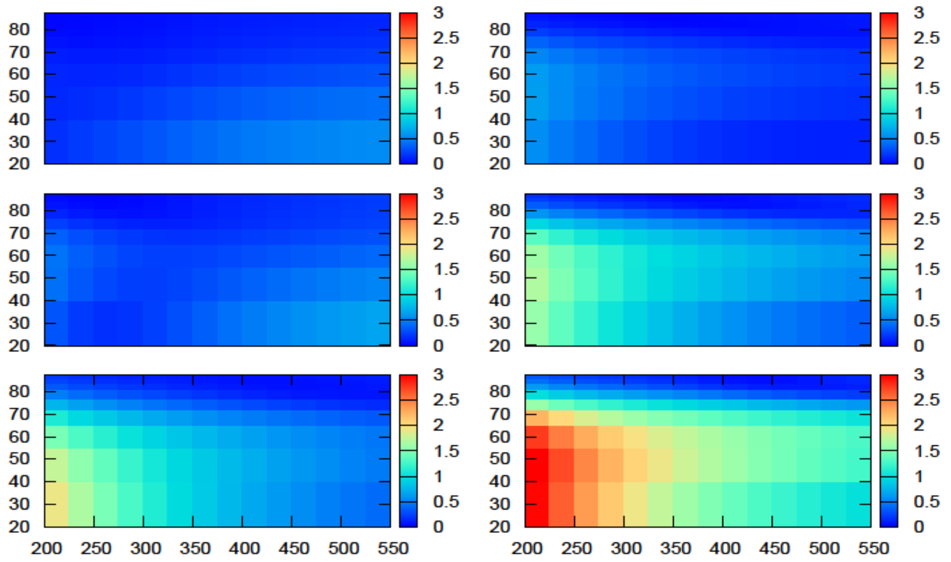


Figure 7.4: Forward model errors (in %) due to the use of the CRB assumption. (Left panels) Cloud-top at 750 hPa with optical depth 20, (right panels) cloud-top at 550 hPa, optical depth 5. Top row: cloud fraction 20%; middle: fraction 50%; bottom: fraction 80%. X-axis amounts are total ozone in [DU], Y-axis denotes solar zenith angles.

For the forward model error (#8), we are looking at much larger sources of error (Figure 7.4). For the lower and denser cloud (left panels), the CRB assumption leads to smaller errors below 1% for moderate cloudiness, but up to 2% for high-cloud situations with low ozone and low SZA (bottom left). The situation for the extensive but optically thin cloud (right panels) is particularly poor, with errors in excess of 3% for low ozone and SZA < 65°. For all levels of ozone and for moderate SZAs between 40° and 70°, the error is still in excess of 1%. It should be noted that in this test, no provision was made for intra-cloud ozone correction.

8 Validation

The geophysical validation of all S5P L2 products will be organized by ESA. A validation plan is outside the scope of the present document. However, a two-year campaign starting in the commissioning phase is strongly recommended.

Validation campaigns for the two S5P cloud algorithms will follow along similar lines to those already carried out for GOME, SCIAMACHY and GOME-2 data. In section 8.1, we report on some existing work that has been done on the validation of OCRA and ROCINN-CRB (Version 2.0), and we discuss options for S5P in section 8.2.

For initial verification results, we would like to refer to the S5P Science Verification Report.

8.1 Examples of heritage-algorithm validations

Here we report on some validations of OCRA and ROCINN-CRB that were carried out as part of the GOME GDP 4.0 total ozone reprocessing in 2005.

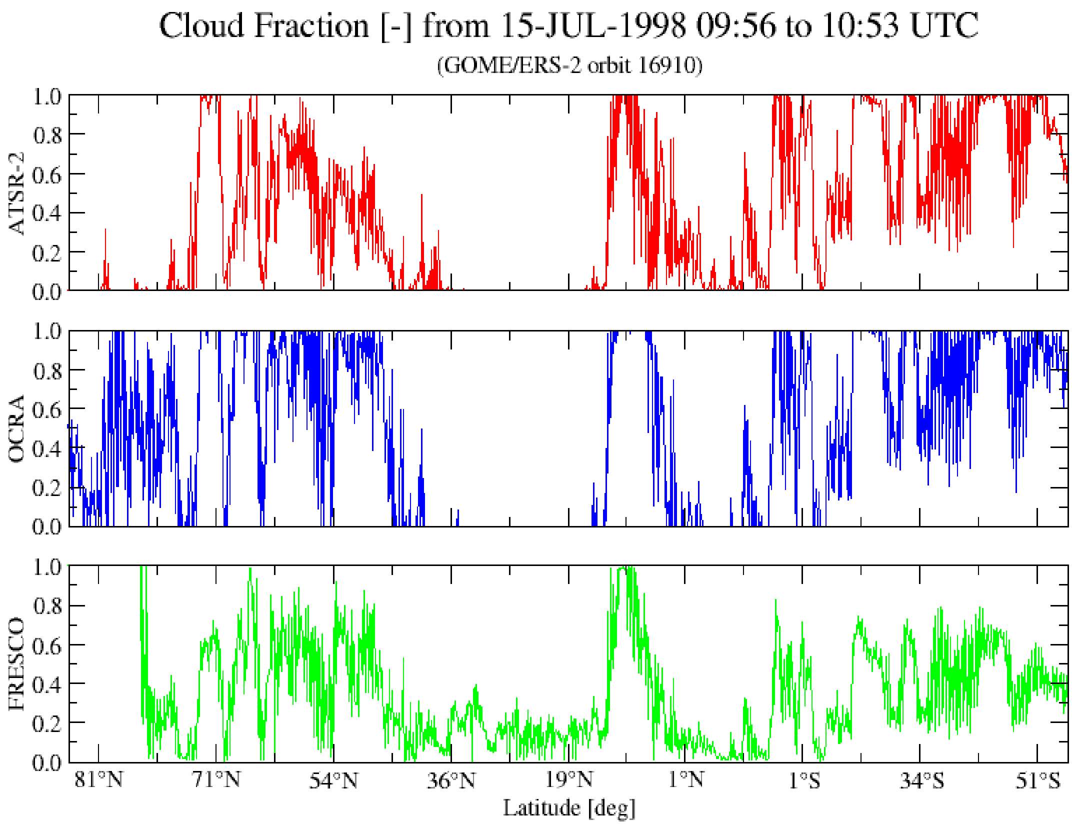


Figure 8.1: Comparison of cloud fraction retrieved from GOME using OCRA (middle panel), FRESCO algorithm (lower panel) and the ATSR-2 sensor (upper panel).

For OCRA working in the DOAS-ozone GDP 4.0 environment, OCRA was validated by comparing cloud fractions to values derived from collocated measurements from the ATSR-2 instrument (which like GOME, was part of the ERS-2 payload). The cloud fraction determined with OCRA, FRESCO and ATSR-2 on July 15th 1998 is shown in Figure 8.1. In most cases, OCRA results are close to those for ATSR-2, while FRESCO has a tendency to underestimate the cloud fraction. OCRA performs well over desert and snow/ice conditions, while FRESCO has problems over desert regions (e.g. between 35°N to 20°N in Figure 8.1).

Cloud-top heights determined with early versions of ROCINN, FRESCO and ATSR-2 for the same orbit are shown in Figure 8.2. The three algorithms provide similar results. ROCINN is smoother and more stable than FRESCO and has fewer spikes.

ROCINN computes realistic cloud-top height values even for pixels with low cloud fraction, whereas for FRESCO, a default value of 5 km is fixed for pixels with a cloud fraction < 0.1. The reason why ROCINN does not produce outliers is due to the robustness of the neural network and its ability to find a global minimal solution to the inverse function; FRESCO must find a local minimal solution for every individual measurements. Note that for cloud-free scenes (e.g. between 35°N to 20°N) both ATSR-2 and ROCINN report a cloud-top height of zero, while FRESCO retrieves the effective ground surface height.

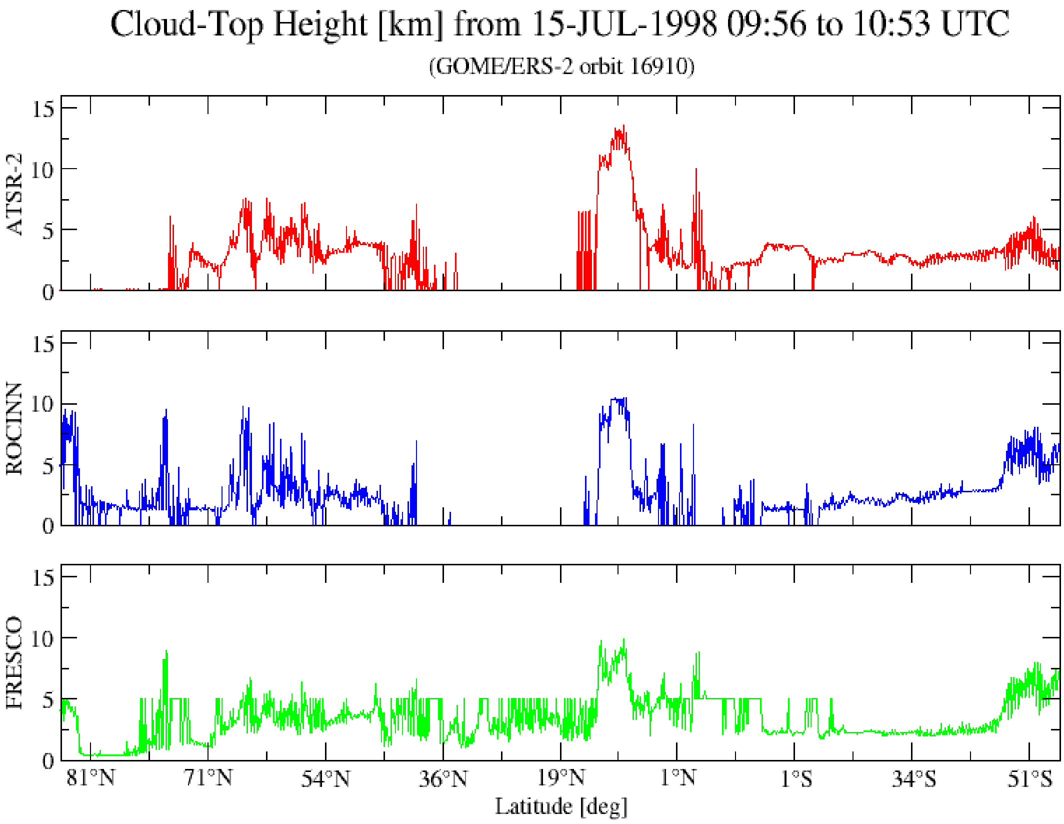


Figure 8.2: Comparison of cloud-top height retrieved from GOME using ROCINN (middle panel), FRESCO (lower panel) and the ATSR-2 sensor (upper panel).

An inter-comparison of GOME and ATSR cloud-top heights was performed for the ROCINN and the SACURA algorithms [*Rozanov et al.*, 2006]. Cloud fractions from OCRA and MERIS were also compared in the work of [*Casadio et al.*, 2006]. MSG (Meteosat Second Generation) comparisons have been done with ROCINN Versions 1.0 and 2.0 [*Loyola et al.*, 2010]. The ATSR-2 comparison (Figure 8.1) confirms the results reported in [*Tuinder et al.*, 2004] where several algorithms for retrieving cloud fraction using GOME data were compared against synoptic surface observations: in this work, OCRA has a mean difference of only ±10% compared with synoptic data, followed by FRESCO (±19.7%) and ICFA (±38.9%).

8.1.1 Comparisons with cloud climatology

There are several cloud climatology datasets derived from surface and satellite observations. The ISCCP data [*Rossow and Schiffer*, 1999] is perhaps the best known and has the longest history. This set been used in a number of satellite comparison studies, for example [*Jin et al*, 1996, *Steubenrauch et al*., 1999], and also for GOME data, starting with the initial study by [*Koelemeijer et al*., 2003] and more recently for a longer multi-year data set of cloud properties derived from OCRA/ROCINN on GOME [*Loyola et al*., 2010].

Other cloud climatology data sets have been created on a global scale, e.g. [*Jacobowitz et al.*, 2003; *Steubenrauch et al.*, 2013] and on regional scales, e.g. [*Karlsson*, 2003; *Meerkotter et al.*, 2004]. Global yearly-averaged GOME-derived cloud properties retrieved with OCRA/ROCINN (the latter in Version 2.0) were compared with ISCCP data in the study by [*Loyola et al*., 2010]. Figure 8.3 is taken from this study and compares zonal mean variations for the two sources, looking at cloud amounts (left panel), cloud-top pressures (middle panel) and cloud optical thickness values (right panel).

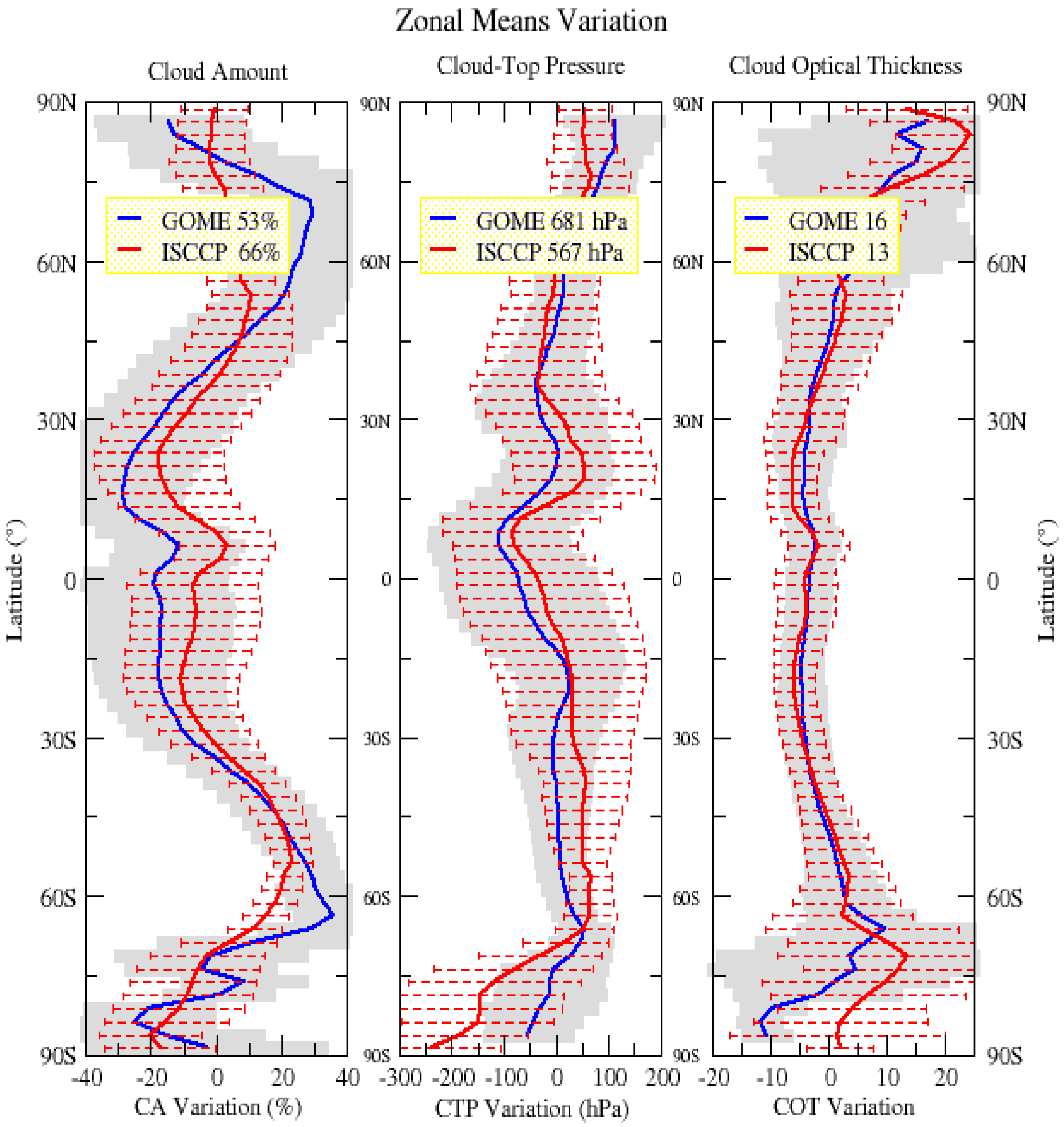
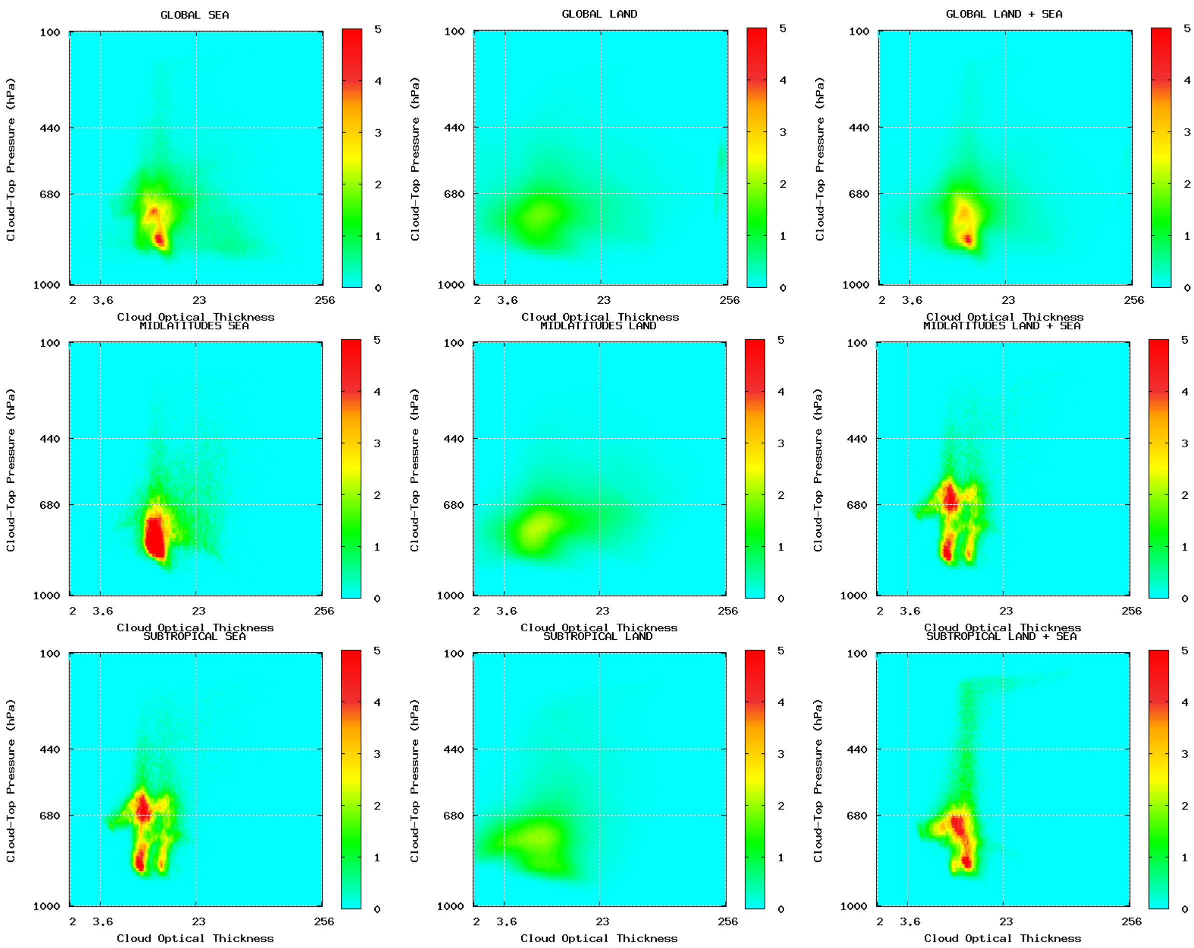


Figure 8.3: Latitudinal cloud zonal mean variation from April 1996 to June 2003 for GOME in blue and ISCCP in red. Global mean values are indicated in the insets. Standard deviations of the GOME and ISCCP parameters are shown as gray surfaces in the background, and dotted red lines in the foreground, respectively.

Figure 8.4 is also taken from this study, and contains two-dimensional histograms of cloud top pressure and cloud optical thickness. Histograms are presented for land and sea surfaces as indicated. Such plots are useful for the analysis of correlations in cloud properties. Notice the double-peak structure over the ocean.

8.2 Ground-based validation and satellite inter-comparison

A ground validation campaign for the two S5P cloud algorithms is difficult to envisage. Validation against ground-based cloud observations is necessarily confined to continental areas where there is a dense cover of surface synoptic observations. There is also a strict requirement on close time coincidences in this kind of validation. The traditional reporting of cloud coverage in "oktas" is not really accurate enough for validation purposes on a continent-wide scale, despite studies that have been done already.



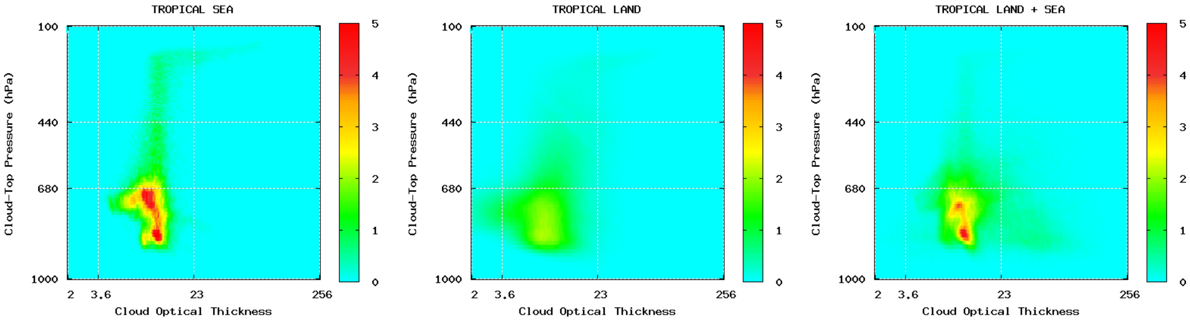


Figure 8.4: GOME-derived cloud optical thickness and cloud-top pressure variation histograms for different geographical regions over land surfaces (left panels) and oceans (middle panels), with the right panels showing the cumulative land and sea histograms. The global distribution is shown in the first row, followed by results for the mid-latitudes (second row), the subtropics (third row) and tropical regions (bottom row).

Polar orbiting and geostationary satellite measurement comparisons offer a more fruitful field for comparisons. Instruments with high spatial resolution are important here. Once again there is a requirement for close time coincidences - something that is not always possible owing to different orbital crossing times. Of particular interest here are high spatial-resolution cloud products from the VIIRS instrument (MODIS heritage) on board the Suomi NPP platform. A detailed comparison with the S5P cloud products is planned, and this will provide some material for validation. In the same way, data from the NPP follow-up mission JPSS should be used for validation.

Space-borne Lidar measurements from the CALIPSO instrument (aerosol profiles) and Radar measurements from CLOUDSAT (cloud profiles) have also proved very useful for the validation of cloud products from OMI and MODIS. CLOUDSAT and CALIPSO are part of NASA's A-train constellation.

One example here is the synergy between OMI (on the Aura platform) and MODIS (Aura and Terra platforms) cloud-field observations. This synergy has been investigated by [*Vasilkov et al.*, 2008]. Synergy is also planned for the S5P - it will be possible to compare the S5P cloud products with those from NPP, as well as with other cloud products generated by instruments on NASA satellite platforms in the A-train.

9 Conclusions

The operational algorithms for the generation of the operational S5P cloud products are presented in this ATBD. Two retrieval algorithms are selected, namely, S5P_CLOUD_OCRA for the retrieval of cloud fraction, and S5P_CLOUD_ROCINN for the retrieval of cloud-top height (pressure) and cloud optical thickness (albedo).

Two types of ROCINN-derived cloud products will be provided: (a) macro-physical cloud properties retrieved using a homogenous single-layer cloud scattering model, and (b) cloud properties based on a Lambertian-reflector cloud model. As with the OCRA and ROCINN products already in use in the operational retrieval of trace-gas products from GOME/ERS-2 and GOME-2 on MetOp-A and MetOp-B, the S5P cloud products are designed for optimal cloud correction in the S5P trace gas algorithms. S5P cloud products will also be used to extend the climate data record of cloud properties derived from O_2 *A*-band measurements started with GOME in 1995.

These operational algorithms for the S5P cloud products will meet the accuracy requirements needed for this mission. Near-real-time products will be available 3 hours after sensing. The products, comprising the cloud parameters themselves and their corresponding errors, will be provided in NetCDF-CF format.

10 References

Acarreta, J.R., J.F. de Haan and P. Stammes, Cloud pressure retrieval using the O2-O2 absorption band at 477 nm, *J. Geophys. Res*., **109**, doi:10.1029/2003JD003915, 2004.

Ahmad, Z., P. K. Bhartia, and N. Krotkov, Spectral properties of backscattered UV radiation in cloudy atmospheres, *J. Geophys. Res* , **109**, D01201, doi:10.1029/2003JD003395, 2004.

Boersma K., Eskes H., and Brinksma E., Error analysis for tropospheric NO2 retrieval from space, *J. Geophys. Res*, **109**, no. D4, 4311, DOI:10.1029/2003JD003 962, 2004.

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

Doicu, A., T. Trautmann, and F. Schreier, *Numerical Regularization for Atmospheric Inverse Problems*, Berlin:Springer Verlag, 2010.

Doicu, A, D. Efremenko, D. Loyola, T. Trautmann, Discrete ordinate method with matrix exponential for stochastic radiative transfer in broken clouds, *J. Quant. Spectrosc. Radiat. Trans.,* in press, 2014.

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

Grzegorski, M., Wenig, M., Platt, U., Stammes, P., Fournier, N., and Wagner, T.: The Heidelberg iterative cloud retrieval utilities (HICRU) and its application to GOME data, *Atmos. Chem. Phys*., **6**, 4461-4476, 2006.

Hahn, C.J., W.B. Rossow, and S.G. Warren, ISCCP Cloud Properties Associated with Standard Cloud Types Identified in Individual Surface Observations, *J. Clim*., **14**, Issue 1, 2001.

Hess, M., P. Koepke, and I. Schult, Optical Properties of Aerosols and Clouds: The Software Package OPAC, *Bull. Am. Met. Soc*, **79**, Issue 5, 1998.

Jacobowitz H., L.L. Stowe, G. Ohring, A. Heidinger, K. Knapp, and N.R. Nalli, The Advanced Very High Resolution Radiometer Pathfinder Atmosphere (PATMOS) climate dataset: A resource for climate research, *Bull. Am. Met. Soc*., **84**, 785-793, doi:10.1175/BAMS-84-6-785, 2003.

Jin Y., W.B. Rossow, and D.P. Wylie, Comparison of the climatologies of high-level clouds from HIRS and the ISCCP, *J. Clim*., **9**, 2850–2879, 1996.

Joiner, J., and A. P. Vassilkov, First results from the OMI rotational Raman scattering cloud pressure algo-rithm, IEEE Trans. Geosci. Remote Sens., **44**, 1272–1282, 2006.

Karlsson K.-G., A 10 year cloud climatology over Scandinavia derived from NOAA Advanced Very High Resolution Radiometer Imagery, *Int. J. Clim*., **23**, 1023-1044, 2003.

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

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

Kokhanovsky, A. A., von Hoyningen-Huene, W., Rozanov, V. V., Noël, S., Gerilowski, K., Bovensmann, H., Bramstedt, K., Buchwitz, M., and Burrows, J. P., The semi-analytical cloud retrieval algorithm for SCIAMACHY II. The application to MERIS and SCIAMACHY data, *Atmos. Chem. Phys*., **6**, 4129-4136, doi:10.5194/acp-6-4129-2006, 2006.

Kokhanovsky, A. A., B. Mayer, V. V. Rozanov, K. Wapler, L. N. Lamsal, M. Weber, J. P. Burrows, and U. Schumann, The influence of broken cloudiness on cloud top height retrievals using nadir observations of backscattered solar radiation in the oxygen *A*-band, *J. Quant. Spectrosc. Radiat. Transfer*, **103**, 460–477, doi:10.1016/j.jqsrt.2006.06.003, 2007.

Kokhanovsky, A. A., and V. V. Rozanov, The uncertainties of satellite DOAS total ozone retrieval for a cloudy sky, *Atmos. Res*., **87**, 27–36, doi:10.1016/j.atmosres.2007.04.006, 2008.

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

Lelli, L., Kokhanovsky, A.A., Rozanov, V.V., Vountas, M., Sayer, A.M., and Burrows, J.P.: Seven years of global retrieval of cloud properties using space-borne data of GOME, Atmos. Meas. Tech., 5, 1551-1570, doi:10.5194/amt-5-1551-2015, 2012.

Liu, X., M. Newchurch, R. Loughman, and P.K. Bhartia, Errors resulting from assuming opaque Lambertian clouds in TOMS ozone retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, **85**, 337-365, 2004.

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

Loyola, D., and T. Ruppert, A new PMD cloud-recognition algorithm for GOME, *ESA Earth Observation Quarterly*, **58**, 45-47, 1998.

Loyola, D., Cloud retrieval for SCIAMACHY, http://wdc.dlr.de/sensors/gome/gdp4/loyola_2000.pdf, 2000.

Loyola, D., W. Thomas, Y. Livschitz, T. Ruppert, P. Albert, and R. Hollmann, Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval, *IEEE Trans. Geosci. Remote Sens.*, **45**(9), 2747–2758, 2007.

Loyola, D., W. Thomas, R. Spurr, B. Mayer, Global patterns in daytime cloud properties derived from GOME backscatter UV-VIS measurements, *Int. J. Remote Sensing*, **31**(16), 4295-4318, doi: 10-1080/01431160903246741, 2010.

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

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

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11, 409-427, https://doi.org/10.5194/amt-11-409-2018, 2018.

Lutz, R., Loyola, D., Gimeno García, S., and Romahn, F.: OCRA radiometric cloud fractions for GOME-2 on MetOp-A/B, Atmos. Meas. Tech. Discuss., 8, 13471-13524, doi:10.5194/amtd-8-13471-2015, 2015.

Marshak, A., and A. B. Davis (eds.), 3D Radiative Transfer in Cloudy Atmospheres, Series Physics of Earth and Space Environments. Springer-Verlag, Berlin, Heidelberg, 2005.

Meerkötter R., C. König, P. Bissolli, G. Gesell, and H. Mannstein, A 14-year European Cloud Climatology from NOAA//AVHRR data in comparison to surface observations, *Geophys. Res. Lett*., **31**, L15103, doi: 10.1029/2004GL020098, 2004.

Menzel, W. P., F. Richard, H. Zhang, D. P. Wylie, C. Moeller, R. E. Holz, B. Maddux, K. I. Strabala, and L. E. Gumley, MODIS global cloud-top pressure and amount estimation: Algorithm description and results, *J. Appl. Meteorol. Climatol*., **47**, 1175–1198, doi:10.1175/2007JAMC1705.1., 2008.

Popp, C., Wang, P., Brunner, D., Stammes, P., Zhou, Y., and Grzegorski, M., MERIS albedo climatology for FRESCO+ O2 A-band cloud retrieval, Atmos. Meas. Tech., **4**, 463-483, 2011.

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

Rossow W.B. and R.A. Schiffer, Advances in understanding clouds from ISCCP, *Bull. Amer. Meteor. Soc.*, **80**, 2261-2287, 1999.

Rothman, L., et al., The HITRAN 2008 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, **110**, 533-572, 2009.

Rozanov, V., and A. Kokhanovsky. Semi-analytical cloud retrieval algorithm as applied to the cloud top altitude and the cloud geometrical thickness determination from top-of-atmosphere reflectance measurements in the oxygen *A-band*, *J. Geophys. Res*., **109**, D05202, doi: 10.1029/2003JD004104, 2004.

Rozanov, A., A. Kokhanovsky, D. Loyola, R. Siddans, B. Latter, A. Stevens, and J. Burrows, Intercomparison of cloud top altitudes as derived using GOME and ATSR-2 instruments onboard ERS-2, *Remote Sensing of Environment*, **102**, 186-193, 2006.

Schiffer, R. A., and W. B. Rossow, The international satellite cloud climatology project ISCCP: The first project of the world climate research program, *Bull. Am. Meteorol. Soc.*, **54**, 779-784, 1983.

Schreier F. and B. Schimpf, A new efficient line-by-line code for high resolution atmospheric radiation computations incl. derivatives, in *IRS 2000: Current Problems in Atmospheric Radiation*, W. L. Smith and Y. Timofeyev, Eds. Hampton, VA, USA: A. Deepak Publishing, pp. 381–384, 2001.

Schreier F., Optimized implementations of rational approximations for the Voigt and complex error function. *J. Quant. Spectrosc. Radiat. Transf.*, **112**, 1010–1025, 2011.

Schuessler, O., D. Loyola, A. Doicu, and R. Spurr, Information Content in the Oxygen *A*-band for the Retrieval of Macrophysical Cloud Parameters, IEEE *Trans. Geosci. Remote Sens.*, **52**(6), 2014.

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

Sneep M., J. F. de Haan, P. Stammes et al.; Three way comparison between OMI and PARASOL cloud pressure products. *J. Geophys. Res*., **113**, 10.1029/2007JD008694, 2008.

Spurr, R. J. D. VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, *J. Quant. Spectrosc. Radiat. Transf.*, **102**, 316-42, 10.1016/j/jqsrt.2006.05.005, 2006.

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

Stammes, P., M. Sneep, J. F. de Haan, J. P. Veefkind, P. Wang, and P. F. Levelt, Effective cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and validation, *J. Geophys. Res*., **113**, D16S38, doi:10.1029/2007JD008820, 2008.

Stubenrauch C.J., W.B. Rossow, N.A. Scott, and A. Chédin, Clouds as seen by satellite sounders (3I) and imagers (ISCCP). Part III: Spatial heterogeneity and radiative effects, *J. Clim*., **12**, 3419-3442, 1999.

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H. Getzewich, B., Di Girolamo, L., Guignard, A., Heidinger, A., Maddux, B., Menzel, P Minnis, P Pearl, C., Platnick, S., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., and Zhao, G., Assessment of Global Cloud Datasets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel, Bull. Amer. Meteorol. Soc., pp. 0003–0007, doi:10.1175/BAMS-D-12-00117, 2013.

Tanskanen, A. Lambertian Surface Albedo Climatology at 360 nm from TOMS Data Using Moving Time-Window Technique. In: Proceedings of the XX Quadrennial Ozone Symposium, 1-8 June 2004, Kos, Greece.

Tuinder, O. N. E., R. de Winter-Sorkina, and P. J. H. Builtjes, Retrieval methods of effective cloud cover for the GOME instrument: an intercomparison, *Atmos. Chem. Phys*., **4**, 255-273, 2004.

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

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

Vasilkov, A., J. Joiner, R. Spurr, P. K. Bhartia, P. Levelt, and G. Stephens, Evaluation of the OMI cloud pressures derived from rotational Raman scattering by comparisons with other satellite data and radiative transfer simulations, *J. Geophys. Res.*, **113**, D15S19, doi:10.1029/2007JD008689, 2008.

Wagner, T., S. Beirle, T. Deutschmann, M. Grzegorski, and U. Platt, Wagner, T., Beirle, S., Deutschmann, T., Grzegorski, M., and Platt, U., Dependence of cloud properties derived from spectrally resolved visible satellite observations on surface temperature, *Atmos. Chem. Phys.*, **8**, 2299-2312, doi:10.5194/acp-8-2299-2008, 2008.

Wang P., P. Stammes, R. van der A, G. Pinardi, M. van Roozendael, FRESCO+: an improved O2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, *Atmos. Chem. Phys.*, **8**, 6565-6576, 2008.

Wang, J., W.B. Rossow, and Y. Zhang, Cloud Vertical Structure and Its Variations from a 20-Yr Global Rawinsonde Dataset, *J. Clim*, **13**, Issue 17, 2000.

Wang, P., P. Stammes, R. van der A, G. Pinardi, and M. van Roozendael, FRESCO+: an improved O_2 *A*-band cloud retrieval algorithm for tropospheric trace gas retrievals, *Atmos. Chem. Phys. Discuss*., **8**, 9697-9729, 2008.