ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Algorithm Theoretical Basis Document (ATBD) – ANNEX C for product CH4_GOS_SRPR

C3S_312a_Lot6_IUP-UB – Greenhouse Gases

Issued by: R. G. Detmers, SRON, The Netherlands

Date: 20/10/2017

Ref: C3S_D312a_Lot6.2.1.2-v1_ATBD_ANNEX-C_v1.1

Official reference number service contract: 2016/C3S_312a_Lot6_IUP-UB/SC1

IMPLEMENTED BY ECMWF

This document has been produced in the context of the Copernicus Climate Change Service (C3S).
The activities leading to these results have been contracted by the European Centre for Medium-Range Weather Forecasts, operator of C3S on behalf of the European Union (Delegation Agreement signed on 11/11/2014). All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission and the European Centre for Medium-Range Weather Forecasts has no liability in respect of this document, which is merely representing the authors view.

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
R. G. Detmers
O. P. Hasekamp

Table of Contents

History of modifications

Version	Date	Description of modification	Chapters / Sections
1.0	21-August-2017	New document	All
1.0b	11-October-2017	Logo replaced on page header	Page header
1.1	20-October-2017	Reference to main ATBD updated	Page 7

Related documents

Reference ID	Document
D1	Main ATBD: Buchwitz, M., et al., Algorithm Theoretical Basis Document (ATBD) – Main document, C3S project C3S_312a_Lot6_IUP-UB – Greenhouse Gases, v1.1, 2017. *(this document is an ANNEX to the Main ATBD)*

Acronyms

Acronym	Definition
ATBD	Algorithm Theoretical Basis Document
CAMS	Copernicus Atmosphere Monitoring Service
C3S	Copernicus Climate Change Service
CDR	Climate Data Record
CDS	(Copernicus) Climate Data Store
CPU	Core Processing Unit
DEM3	Digital Elevation Map 3
DFS	Degrees of Freedom for Signal
DHF	Data Handling Facility
ECMWF	European Centre for Medium Range Weather Forecasting
ESA	European Space Agency
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FP	Full Physics retrieval method
FTIR	Fourier Transform InfraRed
FTS	Fourier Transform Spectrometer
GCOS	Global Climate Observing System
GHG	GreenHouse Gas
GOSAT	Greenhouse Gases Observing Satellite
IUP	Institute of Environmental Physics (IUP) of the University of Bremen, Germany
JAXA	Japan Aerospace Exploration Agency
KIT	Karlsruhe Institute of Technology
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
NIES	National Institute for Environmental Studies

NIR	Near Infra Red
NOAA	National Oceanic and Atmospheric Administration
Obs4MIPs	Observations for Climate Model Intercomparisons
ppb	Parts per billion
ppm	Parts per million
PR	(light path) PRoxy retrieval method
PVIR	Product Validation and Intercomparison Report
RAA	Relative Azimuth Angle
RMS	Root-Mean-Square
RTM	Radiative transfer model
SNR	Signal-to-Noise Ratio
SRON	SRON Netherlands Institute for Space Research
SRTM	Shuttle Radar Topography Mission
SWIR	Short Wave Infra Red
SZA	Solar Zenith Angle
TANSO	Thermal And Near infrared Sensor for carbon Observation
TANSO-FTS	Fourier Transform Spectrometer on GOSAT
TBC	To be confirmed
TBD	To be defined / to be determined
TCCON	Total Carbon Column Observing Network
TIR	Thermal Infra Red
TM5	Transport Model 5
TR	Target Requirements
TRD	Target Requirements Document
VZA	Viewing Zenith Angle

General definitions

Table 1 lists some general definitions relevant for this document.

Table 1: General definitions.

Item	Definition
XCO_2	Column-averaged dry-air mixing ratios (mole fractions) of CO_2
XCH_4	Column-averaged dry-air mixing ratios (mole fractions) of CH_4
L1	Level 1 satellite data product: geolocated radiance (spectra)
L2	Level 2 satellite-derived data product: Here: CO_2 and CH_4 information for each ground-pixel
L3	Level 3 satellite-derived data product: Here: Gridded CO_2 and CH_4 information, e.g., 5 deg times 5 deg, monthly
L4	Level 4 satellite-derived data product: Here: Surface fluxes (emission and/or uptake) of CO_2 and CH_4

Scope of document

This document is an Algorithm Theoretical Basis Document (ATBD) for the Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/) component as covered by project C3S_312a_Lot6 led by University of Bremen, Germany.

Within project C3S_312a_Lot6 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 C3S_312a_Lot 6 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_GOS_SRPR.

These product is the XCH_4 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 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 2 is calculated. XCH_4 is then calculated by multiplying the ratio with XCO_2 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 vapor, 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 CH_4 and CO_2 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 CH_4 (XCH_4)

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 CH_4, denoted XCH_4 (in parts per billion, ppb).

In the following, several satellite instruments are shortly described which are used / can be used to generate the XCH_4 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 CO_2, CH_4 and O_2 absorption bands in the NIR and SWIR spectral region as needed for accurate XCO_2 and XCH_4 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 has been 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 via performing 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 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 (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

channel	wavelength range [nm]	resolution [cm^{-1}]
1	758-775	0.2
2	1460-1720	0.2
3	1920-2080	0.2
4	5560-14300	0.2

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

GOSAT Level 1 Product Description Document, Japan Aerospace Exploration Agency, March 2010, NEB-080035D, available through https://data.gosat.nies.go.jp/productpubinfo/productpubinfo/ProductPubInfoPage/open.do

Data quality and reliability

The quality of the retrieved CO2 and CH4 columns has been tested against ground-based observations (i.e. the TCCON network) 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.

Data class: Model

Required ECMWF data:

Class: ERA interim

Stream: Atmospheric model

Type: Analysis

Dates: 01/04/2009 to 01/01/2017

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

Spatial grid: N128 Quasi-regular Gaussian grid (~0.7^{o})

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 LMD inversions

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 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 v15r4 product that exclusively assimilates about 130 sites of surface air sample measurements from the Global Atmosphere Watch programme

Data class: Model

Data availability & coverage:

CarbonTracker 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 v15r4 inversion product has been validated with many independent measurements and is considered to provide be an accurate global CO2 data set.

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 CO_2 and CH_4 is available publically 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 validaton (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).

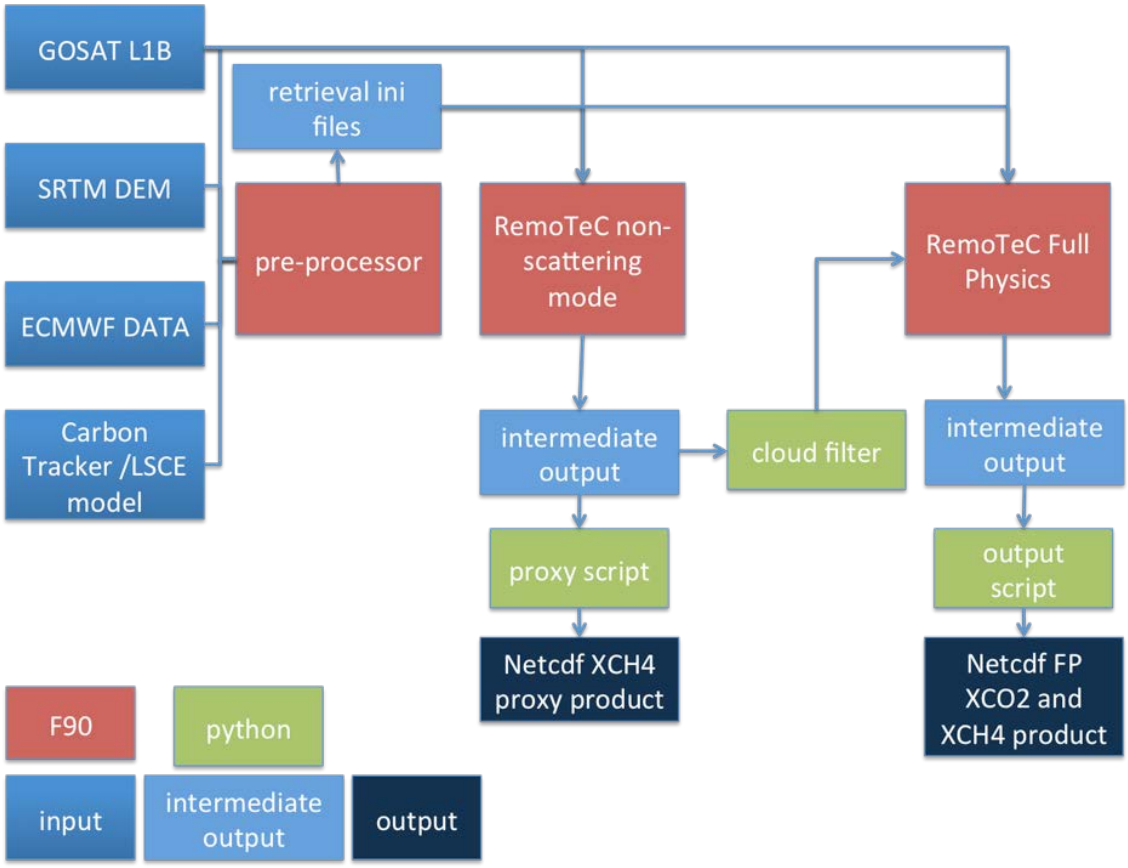
2.2.6 Additional input data

- Absorption cross sections: For the retrieval lookup-tables are used with pre-calculated absorption cross sections of the species of interest (O_2, CO_2, CH_4, H_2O) 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

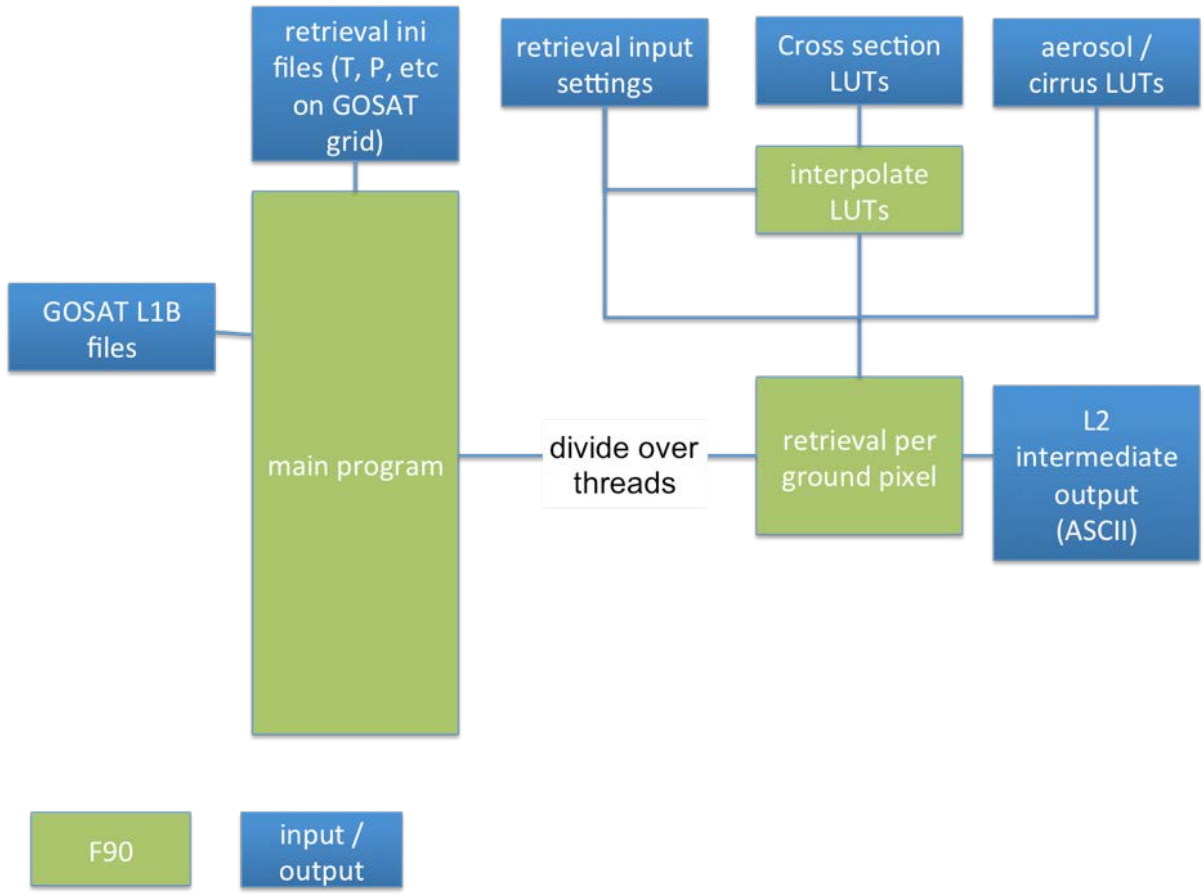
Figure 1 provides a schematic overview of the RemoteC GHG-CCI processing sub system at SRON The first step is to download the required data from the respective data servers to SRON (GOSAT data and ECMWF data are dynamic datasets that are continuously updated, SRTM topography is a static dataset). In 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.



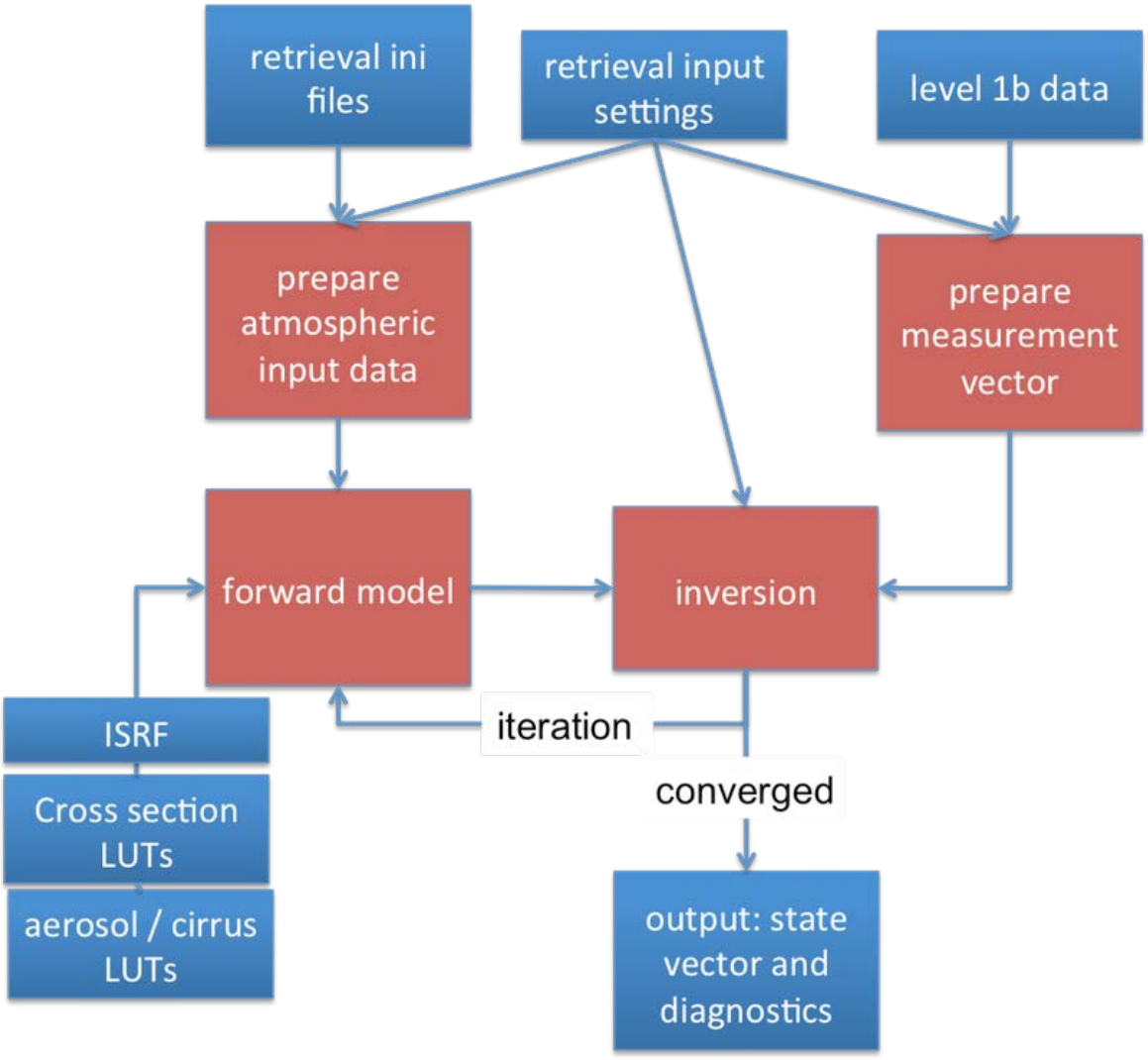
The pre-processor produces for each GOSAT L1B file an auxiliary input file, hereafter referred to as the 'retrieval ini ' file, that contains this information. In the next step columns of CO_2, CH_4, H_2O, and O_2 are retrieved under the assumption of an atmosphere without aerosol/cirrus/cloud scattering ('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 posterior* 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,

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

where $\mathbf{e}_y$ 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 $\mathbf{e}_F$,

$$\mathbf{y} = \mathbf{F}(x, b) + \mathbf{e}_\mathbf{y} + \mathbf{e}_\mathbf{F} \tag{2}$$

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

band	used spectral range
1 (CO_2)	6170 – 6277 cm^{-1}
2 (CH_4)	6045 – 6138 cm^{-1}

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

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

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

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

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

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

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_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:

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

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

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:

$$DV\_CH4_k = XCH4_k \times DV\_AIR_k \quad (7)$$

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

$$DV\_AIR_k = (p_{lev,k+1} - p_{lev,k}) \times R/(M \times g_k \times \left(1 + \frac{XH_2O_k}{1.60855}\right)) \quad (8)$$

where R is Avogado'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 $m = m_r + im_i$
- 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 modeled using Lambert Beers's law and surface reflection:

$$I_{TOA} = R_{surf} F_0 e^{-(\frac{\tau}{\mu_0} + \frac{\tau}{\mu_v})} \quad (9)$$

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 (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).

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

State vector element	**A priori value**
CH_4 sub-columns	TM4
CO2 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

3.4 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 γ according to

$$\hat{\boldsymbol{x}} = \min_{x}(||\mathbf{S}_y^{-\frac{1}{2}}(\mathbf{F}(\mathbf{x}) - \mathbf{y})||^2 + \gamma||\mathbf{W}(\mathbf{x} - \mathbf{x}_a)||^2) \tag{10}$$

where $\boldsymbol{S_y}$ is the diagonal measurement error covariance matrix, which contains the noise estimate. $\mathbf{x}_a$ is an a priori state vector (see Table 3-2), 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

$$\mathbf{x}_{n+1} = \min(\| \mathbf{K}'(\mathbf{x}'-\mathbf{x}_n') - \mathbf{y}'\|^2 + \gamma \| \mathbf{x}'-\mathbf{x}_a' \|^2) \tag{11}$$

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

The solution reads

$$\mathbf{x}_{n+1} = \mathbf{G}'\mathbf{y}' + \mathbf{A}'\mathbf{x}_n' + (I - \mathbf{A}')\mathbf{x}_{apr}' \tag{12}$$

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

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

$$\mathbf{x}_{retr} = \mathbf{A}\mathbf{x}_{true} + (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{G}\mathbf{e}_y + \mathbf{G}\mathbf{e}_F \quad (13)$$

The covariance matrix $\mathbf{S}_x$ describing the retrieval noise ($\mathbf{Ge}_y$) is given by

$$\mathbf{S}_x = \mathbf{G}\mathbf{S}_y\boldsymbol{G}^{\boldsymbol{T}} \quad (14)$$

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

$$XCH_4 = \frac{V_{ch4}}{V_{co2}} \times XCO_{2,prior} \quad (15)$$

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

$$V_{ghg} = \boldsymbol{h_{ghg}}^T x_{retr} \quad (16)$$

where the substract 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

$$\Delta XCH_4 = \frac{\sum_{i=1}^{12}\sum_{j=1}^{12} S_{x,i,j}}{V_{air,dry}} \quad (17)$$

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 $\mathbf{ax}_{true}$ / $V_{air,dry}$, where the quantity

$$\mathbf{a} = \mathbf{h}^T\mathbf{A} \quad (18)$$

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 ***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 ***W*** is given by

$$\mathbf{W} = \mathbf{L}\mathbf{W}' \tag{19}$$

with

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & \cdots & & & & \\ -1 & 2 & -1 & & & & \\ \vdots & & \ddots & & \vdots & & \\ & & -1 & 1 & & & \\ & & \cdots & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \tag{20}$$

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

$$\mathbf{L}_1^T\mathbf{L}_1 \tag{21}$$

where $\mathbf{L}_1$ 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.

3.5 Cloud Filtering

Cloud filtering in RemoTeC proxy is based on retrieved columns of VO2, VCO2 and 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<VO2_retrieved/VO2_ecmwf<1.035, 0.98<VCO2_1.6micron/VCO2_2.0micron<1.15, 0.90<VH2O_1.6micron / 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 XCH_4 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

Buchwitz et al., 2013a: Buchwitz, M., M. Reuter, O. Schneising, H. Boesch, S. Guerlet, B. Dils, I. Aben, R. Armante, P. Bergamaschi, T. Blumenstock, H. Bovensmann, D. Brunner, B. Buchmann, J. P. Burrows, A. Butz, A. Chédin, F. Chevallier, C. D. Crevoisier, N. M. Deutscher, C. Frankenberg, F. Hase, O. P. Hasekamp, J. Heymann, T. Kaminski, A. Laeng, G. Lichtenberg, M. De Mazière, S. Noël, J. Notholt, J. Orphal, C. Popp, R. Parker, M. Scholze, R. Sussmann, G. P. Stiller, T. Warneke, C. Zehner, A. Bril, D. Crisp, D. W. T. Griffith, A. Kuze, C. O'Dell, S. Oshchepkov, V. Sherlock, H. Suto, P. Wennberg, D. Wunch, T. Yokota, Y. Yoshida, The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparison and quality assessment of near-surface-sensitive satellite-derived CO_2 and CH_4 global data sets, *Remote Sensing of Environment*, doi:10.1016/j.rse.2013.04.024, http://authors.elsevier.com/sd/article/S0034425713003520, 2013.

Buchwitz et al., 2015: Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Guerlet, S., Dils, B., Aben, I., Armante, R., Bergamaschi, P., Blumenstock, T., Bovensmann, H., Brunner, D., Buchmann, B., Burrows, J.P., Butz, A., Chédin, A., Chevallier, F., Crevoisier, C.D., Deutscher, N.M., Frankenberg, C., Hase, F., Hasekamp, O.P., Heymann, J., Kaminski, T., Laeng, A., Lichtenberg, G., De Mazière, M., Noël, S., Notholt, J., Orphal, J., Popp, C., Parker, R., Scholze, M., Sussmann, R., Stiller, G.P., Warneke, T., Zehner, C., Bril, A., Crisp, D., Griffith, D.W.T., Kuze, A., O'Dell, C., Oshchepkov, S., Sherlock, V., Suto, H., Wennberg, P., Wunch, D., Yokota, T., Yoshida, Y., The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparison and quality assessment of near-surface-sensitive satellite-derived CO2 and CH4 global data sets. Remote Sens. Environ. 162:344–362, http://dx.doi.org/10.1016/j.rse.2013.04.024, 2015.

Buchwitz et al., 2016: Buchwitz, M., Reuter, M., Schneising, O., Hewson, W., Detmers, R. G., Boesch, H., Hasekamp, O. P., Aben, I., Bovensmann, H., Burrows, J. P., Butz, A., Chevallier, F., Dils, B., Frankenberg, C., Heymann, J., Lichtenberg, G., De Mazière, M., Notholt, J., Parker, R., Warneke, T., Zehner, C., Griffith, D. W. T., Deutscher, N. M., Kuze, A., Suto, H., and Wunch, D.:, Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO_2 and XCH_4 CRDP3 data, Remote Sensing of Environment (in press), Special Issue on Essential Climate Variables, DOI: 10.1016/j.rse.2016.12.027, (link: http://dx.doi.org/10.1016/j.rse.2016.12.027), 2016.

Buchwitz et al., 2017: ESA Climate Change Initiative (CCI) Product Validation and Intercomparison Report (PVIR) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG) for data set Climate Research Data Package No. 4 (CRDP#4), Version 5.0, 9. Feb. 2017, link: http://www.esa-ghg-cci.org/?q=webfm_send/352, 2017.

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., 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 et al., 2010: Chevallier, F., P. Ciais, T. J. Conway, T. Aalto, B. E. Anderson, P. Bousquet, E. G. Brunke, L. Ciattaglia, Y. Esaki, M. Fröhlich, A.J. Gomez, A.J. Gomez-Pelaez, L. Haszpra, P. Krummel, R. Langenfelds, M. Leuenberger, T. Machida, F. Maignan, H. Matsueda, J. A. Morguí, H. Mukai, T. Nakazawa, P. Peylin, M. Ramonet, L. Rivier, Y. Sawa, M. Schmidt, P. Steele, S. A. Vay, A. T. Vermeulen, S. Wofsy, D. Worthy, 2010: CO_2 surface fluxes at grid point scale estimated from a global 21-year reanalysis of atmospheric measurements. J. Geophys. Res., 115, D21307, doi:10:1029/2010JD013887

Detmers et al., 2015: Detmers, R. G., O. Hasekamp, I. Aben, S. Houweling, T. T. van Leeuwen, A. Butz, J. Landgraf, P. Koehler, L. Guanter, and B. Poulter, Anomalous carbon uptake in Australia as seen by GOSAT, Geophys. Res. Lett., 42, doi:10.1002/2015GL065161, 2015.

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.

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.

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

TRD GHG, 2017: Buchwitz, M., Aben, I., Anand, J., Armante, R., Boesch, H., Crevoisier, C., Detmers, R. G., Hasekamp, O. P., Reuter, M., Schneising-Weigel, O., Target Requirement Document, Copernicus Climate Change Service (C3S) project on satellite-derived Essential Climate Variable (ECV) Greenhouse Gases (CO_2 and CH_4) data products (project C3S_312a_Lot6), Version 1, 28-March-2017, pp. 52, 2017.

Velazco et al. 2011: Velazco, V. A., Buchwitz, M., Bovensmann, H., Reuter, M., Schneising, O., Heymann, J., Krings, T., Gerilowski, K., and Burrows, J. P.: Towards space based verification of CO2

emissions from strong localized sources: fossil fuel power plant emissions as seen by a CarbonSat constellation, Atmos. Meas. Tech., 4, 2809-2822, 2011.

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: info@copernicus-climate.eu

climate.copernicus.eu copernicus.eu ecmwf.int