ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Algorithm Theoretical Basis Document (ATBD) – ANNEX C for product CH4_GO2_SRPR (v2.0.1, 2019-2022)

C3S2_312a_Lot2_DLR – Atmosphere

Issued by: Andrew Barr and Tobias Borsdorff, SRON, The Netherlands

Date: 01/02/2024

Ref: C3S2_312a_Lot2_D-WP1_ATBD-2023-GHG_ANNEX-C_v7.1b

Official reference number service contract: 2021/C3S2_312a_Lot2_DLR/SC1

Contributors

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

SRON NETHERLANDS INSTITUTE FOR SPACE RESEARCH, LEIDEN, THE NETHERLANDS
(SRON)
T. Borsdorff
A. G. Barr
O. P. Hasekamp

History of modifications

Version	Date	Description of modification	Chapters / Sections
1.1	20-October-2017	New document for data set CDR1 (2009-2016)	All
2.0	4-October-2018	Update for CDR2 (2009-2017)	All
3.0	12-August-2019	Update for CDR3 (2009-2018)	All
3.1	03-November-2019	Update after review by Assimila: Correction of typos, some links added and some additional explanations given.	All
4.0	18-August-2020	Update for CDR4 (2009-2019)	All
5.0	18-February-2021	Update for CDR5 (2009-mid2020)	All
6.0 Draft	18-February-2022	Update for data set CDR6: First draft	All
6.0	4-August-2022	Update for CDR6: GOSAT-2 retrievals (2019 - 2021)	All
6.1	25-November-2022	Update after review (use of new template, various mostly minor improvements at several places)	All
6.2	31-January-2023	Minor improvements at various places taking into account feedback from 2nd review	All
7.0 (Draft 1)	23-February-2023	Initial draft for data set CDR7	All
7.0 (Draft 2)	27-April-2023	Improvements based on review	All
7.0 (Draft 3)	29-June-2023	Improvements based on review	All
7.0 (Draft 4)	03-August-2023	Improvements based on review	All
7.0	28-August-2023	Minor improvements	All
7.1	13-September-2023	Minor improvements	All
7.1b	1-February-2024	Minor improvements	All

List of datasets covered by this document

Deliverable ID	Product title	Product type (CDR, ICDR)	Version number	Delivery date
WP2-FDDP-GHG-v2	CH4_GO2_SRPR	CDR 7	2.0.1	31-Aug-2023

Related documents

Reference ID	Document
D1	Main ATBD: Buchwitz, M., et al., Algorithm Theoretical Basis Document (ATBD) – Main document for Greenhouse Gas (GHG: CO_2 & CH_4) data set CDR 7 (2003-2022), project C3S2_312a_Lot2_DLR – Atmosphere, v7.1, 2023. *(this document is an ANNEX to the Main ATBD)*
D2	Barr, A. G., et al., Product User Guide and Specification (PUGS) – ANNEX C for product CH4_GO2_SRPR (v2.0.1, 2019-2022), Technical Report C3S project C3S2_312a_Lot2_DLR – Atmosphere, 2023.

Acronyms

Acronym	Definition
AIRS	Atmospheric Infrared Sounder
AMSU	Advanced Microwave Sounding Unit
ATBD	Algorithm Theoretical Basis Document
BESD	Bremen optimal Estimation DOAS
CAR	Climate Assessment Report
C3S	Copernicus Climate Change Service
CCDAS	Carbon Cycle Data Assimilation System
CCI	Climate Change Initiative
CDR	Climate Data Record
CDS	(Copernicus) Climate Data Store
CMUG	Climate Modelling User Group (of ESA's CCI)
CRG	Climate Research Group
D/B	Data base
DOAS	Differential Optical Absorption Spectroscopy
EC	European Commission
ECMWF	European Centre for Medium Range Weather Forecasting
ECV	Essential Climate Variable
EMMA	Ensemble Median Algorithm
ENVISAT	Environmental Satellite (of ESA)
EO	Earth Observation
ESA	European Space Agency
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
FoM	Figure of Merit
FP	Full Physics retrieval method
FTIR	Fourier Transform InfraRed
FTS	Fourier Transform Spectrometer
GCOS	Global Climate Observing System
GEO	Group on Earth Observation
GEOSS	Global Earth Observation System of Systems
GHG	GreenHouse Gas
GOS	GOSAT
GO2	GOSAT-2
GOME	Global Ozone Monitoring Experiment
GMES	Global Monitoring for Environment and Security
GOSAT	Greenhouse Gases Observing Satellite
GOSAT-2	Greenhouse Gases Observing Satellite 2

IASI	Infrared Atmospheric Sounding Interferometer
IMAP-DOAS (or IMAP)	Iterative Maximum A posteriori DOAS
IPCC	International Panel in Climate Change
IUP	Institute of Environmental Physics (IUP) of the University of Bremen, Germany
JAXA	Japan Aerospace Exploration Agency
JCGM	Joint Committee for Guides in Metrology
L1	Level 1
L2	Level 2
L3	Level 3
L4	Level 4
LMD	Laboratoire de Météorologie Dynamique
MACC	Monitoring Atmospheric Composition and Climate, EU GMES project
NA	Not applicable
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Format
NDACC	Network for the Detection of Atmospheric Composition Change
NIES	National Institute for Environmental Studies
NIR	Near Infra-Red
NLIS	LMD/CNRS *neuronal* network mid/upper tropospheric CO_2 and CH_4 retrieval algorithm
NOAA	National Oceanic and Atmospheric Administration
Obs4MIPs	Observations for Climate Model Intercomparisons
OCFP	OCO-2 Full Physics (FP) algorithm (used by Univ. Leicester)
OCO	Orbiting Carbon Observatory
OCPR	OCO-2 Proxy (PR) algorithm (used by Univ. Leicester)
OE	Optimal Estimation
PBL	Planetary Boundary Layer
ppb	Parts per billion
ppm	Parts per million
PQAD	Product Quality Assurance Document
PQAR	Product Quality Assessment Report
PR	(light path) Proxy retrieval method
PVIR	Product Validation and Intercomparison Report
QA	Quality Assurance
QC	Quality Control
RemoTeC	Retrieval algorithm developed by SRON
REQ	Requirement
RMS	Root-Mean-Square
RTM	Radiative transfer model
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY
SCIATRAN	SCIAMACHY radiative transfer model

SRON	SRON Netherlands Institute for Space Research
SRFP	SRON's Full Physics (FP) algorithm (also referred to a RemoTeC FP)
SRPR	SRON's Proxy (PR) algorithm (also referred to a RemoTeC PR)
SWIR	Short Wave Infra-Red
TANSO	Thermal And Near infrared Sensor for carbon Observation
TANSO-FTS	Fourier Transform Spectrometer on GOSAT
TANSO-FTS-2	Fourier Transform Spectrometer on GOSAT-2
TBC	To be confirmed
TBD	To be defined / to be determined
TCCON	Total Carbon Column Observing Network
TIR	Thermal Infra-Red
TR	Target Requirements
TRD	Target Requirements Document
WFM-DOAS (or WFMD)	Weighting Function Modified DOAS
UoL	University of Leicester, United Kingdom
URD	User Requirements Document
WMO	World Meteorological Organization
Y2Y	Year-to-year (bias variability)

General definitions

Essential climate variable (ECV)

An ECV is a physical, chemical, or biological variable or a group of linked variables that critically contributes to the characterization of Earth's climate.

Climate data record (CDR)

The US National Research Council (NRC) defines a CDR as a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change.

Fundamental climate data record (FCDR)

A fundamental climate data record (FCDR) is a CDR of calibrated and quality-controlled data designed to allow the generation of homogeneous products that are accurate and stable enough for climate monitoring.

Thematic climate data record (TCDR)

A thematic climate data record (TCDR) is a long time series of an essential climate variable (ECV).

Intermediate climate data record (ICDR)

An intermediate climate data record (ICDR) is a TCDR which undergoes regular and consistent updates, for example because it is being generated by a satellite sensor in operation.

Satellite data processing levels

The NASA Earth Observing System (EOS) distinguishes six processing levels of satellite data, ranging from Level 0 (L0) to Level 4 (L4) as follows.

L0	Unprocessed instrument data
L1A	Unprocessed instrument data alongside ancillary information
L1B	Data processed to sensor units (geo-located calibrated spectral radiance and solar irradiance)
L2	Derived geophysical variables (e.g., XCO_2) over one orbit
L3	Geophysical variables averaged in time and mapped on a global longitude/latitude horizontal grid
L4	Model output derived by assimilation of observations, or variables derived from multiple measurements (or both)

Table of Contents

Scope of the 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 C3S2_312a_Lot2.

Within this project satellite-derived atmospheric carbon dioxide (CO_2) and methane (CH_4) 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 CO_2 and CH_4, denoted XCO_2 (in parts per million, ppm) and XCH_4 (in parts per billion, ppb), respectively.
- Mid/upper tropospheric mixing ratios of CO_2 (in ppm) and CH_4 (in ppb).

This document describes the retrieval algorithm used to generate the C3S product CH4_GO2_SRPR v2.0.1.

This product is the XCH_4 Level 2 product as retrieved from GOSAT-2 using algorithms developed at SRON, The Netherlands.

Executive summary

In this document the retrieval algorithms are described, which are used to generate satellite-derived atmospheric carbon dioxide (CO_2) and methane (CH_4) Climate Data Record (CDR) data products as generated via the C3S2_312a_Lot2 project of the Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/).

This document describes the RemoTeC algorithm for GHG retrieval from the GOSAT-2 instrument. It should be read in conjunction with the MAIN ATBD document, defined in the Related Documents section at the beginning of this document (D2). 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 real GOSAT data (Butz et al., 2011; Schepers et al., 2012; Guerlet et al., 2012). Here we apply the retrieval to GOSAT-2 data.

The RemoReC algorithm developed by SRON is a linearized vector radiative transfer model which solves the equation of radiative transfer for back-scattered sunlight. It adopts Philips-Tikonov regularization to apply the inversion, and the Gauss-Sidel method to solve the equation of radiative transfer. The retrieval algorithm aims at inferring an atmospheric state vector from a measurement vector. The state vector is linked to the measurement vector through the true forward model.

To account for the effect of aerosols and cirrus cloud, the proxy method retrieves both the CH_4 and CO_2 column under the assumption of a non-scattering atmosphere. Assuming that scattering effects cancel by taking the ratio of the two species, XCH_4 is then calculated by multiplying that ratio with a value for XCO_2 obtained from a model. Additional parameters included in the retrieval state vector are the surface albedo and its 1st order spectral dependence in all bands, and the total column of water vapour.

To obtain a proper characterization of the retrieved XCH_4, 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_4 and CO_2 profile is about one. The inversion is performed using Phillips-Tikhonov regularization in combination with a reduced step size Gauss-Newton iteration scheme.

The greenhouse gas (GHG) activities of this C3S project and its C3S pre-cursor projects are essentially the operational continuation of the research and development (R&D) pre-cursor projects GHG-CCI and GHG-CCI+ of ESA's Climate Change Initiative (CCI). R&D for the GOSAT-2 products is currently an ongoing activity of the ESA GHG-CCI+ project (see https://climate.esa.int/en/projects/ghgs/).

Section 1 gives an overview of the data product including a description of the GOSAT-2 satellite. Section 2 introduces the various input data which enter into the retrieval algorithm such as meteorological data, satellite level 1 data and a priori CH_4/CO_2 profiles. Sections 3 and 4 describe in detail the retrieval algorithm and its output, respectively.

1. Instruments

1.1 TANSO-FTS-2 on GOSAT-2

As explained in the Main ATBD (D1), GOSAT-2 is the follow-on satellite to the Greenhouse Gases Observing Satellite (GOSAT) and is thus very similar to its predecessor. The information in D1 provides an overview of GOSAT and GOSAT-2. In this section we provide a more detailed summary of GOSAT-2. We emphasise that the products described in this document are derived only from GOSAT-2.

The Japanese Greenhouse gases Observing SATellite-2 (GOSAT-2) was launched on 29th October 2018 and started operational observations form February 2019. GOSAT-2 provides dedicated global measurements of total column CO_2 and CH_4 from its Short Wave Infra-red (SWIR) bands. It is equipped with two instruments, the Thermal And Near Infrared Sensor for carbon Observations - Fourier Transform Spectrometer-2 (TANSO-FTS2) as well as a dedicated Cloud and Aerosol Imager-2 (TANSO-CAI-2). For the products described in this document, cloudy scenes are identified via the retrieved columns of oxygen, carbon dioxide and water vapour (section 3.4) thus we do not use data from TANSO-CAI.

The TANSO-FTS2 instrument (Nakajima et al., 2017) has five spectral bands, given in Table 1, with a high spectral resolution 0.2 cm^{-1}. Three operate in the SWIR at 0.75-0.77 µm, 1.56-1.69 µm 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 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. More information on GOSAT-2 can be found here: http://www.gosat-2.nies.go.jp/.

Table 1: TANSO-FTS-2 bands.

Channel	Wavelength range [nm]	Resolution [cm^{-1}]
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.**

2. Input and auxiliary data

The following section describes the various data that are associated with the full physics retrieval. The input measurement for the forward model consists of the GOSAT-2 L1B TANSO-FTS-2 data (section 2.1) and the other data used is auxiliary data such as ECMWF model data, the SRTM DEM, etc. (section 2.2).

2.1 Level 1b data of TANSO-FTS-2 onboard GOSAT-2

Level 1b data of the TANSO-FTS-2 – onboard GOSAT-2 are needed in the project to produce the total column CO_2 and CH_4 products. They serve as input for the retrieval algorithms to be used in this project.

TANSO-FTS-2 level-1b data are available since February 2019. The TANSO FTS-2 L1A/L1B Product Description Document (Japan Aerospace Exploration Agency, July 2019, GST-18005) is available through https://prdct.gosat-2.nies.go.jp/documents/pdf/NIES_GOSAT-2_Product_File_Format_Descriptions_en_03.pdf

The quality of the retrieved CO_2 and CH_4 columns has been tested against ground-based observations (section 2.2.2) and has shown to be of good quality. The spectral bands showed some irregularities which required a shortened retrieval window (O_2 A-band) and spectral intensity offsets for each of the bands. This in effect worsened the cost function χ^2 of the fits compared to similar GOSAT-1 retrievals and the estimated uncertainties of the species. However, validation with respect to ground based observations showed that the retrieval products of both the GOSAT-1 and GOSAT-2 are of similar quality.

2.2 Other data

2.2.1 ECMWF model data

The retrieval algorithms to produce vertical columns of CO_2 and CH_4 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.

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. The ECMWF model data sets are widely considered to be among the best available data sets for meteorological parameters. We have compared retrievals for an overlapping month (August 2019) using both ERA-5 and ERA-interim and no major effects were observed in the retrieved parameters. The source and description of the ECMWF data used can be found in Table 2.

Table 2: Summary of ECMWF atmospheric data used for trace gas retrieval.

Model	ECMWF ERA-5 atmospheric model
Type	Reanalysis complete
Stream	oper
class	ea
Availability	01/08/2019 to present
Coverage	Global scale
Spatial grid	0.75x0.75 deg
Time grid	6 hours
Number of levels	137
Parameters (levels)	• temperature, specific humidity (all levels) • logarithm of surface pressure, geopotential (lowest level)
Parameters (surface)	• 10 meter U wind component • 10 meter V wind component

2.2.2 CO_2 correction data

The retrieval algorithm for CH_4 columns that are based on the “proxy approach” retrieve the ratio of the CH_4 and CO_2 columns, where the CO_2 column serves as a proxy for the light path. In order to obtain the CH_4 column, the retrieved ratio needs to be multiplied by the best estimate of the CO_2 column. For this purpose, we use CO_2 from the Copernicus Atmosphere Monitoring Service (CAMS).

The inversion product used here is the official CAMS v21r1 product that assimilates about 130 sites of surface air sample measurements from the Global Atmosphere Watch programme. The CAMS v21r1 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 (Chevalier et al 2021).

The CAMS v21r1 inversion product has been validated with many independent measurements and is considered to provide be an accurate global CO_2 data set. For the year 2022, we use the CAMS v21r1 product for the year 2021 by taking into account an annual increase of 2 ppm.

2.2.3 SRTM DEM

The RemoTeC retrieval algorithm for CO_2 and CH_4 columns from GOSAT-2 use information about the surface elevation from an extended SRTM digital elevation map available from (http://www.viewfinderpanoramas.org/dem3.html).

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.

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.

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_2, CO_2, CH_4, and H_2O) as a function of wavenumber, temperature, and pressure. One lookup-table per species and per spectral band is used.

- Aerosol optical properties: A lookup table is 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 is downloaded from https://lasp.colorado.edu/lisird/data/tsis1_hsrs.
- Retrieval settings: A file which contains retrieval settings covering instrument characteristics (like the full width half maximum (FWHM) of the instrument profile), spectral window information and spectral sampling, absorbing molecules and profile inversion methods.

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,

$$y = f(x, b) + e_y \tag{1}$$

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,

$$\mathrm{y} = \mathrm{F}(x, b) + \mathrm{e_y} + \mathrm{e_F} \tag{2}$$

For full physics retrieval from the GOSAT-2 FTS instrument the measurement vector contains the measured intensities in the NIR and SWIR (see Table 3).

Table 3: Spectral ranges from the NIR and SWIR bands included in the measurement vector.

Band	Used spectral range (nm)
1 (O_2 a)	757.863 – 772.201
2 (CO_2)	1593.118 – 1620.746
3 (CH_4)	1629.195 – 1654.260
4 (CO_2)	2042.484 – 2080.732

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

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

Where x_n is the state vector for the n-th iteration step and *K* is the Jacobian matrix

$$\mathrm{K} = \frac{\partial \mathrm{F}}{\partial \mathrm{x}} \tag{4}$$

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

3.2 Forward Model

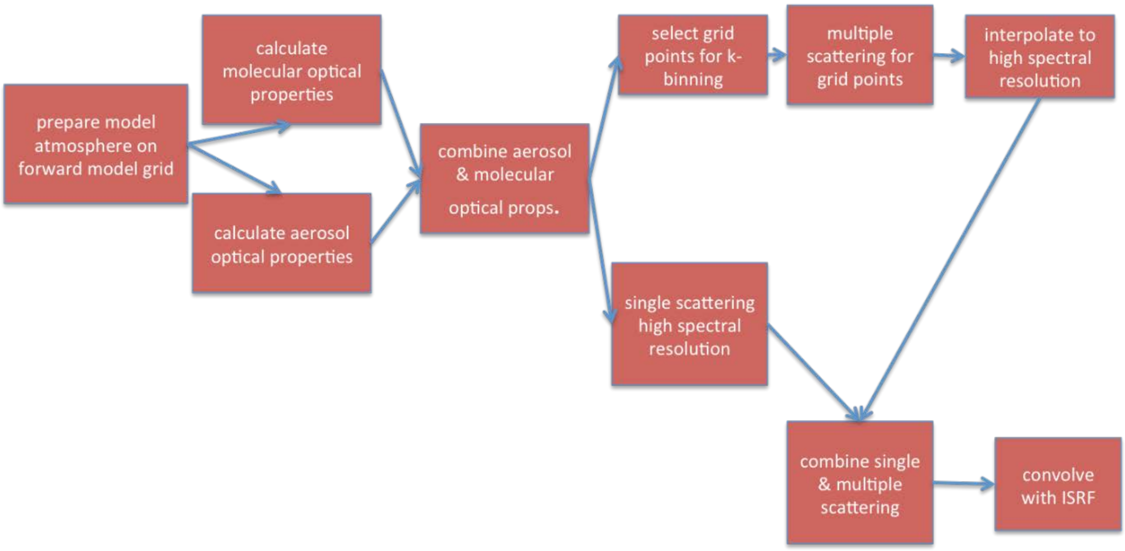
The retrieval forward model *F* simulates the measurement vector *y* for a given model atmosphere defined by the state vector *x* and the ancillary parameter vector *b*. The simulated intensity for a given spectral pixel *I* is given by

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

Where $S_i(\lambda)$ is the Instrument Spectral Response Function (ISRF) for spectral pixel *I* and $I(\lambda)$ is the modelled 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).

The different steps of the forward model calculation are summarised in Figure 1.

Figure 1: Overview of forward model.



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_2O, 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_2O (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:

$$p_{lev,k} = p_{min} + \Delta p * k \tag{6}$$

Where:

$$\Delta p = \frac{p_{surf} - p_{min}}{NLAY} \tag{7}$$

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_CH4_k for the layer bounded by pressure levels $p_{lev,k}$ and $p_{lev,k+1}$ is given by:

$$DV\_CH4_k = XCH4_k \times DV\_AIR_k \qquad (8)$$

Where $XCH4_k$ 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_AIR_k is the sub-column of air in layer k, given by

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

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_2O 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_2O, 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 $m = m_r + i m_i$
- A high spectral resolution solar reference spectrum.
- Lookup tables with absorption cross-sections of CH_4, CO, H_2O, 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 modelled using the Lambert Beers law and surface reflection:

$$I_{TOA} = R_{surf} F_0 e^{-\left(\frac{\tau}{\mu_0}+\frac{\tau}{\mu_v}\right)} \tag{10}$$

where F_0 is the incoming total flux, $\tau = \tau_{abs}$, μ_0 is the cosine of the solar zenith angle, μ_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 listed in Table 4 (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_2O 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).

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 b** contains the following parameters:

- H_2O 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 4: A priori values for the different state vector elements.

State vector element	A priori value
CH_4 sub-columns	TM4
CO_2 sub-columns	Carbontracker
H_2O total column	ECMWF
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

3.3.1 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, 1962; Tikhonov, 1963; Hasekamp and Landgraf, 2005). 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 γ according to

$$\hat{x} = \min_{x}(||\mathrm{S}_y^{-\frac{1}{2}}(\mathrm{F(x) - y})||^2 + \gamma||\mathrm{W}(\mathrm{x} - \mathrm{x}_a)||^2) \tag{11}$$

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 4), and *W* is a weighting matrix (see below).

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

$$x_{n+1} = \min\,(||K'\ (x' - x'_n - y'\ ||^2 + \gamma||x' - x'_n||^2) \tag{12}$$

With the weighted quantities

x'=Wx, y'=$S_y^{-\frac{1}{2}}(y - F(x_n))$ (13)

and

K' = $S_y^{-1/2} K W^{-1}$ (14)

With *K* the Jacobian. The solution reads

$$x_{n+1} = G'y' + A'x'_n + (I - A')x'_{apr} \tag{15}$$

With A' the averaging kernel matrix and G' the contribution function matrix given by

$A'=G'K'$ (16)

and

$G'=(K'^{T}K'+\gamma I)^{-1}K'^{T}$ (17)

If the retrieval converges after a given number of steps N (typically 7-8), the final state vector x_{retr} = x_N is related to the true state vector and to the prior via

$$x_{retr} = Ax_{true} + (I - A)x_n + Ge_y + Ge_F \tag{18}$$

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

$$\mathrm{S}_x = \mathrm{GS}_y G^T \tag{19}$$

The target retrieval quantity is the column averaged dry air CH_4 mixing ratio, XCH_4, This quantity is obtained from the CH_4 and CO_2 entries of the retrieved state vector through

$$XCH_4 = \frac{V_{ch4}}{V_{co2}} \times XCO_{2,prior} \tag{20}$$

where V_{ch4} and V_{co2} are the total columns of CO_2 and CH4, respectively, obtained from the retrieval state vector by

$$V_{ghg} = h_{ghg}{}^{T} x_{retr} \tag{21}$$

where the subscript ghg refers to either CO_2 or CH_4, h is the total column operator (summing up the partial columns in the state vector). The retrieval noise ΔXCH_4 on XCH_4 is given by

$$\Delta XCH_4 = \frac{\sum_{i=1}^{12}\sum_{j=1}^{12} S_{x,i,j}}{V_{air,dry}} \tag{22}$$

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 or XCO_2 is in fact a representation of $ax_{true} / V_{air,dry}$, where the quantity

$$a = h^T \mathrm{A} \tag{23}$$

Is referred to as the column averaging kernel (Rodgers and Connor, 2003).

3.3.2 Regularization of state vector and iteration strategy

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

$$\mathrm{W} = \mathrm{L}\mathrm{W}' \tag{24}$$

With

$$\mathrm{L} = \begin{pmatrix} 1 & -1 & \cdots & & & & \\ -1 & 2 & -1 & & & & \\ \vdots & & \ddots & & \vdots & & \\ & & -1 & 1 & & & \\ & & \cdots & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \tag{25}$$

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.

$$\mathrm{L}_1^T \mathrm{L}_1 \tag{26}$$

Where L_1 is the first derivative matrix. The W'_{jj} is given by 1/MAX[K(1:12,1:12)] for the state vector elements corresponding to the 12 sub-columns of methane,1/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 γ 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.

3.4 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 µm, 1.6 µm, and 2.0 µm bands, respectively, under the assumption of a non-scattering atmosphere:

$0.88 < VO2_{retrieved}/VO2_{ecmwf} < 1.035$

$0.98 < VCO2(1.6\mu m) / VCO2(2.0\mu m) < 1.15$

$0.90 < VH2O(1.6\mu m) / VH2O(2.0\mu m) < 1.5$

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 CO_2 and H_2O columns retrieved (without scattering) from the 1.6 and 2.0 µm bands due to different light path sensitivities in the two bands. Also, the retrieved O_2 column will deviate more from the ECMWF O_2 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 1 and 5 Mb.

The files are named as follows:

ESACCI-GHG-L2-CH4-GOSAT2-SRFP-YYYYMMDD-fv1.nc

Where YYYY, MM and DD denote the year, month and day respectively. The formatting of the files is in accordance with CF-1.6 convention standards and the variables names correspond to CF Standard Name Table v79. There is a quality flag “xch4_quality_flag” included in the data files. The quality flag can have 2 values:

- 0: retrieval quality has been checked
- 1: data should not be used (e.g. bad fit to data, residual cloud contamination)

Pixels flagged with 1 include those pixels rejected by post-processing quality filtering steps. After the retrieval step the data that fulfil the following criteria are flagged as '0' for land and ‘1’ for ocean (glint).

An important step in the post-processing is the bias correction (see the Product User Guide and Specification (PUGS) [D2] for details). As a result, two variables are provided for CH_4, defined as xch4 and raw_xch4 for the bias corrected and non-bias corrected columns, respectively.

Note that the format of the main output data, which are the Level 2 data products, is described in the associated Product User Guide and Specification (PUGS) document [D2].

5. References

Butz et al., 2009: Butz, A., Hasekamp, O. P., Frankenberg, C., Aben, I. (2009). Retrievals of atmospheric CO2 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.: CH4 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_2 and CH_4 observations from GOSAT, *Geophys. Res. Lett.*, doi:10.1029/2011GL047888, 2011.

Butz et al., 2012: Butz, A., Galli, A., Hasekamp, O., Landgraf, J., Tol, P., and Aben, I.: Remote Sensing of Environment, TROPOMI aboard Sentinel-5 Precursor : Prospective performance of CH_4 retrievals for aerosol and cirrus loaded atmospheres, 120, 267-276, doi:10.1016/j.rse.2011.05.030, 2012.

Chevallier, 2021: Chevallier, F.: Validation report for the CO2 fluxes estimated by atmospheric inversion, v21r1Version 2.0, Copernicus Atmosphere Monitoring Service, ECMWF Copernicus report, CAMS255_2021SC1_D55.1.4.2-2022_202208_Validation inversion CO2 fluxes_v2.

Coddington et al., 2021: Coddington, O. M., Richard, E. C., Harber, D., Pilewskie, P., Woods, T. N., Chance, K.: The TSIS-1 Hybrid Solar Reference Spectrum. Geophysical Research Letters, 48, e2020GL091709. https://doi.org/10.1029/2020GL091709, 2021.

Guerlet et al., 2013: Guerlet, S., S. Basu, A. Butz, M. Krol, P. Hahne, S. Houweling, O. P. Hasekamp and I. Aben, Reduced carbon uptake during the 2010 Northern Hemisphere summer from GOSAT, Geophys. Res. Lett., doi: 10.1002/grl.50402, 2013.

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

Kuze et al., 2009: Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T. (2009), Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Appl. Opt., 48, 6716–6733, 2009.

Kuze et al., 2014: Kuze, A., Taylor, T., Kataoka, F., Bruegge, C., Crisp, D., Harada, M., Helmlinger, M., Inoue, M., Kawakami, S., Kikuchi, N., Mitomi, Y., Murooka, J., Naitoh, M., O'Brien, D., O'Dell, C., Ohyama, H., Pollock, H., Schwandner, F., Shiomi, K., Suto, H., Takeda, T., Tanaka, T., Urabe, T., Yokota, T., and Yoshida, Y. (2014), Long-term vicarious calibration of GOSAT short-wave sensors: techniques for error reduction and new estimates of radiometric degradation factors, IEEE T. Geosci. Remote, 52, 3991–4004, doi:10.1109/TGRS.2013.2278696, 2014.

Kuze et al., 2016: Kuze, A., Suto, H., Shiomi, K., Kawakami, S., Tanaka, M., Ueda, Y., Deguchi, A., Yoshida, J., Yamamoto, Y., Kataoka, F., Taylor, T. E., and Buijs, H. L.: Update on GOSAT TANSO-FTS performance, operations, and data products after more than 6 years in space, Atmos. Meas. Tech., 9, 2445-2461, doi:10.5194/amt-9-2445-2016, 2016.

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

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.

O'Brien et al. 2013: O'Brien, CJ, Peloquin, JA, Vogt, M, Heinle, M, Gruber, N, Ajani, P, Andruleit, H, Arístegui, J, Beaufort, L, Estrada, M, Karentz, D, Kopczyńska, E, Lee, R, Poulton, AJ, Pritchard, T, and Widdicombe, C. Global marine plankton functional type biomass distributions: coccolithophores. Earth Syst. Sci. Data, 5, 259-276, 2013 doi:10.5194/essd-5-259-2013

Peters et al., 2007: Peters, Wouter and Jacobson, Andrew R. and Sweeney, Colm and Andrews, Arlyn E. and Conway, Thomas J. and Masarie, Kenneth and Miller, John B. and Bruhwiler, Lori M. P. and Pétron, Gabrielle and Hirsch, Adam I. and Worthy, Douglas E. J. and van der Werf, Guido R. and Randerson, James T. and Wennberg, Paul O. and Krol, Maarten C. and Tans, Pieter P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, PNAS, 104, 48, doi: 10.1073/pnas.0708986104

Philips, 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.

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

Schepers et al., 2012: Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., Blavier, J.F., Deutscher, N.M., Griffith, D.W.T., Hase, F. and Kyro, E.: 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, 2012.

Tikhonov, 1963: Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Math., 4, pp.1035-1038, 1963.

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. 2011: Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network (TCCON), Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011.

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/tccon.ggg2014.documentation.R0/1221662), 2015.

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: https://support.ecmwf.int/