Algorithm Theoretical Basis Document (ATBD) – ANNEX C for product CH4_GOS_SRPR (v2.3.9, 2009-mid2020)

C3S_312b_Lot2_DLR – Atmosphere

Issued by: Lianghai Wu, SRON, The Netherlands
Date: 18/12/2021
Ref: C3S_D312b_Lot2.1.3.2-v3.0_ATBD-GHG ANNEX-C_v5.0
Official reference number service contract: 2018/C3S_312b_Lot2_DLR/SC1
Contributors

INSTITUTE OF ENVIRONMENTAL PHYSICS (IUP), UNIVERSITY OF BREMEN, BREMEN, GERMANY (IUP)
M. Buchwitz

SRON NETHERLANDS INSTITUTE FOR SPACE RESEARCH, UTRECHT, THE NETHERLANDS (SRON)
I. Aben
L. Wu
O. P. Hasekamp
Table of Contents

History of modifications 5
Related documents 6
Acronyms 7
General definitions 9
Scope of document 10
Executive summary 11

1. Data product overview 12
1.1 Column-averaged mixing ratios of CH₄ (XCH₄) 12

2. Input and auxiliary data 12
2.1 Satellite instrument 12
   2.1.1 GOSAT TANSO-FTS level 1b 12
2.2 Other 14
   2.2.1 ECMWF model data 14
   2.2.2 Carbon Tracker Data 15
   2.2.3 CO2 correction 15
   2.2.4 SRTM DEM 16
   2.2.5 TCCON FTS CO₂ and CH₄ data 16
   2.2.6 Additional input data 17
2.3 Overview of Processing Sub-System 18

3. Algorithms 21
3.1 Algorithm description 21
3.2 Forward Model 22
   3.2.1 Model Atmosphere and Optical Properties 22
   3.2.2 Modeling the top-of-atmosphere radiances 23
3.3 Inverse algorithm 24
3.4 Inversion Procedure 25
   3.4.1 Regularization of state vector and iteration strategy 27
3.5 Cloud Filtering 28

4. Output data 28

References 29
## History of modifications

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description of modification</th>
<th>Chapters / Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>4-October-2018</td>
<td>Update for CDR2 (2009-2017)</td>
<td>All</td>
</tr>
<tr>
<td>3.1</td>
<td>03-November-2019</td>
<td>Update after review by Assimila: Correction of typos, some links added and some additional explanations given.</td>
<td>All</td>
</tr>
<tr>
<td>4.0</td>
<td>18-August-2020</td>
<td>Update for CDR4 (2009-2019)</td>
<td>All</td>
</tr>
<tr>
<td>5.0</td>
<td>18-February-2021</td>
<td>Update for CDR5 (2009-mid2020)</td>
<td>All</td>
</tr>
</tbody>
</table>
## Related documents

<table>
<thead>
<tr>
<th>Reference ID</th>
<th>Document</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Main ATBD: Buchwitz, M., et al., Algorithm Theoretical Basis Document (ATBD) – Main document for Greenhouse Gas (GHG: CO₂ &amp; CH₄) data set CDR 5 (01.2003-06.2020), project C3S_312b_Lot2_DLR – Atmosphere, v5.0, 2021. <em>(this document is an ANNEX to the Main ATBD)</em></td>
</tr>
</tbody>
</table>
## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATBD</td>
<td>Algorithm Theoretical Basis Document</td>
</tr>
<tr>
<td>CAMS</td>
<td>Copernicus Atmosphere Monitoring Service</td>
</tr>
<tr>
<td>C3S</td>
<td>Copernicus Climate Change Service</td>
</tr>
<tr>
<td>CDR</td>
<td>Climate Data Record</td>
</tr>
<tr>
<td>CDS</td>
<td>(Copernicus) Climate Data Store</td>
</tr>
<tr>
<td>CPU</td>
<td>Core Processing Unit</td>
</tr>
<tr>
<td>DEM3</td>
<td>Digital Elevation Map 3</td>
</tr>
<tr>
<td>DFS</td>
<td>Degrees of Freedom for Signal</td>
</tr>
<tr>
<td>DHF</td>
<td>Data Handling Facility</td>
</tr>
<tr>
<td>ECMWF</td>
<td>European Centre for Medium Range Weather Forecasting</td>
</tr>
<tr>
<td>ESA</td>
<td>European Space Agency</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>EUMETSAT</td>
<td>European Organisation for the Exploitation of Meteorological Satellites</td>
</tr>
<tr>
<td>FP</td>
<td>Full Physics retrieval method</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier Transform InfraRed</td>
</tr>
<tr>
<td>FTS</td>
<td>Fourier Transform Spectrometer</td>
</tr>
<tr>
<td>GCOS</td>
<td>Global Climate Observing System</td>
</tr>
<tr>
<td>GHG</td>
<td>GreenHouse Gas</td>
</tr>
<tr>
<td>GOSAT</td>
<td>Greenhouse Gases Observing Satellite</td>
</tr>
<tr>
<td>IUP</td>
<td>Institute of Environmental Physics (IUP) of the University of Bremen, Germany</td>
</tr>
<tr>
<td>JAXA</td>
<td>Japan Aerospace Exploration Agency</td>
</tr>
<tr>
<td>KIT</td>
<td>Karlsruhe Institute of Technology</td>
</tr>
<tr>
<td>L1</td>
<td>Level 1</td>
</tr>
<tr>
<td>L2</td>
<td>Level 2</td>
</tr>
<tr>
<td>L3</td>
<td>Level 3</td>
</tr>
<tr>
<td>L4</td>
<td>Level 4</td>
</tr>
<tr>
<td>LMD</td>
<td>Laboratoire de Météorologie Dynamique</td>
</tr>
<tr>
<td>MACC</td>
<td>Monitoring Atmospheric Composition and Climate, EU GMES project</td>
</tr>
<tr>
<td>NA</td>
<td>Not applicable</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NetCDF</td>
<td>Network Common Data Format</td>
</tr>
<tr>
<td>NIES</td>
<td>National Institute for Environmental Studies</td>
</tr>
<tr>
<td>NIR</td>
<td>Near Infra Red</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration</td>
</tr>
<tr>
<td>Obs4MIPs</td>
<td>Observations for Climate Model Intercomparisons</td>
</tr>
<tr>
<td>ppb</td>
<td>Parts per billion</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>PR</td>
<td>(light path) PRoxy retrieval method</td>
</tr>
<tr>
<td>PVIR</td>
<td>Product Validation and Intercomparison Report</td>
</tr>
<tr>
<td>RAA</td>
<td>Relative Azimuth Angle</td>
</tr>
<tr>
<td>RMS</td>
<td>Root-Mean-Square</td>
</tr>
<tr>
<td>RTM</td>
<td>Radiative transfer model</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SRON</td>
<td>SRON Netherlands Institute for Space Research</td>
</tr>
<tr>
<td>SRTM</td>
<td>Shuttle Radar Topography Mission</td>
</tr>
<tr>
<td>SWIR</td>
<td>Short Wave Infra Red</td>
</tr>
<tr>
<td>SZA</td>
<td>Solar Zenith Angle</td>
</tr>
<tr>
<td>TANSO</td>
<td>Thermal And Near infrared Sensor for carbon Observation</td>
</tr>
<tr>
<td>TANSO-FTS</td>
<td>Fourier Transform Spectrometer on GOSAT</td>
</tr>
<tr>
<td>TBC</td>
<td>To be confirmed</td>
</tr>
<tr>
<td>TBD</td>
<td>To be defined / to be determined</td>
</tr>
<tr>
<td>TCCON</td>
<td>Total Carbon Column Observing Network</td>
</tr>
<tr>
<td>TIR</td>
<td>Thermal Infra Red</td>
</tr>
<tr>
<td>TMS</td>
<td>Transport Model 5</td>
</tr>
<tr>
<td>TR</td>
<td>Target Requirements</td>
</tr>
<tr>
<td>TRD</td>
<td>Target Requirements Document</td>
</tr>
<tr>
<td>VZA</td>
<td>Viewing Zenith Angle</td>
</tr>
</tbody>
</table>
General definitions

Table 1 lists some general definitions relevant for this document.

Table 1: General definitions.

<table>
<thead>
<tr>
<th>Item</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>XCO₂</td>
<td>Column-averaged dry-air mixing ratios (mole fractions) of CO₂</td>
</tr>
<tr>
<td>XCH₄</td>
<td>Column-averaged dry-air mixing ratios (mole fractions) of CH₄</td>
</tr>
<tr>
<td>L₁</td>
<td>Level 1 satellite data product: geolocated radiance (spectra)</td>
</tr>
<tr>
<td>L₂</td>
<td>Level 2 satellite-derived data product: Here: CO₂ and CH₄ information for each ground-pixel</td>
</tr>
<tr>
<td>L₃</td>
<td>Level 3 satellite-derived data product: Here: Gridded CO₂ and CH₄ information, e.g., 5 deg times 5 deg, monthly</td>
</tr>
<tr>
<td>L₄</td>
<td>Level 4 satellite-derived data product: Here: Surface fluxes (emission and/or uptake) of CO₂ and CH₄</td>
</tr>
</tbody>
</table>
Scope of document

This document is an Algorithm Theoretical Basis Document (ATBD) for the Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/) greenhouse gas (GHG) component as covered by project C3S_312b_Lot2.

Within this project satellite-derived atmospheric carbon dioxide (CO₂) and methane (CH₄) Essential Climate Variable (ECV) data products will be generated and delivered to ECMWF for inclusion into the Copernicus Climate Data Store (CDS) from which users can access these data products and the corresponding documentation.

The GHG satellite-derived data products are:
- Column-averaged dry-air mixing ratios (mole fractions) of CO₂ and CH₄, denoted XCO₂ (in parts per million, ppm) and XCH₄ (in parts per billion, ppb), respectively.
- Mid/upper tropospheric mixing ratios of CO₂ (in ppm) and CH₄ (in ppb).

This document describes the retrieval algorithm used to generate the C3S product CH4_GOS_SRPR.

These product is the XCH₄ Level 2 product as retrieved from GOSAT using algorithms developed at SRON, The Netherlands.
Executive summary

This document describes the RemoTeC algorithm for GHG retrieval from the GOSAT instrument. The 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 real GOSAT data (Butz et al., 2011; Schepers et al., 2012; Guerlet et al., 2012, Detmers et al. 2015).

In order to account for the effect of aerosols and cirrus, the proxy method retrieves both the CH₄ and CO₂ column under the assumption of a non-scattering atmosphere assuming that scattering effects cancel by taking the ratio of the 2 is calculated. XCH₄ is then calculated by multiplying the ratio with XCO₂ obtained from a model (CarbonTracker 2013 or LMD CO2 flask inversions). Additional fit parameters are the surface albedo and its 1st order spectral dependence in all bands, and the total column of water vapour, respectively.

In order to obtain a proper characterization of the retrieved XCH₄, 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 CH₄ and CO₂ profile is about 1. The inversion is performed using Phillips-Tikhonov regularization in combination with a reduced step size Gauss-Newton iteration scheme.
1. Data product overview

1.1 Column-averaged mixing ratios of CH₄ (XCH₄)

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

The data products
- Column-averaged dry-air mixing ratios (mole fractions) of CH₄, denoted XCH₄ (in parts per billion, ppb).

In the following, several satellite instruments are briefly described which are used / can be used to generate the XCH₄ data product CH4_GOS_SRPR.

TANSO-FTS is a Fourier-Transform-Spectrometer (FTS) onboard the Japanese GOSAT satellite (Kuze et al., 2009, 2014, 2016). The Greenhouse Gases Observing Satellite "IBUKI" (GOSAT) is the world's first spacecraft in orbit dedicated to measure the concentrations of carbon dioxide and methane from space. The spacecraft was launched successfully on January 23, 2009, and has been operating properly since then. GOSAT covers the relevant CO₂, CH₄ and O₂ absorption bands in the NIR and SWIR spectral region as needed for accurate XCO₂ and XCH₄ retrieval (in addition GOSAT also covers a large part of the Thermal Infrared (TIR) spectral region). The spectral resolution of TANSO-FTS is much higher compared to SCIAMACHY and also the ground pixels are smaller (10 km compared to several 10 km for SCIAMACHY). However, in contrast to SCIAMACHY, the GOSAT 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 see also http://www.gosat.nies.go.jp/en/.

2. Input and auxiliary data

2.1 Satellite instrument

2.1.1 GOSAT TANSO-FTS level 1b

Version V201201 / 201202 of the Level 1b data of the TANSO-FTS (Thermal And Near-infrared Sensor for carbon Observation -Fourier Transform Spectrometer) onboard GOSAT (Greenhouse gas Observing SATtellite) are needed in the project to produce the total column CO₂ and CH₄ products. They serve as input for the retrieval algorithms to be used in this project.

Originating system:

The GOSAT satellite was launched in January 2009. The GOSAT observational data are routinely processed at the GOSAT Data Handling Facility. The development of the GOSAT Data Handling Facility (GOSAT DHF) was completed in late 2008, and NIES has been maintaining it for the routine processing of the GOSAT data.
Data class: Earth Observation

Sensor type and key technical characteristics:

TANSO-FTS is an instrument that utilizes optical interference. Within the instrument the incoming light is split into two beams which propagate in separate optical paths to create an optical path difference between the two. These beams are then recombined to cause interference. FTS measures the intensity of the interference by continuously changing the optical path difference. A spectrum, which is distribution of light intensity over a span of wavelength, is obtained by performing a mathematical operation called the Fourier transform on that measured data.

FTS 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 in the daytime, and the latter is captured in band 4 during both the day and the night. Within this project only level 1 data from the SWIR channels 1-3 is used. Prior to reaching the detectors of the instrument, the light in the bands 1 through 3 is split into two orthogonally-polarized components (P and S components) and measured independently. The light in the band 4, however, is not split. The instrument thereby observes the incoming light in seven different channels.

The TANSO-FTS Level 1b data are radiance spectra that are obtained by performing the Fourier transformation on the signals detected by FTS. A single data file of the FTS Level 1B data contains the radiance spectra obtained during 1/60 of an orbital revolution (defined as "one scene").

Table 1: GOSAT-FTS bands

<table>
<thead>
<tr>
<th>channel</th>
<th>wavelength range [nm]</th>
<th>resolution [cm(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>758-775</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>1460-1720</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>1920-2080</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>5560-14300</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Data availability & coverage:

TANSO-FTS level-1b data are available since April 2009. Over a three-day period, GOSAT-FTS takes fifty-six thousand measurements, covering the entire globe.

Source data product name & reference to product technical specification documents


Data quality and reliability

The quality of the retrieved CO2 and CH4 columns has been tested against ground-based observations (i.e. the Total Carbon Column Observing Network (TCCON) network, [http://www.tccon.caltech.edu/](http://www.tccon.caltech.edu/)) and has shown to be of good quality. The L1B data are updated yearly with a new version to improve calibration.
2.2 Other

2.2.1 ECMWF model data

The retrieval algorithms to produce vertical columns of CO2 and CH4 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 is uses the '4D-Var' data assimilation approach, which provides a physically consistent best fit to observations. For this project the ERA-interim archive as well as the operational archive is of importance. Here, the ERA interim data will be used to produce the C3S time series as this dataset uses one model version for the entire period. For sensitivity studies it is planned to use one year of data from the operational archive. From August 2019, the ERA interim data has been superseded by the ERA5 reanalysis data. From then on, we use the ERA5 reanalysis data to produce the GOSAT level 2 products.

**Data class:** Model

**Required ECMWF data:**

**Class:** ERA interim and ERA 5

**Stream:** Atmospheric model

**Type:** Analysis

**Dates:** (1) ERA interim: 01/04/2009 to 01/08/2019; (2) ERA 5: 01/08/2019 to 01/01/2020

**Time:** 00:00:00, 06:00:00, 12:00:00, 18:00:00

**Spatial grid:** (1) ERA interim: 79 km globally, 60 levels to 0.1 hPa; (2) ERA 5: 31 km globally, 62km for the Ensemble of Data Assimilations (EDA), 137 levels to 0.01 hPa

Parameters at model levels:

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

Parameters at surface:

- 10 metre U wind component
- 10 metre 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 considered to be among the best available data sets for meteorological parameters.

2.2.2 Carbon Tracker Data

The retrieval algorithms for CH4 columns from SCIAMACHY and GOSAT that are based on the "proxy approach" retrieve the ration of the CH4 and CO2 columns, where the CO2 column serves as a proxy for the light path. In order to obtain the CH4 column, the retrieved ratio needs to be multiplied by the best estimate of the CO2 column. It is assumed that data from the Carbon Tracker model can provide the best estimate of the true CO2 column.

Originating system:
Carbon Tracker is a data assimilation system developed at the National Oceanic and Atmospheric Administration's (NOAA's) Earth System Research Laboratory (ESRL). It is based on a state-of-the-art atmospheric transport model coupled to an ensemble Kalman filter. CarbonTracker assimilates atmospheric CO2 mole fractions, using a variety of in situ measurements worldwide.

Data class: Model

Data availability & coverage:
CarbonTracker provides global daily data at a spatial resolution of 3 by 2 degree. Data are available with a delay typically of one or more years.

Source data product name & reference to product technical specification documents
CarbonTracker 3D Mole Fractions
Peters, W. et al. (2007)

Data quality and reliability
CarbonTracker has been validated with many independent measurements and is considered to provide be an accurate global CO2 data set.

2.2.3 CO2 correction

In order to obtain the CH4 column, the retrieved ratio needs to be multiplied by the best estimate of the CO2 column. Instead of using CO2 column values from CarbonTracker, we can use the data from the LMD CO2 flask inversions as alternative choice. It is assumed that data from the LMD flask inversion can provide the best up-to-date estimate of the true CO2 column.

Originating system:
The inversion product used here is the official CAMS v18r1 product that exclusively assimilates about 130 sites of surface air sample measurements from the Global Atmosphere Watch programme

Data class: Model
**Data availability & coverage:**
The CAMS v18r1 inversion product provides global 3-hourly data at a spatial resolution of 3.75 by 1.89 degree. Data are typically available within one year.

*Source data product name & reference to product technical specification documents*

Chevalier et al 2010.

**Data quality and reliability**
The CAMS v16r1 inversion product has been validated with many independent measurements and is considered to provide an accurate global CO2 data set. For the year 2019, we use the CAMS v18r1 product for the year 2018 by taking into account an annual increase of 2.75 ppm.

### 2.2.4 SRTM DEM

The RemoTeC retrieval algorithm for CO2 and CH4 columns from GOSAT 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

### 2.2.5 TCCON FTS CO₂ and CH₄ data

TCCON data for CO₂ and CH₄ is available publicly for all TCCON stations (https://tccon-wiki.caltech.edu/). We use the GGG2014 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 a Fourier Transform Spectrometer (FTS).
Data availability & coverage:
Coverage is limited to the locations of the TCCON stations themselves. Depending on the instrument setting (gain H, gain M or sunglint), we use different TCCON stations for validation (see Product User Guide). 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., 2011

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.

Data quality and reliability:
For XCO₂ the precision is 0.25% (1ppm) and the systematic error (bias) is 0.2% (0.8 ppm). For XCH₄ the precision is 0.40% (7ppb) and the systematic error (bias) is also 0.40% (7 ppb).

2.2.6 Additional input data

- Absorption cross sections: For the retrieval lookup-tables are used with pre-calculated absorption cross sections of the species of interest (O₂, CO₂, CH₄, H₂O) 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 (Fig. 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.

- 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 data and ECMWF data are dynamic datasets that are continuously updated, SRTM topography is a static dataset). In the next step a pre-processing program is combining all relevant information per GOSAT ground pixel. This includes interpolation of ECMWF data in space and time to the coordinates of the GOSAT ground pixel, calculating the average height of a GOSAT ground pixel and its standard deviation from the topography database.

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

The pre-processor produces for each GOSAT L1B file an auxiliary input file, hereafter referred to as the ‘retrieval ini’ file, that contains this information. In the next step columns of CO₂, CH₄, H₂O, and O₂ are retrieved under the assumption of an atmosphere without aerosol/cirrus/cloud scattering (‘RemoTeC non-scattering mode’ in Figure 1). The outcome of these retrievals (in an intermediate ACII output file) is used to create the RemoTeC XCH₄ proxy product, but also for cloud filtering to select cloud free scenes to be processed by the RemoTeC Full Physics algorithm. 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.

Figure 2 gives an 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.

Figure 2: Schematic overview of the RemoTeC retrieval procedure including multi-threading.
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 ground pixel to be processed, together with the GOSAT-FTS level 1 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 3: Overview of RemoTeC processing per ground pixel
3. Algorithms

3.1 Algorithm description

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

\[
y = f(x, b) + e_y
\]  

where \( e_y \) represents the measurement noise vector. A retrieval method approximates the true forward model \( f \) by a retrieval forward model \( F \), with a forward model error vector \( e_F \),

\[
y = F(x, b) + e_y + e_F
\]  

For proxy retrieval from the GOSAT FTS instrument the measurement vector contains the measured intensities in the SWIR (see Table 2).

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

<table>
<thead>
<tr>
<th>band</th>
<th>used spectral range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (CO(_2))</td>
<td>6170 – 6277 cm(^{-1})</td>
</tr>
<tr>
<td>2 (CH(_4))</td>
<td>6045 – 6138 cm(^{-1})</td>
</tr>
</tbody>
</table>

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

\[
F(x, b) \approx F(x_n, b) + K(x_n - x)
\]  

where \( x_n \) is the state vector for the \( n \)-th iteration step and \( K \) is the Jacobian matrix

\[
K = \frac{\partial F}{\partial x}
\]  

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

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

\[
I_i = \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} I(\lambda) S_i(\lambda) d\lambda
\]

where \( S_i(\lambda) \) is the Instrument Spectral Response Function (ISRF) for spectral pixel \( i \) and \( I(\lambda) \) is the modeled intensity at high spectral resolution. 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 cm\(^{-1}\) in the NIR band and 0.02 cm\(^{-1}\) in the SWIR).

3.2.1 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 O\(_2\) (in the NIR band) and CH\(_4\), H\(_2\)O, and 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 CH\(_4\), CO (TM5) and H\(_2\)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 \( p_{\text{surf}} \) 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:

\[
p_{\text{lev},k} = p_{\text{min}} + \Delta p * k
\]

\[
\Delta p = (p_{\text{surf}} - p_{\text{min}})/\text{NLAY}
\]

where \( p_{\text{min}} \) 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 \( DV_{\text{CH}4,k} \) for the layer bounded by pressure levels \( p_{\text{lev},k} \) and \( p_{\text{lev},k+1} \) is given by:

\[
DV_{\text{CH}4,k} = X \text{CH}4_k \times DV_{\text{AIR}k}
\]

where \( X \text{CH}4_k \) 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 \((p_k+p_{k+1})/2\). \( DV_{\text{AIR}k} \) is the sub-column of air in layer \( k \), given by
\[ DV_{AIR_k} = \left( p_{\text{lev},k+1} - p_{\text{lev},k} \right) \times \frac{R}{M} \times g_k \times \left( 1 + \frac{XH_2O_k}{1.60855} \right) \]  

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\(_2\)O are calculated in the same manner as for CH\(_4\), and the O\(_2\) sub-column is obtained by multiplying the air sub-column by the O\(_2\) mixing ratio (=0.2095).

In the proxy method scattering is neglected in the forward model and hence atmospheric scattering properties do not need to be calculated.

To summarize, the forward model needs the following inputs:

- Surface pressure to define the equidistant pressure grid
- Sub-columns of CH\(_4\), CO, H\(_2\)O, O\(_2\), 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.
- Solar Zenith Angle (SZA).
- Viewing Zenith Angle (VZA).
- Relative Azimuth Angle (RAA).
- The aerosol complex refractive index \( m = m_r + im_i \)
- A high spectral resolution solar reference spectrum.
- Lookup tables with absorption cross-sections of CH\(_4\), CO, H\(_2\)O, and O\(_2\) as function of pressure, temperature, and wavenumber.

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

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

For the proxy method the top of the atmosphere radiance can be modeled using Lambert Beers’ law and surface reflection:

\[ I_{TOA} = R_{surf} F_0 e^{-\left( \mu_0 - \frac{\tau}{\mu_v} \right)} \]  

where \( F_0 \) is the incoming total flux, \( \tau = \tau_{abs} \), \( \mu_0 \) is the cosine of the solar zenith angle, \( \mu_v \) is the cosine of the viewing zenith angle, and \( R_{surf} \) is the surface reflection for the specific solar and viewing geometry under consideration.
3.3 Inverse algorithm

Definition of state vector and ancillary parameters

The state vector $x$ contains the following elements (between brackets are optional elements):

- CO$_2$ sub-columns in 12 vertical layers (layer interfaces coincide with NLAY layers of forward model grid).
- CH$_4$ sub-columns in 12 vertical layers (layer interfaces coincide with NLAY layers of forward model grid).
- H$_2$O total column.
- 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).
- Spectral shift of Earth radiances in all bands (higher orders optional).
- (offset in input temperature profile).
- (surface pressure).

The ancillary parameter vector $b$ contains the following parameters:

- H$_2$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.

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

<table>
<thead>
<tr>
<th>State vector element</th>
<th>A priori value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_4$ sub-columns</td>
<td>TM4</td>
</tr>
<tr>
<td>CO$_2$ sub-columns</td>
<td>Carbontracker</td>
</tr>
<tr>
<td>H$_2$O total column</td>
<td>ECMWF</td>
</tr>
<tr>
<td>surface albedo (NIR + SWIR)</td>
<td>no prior value needed (first guess at maximum of measured reflectance)</td>
</tr>
<tr>
<td>spectral shifts</td>
<td>no prior needed (first guess = 0)</td>
</tr>
<tr>
<td>temperature offset</td>
<td>no prior needed (first guess = 0)</td>
</tr>
<tr>
<td>surface pressure</td>
<td>ECMWF + SRTM DEM</td>
</tr>
</tbody>
</table>
3.4 Inversion Procedure

The inverse method optimizes the state vector $x$ with respect to the measurements $y$ after applying the forward model $F$ to $x$. The inverse method is based by default on a Phillips-Tikhonov regularization scheme [Phillips et al., 1962; Tikhonov et al., 1963; Hasekamp and Landgraf, 2005a]. Regularization is required because the inverse problem is ill-posed, i.e., the measurements $y$ typically contain insufficient information to retrieve all state vector elements independently. The inverse algorithm finds $x$ 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

$$\hat{x} = \min_{x} \left( \frac{1}{2} S_y^{-1} (F(x) - y)^2 + \gamma \| W(x - x_a) \|^2 \right)$$

(10)

where $S_y$ is the diagonal measurement error covariance matrix, which contains the noise estimate. $x_a$ is an a priori state vector (see Table 3-2), and $W$ is a weighting matrix (see below).

For the linearized forward model for iteration step $n$, the equation for the updated state vector $x_{n+1}$ reduces to

$$x_{n+1} = \min_{x} \left( \| K'(x' - x_a) - y' \|^2 + \gamma \| x' - x_a \|^2 \right)$$

(11)

with the weighted quantities $x' = Wx$, $y' = S_y^{-1/2} (y - F(x_a))$ and $K' = S_y^{-1/2} KW^{-1}$.

The solution reads

$$x_{n+1} = G'y' + A'x_n' + (I - A')x_{aprior}$$

(12)

with $A'$ the averaging kernel matrix and $G'$ the contribution function matrix given by $A' = G'K'$ and $G' = (K'TK' + \gamma I)^{-1} K'T$. 

C3S_312b_Lot2_DLR_2018SC1 - ATBD ANNEX-C v5.0

25 of 31  2/18/2021
If the retrieval converges after a given number of steps $N$ (typically 7-8), the final state vector $x_{retr}$ is related to the true state vector and to the prior via

$$x_{retr} = Ax_{true} + (I - A)x_a + Ge_y + Ge_F$$

(13)

The covariance matrix $S_x$ describing the retrieval noise ($Ge_y$) is given by

$$S_x = GS_yG^T$$

(14)

The target retrieval quantity is the column averaged dry air CH4 mixing ratio, $XCH_4$. This quantity is obtained from the CH4 and CO2 entries of the retrieved state vector through

$$XCH_4 = \frac{V_{ch4}}{V_{co2}} \times XCO_{2,prior}$$

(15)

where $V_{ch4}$ and $V_{co2}$ are the total columns of CO2 and CH4, respectively, obtained from the retrieval state vector by

$$V_{ghg} = h_{ghg}^T x_{retr}$$

(16)

where the subscript ghg refers to either CO2 or CH4, $h$ is the total column operator (summing up the partial columns in the state vector). The retrieval noise $\Delta CH_4$ on $XCH_4$ is given by

$$\Delta XCH_4 = \frac{\sum_{i=1}^{12} \sum_{j=1}^{12} S_{x,i,j}}{V_{air,dry}}$$

(17)

This is the error estimate that will be given in the output together with $XCH_4$.

For validation and application purposes it is important to realize that the retrieved $XCH_4$ is in fact a representation of $ax_{true} / V_{air,dry}$, where the quantity

$$a = h^T A$$

(18)

is referred to as the column averaging kernel (Rodgers and Connor, 2003).
3.4.1 Regularization of state vector and iteration strategy

To retrieve a meaningful state vector $\mathbf{x}$, the side-constraint 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

$$\mathbf{W} = \mathbf{LW}'$$

with

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

The upper left 12 by 12 sub matrix works on the state vector elements that contain the CH$_4$ sub-columns in the 12 altitude layers of the retrieval vertical grid. This sub-matrix corresponds to the matrix product

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

where $\mathbf{L}_1$ is the first derivative matrix. The $W'_{jj}$ is given by $1/\text{MAX}[\mathbf{K}(1:12,1:12)]$ for the state vector elements corresponding to the 12 sub-columns of methane, $1/\text{MAX}[\mathbf{K}(:,j)]$ 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.
3.5 Cloud Filtering

Cloud filtering in RemoTeC proxy is based on retrieved columns of oxygen (VO2), carbon dioxide (VCO2) and water vapor (VH2O) retrieved independently from the 0.75 micron, 1.6 micron, and 2.0 micron bands, respectively, under the assumption of a non-scattering atmosphere:

\[ 0.88 < \frac{\text{VO2} \text{ retrieved}}{\text{VO2 ecmwf}} < 1.035, \quad 0.98 < \frac{\text{VCO2} \text{ 1.6micron}}{\text{VCO2 2.0micron}} < 1.15, \]
\[ 0.90 < \frac{\text{VH2O 1.6micron}}{\text{VH2O 2.0micron}} < 1.5. \]

The rationale for these cloud filters is that scenes with a large light path deviation with respect to a non scattering atmosphere will result in different CO2 and H2O columns retrieved (without scattering) from the 1.6 and 2.0 micron band due to different light path sensitivities in the two bands. Also, the retrieved O2 column will deviate more from the ECMWF O2 column for large light path differences with a non-scattering atmosphere. For the proxy the cloud filters are less strict than for the Full Physics Algorithm as the ratio retrieved in the proxy method is less sensitive to cloud contamination.

4. Output data

The output data are stored in one netCDF file per day. The file size varies between 3 and 5 Mb. A separate output file is being generated for the XCH4 proxy product.

Note that the format of the main output data, which are the Level 2 data products, is described in the separate Product User Guide (PUG) document.
References


