 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 1
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

ESA Climate Change Initiative “Plus” (CCI+)

# Algorithm Theoretical Basis Document (ATBD) Version 1.0

—

## For the RemoTeC XCO<sub>2</sub> and XCH<sub>4</sub> GOSAT-2 SRON Full Physics Products (CO<sub>2</sub>\_GO<sub>2</sub>\_SRFP and CH<sub>4</sub>\_GO<sub>2</sub>\_SRFP) Version 2.0.2

for the Essential Climate Variable (ECV)  
**Greenhouse Gases (GHG)**

Written by:

GHG-CCI+ project team

Lead author:


Andrew Gerald Barr

(SRON Netherlands Institute for Space Research, Leiden, The Netherlands)

Co-authors:


Tobias Borsdorff, Jochen Landgraf

(SRON Netherlands Institute for Space Research, Leiden, The Netherlands)

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document (ATBD)</b> for the <b>Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 2
		Version 1.0
		10. Feb 2023


## Change log

<b>Version Nr.</b>	<b>Date</b>	<b>Status</b>	<b>Reason for change</b>
Version 1.0	10. Feb. 2023	Draft	New document for new product version

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 3
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

## Table of Contents

List of acronyms and abbreviations .....	4
Scope of document.....	6
Executive summary .....	7
1 Data product overview .....	8
1.1 Column-averaged mixing ratios of CH <sub>4</sub> (XCH <sub>4</sub> ) and CO <sub>2</sub> (XCO <sub>2</sub> ).....	8
2 Input and auxiliary data .....	9
2.1 Satellite instrument.....	9
2.1.1 GOSAT-2 TANSO-FTS-2 Level 1b .....	9
2.2 Other .....	10
2.2.1 ECMWF model data.....	10
2.2.2 SRTM DEM .....	11
2.2.3 TCCON FTS CO <sub>2</sub> and CH <sub>4</sub> data.....	11
2.2.4 Additional input data.....	12
2.3 Overview of Processing Sub-System .....	12
3 Algorithm .....	16
3.1 Forward Model .....	17
3.1.1 Model Atmosphere and Optical Properties .....	17
3.1.2 Modeling the top-of-atmosphere radiances.....	22
3.1.3 Summary of Forward Model.....	25
3.1.4 Inverse algorithm.....	26
3.1.5 Inversion Procedure .....	27
3.1.6 Regularization of state vector and iteration strategy .....	29
3.1.7 Convergence criteria .....	30
3.1.8 Scaling of O <sub>2</sub> Cross Sections .....	30
4 Output data.....	33
References .....	34

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document (ATBD)</b> for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Page 4
		Version 1.0
		10. Feb 2023

## List of acronyms and abbreviations

Acronym	Definition
AOT	Aerosol Optical Thickness
ATBD	Algorithm Theoretical Basis Document
CAMS	Copernicus Atmosphere Monitoring Service
CCI+	Climate Change Initiative Plus
CH <sub>4</sub>	Methane
CO	Carbon Monoxide
CO <sub>2</sub>	Carbon Dioxide
CPU	Core Processing Unit
DEM	Digital Elevation Map
DFS	Degrees of Freedom for Signal
ECMWF	European Centre for Medium Range Weather Forecasting
ESA	European Space Agency
FP	Full Physics retrieval method
FTIR	Fourier Transform InfraRed
FTS	Fourier Transform Spectrometer
GHG	GreenHouse Gas
GOSAT	Greenhouse Gases Observing Satellite
HSRS	Hybrid Solar Reference Spectrum
ISRF	Instrument Spectral Response Function
JAXA	Japan Aerospace Exploration Agency
L1	Level 1
L2	Level 2
LMD	Laboratoire de Météorologie Dynamique
NA	Not applicable
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Format
NIR	Near Infrared
O <sub>2</sub>	Oxygen
ppb	Parts per billion
ppm	Parts per million
PUG	Product User Guideline
RAA	Relative Azimuth Angle
RemoTeC	Remote Sensing of Greenhouse Gases for Carbon Cycle Modeling
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric CHartography
SNR	Signal-to-Noise Ratio
SRFP	SRON Full Physics
SRPR	SRON Proxy



ghg  
cci

GHG-CCI+ project


ESA Climate Change Initiative "Plus" (CCI+)  
**Algorithm Theoretical Basis Document  
(ATBD)**  
for the Essential Climate Variable (ECV)  
Greenhouse Gases (GHG)

Page 5

Version 1.0

10. Feb 2023

SRON	SRON Netherlands Institute for Space Research
SRTM	Shuttle Radar Topography Mission
SWIR	Short Wave Infrared
SZA	Solar Zenith Angle
TANSO	Thermal And Near infrared Sensor for carbon Observation
TCCON	Total Carbon Column Observing Network
TIR	Thermal Infrared
TM	Transport Model
TSIS	Total and Spectral Solar Irradiance Sensor
VZA	Viewing Zenith Angle
XCO <sub>2</sub>	Column-averaged dry-air mixing ratios (mole fractions) of CO <sub>2</sub>
XCH <sub>4</sub>	Column-averaged dry-air mixing ratios (mole fractions) of CH <sub>4</sub>


 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document (ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 6
		Version 1.0
		10. Feb 2023

## Scope of document

This Algorithm Theoretical Basis Document (ATBD) describes the SRON-RemoTeC algorithm used to generate the CO<sub>2</sub>\_GO<sub>2</sub>\_SRFP and CH<sub>4</sub>\_GO<sub>2</sub>\_SRFP products. GO<sub>2</sub> stands for GOSAT-2 and SRFP is an abbreviation for the “SRON Full Physics” products.

The RemoTeC algorithm is used to retrieve column averaged dry air mole fraction of carbon dioxide (CO<sub>2</sub>) and methane (CH<sub>4</sub>), denoted as XCO<sub>2</sub> and XCH<sub>4</sub>, as well as other parameters included in the Level 2 CO<sub>2</sub>\_GO<sub>2</sub>\_SRFP and CH<sub>4</sub>\_GO<sub>2</sub>\_SRFP products generated from GOSAT-2 Level 1b spectra.

This document details various input data required for retrievals, physical theory and mathematical background underlying retrieval assumptions, and outlines retrieval implementation and limitations of the approach used.

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 7
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023


## Executive summary

This document describes the RemoTeC algorithm for GHG retrieval from the GOSAT-2 instrument. The original algorithm is based on the paper of *Butz et al., 2009*. Tests of the retrieval algorithm have been performed on synthetic GOSAT data (*Butz et al., 2010*), and applied to real GOSAT data (*Butz et al., 2011; Schepers et al., 2012; Guerlet et al., 2013*). Here we apply the retrieval for the first time to GOSAT-2 data.

In order to account for the effect of aerosols and cirrus, the developed algorithm retrieves the methane column simultaneously with the aerosol/cirrus amount (column integrated particle number concentration), a parameter related to the particle size distribution, and a parameter describing the height distribution. Here, the particle size distribution is described by a power-law function, which only has two free parameters (related to amount and size). The choice of aerosol/cirrus parameters reflects the information content of the measurements as close as possible. The retrieval algorithm uses measured radiances in the Short Wave Infra-Red (SWIR) band and additionally in the Near Infra-Red (NIR, O<sub>2</sub> A-band). Additional fit parameters are the surface albedo and its 1st order spectral dependence in all bands, and the total column of water vapor, respectively.

In order to obtain a proper characterization of the retrieved XGHG, it is important to first retrieve a vertical profile (layer averaged number density in different layers of the model atmosphere) and use this retrieved vertical profile to calculate the vertical column. Here, we choose to provide the vertical column as a product, and not the full profile, because the Degrees of Freedom for Signal (DFS) of the retrieved CH<sub>4</sub> and CO<sub>2</sub> profile is about 1. The inversion is performed using Phillips-Tikhonov regularization in combination with a reduced step size Gauss-Newton iteration scheme.

The forward model of the retrieval algorithm uses online radiative transfer calculations, fully including multiple scattering. Here, the radiative transfer model developed by *Landgraf et al., 2001; Hasekamp and Landgraf, 2002, 2005* is being used. This model uses the Gauss-Seidel iterative method to solve the radiative transfer equation in a plane-parallel, vertically inhomogeneous atmosphere. To avoid time consuming line-by-line calculations we employ the linear-k method developed by *Hasekamp and Butz 2008*. Absorption cross sections of the relevant atmospheric trace gases are tabulated in a lookup table as function of pressure and temperature. Optical properties of aerosols are also calculated from lookup tables as described in *Dubovik et al., 2006*. The linear k-binning method in combination with other speed optimizations allow us to perform the GHG retrievals from GOSAT-2 with online RT calculations within ~60 seconds for a single retrieval. This makes the algorithm feasible for global processing.

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 8
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

## 1 Data product overview

### 1.1 Column-averaged mixing ratios of CH<sub>4</sub> (XCH<sub>4</sub>) and CO<sub>2</sub> (XCO<sub>2</sub>)

In this section an overview of the data products (specified in terms of variable, its property, processing level(s) and instrument(s)) are given.


The data products:

- Column-averaged dry-air mixing ratios (mole fractions) of CH<sub>4</sub>, denoted XCH<sub>4</sub> (in parts per billion, ppb).
- Column-averaged dry-air mixing ratios (mole fractions) of CO<sub>2</sub>, denoted XCO<sub>2</sub> (in parts per million, ppm).

In the following, several satellite instruments are shortly described which are used or can be used to generate the XCH<sub>4</sub> and XCO<sub>2</sub> data products CH<sub>4</sub>\_GO<sub>2</sub>\_SRFP and CO<sub>2</sub>\_GO<sub>2</sub>\_SRFP.

TANSO-FTS-2 is a Fourier-Transform-Spectrometer (FTS) onboard the Japanese GOSAT-2 satellite (*Nakajima et al., 2017*). The Japanese Greenhouse gases Observing SATellite-2 (GOSAT-2) was launched on 29th October 2018 and started operational observations from February 2019. GOSAT-2 provides dedicated global measurements of total column CO<sub>2</sub> and CH<sub>4</sub> from its SWIR bands. GOSAT-2 covers the relevant CO<sub>2</sub>, CH<sub>4</sub> and O<sub>2</sub> absorption bands in the NIR and SWIR spectral region as needed for accurate XCO<sub>2</sub> and XCH<sub>4</sub> retrieval (in addition GOSAT-2 also covers a large part of the Thermal Infrared (TIR) spectral region). The spectral resolution of TANSO-FTS-2 is much higher compared to e.g. SCIAMACHY and also the ground pixels are smaller (9.7 km compared to several 10 km for SCIAMACHY). However, in contrast to SCIAMACHY, the GOSAT-2 scan pattern consists of non-consecutive individual ground pixels, i.e., the scan pattern is not gap-free. For a good general overview about [GOSAT-2](http://www.GOSAT-2.nies.go.jp/) see also <http://www.GOSAT-2.nies.go.jp/>.



 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Page 9
		Version 1.0
		10. Feb 2023

## 2 Input and auxiliary data

### 2.1 Satellite instrument

#### 2.1.1 GOSAT-2 TANSO-FTS-2 Level 1b

Level 1b data of the TANSO-FTS-2 (Thermal And Near-infrared Sensor for carbon Observation - Fourier Transform Spectrometer) onboard GOSAT-2 (Greenhouse gas Observing SATellite) are needed in the project to produce the total column CO<sub>2</sub> and CH<sub>4</sub> products. They serve as input for the retrieval algorithms to be used in this project.


The TANSO-FTS-2 instrument (*Nakajima et al., 2017*) has five spectral bands with a high spectral resolution 0.2 cm<sup>-1</sup>. Three operate in the SWIR at 0.75-0.77, 1.56-1.69 and at the extended 1.92-2.33 μm range, providing sensitivity to the near-surface absorbers. The fourth and fifth channels operating in the thermal infrared between 5.5-8.4 and 8.4-14.3 μm providing mid-tropospheric sensitivity. FTS-2 observes sunlight reflected from the earth's surface and light emitted from the atmosphere and the surface. The former is observed in the spectral bands 1 through 3 of FTS-2 in the daytime, and the latter is captured in band 4 and 5 during both the day and the night. Within this project only level 1 data from the SWIR channels 1-3 will be used. Prior to reaching the detectors of the instrument, the light in the bands 1 through 3 is split into two orthogonally-polarized components and measured independently. The intensity component of Stokes vector is approximated by the mean of parallel (P) and perpendicular (S) components (*O'Brien et al., 2013*).

The measurement strategy of TANSO-FTS-2 is optimized for the characterization of continental-scale sources and sinks. TANSO-FTS-2 utilizes a pointing mirror to perform off-nadir measurements at the same location on each 6-day repeat cycle. The pointing mirror allows TANSO-FTS-2 to observe up to ±35° across track and ±40° along-track. These measurements nominally consist of 5 across track points spaced ~160km apart with a ground footprint diameter of approximately 9.7 km and a 4 second exposure duration. The satellite has an Intelligent pointing Monitor camera which makes it possible to adjust the line of sight of the FTS-2 to steer away from cloud contaminated areas. Whilst the majority of data is limited to measurements over land where the surface reflectance is high, TANSO-FTS-2 also observes in sunglint mode over the ocean.

Table 1: GOSAT-FTS-2 bands

Channel	Wavelength range [nm]	Resolution [cm <sup>-1</sup> ]
1	758-775	0.2
2	1460-1720	0.2
3	1920-2330*	0.2
4	5560-8400	0.2
5	8400-14300	0.2

\*GOSAT-1 only had a spectral range up to 2080nm.

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 10
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

*Data availability & coverage:* TANSO-FTS-2 level-1b data are available since February 2019.

*Source data product name & reference to product technical specification documents:* GOSAT FTS-2 L1A/L1B Product Description Document, Japan Aerospace Exploration Agency, July 2019, GST-18005, available through [https://prdct.gosat-2.nies.go.jp/en/documents/JAXA\\_GOSAT-2\\_FTS-2\\_L1\\_Data\\_Description\\_Document\\_101101\\_en.pdf](https://prdct.gosat-2.nies.go.jp/en/documents/JAXA_GOSAT-2_FTS-2_L1_Data_Description_Document_101101_en.pdf)

*Data quality and reliability:* The quality of the retrieved CO<sub>2</sub> and CH<sub>4</sub> columns has been tested against ground-based observations (i.e. the TCCON network) and has shown to be of good quality. The spectral bands showed some irregularities which required a shortened retrieval window (O<sub>2</sub> A-band) and spectral intensity offsets for each of the bands. This in effect worsened the cost function  $\chi^2$  of the fits compared to similar GOSAT-1 retrievals and the estimated uncertainties of the species. However, comparison to TCCON observations showed that the retrieval products of both the GOSAT-1 and GOSAT-2 are of similar quality.

## 2.2 Other

### 2.2.1 ECMWF model data

The retrieval algorithms to produce vertical columns of CO<sub>2</sub> and CH<sub>4</sub> need as input for each scene the temperature vertical profile, pressure vertical profile, specific humidity vertical profile, and wind speed. Here, temperature and pressure are needed to calculate absorption cross sections, the specific humidity vertical profile is needed to account for water vapor absorption, and the wind speed is needed to calculate the Fresnel reflection contribution on a rough ocean surface. The meteorological data mentioned above will be taken from the ECMWF model.

*Originating system:* ECMWF has developed one of the most comprehensive earth-system models available anywhere. The ECMWF model uses the '4D-Var' data assimilation approach, which provides a physically consistent best fit to observations. For this project the ERA-5 data are used.

*Data class:* Model

*Required ECMWF data:*

*Class:* ERA-5

*Stream:* Atmospheric model

*Type:* Analysis

*Dates:* 05/02/2019 to 31/12/21


*Time:* 00:00:00, 03:00:00, 06:00:00, 09:00:00, 12:00:00, 15:00:00, 18:00:00, 21:00:00

*Spatial grid:* T639 Quasi-uniform Gaussian grid (~0.28°, 31km)

Parameters at model levels:

- temperature, specific humidity (all levels)
- logarithm of surface pressure, geopotential (lowest level)

Parameters at surface:

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+) <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Page 11
		Version 1.0
		10. Feb 2023

- 10 meter U wind component
- 10 meter V wind component

*Data availability & coverage:* All data are required on a global scale, with a typical delay of three months.

*Data quality and reliability:* The ECMWF model data sets are widely considered to be among the best available data sets for meteorological parameters.

### 2.2.2 SRTM DEM

The RemoTeC retrieval algorithm for CO<sub>2</sub> and CH<sub>4</sub> columns from GOSAT-2 use information about the surface elevation from an extended SRTM digital elevation map.

*Originating system:* The original Shuttle Radar Telemetry Mission (SRTM) was provided by the United States National Aeronautics and Space Administration (NASA). The dataset used (DEM3) is based on the SRTM dataset and includes extrapolation and gap filling from various sources.

*Data class:* Model

*Sensor type and key technical characteristics:* n/a

*Data availability & coverage:* The original SRTM dataset provides elevation data ranging from 56 degrees south to 60 degrees north at a 90 meter resolution. The adjusted DEM3 dataset extends the coverage, while keeping the 90 meter resolution.

*Source data product name & reference to product technical specification documents:* <http://www.viewfinderpanoramas.org/dem3.html>.

### 2.2.3 TCCON FTS CO<sub>2</sub> and CH<sub>4</sub> data

TCCON data for CO<sub>2</sub> and CH<sub>4</sub> is available for public for all TCCON stations (<https://tccon-wiki.caltech.edu/>). We use the GGG2020 official release of the data product.

*Originating system:* Ground based


*Data class:* Ground based

*Sensor type and key technical characteristics:* The measurements are performed using the solar absorption spectroscopy in the near infrared using an FTS.

*Data availability & coverage:* Coverage is limited to the locations of the TCCON stations themselves. Depending on the instrument setting (land or sunglint), we use different TCCON stations for validation. Data is typically available only after one year, although some stations deliver on a more regular interval (3 months).

*Source data product name & reference to product technical specification documents:* Wunch et al., 2015, Laughner et al., 2021

*Data quantity:* Individual measurements can be taken in intervals of about 20 min. The observations can only be taken with the direct sunlight. This limits the amount of data, which is different from site to site.

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b>	Page 12
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	<b>for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)</b>	Version 1.0
		10. Feb 2023

*Data quality and reliability:* For XCO<sub>2</sub> the precision is 0.25% (1ppm) and the systematic error (bias) is 0.2% (0.8 ppm). For XCH<sub>4</sub> the precision is 0.40% (7ppb) and the systematic error (bias) is also 0.40% (7 ppb).

## 2.2.4 Additional input data

- Absorption cross sections: For the retrieval lookup-tables are used with pre-calculated absorption cross sections of the species of interest (O<sub>2</sub>, CO<sub>2</sub>, CH<sub>4</sub>, and H<sub>2</sub>O) as a function of wavenumber, temperature, and pressure. One lookup-table per species and per spectral band is being used. At the start of processing at a given CPU (Figure 3) the cross-section lookup table is read into memory.
- Aerosol optical properties: A lookup table is being used with pre-calculated aerosol optical properties (Mie and t-Matrix theory) as a function of size parameter and refractive index.
- Solar source: The solar source for the forward simulation uses data of extraterrestrial irradiance from the Total and Spectral Solar Irradiance Sensor-1 Hybrid Solar Reference Spectrum (TSIS-1 HSRS). It covers a temporally constant irradiance spectrum between 200 and 2700 nm with a spectral resolution of 0.001 nm (Coddington *et al.*, 2021). The TSIS-1 HSRS data are downloaded from [https://lasp.colorado.edu/lisird/data/tsis1\\_hsr](https://lasp.colorado.edu/lisird/data/tsis1_hsr).
- Retrieval settings: A file is read in with retrieval settings such as fit parameters, spectral range, etc.

## 2.3 Overview of Processing Sub-System

Figure 1 provides a schematic overview of the RemoTeC GHG-CCI+ processing sub system at SRON. The first step is to download the required data from the respective data servers to SRON. GOSAT-2 L1b and ECMWF data are dynamic datasets that are continuously updated, while the SRTM topography is a static dataset). In the next step a pre-processing program is combining all relevant information per GOSAT-2 ground pixel. This includes interpolation of ECMWF data in space and time to the coordinates of the GOSAT-2 ground pixel, calculating the average height of a GOSAT-2 ground pixel and its standard deviation from the topography database.

The pre-processor produces for each GOSAT-2 L1B file an auxiliary input file, hereafter referred to as the ‘retrieval ini’ file that contains this information. In the next step columns of CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>O, and O<sub>2</sub> are retrieved under the assumption of an atmosphere without aerosol/cirrus/cloud scattering (see ‘RemoTeC non scattering mode’ in Figure 1). The outcome of these retrievals is written in an ASCII file (intermediate output) and is used to generate the RemoTeC XCH<sub>4</sub> proxy product. Additionally, it is used for cloud filtering to select scenes to be processed by the RemoTeC Full Physics algorithm (ATBD\_CH4\_GO2\_SRPR, 2023). The Full Physics retrieval produces intermediate (ASCII) output files which go into an *a posteriori* filtering procedure, quality check (based on non-convergence, parameter boundary hits,

retrieved aerosol parameters), and bias correction and finally a NetCDF output file is created (*PUG\_CH4\_GO2\_SRFP, 2023; PUG\_CO2\_GO2\_SRFP, 2023*).

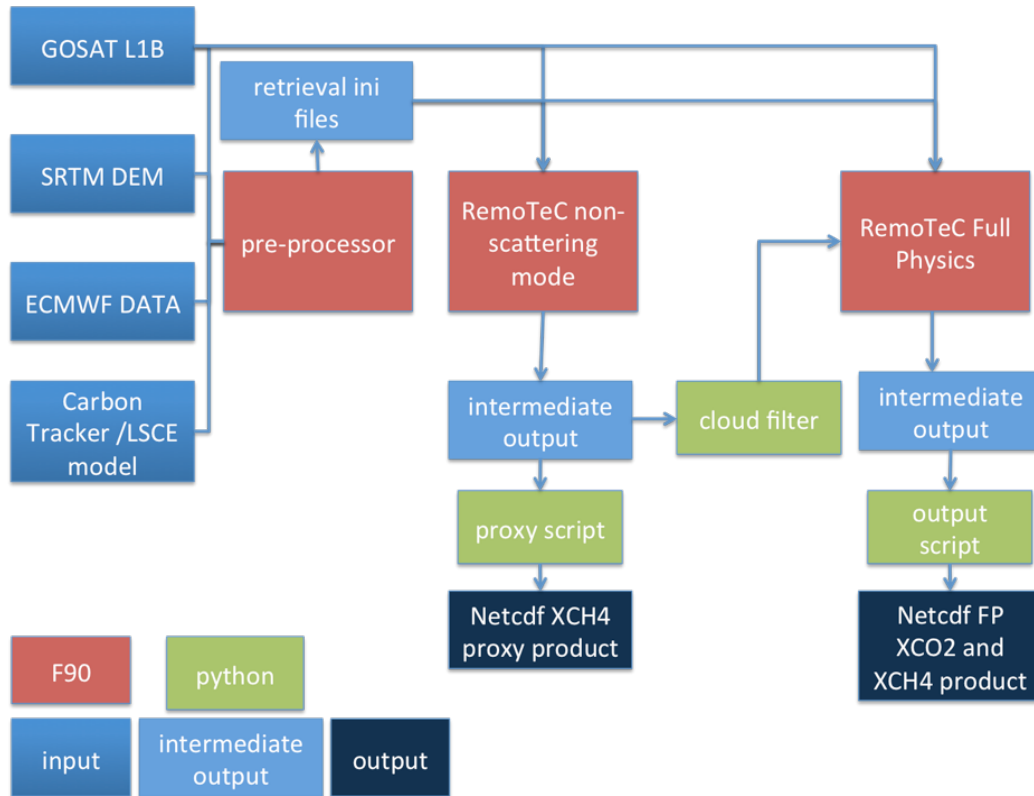


Figure 1: Schematic overview of the RemoTeC algorithm processing sub-system.

Figure 1 gives a schematic overview of the core RemoTeC retrieval algorithm (same for non-scattering and Full Physics). Here, multi-threading capability is implemented using *openMP*, where different ground pixels are divided over multiple threads. Figure 2 shows the processing per ground pixel (i.e. for a single thread) in more detail.

The static input that is required is a lookup table with the relevant absorption cross sections (read into memory at beginning of processing), a lookup table with aerosol optical properties, and a file indicating the retrieval settings (i.e. fit parameters, spectral range, etc.). Further, the auxiliary retrieval input files that are produced by the pre-processor are needed for each GOSAT-2 ground pixel to be processed, together with the GOSAT-2-FTS-2 Level 1b data. The retrieval per pixel is then run (iterative scheme with forward model and inversion module) and after convergence an intermediate (ASCII) output file is created that is used in the a posteriori filtering and quality check (see Figure 3) and the processing of the next ground pixel starts.

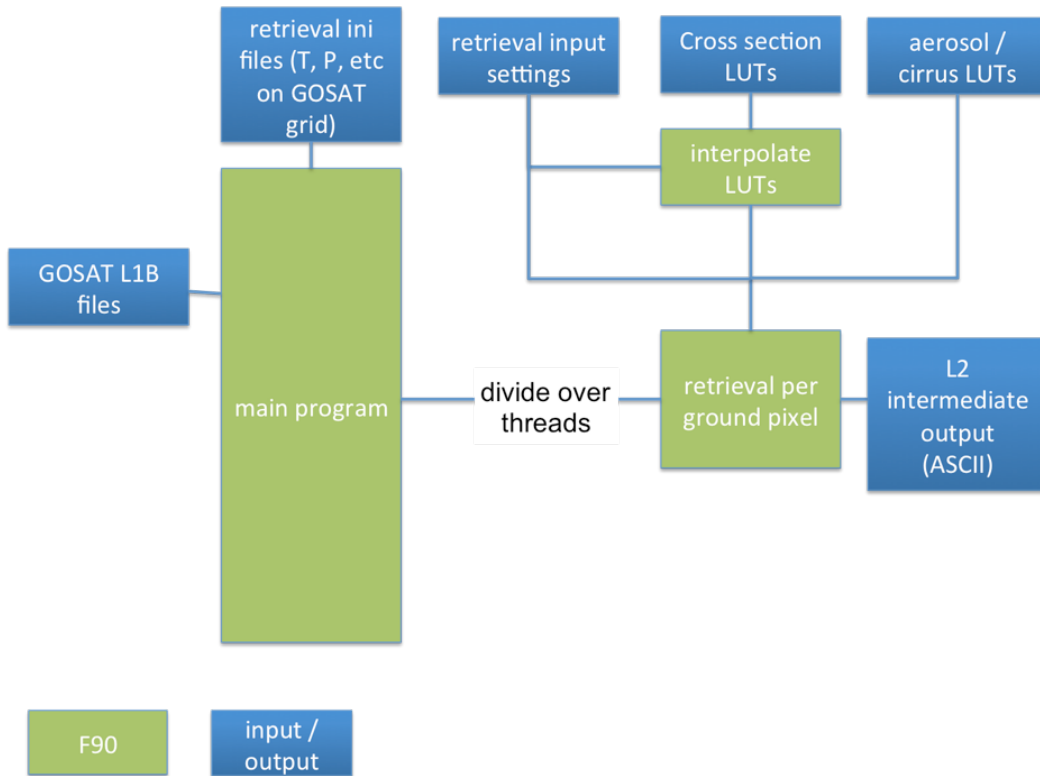


Figure 2: Schematic overview of the RemoTeC retrieval procedure including multi-threading.

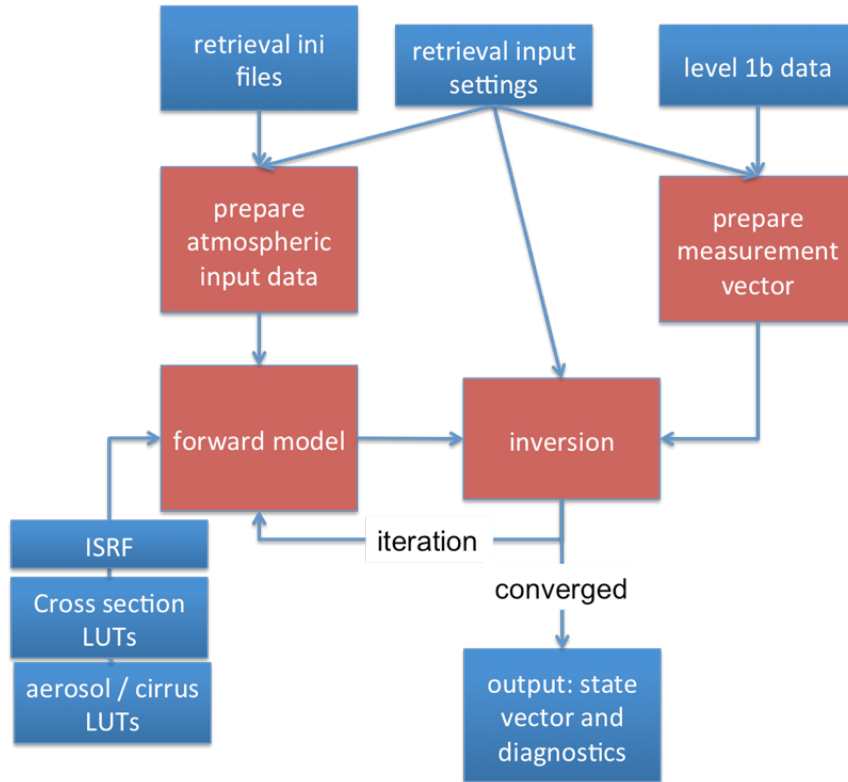



Figure 3: Overview of RemoTeC processing per ground pixel.

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 16
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

### 3 Algorithm

Any retrieval algorithm aims at inferring an atmospheric state vector  $\mathbf{x}$  from a measurement vector  $\mathbf{y}$ . The state vector is linked to the measurement vector through the true forward model  $\mathbf{f}(\mathbf{x}, \mathbf{b})$  that depends on the state vector  $\mathbf{x}$  and the vector  $\mathbf{b}$  containing ancillary parameters that are not retrieved,

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \mathbf{b}) + \mathbf{e}_y \quad (1)$$

Where  $\mathbf{e}_y$  represents the measurement noise vector. A retrieval method approximates the true forward model  $\mathbf{f}$  by a retrieval forward model  $\mathbf{F}$ , with a forward model error vector  $\mathbf{e}_F$ ,

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \mathbf{e}_y + \mathbf{e}_F \quad (2)$$

For full physics retrieval from the GOSAT-2 FTS-2 instrument the measurement vector contains the measured intensities in the NIR and SWIR (see Table 2).

Table 2: Spectral ranges from the NIR and SWIR band included in the measurement vector.

Band	Used spectral range
1 (O <sub>2</sub> a)	12950* – 13195 cm <sup>-1</sup>
2 (CO <sub>2</sub> )	6170 – 6277 cm <sup>-1</sup>
3 (CH <sub>4</sub> )	6045 – 6138 cm <sup>-1</sup>
4 (CO <sub>2</sub> )	4806 – 4896 cm <sup>-1</sup>

\* This is slightly different compared to the GOSAT-1 retrieval as it has been shortened from 12920 to 12950.

For the retrieval procedure it is needed that the non-linear forward model is linearized so that the retrieval problem can be solved iteratively. For iteration step  $n$  the forward model is approximated by


$$\mathbf{F}(\mathbf{x}, \mathbf{b}) \approx \mathbf{F}(\mathbf{x}_n, \mathbf{b}) + \mathbf{K}(\mathbf{x}_n - \mathbf{x}) \quad (3)$$

where  $\mathbf{x}_n$  is the state vector for the  $n$ -th iteration step and  $\mathbf{K}$  is the Jacobian matrix

$$\mathbf{K} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \quad (4)$$

Below, we will describe the retrieval forward model, state vector, ancillary parameter vector, and the inversion method in more detail.



 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 17
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

### 3.1 Forward Model

The retrieval forward model  $\mathbf{F}$  simulates the measurement vector  $\mathbf{y}$  for a given model atmosphere defined by the state vector  $\mathbf{x}$  and the ancillary parameter vector  $\mathbf{b}$ . The simulated intensity for a given spectral pixel  $i$  is given by

$$I_i = \int_{\lambda_{min}}^{\lambda_{max}} I(\lambda) S_i(\lambda) d\lambda \quad (5)$$

where  $I(\lambda)$  is the modeled intensity at high spectral resolution and  $S_i(\lambda)$  is the Instrument Spectral Response Function (ISRF) for spectral pixel  $i$ . In the NIR and SWIR channel  $I(\lambda)$  contains many fine spectral structures due to molecular absorption, so it has to be calculated at fine spectral resolution ( $0.1 \text{ cm}^{-1}$  in the NIR band and  $0.02 \text{ cm}^{-1}$  in the SWIR).

#### 3.1.1 Model Atmosphere and Optical Properties

For the RemoTeC algorithm described here the model atmosphere is defined for  $NLAY = 36$  homogeneous vertical layers that are equidistant in pressure, the lowest pressure level being defined by the surface pressure. The absorbing trace gases of interest are  $\text{O}_2$  (in the NIR band) and  $\text{CH}_4$ ,  $\text{H}_2\text{O}$ , and  $\text{CO}$  in the SWIR band. The layer sub-columns of these gases are for the first iteration step of each retrieval calculated from the input profiles of  $\text{CH}_4$ ,  $\text{CO}$  (TM5) and  $\text{H}_2\text{O}$  (ECMWF) and the temperature and pressure profiles (ECMWF). They are obtained on the grid of the model atmosphere by linear interpolation. Here, first the surface pressure  $p_{surf}$  is obtained by interpolation of the input pressure profile as function of height to the surface height (input) for the corresponding ground pixel. Next the pressure values at the layer boundaries are calculated, with the pressure  $p_k$  at the lower boundary of layer  $k$  (counting from top to bottom) is given by:


$$p_{lev,k} = p_{min} + \Delta p \cdot k \quad (6)$$

$$\Delta p = (p_{surf} - p_{min}) / NLAY \quad (7)$$

where  $p_{min}$  is the pressure value of the upper boundary of the input (ECMWF) atmosphere. The different atmospheric profiles are constructed on this pressure grid. For example, the methane sub-column  $DV_{CH_{4k}}$  for the layer bounded by pressure levels  $p_{lev,k}$  and  $p_{lev,k+1}$  is given by:

$$DV_{CH_{4k}} = XCH_{4k} DV_{AIR_k} \quad (8)$$

where  $XCH_{4k}$  is the methane dry air mixing ratio linearly interpolated from the input pressure grid to the pressure at the ‘middle’ of layer  $k$  defined by  $(p_k + p_{k+1})/2$ .  $DV_{AIR_k}$  is the sub-column of air in layer  $k$ , given by

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 18
		Version 1.0
		10. Feb 2023

$$DV\_AIR_k = (p_{lev,k+1} - p_{lev,k}) R / (M g_k \left(1 + \frac{XH_2O_k}{1.60855}\right)) \quad (9)$$

where  $R$  is Avogadro's number,  $M$  is the molecular mass of air,  $g_k$  is the gravity constant in altitude layer  $k$ , and 1.60855 is the mass of air relative to the mass of water (*Wunch et al., 2010*). The sub columns of CO and H<sub>2</sub>O are calculated in the same manner as for CH<sub>4</sub>, and the O<sub>2</sub> sub-column is obtained by multiplying the air sub-column by the O<sub>2</sub> mixing ratio (=0.2095).

For a radiative transfer calculation at a given wavelength the layer absorption optical thickness, scattering optical thickness, and scattering phase function for each model layer are needed. For layer  $k$  of the model atmosphere the CH<sub>4</sub> absorption optical thickness at wavelength  $\lambda_j$  is calculated by:


$$\tau_{abs,CH_4}(\lambda_j) = \frac{1}{N} \sum_{i=1}^N \sigma(p_i, T_i, \lambda_j) DV\_CH_4 \quad (10)$$

where  $N$  is the number of sub-layers in which the model atmosphere layers are divided (set at  $N = 2$ ) and  $\sigma$  is the absorption cross-section of CH<sub>4</sub> at wavelength  $\lambda_j$  and the pressure  $p_i$  and temperature  $T_i$  for the middle of model sub-layer  $i$ . The absorption optical thickness for the other trace gases is calculated in the same way. Pre-calculated absorption cross-sections for CH<sub>4</sub>, CO, H<sub>2</sub>O, and O<sub>2</sub> are stored as lookup-table as a function of pressure, temperature, and wavenumber. These cross-section lookup-tables are calculated from the latest spectroscopic databases (*Jenouvrier et al., 2007; Rothman et al., 2009*) assuming Voigt line shapes. For water vapor, an updated spectroscopic line list has been developed by *Scheepmaker et al., 2012*. In the NIR spectral range the absorption cross sections of O<sub>2</sub> are calculated according to *Tran et al., 2006* taking into account line mixing and collision induced absorption. The cross-section for pressure  $p_i$ , temperature  $T_i$  and wavelength  $\lambda_j$  are obtained by linear interpolation from the tabulated values. The reason that each of the  $NLAY$  layers of the model atmosphere is further divided into  $N$  sub-layers, is to properly account for the strong dependence of temperature and pressure of the absorption cross-sections. So, the absorption cross-sections are needed for  $NLAY \times N = 36 \times 2 = 72$  vertical layers equidistant in pressure. The total molecular absorption optical thickness is obtained by summing the contribution of the different trace gases. The Rayleigh scattering optical thickness for layer  $k$  and wavelength  $\lambda_j$  is given by

$$\tau_{ray,k}(\lambda_j) = \sigma_{ray}(\lambda_j) DV\_AIR_k \quad (11)$$

where  $\sigma_{ray}$  is the Rayleigh scattering cross section given by (*Buchholtz, 1995*)

$$\sigma_{ray}(\lambda) = A\lambda^{-(4+X)} \quad (12)$$

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 19
		Version 1.0
		10. Feb 2023

$$X = B\lambda + \frac{C}{\lambda} - D \quad (13)$$

with  $A = 4.02E-28$ ,  $B = 0.389$ ,  $C = 0.04926$ ,  $D = 0.3228$ . The Rayleigh scattering phase function is given by (e.g. *Hansen and Travis, 1974*)

$$P(\theta) = \frac{3}{4} (1 + \cos^2\theta) \frac{(1 - \delta)}{(1 + \delta/2)} \quad (14)$$

where  $\delta$  is the depolarization ratio and  $\theta$  is the scattering angle defined by

$$\cos(\theta) = -u_0 u_v + \sqrt{(1 - u_v)^2} \sqrt{(1 - u_0)^2} \cos(\phi_0 - \phi_v) \quad (15)$$

where  $u_0$  and  $u_v$  are the cosines of the solar and viewing zenith angle, respectively (absolute values) and  $\phi_0$  and  $\phi_v$  are the solar and viewing azimuth angles.

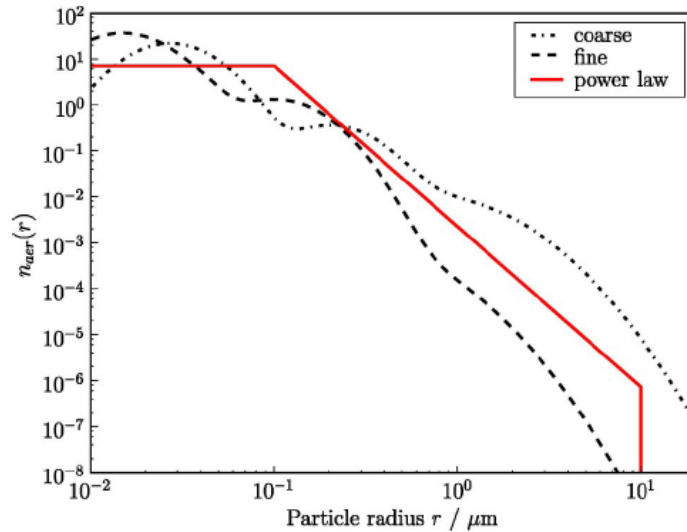



Figure 4: Aerosol size distribution  $n_{\text{aer}}(r)$  as a function of particle radius  $r$ . The retrieval method relies on a power law (red solid) size distribution. Also shown are more realistic Multi-modal lognormal size distributions for a fine mode (black dashed) and a coarse mode (black dotted) dominated aerosol type.

In addition to trace gases, aerosols are present in the model atmosphere. In our algorithm, aerosols are described by the following parameters (*Butz et al., 2009; 2010*):

1. Number of particles in each layer of the model atmosphere. This is provided by the total amount of particles,  $N_{\text{aer}}$ , and a normalized altitude distribution described by a Gaussian function of center height  $z_{\text{aer}}$  and width  $w_0$ . Hence, in model layer  $k$  with central height  $z_k$ , the number of particles is given by

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 20
		Version 1.0
		10. Feb 2023

$$h(z_k) = N_{\text{aer}} \exp \left[ -\frac{4 \ln(z_k - z_{\text{aer}})^2}{(2 w_0)^2} \right] \quad (16)$$

2. A size distribution of  $n_{\text{aer}}(r)$ , described by a power law size function, characterized by power law exponent and upper and lower cut-off (e.g. *Mishchenko et al., 1999*).

$$n(r) = \begin{cases} A & \text{for } r \leq r_1 \\ A(r/r_1)^{-\alpha} & \text{for } r_1 < r \leq r_2 \\ 0 & \text{for } r > r_2 \end{cases} \quad (17)$$

The cut-offs are  $r_1 = 0.1 \mu\text{m}$ ,  $r_2 = 10 \mu\text{m}$  and the constant  $A$  is determined from normalization of the size distribution. Figure 4 illustrates  $n_{\text{aer}}(r)$  and compares it to a more realistic multimodal lognormal size distribution (*Stier et al., 2005*). Through its parameter  $\alpha$ , the aerosol size distribution controls the spectral dependence of aerosol optical properties among the considered retrieval windows.

3. The complex refractive index  $m = m_r + i m_i$ , which is assumed independent of wavelength within a retrieval window.

From the aerosol size distribution, refractive index, and number of particles of each layer the aerosol scattering optical thickness  $\tau_{\text{scat,aer}}$  and aerosol absorption optical thickness  $\tau_{\text{abs,aer}}$  are calculated

$$\tau_{\text{scat,aer}}(z_k) = \sigma_{\text{scat,aer}} h(z_k) \quad (18)$$


$$\tau_{\text{abs,aer}}(z_k) = \sigma_{\text{abs,aer}} h(z_k) \quad (19)$$

where  $\sigma_{\text{scat,aer}}$  and  $\sigma_{\text{abs,aer}}$  are the aerosol scattering and absorption cross-section, respectively. They are obtained by:

$$\sigma_{\text{scat,aer}} = \sum_{i=1}^M K_{\text{scat},i}(m) r_i n_{\text{aer}}(r_i) v(r_i) \quad (20)$$

$$K_{\text{scat},i}(m) = \int_{\Delta \ln r_i} \frac{\sigma_{\text{scat}}(r, m)}{v(r)} d \ln r \quad (21)$$

where  $v$  denotes particle volume and  $K_{\text{scat},i}$ , representative for particle radius  $r_i$ , is pre-calculated for  $M$  size bins according to *Dubovik et al., 2006* and stored in a lookup table as function of aerosol size parameter  $x = 2\pi r/\lambda$ , real refractive index, and imaginary refractive index. The values for the actual aerosol refractive index are obtained by linear interpolation from the tabulated values. The lookup table contains values for spheres (Mie calculations) as

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 21
		Version 1.0
		10. Feb 2023

well as for spheroids with a pre-defined axis ratio distribution (*Dubovik et al., 2006*) but in our algorithm baseline we only consider spherical particles. Similar expressions hold for the absorption cross-section and the aerosol scattering phase function.

Finally, the total optical properties per layer in the model atmosphere are obtained by combining the contribution of gases and aerosols:

$$\tau_{\text{abs}} = \tau_{\text{abs,mol}} + \tau_{\text{abs,aer}} \quad (22)$$

$$\tau_{\text{sca}} = \tau_{\text{sca,mol}} + \tau_{\text{sca,aer}} \quad (23)$$

$$P(\theta) = \frac{\tau_{\text{sca,mol}}P_{\text{mol}}(\theta) + \tau_{\text{sca,aer}}P_{\text{aer}}(\theta)}{\tau_{\text{sca,mol}} + \tau_{\text{sca,aer}}} \quad (24)$$


For multiple scattering calculations the scattering phase function is needed in the form of expansion coefficients for generalized spherical functions, where expansion coefficient  $\alpha^l$  with index  $l$  is given by:

$$\alpha^l = \frac{2l+1}{2} \int_{-1}^1 P_{0,0}^l(\cos\theta)P(\theta)d\cos\theta \quad (25)$$

Where  $P_{0,0}^l$  is element (1,1) of the Generalized Spherical Function matrix (e.g. *de Haan et al., 1987*).

To summarize, the forward model needs the following inputs:

- Surface pressure to define the equidistant pressure grid
- Sub-columns of CH<sub>4</sub>, CO, H<sub>2</sub>O, O<sub>2</sub>, and air for the vertical layers of the model atmosphere.
- Pressure and temperature at the middle of the model sub-layers for absorption cross-sections.
- The aerosol column  $N_{\text{aer}}$ .
- The aerosol size parameter  $\alpha$  (power of the power law size distribution function).
- The central height  $z_{\text{aer}}$  and width  $w_0$  of the Gaussian aerosol altitude distribution.
- Solar Zenith Angle (SZA).
- Viewing Zenith Angle (VZA).
- Relative Azimuth Angle (RAA).
- The aerosol complex refractive index  $m = m_r + i m_i$ .
- A high spectral resolution solar reference spectrum.
- Lookup tables with absorption cross-sections of CH<sub>4</sub>, CO, H<sub>2</sub>O, and O<sub>2</sub> as function of pressure, temperature, and wavenumber.
- Lookup tables with pre-calculated aerosol properties as function of aerosol size parameter, real refractive index, and imaginary refractive index (cross sections and phase functions integrated over each size parameter bin).

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 22
		Version 1.0
		10. Feb 2023

Based on these inputs the optical properties can be calculated for each layer of the model atmosphere.

### 3.1.2 Modeling the top-of-atmosphere radiances

Based on the optical properties ( $\tau_{\text{abs}}$ ,  $\tau_{\text{sca}}$ ,  $P(\Theta)$ ) defined for each wavelength and layer of the model atmosphere we can compute the top-of-atmosphere radiance as measured by the instrument. The first step is to separate the radiation field in a singly scattered component  $I_{\text{ss}}$  and a multiply scattered component  $I_{\text{ms}}$ , respectively:

$$I = I_{\text{ss}} + I_{\text{ms}} \quad (26)$$

The computation of  $I_{\text{ss}}$  for a given wavelength is straightforward:

$$I_{\text{ss}} = F_0 \sum_{k=1}^{NLAY} \omega_k P_k(\Theta) \left[ 1 - e^{-\tau_{\text{tot},k} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \right] \frac{u_o}{4\pi(u_o + u_v)} e^{-\sum_{i=1}^k \tau_{\text{tot},i} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} + e^{-\frac{1}{u_o} \sum_{i=1}^{NLAY} \tau_{\text{tot},i} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} R_{\text{surf}} \quad (27)$$


where  $F_0$  is the incoming total flux,  $\tau_{\text{tot}} = \tau_{\text{abs}} + \tau_{\text{sca}}$ ,  $\omega = \tau_{\text{sca}}/\tau_{\text{tot}}$ ,  $u_o$  is the cosine of the solar zenith angle,  $u_v$  is the cosine of the viewing zenith angle, and  $R_{\text{surf}}$  is the surface reflection for the specific solar and viewing geometry under consideration. Besides the  $I_{\text{ss}}$  itself, also the derivatives with respect to  $\tau_{\text{sca},k}$ ,  $\tau_{\text{abs},k}$ ,  $\omega_k$ , and  $P_k$  are needed:

$$\frac{\partial I_{\text{ss}}}{\partial \tau_{\text{tot},k}} = F_0 \omega_k P_k(\Theta) \left[ e^{-\tau_{\text{tot},k} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \right] \frac{u_o}{4\pi(u_o + u_v)} e^{-\sum_{i=1}^k \tau_{\text{tot},i} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} - \sum_{i=1}^k \omega_k P_k(\Theta) \left[ 1 - e^{-\tau_{\text{tot},k} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \right] \frac{u_o}{4\pi(u_o + u_v)} e^{-\sum_{i=1}^k \tau_{\text{tot},i} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \quad (28)$$

$$\frac{\partial I_{\text{ss}}}{\partial P_k} = F_0 \frac{u_o}{4\pi(u_o + u_v)} e^{-\sum_{i=1}^k \tau_{\text{tot},i} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \left[ 1 - e^{-\tau_{\text{tot},k} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \right] \omega_k \quad (29)$$

$$\frac{\partial I_{\text{ss}}}{\partial \omega_k} = F_0 \frac{u_o}{4\pi(u_o + u_v)} e^{-\sum_{i=1}^k \tau_{\text{tot},i} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \left[ 1 - e^{-\tau_{\text{tot},k} \left( \frac{1}{u_o} + \frac{1}{u_v} \right)} \right] P_k \quad (30)$$

$$\frac{\partial I_{\text{ss}}}{\partial \tau_{\text{sca},k}} = \frac{\partial I_{\text{ss}}}{\partial \omega_k} \frac{1}{\omega_k} - \frac{\partial I_{\text{ss}}}{\partial \tau_{\text{tot},k}} \quad (31)$$

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b>	Page 23
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	<b>for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)</b>	Version 1.0
		10. Feb 2023

$$\frac{\partial I_{ss}}{\partial \tau_{abs,k}} = \frac{\partial I_{ss}}{\partial \tau_{tot,k}} - \frac{\omega_k}{\tau_{tot,k}} \frac{\partial I_{ss}}{\partial \omega_k} \quad (32)$$

The computation of the multiply scattered radiation for a given wavelength involves the solution of the plane-parallel radiative transfer equation. We solve this equation using the Gauss-Seidel iterative method. The solution is described in detail by *Landgraf et al., 2001* and will not be repeated here. The calculation of the derivatives with respect to optical properties per layer  $k$  of the model atmosphere is performed using the forward adjoint perturbation theory and is described in detail by *Hasekamp and Landgraf 2005*.

The radiative transfer calculations yield the derivatives of the radiance with respect to the optical parameters  $\tau_{abs}$ ,  $\tau_{sca}$ , and  $P$  or its expansion coefficients  $\alpha^l$ . From these derivatives, the derivatives with respect to a physical parameter  $x$  can be calculated in a straightforward manner by the derivative chain rule

$$\frac{\partial I_{ss}}{\partial x} = \sum_{k=1}^{NLAY} \frac{\partial I_{ss}}{\partial \tau_{sca,k}} \frac{\partial \tau_{sca,k}}{\partial x} + \frac{\partial I_{ss}}{\partial \tau_{abs,k}} \frac{\partial \tau_{abs,k}}{\partial x} + \frac{\partial I_{ss}}{\partial P_k} \frac{\partial P_k}{\partial x} \quad (33)$$

$$\frac{\partial I_{ms}}{\partial x} = \sum_{k=1}^{NLAY} \frac{\partial I_{ms}}{\partial \tau_{sca,k}} \frac{\partial \tau_{sca,k}}{\partial x} + \frac{\partial I_{ms}}{\partial \tau_{abs,k}} \frac{\partial \tau_{abs,k}}{\partial x} + \sum_{l=0}^M \frac{\partial I_{ms}}{\partial \alpha_k^l} \frac{\partial \alpha_k^l}{\partial x} \quad (34)$$

In order to avoid time consuming multiple scattering calculations on the high spectral resolution line-by-line grid we aim at reducing the number of spectral calculations, following the approach of *Hasekamp and Butz, 2008*. For this purpose, we consider the intensity  $I_{ms}$  as a function of total absorption optical thickness  $\tau_{abs}$  and its normalized vertical distribution  $\mathbf{n}$  (similar to  $k$ -distribution and spectral mapping methods),


$$I_{ms}(\lambda) = I_{ms}(\tau_{abs}(\lambda), \mathbf{n}(\lambda)) \quad (35)$$

where we assume that the atmospheric scattering properties and surface reflection properties are constant over the spectral range under consideration. Here, the explicit separation between total absorption optical thickness and its vertical distribution is chosen for later convenience. Element  $n_k$  of the vector  $\mathbf{n}$  represents the relative contribution of the absorption optical thickness of altitude layer  $k$  of the model atmosphere to the total absorption optical thickness of the atmosphere, such that

$$\tau_{abs}^k(\lambda) = n_k(\lambda) \tau_{abs}(\lambda) \quad (36)$$

where  $\tau_{abs,k}$  is the absorption optical thickness of layer  $k$  of the model atmosphere.

For a vertically homogeneous atmosphere the advantage of the description as function of absorption optical thickness is obvious. Namely, the intensity depends smoothly on absorption optical thickness, which means that only calculations at a limited number of values of  $\tau_{abs}$  are

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b>	Page 24
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	<b>for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)</b>	Version 1.0
		10. Feb 2023

needed from which the Stokes parameters at other values of  $\tau_{\text{abs}}$  can be obtained by interpolation (see e.g. *van Diedenhoven et al., 2006*). Finally, the Stokes parameters can be mapped back into wavelength space. To apply this procedure for a non-homogeneous atmosphere, one has to assume that the vertical distribution  $\mathbf{n}(z)$  of  $\tau_{\text{abs}}$  can be approximated by a vertical distribution that is independent of wavelength in the spectral interval under consideration. This is the underlying assumption of the correlated  $k$  method. For the simulation of moderate- or high resolution spectra in the near- and shortwave infrared spectral ranges, this assumption may cause errors of several percent in reflectance spectra for realistic inhomogeneous terrestrial atmospheres (see e.g. *Duan et al., 2005*).

Clearly, an efficient radiative transfer model is needed that accounts for the vertical distribution of absorption optical thickness at different wavelengths. For this purpose we use the transformation into absorption optical thickness space, and perform calculations for a limited number  $N$  values of  $\tau_{\text{abs},k}$  of the absorption optical thickness and corresponding vertical distributions  $\mathbf{n}_k$ , with  $k = 1, \dots, N$ . From the reference calculations  $I_{\text{ms}}(\tau_{\text{abs},k})$  we want to obtain the multiply scattered intensity vector  $I_{\text{ms}}(\lambda_j)$  at any wavelength  $\lambda_j$  in the spectral range under consideration with absorption optical thickness  $\tau_{\text{abs}}(\lambda_j)$  and its vertical distribution  $\mathbf{n}(\lambda_j)$ .

The basic principle of our linear- $k$  method is to account for differences between the actual vertical distribution  $\mathbf{n}(\lambda_j)$  and the vertical distributions  $\mathbf{n}^k$  used in the reference calculations, by employing the linear approximation:

$$I_{\text{ms}}(\tau_{\text{abs},k}, \mathbf{n}(\lambda_j)) \approx I_{\text{ms}}(\tau_{\text{abs},k}, \mathbf{n}^k) + \frac{\partial I_{\text{ms}}}{\partial \mathbf{n}} [\mathbf{n}(\lambda_j) - \mathbf{n}^k], \quad (37)$$


where the derivatives with respect to  $\mathbf{n}$  follow from the forward adjoint perturbation theory (*Hasekamp and Landgraf, 2005*). Applying this equation yields the intensity vector of the multiply scattered radiation at the grid points  $\tau_{\text{abs},k}$ , corrected for the actual vertical distribution. In order to obtain the intensity  $I_{\text{ms}}$  at  $\tau_{\text{abs}}(\lambda_j)$  we fit a second order polynomial to the logarithm of the (absorption profile corrected) intensities at the grid points, using the grid points closest to  $\tau_{\text{abs}}(\lambda_j)$  and the two neighbouring points. In order to correct for variation of scattering properties and surface albedo within the spectral range under consideration, also a linear correction is used.

The grid points are chosen equidistant on a logarithmic scale. For grid point  $k$  the total absorption optical thickness is given by:

$$\tau_{\text{grid},k} = e^{\log(\tau_{\text{min}}) + \frac{(k-1)}{N} [\log(\tau_{\text{max}}) - \log(\tau_{\text{min}})]} \quad (38)$$

where  $\tau_{\text{min}}$  and  $\tau_{\text{max}}$  are respectively the minimum and maximum absorption optical thickness in the spectral window under consideration. If  $\tau_{\text{max}}$  is larger than 15 its value is set to 15, because for larger values of the absorption optical thickness the radiation field is dominated



 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 25
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

by single scattering (being calculated exactly) and hence interpolation errors are of minor importance. The rationale of choosing a logarithmic scale is to obtain more grid points at small values of absorption optical thickness, where multiple scattering effects are most important. If a certain spectral range is influenced by considerable absorption by two or more species we use 2 grids: one for the target absorber and one for the total of other absorbers. For the radiative transfer calculations in the methane retrieval algorithm we use 10 grid points in the NIR band and  $5 \times 4 = 20$  grid points (5 for  $\text{CH}_4$  and 4 for  $\text{H}_2\text{O}$  and  $\text{CO}$  combined) in the SWIR band. Figure 5: Relative difference between a spectrum calculated using the linear- $k$  method and a spectrum obtained using line-by-line calculations. Upper panel shows the intensity and the lower panel the degree of linear polarization. The solid and dashed lines correspond to spectral resolutions of S5P and GOSAT-2, respectively. For the calculations a boundary layer aerosol was used with an optical thickness of 0.3 at 765 nm. Furthermore, we used a solar zenith angle (SZA) of  $50^\circ$  and a viewing zenith angle of  $0^\circ$ . illustrates the accuracy of the linear- $k$  method for the NIR spectral range. For more information on the linear- $k$  method we refer to the paper of *Hasekamp and Butz 2008*.

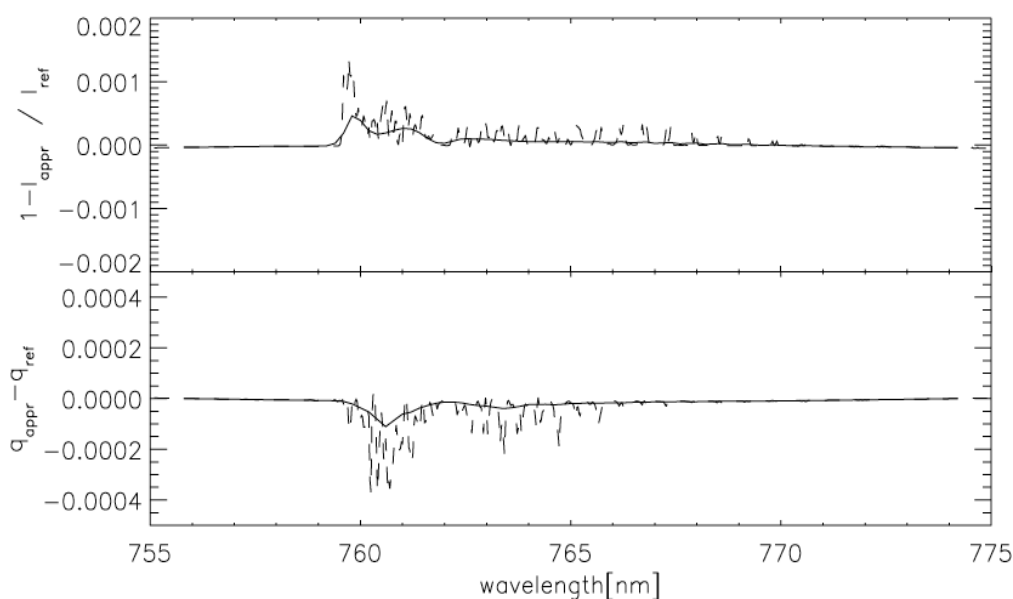



Figure 5: Relative difference between a spectrum calculated using the linear- $k$  method and a spectrum obtained using line-by-line calculations. Upper panel shows the intensity and the lower panel the degree of linear polarization. The solid and dashed lines correspond to spectral resolutions of S5P and GOSAT-2, respectively. For the calculations a boundary layer aerosol was used with an optical thickness of 0.3 at 765 nm. Furthermore, we used a solar zenith angle (SZA) of  $50^\circ$  and a viewing zenith angle of  $0^\circ$ .

### 3.1.3 Summary of Forward Model

The different steps of the forward model calculation are summarized in Figure 6.

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 26
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

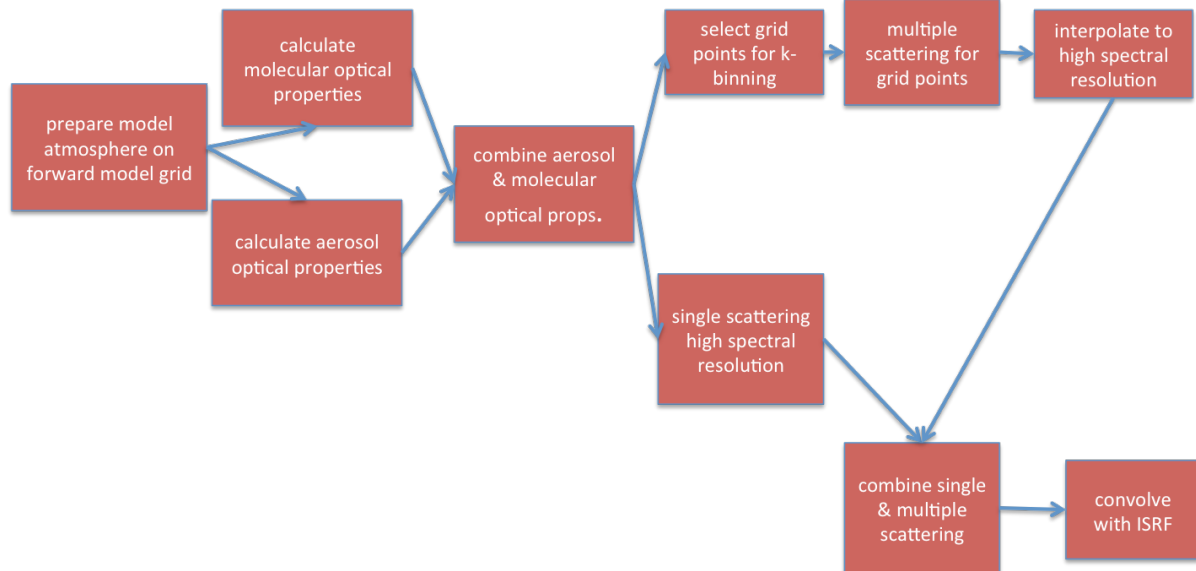


Figure 6: Overview of forward model.


### 3.1.4 Inverse algorithm

Definition of state vector and ancillary parameters are given as follows. The **state vector  $\mathbf{x}$**  contains the following elements (between brackets are optional elements):

- CO<sub>2</sub> sub-columns in 12 vertical layers (layer interfaces coincide with  $N_{LAY}$  layers of forward model grid).
- CH<sub>4</sub> sub-columns in 12 vertical layers (layer interfaces coincide with  $N_{LAY}$  layers of forward model grid).
- H<sub>2</sub>O total column.
- Aerosol parameter  $N_{aer}$  (number column).
- Aerosol parameter  $\alpha$  (size parameter).
- Aerosol parameter  $z_{aer}$  (central height of Gaussian height distribution).
- Lambertian surface albedo in all bands band.
- First order spectral dependence of surface albedo in all bands.
- Spectral shift of Earth radiances in all bands (higher orders optional).
- Intensity offset in each of the bands (in the GOSAT-1 retrieval only the O<sub>2</sub> A-Band has an intensity offset).\*
- (Offset in input temperature profile.)
- (Surface pressure.)

\*For convergence it was essential to include intensity offsets to each of the individual spectral windows. Potentially there are still irregularities in the quality of the individual spectral bands.

The **ancillary parameter vector  $\mathbf{b}$**  contains the following parameters:

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 27
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

- H<sub>2</sub>O sub-columns in 36 vertical layers of forward model grid.
- Temperature vertical profile at 72 layers of cross-section vertical grid.
- Pressure vertical profile at 72 layers of cross-section vertical grid.
- The aerosol complex refractive index, fixed at  $m = 1.4 - 0.01i$  for NIR and  $m = 1.47 - 0.008i$  for SWIR.
- The width  $w_0$  of the Gaussian aerosol height distribution (fixed at 2000 meter).

Table 3: A priori values for the different state vector elements.

State vector element	A priori value
CH <sub>4</sub> sub-columns	TM4
CO <sub>2</sub> sub-columns	Carbontracker
H <sub>2</sub> O total column	ECMWF
$N_{\text{aer}}$	Aerosol Optical Thickness (AOT) at 0.76 $\mu\text{m}$ = 0.1
$\alpha$	fixed at 3.5
$z_{\text{aer}}$	fixed at 5000 meter
surface albedo (NIR + SWIR)	no prior value needed (first guess at maximum of measured reflectance)
spectral shifts	no prior needed (first guess = 0)
temperature offset	no prior needed (first guess = 0)
surface pressure	ECMWF + SRTM DEM


### 3.1.5 Inversion Procedure

The inverse method optimizes the state vector  $\mathbf{x}$  with respect to the measurements  $\mathbf{y}$  after applying the forward model  $\mathbf{F}$  to  $\mathbf{x}$ . The inverse method is based by default on a Phillips-Tikhonov regularization scheme (*Phillips, 1962; Tikhonov, 1963; Hasekamp and Landgraf, 2005*). Regularization is required because the inverse problem is ill-posed, i.e., the measurements  $\mathbf{y}$  typically contain insufficient information to retrieve all state vector elements independently. The inverse algorithm finds  $\mathbf{x}$  by minimizing the cost function that is the sum of the least-squares cost function and a side constraint weighted by the regularization parameter  $\gamma$  according to

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} (\|\mathbf{S}_y^{-\frac{1}{2}}(\mathbf{F}(\mathbf{x}) - \mathbf{y})\|^2 + \gamma \|\mathbf{W}(\mathbf{x} - \mathbf{x}_a)\|^2) \quad (39)$$

where  $\mathbf{S}_y$  is the diagonal measurement error covariance matrix, which contains the noise estimate.  $\mathbf{x}_a$  is an a priori state vector (see Table 3), and  $\mathbf{W}$  is a weighting matrix (see below). For the linearized forward model for iteration step  $n$ , the equation for the updated state vector  $\mathbf{x}_{n+1}$  reduces to

$$\mathbf{x}_{n+1} = \min_{\mathbf{x}} \{ \|\mathbf{K}'(\mathbf{x}' - \mathbf{x}'_n) - \mathbf{y}'\|^2 + \gamma \|\mathbf{x}' - \mathbf{x}'_a\|^2 \} \quad (40)$$

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 28
		Version 1.0
		10. Feb 2023

with the weighted quantities  $\mathbf{x}' = \mathbf{W}\mathbf{x}$ ,  $\mathbf{y}' = \mathbf{S}_y^{-\frac{1}{2}}(\mathbf{y}-\mathbf{F}(\mathbf{x}_n))$ , and  $\mathbf{K}' = \mathbf{S}_y^{-\frac{1}{2}}\mathbf{K}\mathbf{W}^{-1}$ .

The solution reads

$$\mathbf{x}_{n+1} = \mathbf{G}'\mathbf{y}' + \mathbf{A}'\mathbf{x}'_n + (\mathbf{I} - \mathbf{A})\mathbf{x}'_{\text{apr}} \quad (41)$$

with  $\mathbf{A}'$  the averaging kernel matrix and  $\mathbf{G}'$  the contribution function matrix given by  $\mathbf{A}' = \mathbf{G}'\mathbf{K}'$  and  $\mathbf{G}' = (\mathbf{K}'^T\mathbf{K}' + \gamma\mathbf{I})^{-1}\mathbf{K}'^T$ . If the retrieval converges after a given number of steps  $N$  (typically 7-8), the final state vector  $\mathbf{x}_{\text{ret}} = \mathbf{x}_N$  is related to the true state vector and to the prior via

$$\mathbf{x}_{\text{ret}} = \mathbf{A}\mathbf{x}_{\text{true}} + (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{G}\mathbf{e}_y + \mathbf{G}\mathbf{e}_f \quad (42)$$

The covariance matrix  $\mathbf{S}_x$  describing the retrieval noise  $\mathbf{G}\mathbf{e}_y$  is given by

$$\mathbf{S}_x = \mathbf{G}\mathbf{S}_y\mathbf{G}^T \quad (43)$$

The target retrieval quantity is the column averaged dry air greenhouse gas mixing ratio, XGHG, where GHG denotes either CO<sub>2</sub> or CH<sub>4</sub>. This quantity is obtained from the CO<sub>2</sub> and CH<sub>4</sub> entries of the retrieved state vector through

$$\text{XGHG} = \frac{1}{V_{\text{air,dry}}} \mathbf{h}^T \mathbf{x}_{\text{retr}} \quad (44)$$


where  $\mathbf{h}$  is the total column operator for methane (summing up the partial columns in the state vector) and  $V_{\text{air,dry}}$  is the dry air column calculated from the surface pressure and water vapor profile, both obtained from a meteorological model (required as input). The retrieval noise  $\Delta\text{XCH}_4$  is given by

$$\Delta\text{XCH}_4 = \frac{\sum_{i,j=1}^{12} S_{x,i,j}}{V_{\text{air,dry}}} \quad (45)$$

and  $\Delta\text{XCO}_2$  can be obtained by a similar expression. For validation and application purposes it is important to realize that the retrieved XCH<sub>4</sub> or XCO<sub>2</sub> is in fact a representation of  $\mathbf{a}\mathbf{x}_{\text{true}}/V_{\text{air,dry}}$ , where the quantity

$$\mathbf{a} = \mathbf{h}^T \mathbf{A} \quad (46)$$

is referred to as the column averaging kernel (*Rodgers and Connor, 2003*)

 GHG-CCI+ project	ESA Climate Change Initiative “Plus” (CCI+)	Page 29
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1.0
		10. Feb 2023

### 3.1.6 Regularization of state vector and iteration strategy

To retrieve a meaningful state vector  $\mathbf{x}$ , the side-constrain in the minimization equation should be chosen in a way that contributions from measurement noise are minimized while retaining all valuable information in the first part of the merit-function. The inverse algorithm relies on a regularized Phillips-Tikhonov scheme. The diagonal weighting matrix  $\mathbf{W}$  is given by

$$\mathbf{W} = \mathbf{L}\mathbf{W}' \tag{47}$$

with

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & \dots & & & \\ -1 & 2 & -1 & & & \\ \vdots & & \ddots & & \vdots & \\ & & & -1 & 1 & \\ & & & \dots & & 1 & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix} \tag{48}$$

The upper left 12 by 12 sub matrix works on the state vector elements that contain the CH<sub>4</sub> sub-columns in the 12 altitude layers of the retrieval vertical grid. This sub-matrix corresponds to the matrix product


$$\mathbf{L}_1^T \mathbf{L}_1 \tag{49}$$

where  $\mathbf{L}_1$  is the first derivative matrix. The  $\mathbf{W}'_{jj}$  is given by  $1/\text{MAX}[K(1:12,1:12)]$  for the state vector elements corresponding to the 12 sub-columns of methane,  $1/\text{MAX}[K(:,j)]$  for the aerosol parameters, and 0 for all other parameters (which means they are not constrained by the side constraint and are retrieved in a least-squares sense). The value for  $\gamma$  is fixed such that the Degrees of Freedom for Signal (DFS) for the methane profile is in the range 1.0-1.5. This value is found empirically.

If the true solution is far from the current iteration  $\mathbf{x}_n$  and the forward model  $\mathbf{F}(\mathbf{x}_n)$  far from being linear, the linear approximation may fail and yield a solution  $\mathbf{x}_{n+1}$  that is further away from the true solution than  $\mathbf{x}_n$ . To avoid diverging retrievals, we use a Gauss-Newton scheme (see for instance *Rodgers, 2000*) with reduced step-size by introducing a filter factor that limits the update per iteration step of the state vector. The updated state vector is then given by:

$$\mathbf{x}_{n+1} = \Lambda \mathbf{G}'\mathbf{y}' + \mathbf{A}'\mathbf{x}'_n + (\mathbf{I} - \mathbf{A}')\mathbf{x}'_{\text{appr}} \tag{50}$$

with the filter factor  $\Lambda$  given by

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b>	Page 30
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	<b>for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)</b>	Version 1.0
		10. Feb 2023

$$\Lambda = \frac{1}{1 + \xi}, \xi > 0 \quad (51)$$

If  $\xi$  is large the update of the state vector is small. If  $\xi = 0$ , the equation for the updated state vector is equivalent to the pure Phillips-Tikhonov equation. The iteration is started with a large  $\xi$ , typically on the order of 10. It is then reduced or increased in the following iteration steps according to an empirically found scheme similar to Levenberg-Marquardt strategies (Rodgers, 2000). We accept solution  $\mathbf{x}'_{n+1}$  and decrease  $\xi$  by a factor of 2.5 if the least squares norm of iteration  $n+1$  is smaller than 1.1 times the least squares norm of iteration  $n$ . Otherwise, we discard the solution of iteration  $n+1$ , increase  $\xi$  by a factor of 2.5 and solve again for  $\mathbf{x}'_{n+1}$ . If  $\xi$  is smaller than a threshold value of 0.05, it is set to zero and the iteration is continued without a reduction in step size, assuming that the current state vector is sufficiently close to the true solution to finally converge.


### 3.1.7 Convergence criteria

The iteration is terminated and the retrieval is considered to have converged to a valid solution  $\mathbf{x}_{\text{ret}}$  if the following four conditions are all met:

1. The update of the state vector has become smaller than its theoretical uncertainty.
2. The state vector corresponding to CO<sub>2</sub> and/or CH<sub>4</sub> entries have never reached unrealistic values during the retrieval (negative methane densities for instance).
3. The merit-function has not increased for the current iteration step, and the step-size factor has reached 0.
4. The merit-function divided by the degrees of freedom is smaller than 2.0.

### 3.1.8 Scaling of O<sub>2</sub> Cross Sections

Butz *et al.*, 2011 proposed a method to discriminate between errors related to unaccounted light scattering effects and other sources of error using ocean-glint observations. The method has been demonstrated using measurements of the TANSO-FTS onboard GOSAT. Figure 7 (adapted from Butz *et al.*, 2013) motivates the general concept by comparing the total column concentration of molecular oxygen ([O<sub>2</sub>]) retrieved from TANSO-FTS O<sub>2</sub> A-band (~13 100 cm<sup>-1</sup>) spectra to [O<sub>2</sub>] calculated with high accuracy from meteorological support data. The retrievals neglect any (particulate as well as molecular) light scattering effects and, thus, deviations between meteorological and retrieved [O<sub>2</sub>] are to be expected. Figure 7, however, shows that these deviations exhibit a quite different pattern for land-nadir and ocean-glint observations. For land nadir, the distribution scatters rather symmetrically about the occurrence peak with a long tail to strong overestimation as well as underestimation. For ocean glint, in contrast, the distribution exhibits a sharp “upper edge” and wide lower tail. The occurrence peak for ocean glint is offset to lower retrieved [O<sub>2</sub>] with respect to the land-nadir retrievals. The rationale to explain the observed pattern assumes that, for ocean glint, virtually all scattering effects result in a shortening of the light path while, for land nadir, scattering

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document (ATBD)</b> for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Page 31
		Version 1.0
		10. Feb 2023

effects can have a light path enhancing as well as a light path shortening effect. Light path enhancement typically requires a scattering event at the Earth’s surface, which, however, is dark for the off-glint ocean. Retrievals of  $[O_2]$  from ocean glint observations therefore virtually always suffer from unaccounted net light path shortening and, thus, yield an underestimation of the true  $[O_2]$  if the retrieval algorithm neglects light scattering effects. For land nadir, the trade-off between light path shortening and light path enhancement is largely controlled by surface albedo, which typically varies substantially for land surfaces. This rationale fits the pattern observed in Figure 7, explaining the sharp upper edge and wide lower tail for ocean glint as well as the rather symmetric distribution for land nadir.

If this argument is true, simple “non-scattering”  $[O_2]$  retrievals from the  $O_2$  A-band can be used to find an ensemble of clean ocean-glint scenes where light path modification due to scattering effects is negligible. Following our rationale above, the upper edge of ocean-glint retrievals, i.e. the retrievals which show the least underestimation of the expected meteorological  $O_2$  concentration, are the scenes with the least contamination by scattering effects. For the upper edge ensemble, light scattering as a source of error is negligible, making it a good candidate ensemble to examine other sources of error such as spectroscopic uncertainties or instrumental deficiencies. Using this method, we found the need to scale the  $O_2$  absorption cross sections by a factor 1.03 in order to find agreement with the  $O_2$  column inferred from ECMWF data.

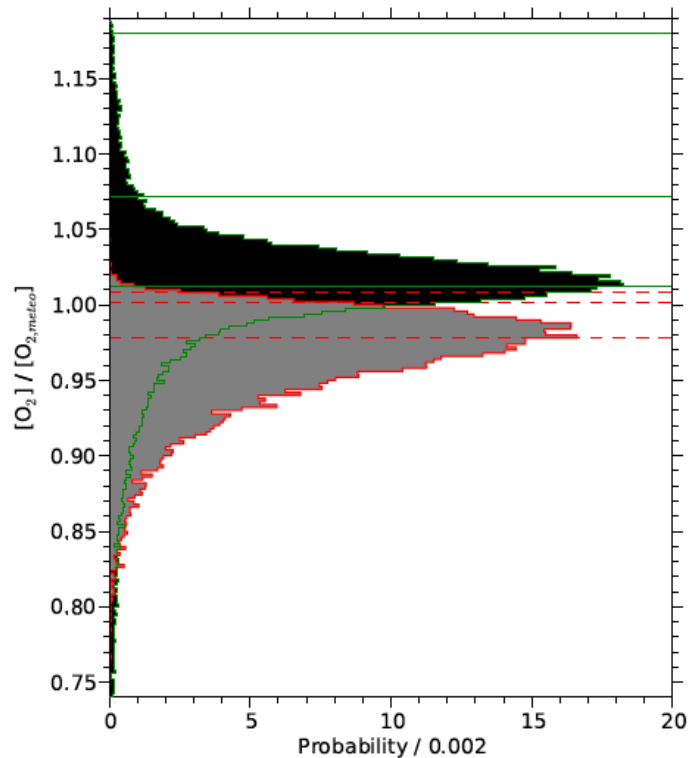




Figure 7: Histogram of the ratio between the total column  $O_2$  concentration retrieved from TANSO-FTS  $O_2$  A-band spectra ( $[O_2]$ ) and the expected meteorological  $O_2$  total column concentration ( $[O_{2,meteo}]$ ). Histograms are shown separately for retrievals from nadir soundings over land (green lines, black fill) and from glint spot soundings over the ocean (red lines, grey fill). The horizontal lines depict the 99th and 95th percentiles and the occurrence peak (from top to bottom). Retrievals neglect molecular as well as particulate light scattering effects and cover more than 3 yr of GOSAT operations from June 2009 to September 2012 (1 million nadir soundings, 0.3 million glint soundings). Basic quality screening is based on convergence of the iterative algorithm, instrument anomaly flagging, goodness of fit, and cirrus contamination detectable at the  $5160\text{ cm}^{-1}$  water vapor absorption band.



 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document (ATBD)</b> for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Page 33
		Version 1.0
		10. Feb 2023

## 4 Output data

The output data are stored in one NetCDF file per day. The file size varies between 1 and 5 Mb. Separate output files are generated for the XCH<sub>4</sub> and the XCO<sub>2</sub> Full Physics products. The format of the main output data, which are the Level 2 data products, is described in the separate document (*PUG\_CH4\_GO2\_SRF*, 2023; *PUG\_CO2\_GO2\_SRF*, 2023).

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b>	Page 34
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	<b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Version 1.0
		10. Feb 2023

## References

**ATBD\_CH4\_GO2\_SRPR, 2023:** Barr, A.G., and the GHG-CCI+ group at SRON: Algorithm Theoretical Basis Document (ATBD) Version 1.0 for the RemoTeC XCH4 GOSAT-2 SRON Proxy Product (CH4\_GO2\_SRPR version 2.0.2) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), 2023.

**Bucholtz et al., 1995:** Bucholtz, A., Rayleigh-scattering calculations for the terrestrial atmosphere," *Appl. Opt.* 34, 2765-2773, 1995.

**Butz et al., 2009:** Butz, A., Hasekamp, O. P., Frankenberg, C., Aben, I. (2009). Retrievals of atmospheric CO<sub>2</sub> from simulated space-borne measurements of backscattered near-infrared sunlight: Accounting for aerosol effects. *Appl. Opt.*, 48, 3322.

**Butz et al., 2010:** Butz, A., Hasekamp, O. P., Frankenberg, C., Vidot, J., and Aben, I.: CH<sub>4</sub> retrievals from space-based solar backscatter measurements: Performance evaluation against simulated aerosol and cirrus loaded scenes, *J. Geophys. Res.*, 115, D24302, doi:10.1029/2010JD014514, 2010.

**Butz et al., 2011:** Butz, A., Guerlet, S., Hasekamp, O., et al., Toward accurate CO<sub>2</sub> and CH<sub>4</sub> observations from GOSAT, *Geophys. Res. Lett.*, doi:10.1029/2011GL047888, 2011.

**Duan et al., 2005:** Duan, M., Min, Q., and Li, J. (2005), A fast radiative transfer model for simulating high-resolution absorption bands, *J. Geophys. Res.*, 110, D15201, doi:10.1029/2004JD005590.


**Van Diedenhoven et al., 2006:** van Diedenhoven, B., Hasekamp, O.P. and Landgraf, J., 2006. Efficient vector radiative transfer calculations in vertically inhomogeneous cloudy atmospheres. *Applied optics*, 45(23), pp.5993-6006.

**Dubovik et al., 2006:** Dubovik O., Sinyuk A., Lapyonok T., Holben B. N., Mishchenko M., Yang P., Eck T. F., Volten H., Muñoz O., Veihelmann B., van der Zande W. J., Leon J.-F., Sorokin M., Slutsker I. 2006. Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust. *J. Geophys. Res. D* 111, D11208.

**Guerlet et al., 2013:** Guerlet, S., A. Butz, D. Schepers, S. Basu, O. P. Hasekamp, A. Kuze, T. Yokota, J.-F. Blavier, N. M. Deutscher, D. W. T. Griffith, F. Hase, E. Kyro, I. Morino, V. Sherlock, R. Sussmann, A. Galli and I. Aben (2013) Impact of aerosol and thin cirrus on retrieving and validating XCO<sub>2</sub> from GOSAT shortwave infrared measurements, *J. Geophys. Res.*, doi: 10.1002/jgrd.50332

**De Haan et al., 1987:** de Haan, J. F., Bosma, P. B., Hovenier, J. W. (1987), The adding method for multiple scattering calculations of polarized light, *Astron. and Astrophys.*, 183, 371 – 391.

**Hansen and Travis, 1974:** Hansen, J.E. and Travis, L.D.: Light scattering in planetary atmospheres. *Space science reviews*, 16(4), pp.527-610, 1974.

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b> <b>Algorithm Theoretical Basis Document</b> <b>(ATBD)</b> <b>for the Essential Climate Variable (ECV)</b> <b>Greenhouse Gases (GHG)</b>	Page 35
		Version 1.0
		10. Feb 2023

**Hasekamp and Landgraf, 2002:** Hasekamp, O. P. and Landgraf, J. (2002), A linearized vector radiative transfer model for atmospheric trace gas retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, 75, 221–238.

**Hasekamp and Landgraf, 2005:** Hasekamp, O. P. and Landgraf, J. (2005), Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing, *J. Geophys. Res.*, 110, D04203.

**Hasekamp and Butz, 2008:** Hasekamp, O. P. and Butz, A. (2008), Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres, *J. Geophys. Res.*, 113, D20309.

**Jenouvrier et al., 2007:** Jenouvrier, A., Daumont, L., Régalia-Jarlot, L., Tyuterev, V. G., Carleer, M., Vandaele, A. C., Mikhailenko, S., Fally, S. (2007), Fourier transform measurements of water vapor line parameters in the 4200 – 6600 cm<sup>-1</sup> region, *J. Quant. Spectrosc. Ra.*, 105, 326 – 355.

**Landgraf et al., 2001:** Landgraf, J., O. Hasekamp, T. Trautmann, and M. Box (2001), A linearized radiative transfer model for ozone profile retrieval using the analytical forward-adjoint perturbation theory, *J. Geophys. Res.*, 106, 27,291 – 27,306.

**Laughner et al., 2021:** Laughner, J., Andrews, A., Roche, S., Kiel, M., and Toon, G. *input v1.0.10: GGG2020 prior profile software (Version 1.0.10)*, <https://doi.org/10.22002/D1.1944>, 2021.

**Mishchenko et al., 1999:** Mishchenko, M. I., Geogdzhayev, I. V. , Cairns, B., Rossow, W. B., Lacis, A. A. (1999), Aerosol retrievals over the ocean using channel 1 and 2 AVHRR data: A sensitivity analysis and preliminary results, *Appl. Opt.*, 38, 7325 – 7341.


**Nakajima et al., 2017:** M. Nakajima, Hiroshi Suto, K. Yotsumoto, K. Shiomi, T. Hirabayashi, "Fourier transform spectrometer on GOSAT and GOSAT-2," *Proc. SPIE 10563, International Conference on Space Optics — ICSO 2014, 105634O (17 November 2017)*; <https://doi.org/10.1117/12.2304062>.

**Phillips et al., 1962:** Phillips, D.L. (1962), A technique for the numerical solution of certain integral equations of the first kind, *J. Assoc. Comput. Mach.*, 9, 84 – 97.

**PUG\_CH4\_GO2\_SRFP, 2023,** Barr, A.G., and the GHG-CCI+ group at SRON: ESA Climate Change Initiative “Plus” (CCI+) Product User Guide (PUG) Version 1.0 for the RemoTeC XCH4 GOSAT-2 Full Physics Product (CH4\_GO2\_SRFP version 2.0.2) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), ESA GHG CCI+, 2023.

**PUG\_CO2\_GO2\_SRFP, 2023,:** Barr, A.G., and the GHG-CCI+ group at SRON: ESA Climate Change Initiative “Plus” (CCI+) Product User Guide (PUG) Version 1.0 for the RemoTeC XCO2 GOSAT-2 Full Physics Product (CO2\_GO2\_SRFP version 2.0.2) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), ESA GHG CCI+, 2023.

**Rodgers, 2000:** Rodgers, C.D. (2000), *Inverse methods for atmospheric sounding: Theory and practice*, World Sc., River Edge, USA.

 GHG-CCI+ project	<b>ESA Climate Change Initiative “Plus” (CCI+)</b>	Page 36
	<b>Algorithm Theoretical Basis Document (ATBD)</b>	
	<b>for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)</b>	Version 1.0
		10. Feb 2023

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

**Rothman et al., 2009:** Rothman, L. S., Gordon, I. E., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Boudon, V., Brown, L. R., Campargue, A., Champion, J. P., et al. (2009), The HITRAN 2008 molecular spectroscopic database, *J. Quant. Spectrosc. Ra.*, 110, 533 – 572.

**Scheepmaker et al., 2012:** Scheepmaker, R. A., Frankenberg, C., Galli, A., Schrijver, H., Fally, S., Deutscher, N. M., Wunch, D., Warneke, T., Aben, I. (2012), Improved water vapour spectroscopy in the 4174 – 4300 cm<sup>-1</sup> region and its impact on SCIAMACHY’s HDO/ H<sub>2</sub>O measurements, *Atmos. Meas. Tech.*, in preparation.

**Schepers et al., 2012:** Schepers, D., et al. (2012), Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, *J. Geophys. Res.*, 117, D10307, doi:[10.1029/2012JD017549](https://doi.org/10.1029/2012JD017549).

**Stier et al., 2005:** Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., Ganzeveld, L., Tegen, I., Werner, M., Balkanski, Y., Schulz, M., Boucher, O., Minikin, A., Petzold, A. (2005), The aerosol-climate model ECHAM5-HAM, *Atmos. Chem. Phys.*, 5, 1125 – 1156.

**Tikhonov et al., 1963:** Tikhonov, A. (1963), On the solution of incorrectly stated problems and a method of regularization, *Dokl., Akad., Nauk SSSR*, 151, 501-504.

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

**Wunch et al., 2010:** Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, *Atmospheric Measurement Techniques*, 3, 1351–1362, doi:10.5194/amt-3-1351-2010, URL <http://www.atmos-meas-tech.net/3/1351/2010/>, 2010.

**Wunch et al., 2015:** Wunch, D., Toon, G.C., Sherlock, V., Deutscher, N.M., Liu, X., Feist, D.G., Wennberg, P.O., The Total Carbon Column Observing Network’s GGG2014 Data Version. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA (available at: doi:10.14291/tcon.ggg2014.documentation.R0/1221662), 2015.