ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Algorithm Theoretical Basis Document (ATBD) – ANNEX A for products CO2_GOS_OCFP (v7.3), CH4_GOS_OCFP (v7.3) & CH4_GOS_OCPR (v9.0) (CDR6, 2009-2021)

C3S2_312a_Lot2_DLR – Atmosphere

Issued by: Hartmut Boesch, Antonio Di Noia, University of Leicester, Leicester, UK

Date: 31/01/2023

Ref: C3S2_312a_Lot2_D-WP1_ATBD-2022-GHG_ANNEX-A_v6.2

Official reference number service contract: 2021/C3S2_312a_Lot2_DLR/SC1

PROGRAMME OF THE EUROPEAN UNION

IMPLEMENTED BY ECMWF

Contributors

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

UNIVERSITY OF LEICESTER, LEICESTER, UK (UoL)
H. Boesch (now at IUP, Univ. Bremen)
P. Somkuti
R. Parker
J. Anand
A. Di Noia

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	Updated to reflect changes made to the aerosol a priori for v7.2 data 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: Primarily correction of typos. Some additional information added.	All
4.0	18-August-2020	Updated for CDR4 Updated to reflect changes in ERA reanalysis data used for meteorological a priori	All
5.0	18-February-2021	Update for CDR 5 (2009-2020) One reference updated	2.2.3
6.0 Draft	18-February-2022	Update for data set CDR6: First draft	All
6.0	4-August-2022	Update for data set CDR6: GOSAT retrievals (2009-2021)	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	Update after 2nd review. Several improvements at various places.	All

List of datasets covered by this document

Deliverable ID	Product title	Product type (CDR, ICDR)	Version number	Delivery date
WP2-FDDP-GHG-v1	CO2_GOS_OCFP	CDR 6	7.3	31-Aug-2022
WP2-FDDP-GHG-v1	CH4_GOS_OCFP	CDR 6	7.3	31-Aug-2022
WP2-FDDP-GHG-v1	CH4_GOS_OCPR	CDR 6	9.0	31-Aug-2022

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 6 (2003-2021), project C3S2_312a_Lot2_DLR – Atmosphere, v6.2, 2023. *(this document is an ANNEX to the Main ATBD)*
D2	PUGS – Annex A: Boesch, H., et al., Product User Guide and Specification (PUGS) – ANNEX A for products CO2_GOS_OCFP, CH4_GOS_OCFP (v7.3, 2009-2021) & CH4_GOS_OCPR (v9.0, 2009-2021), 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
LIDORT	Linearised Discrete Ordinate Radiative Transfer
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 CO2 and CH4 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 ChartographY
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 Wava 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 are 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 satellite-derived GHG 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 algorithms to generate the C3S products CO2_GOS_OCFP (v7.3), CH4_GOS_OCFP (v7.3) and CH4_GOS_OCPR (v9.0).

These products are XCO_2 and XCH_4 Level 2 products as retrieved from GOSAT using algorithms developed at the University of Leicester, UK.

Executive summary

The goal of the greenhouse gas (GHG) activities of the C3S_312b_Lot2 project of the Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/) is the extension of the Atmospheric carbon dioxide (CO2) and methane (CH4) Climate Data Record (CDR) data products. These data products are generated from the satellite instruments SCIAMACHY/ENVISAT, TANSOFTS/GOSAT and OCO-2 (XCO2 and/or XCH4 products) and AIRS and IASI (mid/upper tropospheric CO2 and CH4 products). The MAIN ATBD document (Algorithm Theoretical Basis Document (ATBD) – Main document for Greenhouse Gas (GHG: CO2 & CH4) data set CDR3 (2003-2018) provides a very short overview about the data products and their underlying retrieval algorithms.

This algorithm theoretical basis document (ATBD) as an appendix to the MAIN ATBD describes the University of Leicester (UoL) retrieval algorithms used to retrieve column averaged dry air mole fractions of carbon dioxide (XCO_2) and methane (XCH_4) from spectra measured by the Greenhouse Gases Observing Satellite (GOSAT). This document details the various input data required for retrievals, the physical theory, and the mathematical background underlying retrieval assumptions, and outlines the retrieval implementation and the limitations of the approach used.

This document is divided in four sections. Section 1 describes the TANSO-FTS instrument onboard the GOSAT satellite. Section 2 describes the input data used for generating the UoL XCO_2 and XCH_4 data products. Section 3 describes the algorithms used in order to generate the data products. Section 4 describes the general structure of the output products.

1. Instruments

1.1 TANSO-FTS onboard GOSAT

1.1.1 The GOSAT Mission

GOSAT was developed jointly by the Japan Aerospace Exploration Agency (JAXA), the National Institute for Environmental Studies (NIES) and the Ministry of Environment (MOE) with the primary aim to estimate emissions and absorptions of greenhouse gases with an increased accuracy on a sub-continental scale, and to provide the environmental administration with an assessment of forest carbon balances, regional emissions and absorptions. Research using GOSAT will provide an enhanced understanding of the global distribution and temporal variations of greenhouse gases. This will develop our knowledge of the carbon cycle on a global scale and its influence on the climate, which is essential for the prediction of climate change and its possible impacts. Additionally, it aims to lead to new developments in both Earth observation satellite technologies and the approach of greenhouse gas measurements.

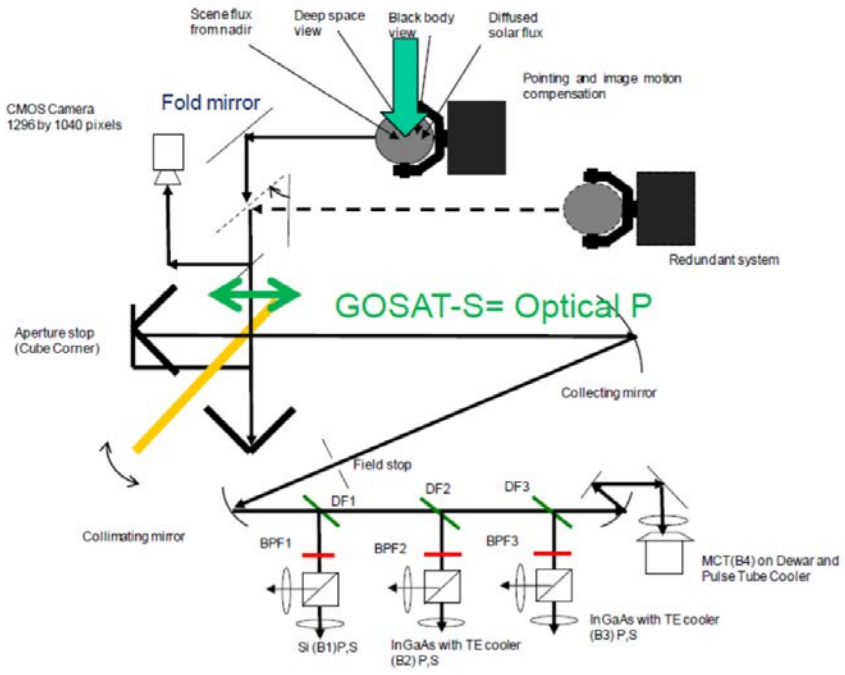
GOSAT was successfully launched on 23 January 2009 by JAXA's H-IIA rocket and is the first dedicated greenhouse gas satellite. Aboard GOSAT is the Thermal And Near-infrared Sensor for carbon Observation (TANSO) instrument which comprises two sensors; a Fourier Transform Spectrometer (FTS) that is used to measure greenhouse gases, and a Cloud and Aerosol Imager (CAI) that provides information about the scene, such as cloud and aerosol quantities. Both sensors are kept looking towards the Earth's geocentric direction by a three-axis attitude control system. The power required for this as well as the sensors operation is obtained from two solar paddles which face the Sun. The satellite has a Sun-synchronous orbit, optimising the power supply, and is quasi-recurrent with 14 + 2/3 revolutions per day. The follow-up of the GOSAT satellite is GOSAT-2, which was launched on 29 October 2018. Similarly to GOSAT, also GOSAT-2 hosts a Fourier Transform Spectrometer (FTS-2) and a cloud and aerosol imager (CAI-2). The main novel elements of GOSAT-2 compared to its predecessor are the extension of FTS-2 band 3 to 2.33 µm, the availability of an additional thermal channel in FTS-2 and the availability of two viewing angles in CAI-2. While the products described in this document are obtained using GOSAT data only, we may consider including products based on GOSAT-2 in the future.

1.1.2 The TANSO-FTS Instrument

TANSO-FTS comprises of three units: an electronic circuit unit, a control unit and an optical unit. The electronic unit is composed of a main and a redundant digital processing circuit that functions as a data processor, data-telemetries-commands-transmission interface with the satellite bus system and includes the temperature controller. The control unit acts as a controller for the cooling machine, the electronic cooling machine, the optical path switcher and solar irradiance diffusing board. It also functions as a driver and controller of the interferometer and the pointing mechanism. The optical unit, a diagram of which is shown in Figure 1, contains the pointing mechanism (with redundancy),

the blackbody diffusing board, the solar irradiance diffusing mechanism, the optical path switcher, monitoring camera, Fourier interferometer, relay and band splitting optics, the detector optics, analogue processing circuit and the cooling machine. The interferometer, based on the original Michelson interferometer, is a device that can divide a beam of radiation into two separate paths and then recombine these after a path difference has been introduced. This creates a condition whereby interference between the two beams can occur. The intensity variations of the resulting beam is measured as a function of path difference by the detector.

Figure 1: TANSO-FTS instrument design showing the major optic components and optical path, *Kuze et al., 2009*.



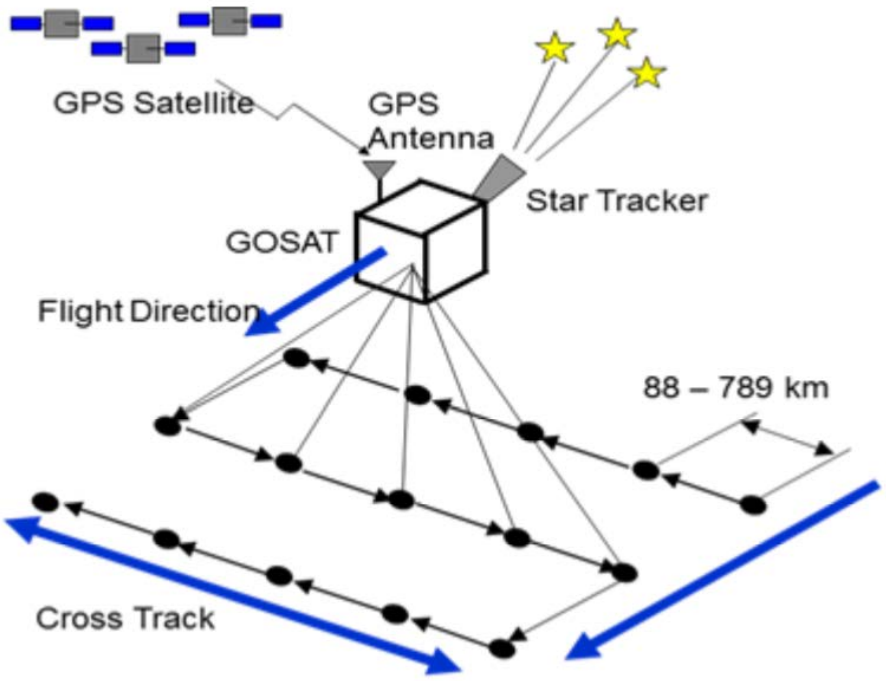
The interferometer used by TANSO-FTS is of double pendulum type as shown in Figure 1. It uses corner cube mirrors, which are statically aligned with each other and the rotary motor to maximise modulation efficiency with the minimum optical shear. These are attached to a swing arm that is moved by the rotary motor with less than 1% speed stability, creating uniform frequency of the output modulated output signal. The rotary speed is required to take four seconds in order to achieve a small enough sampling electrical bandwidth and the necessary signal to noise ratio (SNR). This is used as the nominal time for interferogram acquisition, although TANSO-FTS is also capable of faster scanning modes that allow denser observations but at the expense of SNR. Between the mirrors is a beam splitter, where a beam of radiation from an external source can be partially reflected to one mirror and partially transmitted to the other mirror. Once reflected from the mirrors both beams return to the beam split and are partially reflected, transmitted and interfere. Due to the effect of interference, the intensity of each beam reaching the detector depends on the optical path difference (OPD) of the two beams. The Optical Path Difference (OPD) created are +/- 2.5 cm, four times as long as the mechanical motion, and are obtained to make both the real and imaginary spectra for suitable phase correction. The width of the beam splitter was selected to be larger than the maximum OPD to

try to reduce the effects of channelling. The resulting light is reflected from a collecting mirror towards a circular slit, which due to its physical geometry forces the different spectral bands to be adequately aligned. The circular slit's geometry was designed to provide the maximum efficiency while maintaining within the maximum IFOV divergence for band 2 for 0.2 cm^{-1} spectral resolution. This defines the instantaneous field of view (IFOV) to be 15.8 mrad, equivalent to 10.5 km projected onto the surface of the Earth. The narrowed light passes to a collimating mirror and is then split through dichroic filters for each band in series. To minimize the background TIR, these filters transmit longer wavelength light and reflect shorter wavelength light. This minimizes the polarization sensitivity of the shorter wavelengths which then travel though a narrow band-pass filter. These are divided by a beam splitter into two polarizations that are then measured by two detectors. The bandpass filters are fixed to reduce channelling, to decrease stray light and to discard shorter wavelength light to evade aliasing (indistinguishable signals). While band 1 is measured by two Si detectors, both bands 2 and 3 are measured using InGaAs detectors that are nonbiased and cooled to -40 degrees C using thermoelectric coolers which minimizes dark currents. The TIR light is observed by a Photo-Conductive (PC) HgCdTe (MCT) detector that is cooled to 70 K using a pulse tube cooler with low vibration, which does not influence the interferogram.

1.1.3 Observing modes

GOSAT TANSO-FTS nominally performs a cross-track scanning pattern where the number of lattice points can be varied from one to nine points per one cross-track scan, until August 2010 5 points have been used and afterwards 3 points have been used to reduce pointing errors caused by vibrations.
A diagram illustrating the TANSO-FTS scanning pattern, for the setup using 5 lattice points, is shown in Figure 2. Additionally, it can measure in a specific observation mode which provides targeted observations for validation and experimental purposes. It can also measure over the ocean using the Sun glint mode, whereby the along-track and cross-track mechanism can be used to point at the angle where sunlight is reflected from the ocean causing the radiance to be very high. This overcomes the difficulty of measuring SWIR over the ocean that has a very low albedo, and hence a low radiance that is hard to measure. The fourth instrument channel measures the thermal infrared (TIR) region that allows vertical profiling over both land and ocean. The instrument uses a solar irradiance calibration for the SWIR observations and a blackbody calibration for the TIR observations. It also performs deep-space and lunar calibrations for both SWIR and TIR bands.

Figure 2: Diagram illustrating the observation points of GOSAT in 5-point mode, Hamazaki, 2006.



2. Input and auxiliary data

2.1 Satellite instrument

2.1.1 Data product delivery

The GOSAT L1B data is downloaded via the GOSAT Data Archive Service (GDAS) and stored at the University of Leicester, where it is backed up and processed to provide radiometrically calibrated radiances with derived a priori values. Each observation is then processed further to estimate the column averaged dry air mole fraction of CO_2 and CH_4 (Level 2 Product), along with a number of other retrieved parameters such as surface pressure, temperature, water vapour, albedo, aerosol profiles, cirrus profile, CO_2 column averaging kernels, and a number of quality control products.

2.1.2 L1B processing

All products discussed in this work use GOSAT radiances taken from a combination of the JAXA v201 and v202 L1B product, (Kuze et al., 2016), over which the SWIR channel data remains the same. In this product the radiometric conversion coefficients and the degradation correction are applied by JAXA already and the radiances given in the L1B files are not further manipulated before using them in the Level-2 retrieval algorithm.

The noise for each SWIR band and polarization components P & S is estimated by taking the standard deviation of the out-of-band radiance values of the GOSAT L1B raw spectra and multiplying it with the radiometric calibration and degradation correction. The polarizations of the noise for each band are combined to give the noise for each band, using:

$$Noise = \frac{\sqrt{NoiseP^2 + NoiseS^2}}{2}. \qquad (1)$$

2.1.3 Dispersion

The dispersion is given in the GOSAT L1B data but requires adjusting in all SWIR bands. This adjustment is calculated by the spectral difference of the spectra to a well-known single strong Fraunhofer line (12985.163 cm^{-1}) in the O_2-A band. It should be understood that although all SWIR bands' dispersions are adjusted they are only approximate and require the dispersion to be retrieved.

2.2 Other

2.2.1 Profile of Carbon Dioxide Concentration

The CO_2 a priori profile is obtained from LMDZ MACC-II v14r2 CO_2 model fields (Chevallier et al., 2010). The CO_2 profile is interpolated with latitude, longitude and time to the specific location desired. MACC-II CO_2 model data is available up to December 2014. For later years, the CO_2 is incremented according to yearly global CO_2 increases given by NOAA. The choice of keeping the MACC-II 14r2 as source for the CO_2 a priori is dictated by the need of maintaining stability of the CO_2 time series. In the eventuality of a future reprocessing of the entire GOSAT dataset, the use of CAMS 18r2 model data will be considered.

2.2.2 Profile of Methane Concentration

The a priori profile of CH_4 is obtained by merging the MACC-II Reanalysis (S1-NOAA) dataset with a TOMCAT full chemistry model run, which in turn has been validated against stratospheric observations made by the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS). These atmospheric profiles are then interpolated to the desired latitude, longitude and time.

2.2.3 Surface Pressure

Both under-constrained and over-constrained retrieval scenarios rely on a priori information to constrain the solution to be physically reasonable. In order to perform retrievals, the a priori need to be well constrained. In particular, it is most important that the surface pressure is well constrained since this can have the largest effect on the retrieved spectral line depths. The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim (Dee et al., 2011) is a reanalysis product that uses observations from surface buoy and satellite measurements. It provides atmospheric profiles of pressure, temperature and specific humidity on a 0.75 degree by 0.75 degree global grid with 60 levels. Discontinued in August 2019, ERA-Interim has been replaced by the ERA-5 reanalysis (Hersbach et al., 2020), which has a grid of 0.25 by 0.25 degrees and 137 levels. Given the latitude, longitude and altitude of a site of interest, the surface pressure can be determined from these profiles. ECMWF provide geopotential data for the lowest level of the same grid, which can be used to find the geopotential height of each grid point level as:

$$\text{Height} = \Phi / g \qquad (2)$$

where Φ is the geopotential and g is the gravitational acceleration, which is calculated as a function of latitude and approximate altitude. Taking the four surrounding grid points of the site of interest, the pressure, P, at the site altitude can be found for each grid point by using the hydrostatic equation:

$$P = P_0 e^{\left(-\frac{z}{z_0}\right)} \quad (3)$$

where P_0 is the pressure of the grid point level lower than the site altitude, Z is the difference in altitude between the grid point level and the site altitude and Z_0 is the scale height defined as:

$$Z_0 = \frac{RT}{Mg} \quad (4)$$

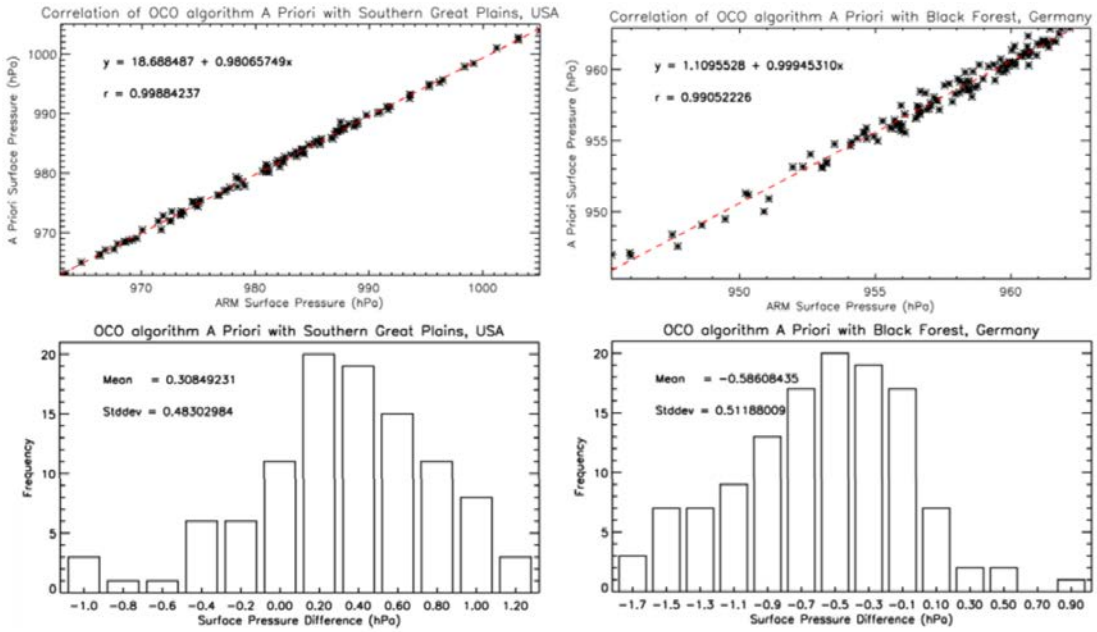
where R is the ideal gas constant, T is the average temperature across the differential, M is the Molar mass of wet air and g is the gravitational acceleration as a function of latitude and altitude. The Molar mass of wet air is calculated by:

$$M = \rho_d(1 - SH) + \rho_w SH \quad (5)$$

where SH is the ECMWF specific humidity, ρ_d is the dry air mass and ρ_w is the mass of wet air. The site altitude is obtained from the Shuttle Radar Topography Mission (SRTM) dataset (http://viewfinderpanoramas.org/dem3.html). Note, that in the case that the site altitude is lower than the lowest level of a grid point the pressure is calculated with respect to the lowest level, where the temperature and molar mass are extrapolated downwards based on the lapse rate and gradient of the 5 lowest levels above, respectively. The surface pressure for the site can then be resolved by interpolating the pressures with latitude, longitude and time.

To validate the surface pressure two sites were chosen: Southern Great Plains (SGP)/USA, and Black Forest/Germany, which are flat and mountainous in topography respectively. The Atmospheric Radiation Measurement (ARM) program has a central facility in SGP which takes in situ ground based measurements of surface pressure using the Temperature, Humidity, Wind and Pressure Sensors (THWAPS) instrument. Figure 3 shows that the surface pressure determined is within 1 hPa of THWAPS observations. Surface Pressure observations were also made by the ARM mobile facility that visited Black Forest, Germany, in 2007 with the Surface Meteorological Instrument (MET). The surface pressure a priori can be constrained to less than 0.1%, although we use a 1- sigma error of 4 hPa on the a priori surface pressure estimate to allow for more difficult topographies.

Figure 3: Validation of a priori surface pressure with in-situ ground based measurements from ARM instruments in Southern Great Plains/USA and Black Forest/Germany.



2.2.4 Temperature and Water Vapour

The ECMWF specific humidity data is used to generate water vapour volume mixing ratio profiles using the equation:

$$\mathrm{H_2O}_{VMR} = 10^6 \left(\frac{SH}{(R_d/R_w) - \left(SH\left((R_d/R_w) - 1\right)\right)} \right) \qquad (6)$$

where R_d and R_w are the gas constants in dry and wet air respectively. ECMWF also provides temperature profiles which, along with the H_2O_{VMR}, are interpolated with latitude, longitude and time to the specific observation. A constant O_2 VMR profile of 0.2095 is also used as a priori. We have assumed the temperature and H_2O profiles are correct but allow the retrieval to scale them.

2.2.5 Aerosols

Previous operational UoL retrievals employed two aerosol particles with mixed optical properties to describe the observed aerosol load. They consisted of a carbonaceous, sooty continental mixture and a carbonaceous, dusty continental mixture as described in *Kahn et al., 2001*. Aerosol type 1 (small) and aerosol type 2 (large) correspond to mixture 5b and 2b in *Kahn et al., 2001* with optical properties displayed in Table 1. Both particles were assigned an a priori extinction of 0.05 with a Gaussian-shaped profile with a height and width of 2 km.

We have replaced this static a priori setup for aerosols with a dynamic (scene-dependent) setup informed by CAMS. The CAMS model calculates five different tropospheric aerosol types: sea salt (SS), dust (DU), organic matter (OM), black carbon (BC) and sulphate aerosols (SU). Based on these general typology, OM and BC are separated into hydrophobic and hydrophilic particles whereas SS and SU are treated as hydrophilic and DU as hydrophobic only. Furthermore, SS and DU are separated into three size bins. The optical properties of all particles were calculated following the original MACC scheme as outlined in Morcrette et al., 2009, and Reddy et al., 2005. Given the assumed size distributions, the optical properties of the MACC aerosols are calculated using Mie theory and the code provided in Mishchenko et al., 2002.

For v7.1 we used a combination of MACC and CAMS to give full coverage of the 2009-2017 observation period. The MACC model reanalysis dataset covered 2009-2012, while the CAMS near-real time model provides analyses from 2013 onwards. However, we found that the two datasets contained very large differences in the quantity of desert dust present over regions like the Sahara, causing significant discontinuities in the retrieved AOD time series around 2013.

To rectify this issue and avoid potential biases in the retrieved full-physics datasets, we now use a reference climatology based on CAMS data during 2014-2016. For each aerosol species we took the median mass mixing ratio for all years at each horizontal, vertical, and temporal grid point. The resulting climatology accounts for seasonal variations in aerosol loading and has the same resolution

as MACC and CAMS, allowing it to be readily integrated into our pre-processing scheme. However, this climatology does not account for fires that would otherwise be included in the CAMS near-real time model dataset.

Table 1: Basic aerosol properties for MACC aerosol types and spheroidal dust aerosols. If the respective particle type depends on rel. humidity, all values are given for a rel. humidity of 80%. Log-normal size distributions are assumed except for SS, which has a bimodal lognormal size distribution at number concentrations for the first and second modes of 70 and 3 cm^{-3}, respectively, *Reddy, 2005*; *O'Dowd et al., 1997*. The refractive index is indicated for 770nm. OM_{phob} and BM_{phi} share the same properties, *Reddy, 2005*. The spheroidal, medium and coarse dust particles consist of log-normal distributions of spheroids with a mixture of oblate and prolate particles.

Type	Shortname	dependent on rel. hum.	r_0	r_{min}	r_{max}	σ	n_{refr} [a]
Sea Salt 1	SS1	yes	0.199, 1.99	0.03	0.5	1.9, 2	1.39 + i1.2e-07
Sea Salt 2	SS2	yes	0.199, 1.99	0.5	5	1.9, 2	1.39 + i1.2e-07
Sea Salt 3	SS3	yes	0.199, 1.99	5	20	1.9, 2	1.39 + i1.2e-07
Dust 1	DU1	no	0.29	0.03	0.55	2	1.53 + i2.9e-03
Dust 2	DU2	no	0.29	0.55	0.9	2	1.53 + i2.9e-03
Dust 3	DU3	no	0.29	0.9	20	2	1.53 + i2.9e-03
Organic Matter	OM_{phob}	no	0.0355	0.002	20	2	1.49 + i5.0e-04
Organic Matter	OM_{phil}	yes	0.0355	0.002	20	2	1.42 + i5.0e-04
Black Matter	BM_{phob}	no	0.0118	0.005	0.5	2	1.75 + i4.3e-01
Black Matter	BM_{phil}	no	0.0118	0.005	0.5	2	1.75 + i4.3e-01
Sulphat	SU	yes	0.0355	0.002	20	2	1.4 + i1.4e-07
medium Dust	P19	no	0.5	0.1	1.0	1.5	1.53 + i2.9e-03
coarse Dust	P21	no	1.0	0.1	6.0	2.0	1.53 + i2.9e-03

These particle types are aggregated into aerosol type 1 (small) and aerosol type 2 (large). To mitigate errors from assuming a spherical shape for dust particles as in the CAMS, the dust particle of CAMS at three size bins is replaced by the medium and coarse dust particle used in the MISR retrievals. A set of calibration retrievals have been performed over the Sahara region to adjust the numbers of medium and coarse dust particles assigned to three size bins as of MACC. In order to keep the total extinction of dust as modelled by CAMS at 550 nm, where the model runs assimilate MODIS aerosol optical depth, it is ensured that total dust AOD remains the same.

CAMS aerosol profiles are interpolated to the individual time and location of GOSAT observations and the number of profile levels is downscaled from 60 to 20 for use in the retrieval. The retrieval uses two different aerosol types to fit observed aerosol scattering; one with a small and one with a large Angstrom coefficient.

Covariance matrixes are calculated for each individual scenario depending on the assigned a priori total AOD. Lower a priori parameters are allowed far greater relative variations than higher a priori values to reflect the lower relative accuracy of predictions for low values. The expected greater variability of aerosol properties is accounted for by allowing a standard deviation with a factor of 50

for profiles with a priori AOD < 0.05, which decreases linearly to 10 for profiles with a priori AODs < 0.2 and remains at a factor of 10 for values above.

2.2.6 Surface Properties

The a priori albedo value is calculated from the spectral continuum from the GOSAT L1B using the reflectivity:

$$\text{Albedo} = \frac{\pi I_{GOSAT}}{F_{Solar}\cos(\text{SZA})} \qquad (7)$$

where *SZA* is the solar zenith angle and the solar irradiance and observed GOSAT radiance are given by F_{Solar} and I_{GOSAT} respectively. The retrieval uses two albedo parameters for each spectral band, giving the albedo for the centre wavelength of the band and the slope of the albedo. The slope of the albedo is set to zero in the a priori. The covariance for albedo is completely open and the slope a priori error is set so that the band edges can vary by 50%.

3. UoL-FP Algorithm

3.1 Overview

The UoL-FP retrieval algorithm is based on the algorithm that was developed to retrieve XCO_2 from a simultaneous fit of the near-infrared O_2-A Band spectrum at 0.76 µm and the CO_2 bands at 1.61 and 2.06 µm as measured by the OCO instrument. While the algorithm was developed to retrieve XCO_2 from OCO and OCO-2 observations, it was designed to be adaptable to analyse data from other instruments for algorithm testing and validation. The algorithm has already been successfully used to retrieve XCO_2 and XCH_4 from observations from SCIAMACHY, GOSAT, TanSat, and ground-based Fourier Transform spectrometers (FTS) (Parker et al., 2011, 2020; Crisp et al., 2012, Boesch et al., 2006, Yang et al., 2020*)* and to carry out retrieval simulations for OCO and OCO-2, GOSAT, Sentinel 5, the ESA CAMELOT project and CNES Minicarb (Boesch et al., 2011; O'Dell et al., 2012).

The retrieval algorithm uses an iterative retrieval scheme based on Bayesian optimal estimation to estimate a set of atmospheric/surface/instrument parameters, referred to as the state vector **x**, from measured, calibrated spectral radiances.

The Forward model describes the physics of the measurement process and relates measured radiances to the state vector **x**. It consists of a radiative transfer (RT) model coupled to a model of the solar spectrum to calculate the monochromatic spectrum of light that originates from the sun, passes through the atmosphere, reflects from the Earth's surface or scatters back from the atmosphere, exits at the top of the atmosphere and enters the instrument. The top of atmosphere (TOA) radiances are then passed through the instrument model to simulate the measured radiances at the appropriate spectral resolution. The Forward model employs the LIDORT radiative transfer model combined with a fast 2-orders-of-scattering vector radiative transfer code, Natraj et al., 2008. In addition, the code uses the low-streams interpolation functionality, O'Dell, 2010, to accelerate the radiative transfer component of the retrieval algorithm.

The inverse method employs the Levenberg-Marquardt modification of the Gauss-Newton method to find the estimate of the state vector $\hat{\mathbf{x}}$ with the maximum a posteriori probability, given the measurement **y** (Connor et al., 2008; Rodgers, 2000). The state vector will typically include a CO_2 (or CH_4) profile together with non-CO_2 state vector. After this iterative retrieval process has converged to a solution, the error covariance matrix $\hat{\mathbf{S}}$

$$\hat{\mathrm{S}} = \left(\mathrm{K}^{\mathrm{T}}\mathrm{S}_{\varepsilon}{}^{-1}\mathrm{K} + \mathrm{S}_{\mathrm{a}}{}^{-1}\right)^{-1} \qquad (8)$$

and the averaging kernel matrix **A**

$$\mathrm{A} = \frac{\partial \hat{\mathrm{x}}}{\partial \mathrm{x}} = \hat{\mathrm{S}}\mathrm{K}^{\mathrm{T}}\mathrm{S}_{\varepsilon}{}^{-1}\mathrm{K} \qquad (9)$$

are calculated using the a priori covariance matrix $\mathbf{S_a}$ and the measurement covariance matrix $\mathbf{S_\varepsilon}$. XCO_2 is inferred by averaging the retrieved CO_2 profile, weighted by the pressure weighting function, **h**, such that

$$\mathrm{XCO_2} = \mathrm{h^T}\hat{\mathrm{x}} \qquad (10)$$

The associated column averaging kernel for a level *j* is then given by

$$(\mathbf{a}_{\mathrm{CO2}})_\mathrm{j} = \frac{\partial \mathrm{XCO_2}}{\partial \mathrm{u_j}} \frac{1}{\mathrm{h_j}} = (\mathrm{h^T A})_\mathrm{j} \frac{1}{\mathrm{h_j}} \qquad (11)$$

and the variance of XCO_2 is given by

$$\sigma_{\mathrm{XCO2}} = \mathrm{h^T \hat{S} h}. \qquad (12)$$

The main parameters for the characterization of the XCO_2 retrieval that are calculated by the retrieval algorithm are the a posteriori XCO_2 retrieval error given by the square root of the variance σ_{XCO2} and the column averaging kernel $\mathbf{a}_{CO2}$.

3.2 Forward model

3.2.1 Solar model

The monochromatic TOA spectrum calculated by the RT code is multiplied with a synthetic solar spectrum, calculated with an algorithm based on an empirical list of solar line parameters (G. Toon, private communication). The solar line list covers the range from 550 to 15,000 cm^{-1} and is derived from FTS solar spectra: Atmospheric Trace Molecule Spectroscopy (ATMOS), MkIV balloon spectra for the range 550–5650 cm^{-1}, and Kitt Peak ground-based spectra for 5000–15,000 cm^{-1}. The solar model includes both disk centre and disk integrated line lists.

3.2.2 Radiative transfer

The radiative transfer (RT) model (RTM) attempts to approximate physics associated with the modification of the solar radiation during its passage through the atmosphere and reflection by the surface. The RTM calculates the top-of-atmosphere Stokes parameters *I*, *Q*, and *U* on a high resolution (0.01 cm^{-1}) wavelength grid. The Stokes parameter *V*, representing circularly polarized radiation, is ignored as it is generally negligible for most instruments. The solar spectrum is multiplied with the

high-resolution Stokes vectors calculated by the RTM, which are initially dimensionless reflectances, to obtain the physical radiance units.

A fully-polarimetric vector calculation of radiative transfer (RT) would be desirable to calculate the Stokes vector at each monochromatic wavelength. However, at 0.01 cm^{-1} resolution, this would lead to tens of thousands of computationally expensive RT calculations per Forward model run. In order to speed up the RT calculations, the code can use an approximate approach called "Low Streams Interpolation" (LSI, O'Dell, 2010), or a method based on principal component analysis (PCA, Somkuti et al., 2017*)*.
The basic idea of these approximate methods is that rather than performing full-accuracy calculations with a large number of angular streams at all monochromatic wavelengths, accurate calculations are only performed at a few tens of wavelengths. Very fast, low accuracy (2-stream) calculations are performed at all the monochromatic wavelengths, and are used in combination with the small number of high accuracy calculations in order to estimate of the Stokes vector at each monochromatic point.

Monochromatic RT calculations are made using a combination of a fast single-scattering model (Nakajima and Tanaka, 1988), the Linearized Discrete Ordinate Radiative Transfer (LIDORT) scalar multiple-scattering model (Spurr et al., 2001), and a second-order-of-scattering polarization model called 2OS (Natraj and Spurr, 2007). Neglecting higher orders of scattering for Q and U is shown to lead to radiance errors on the order of 20% or less of the expected OCO instrument noise, and XCO_2 errors typically on the order of a few tenths of a ppm or less (Natraj et al., 2008). The LSI method has radiance errors typically smaller than a tenth of a percent, O'Dell, 2010.

3.2.3 LIDORT

The Full Physics algorithm uses LIDORT (Spurr et al., 2001*;* Spurr, 2002) to solve the radiative transfer equation (RTE). LIDORT is a linearized discrete ordinate radiative transfer (RT) model that generates radiances and Jacobians (derivatives of the radiance with respect to atmospheric and surface properties) simultaneously. The Jacobians are computed by an internal perturbation analysis of the complete discrete ordinate solution to the RTE. LIDORT is a quasi-spherical model; the direct beam and line of sight attenuations are treated for a curved atmosphere while the multiple scattering is considered to be locally plane parallel.

3.2.4 Two Orders of Scattering (2OS) Model

Since multiple scattering is depolarizing, it is reasonable to expect that the polarization could be accounted for by a low-order scattering approximation. Natraj and Spurr, 2007 extended the scalar, *Kawabata and Ueno, 1988*, model to compute the first two orders of scattering (2OS) for vertically inhomogeneous scattering media with polarization included. To enable accurate computations for a large range of solar viewing angles, atmospheric transmittances for the incoming solar beam are treated for a curved spherical-shell atmosphere (Spurr, 2002). For glint and nadir modes of operation, there is also a correction for the sphericity along the line of sight. Polarization induces a change in the

intensity; to account for this, we compute a correction to the scalar intensity. The 2OS model simultaneously computes both the simulated backscatter radiance (and intensity correction) and any number of associated weighting functions (partial derivatives of the radiance with respect to retrieved and other atmospheric and surface properties). The 2OS computation is an order of magnitude faster than a full multiple scattering scalar calculation and two orders of magnitude faster than a vector multiple scattering computation.

3.2.5 Two stream model

Making use of the PCA-based method allows a significant speed-up of RT calculations when used in combination with a dedicated RT model for the low accuracy, low-stream line-by-line calculations. For the low-stream calculation, we utilise the two-stream model TWOSTR (Spurr and Natraj, 2011), which decreases the computation time compared to LIDORT when using N=2 streams.

3.2.6 Instrument model

The instrument model convolves the monochromatic radiance spectrum with the instrument line shape function (ILS). As described in, Boesch et al., 2006, the instrument model can also simulate continuum intensity scaling, zero-level offsets and channelling effects. The instrument model performs these actions and is described below. Additionally, the steps taken to calibrate the measured spectra are also described.

3.2.6.1 Pixel-Wavelength Mapping

The dispersion gives the pixel-wavelength mapping and consists of two parameters for each band; the start wavenumber *W* and the wavenumber increment *D* needed to reach the next pixel's wavenumber. The wavenumber *v* for detector pixel *i* and spectrometer *k* is given by

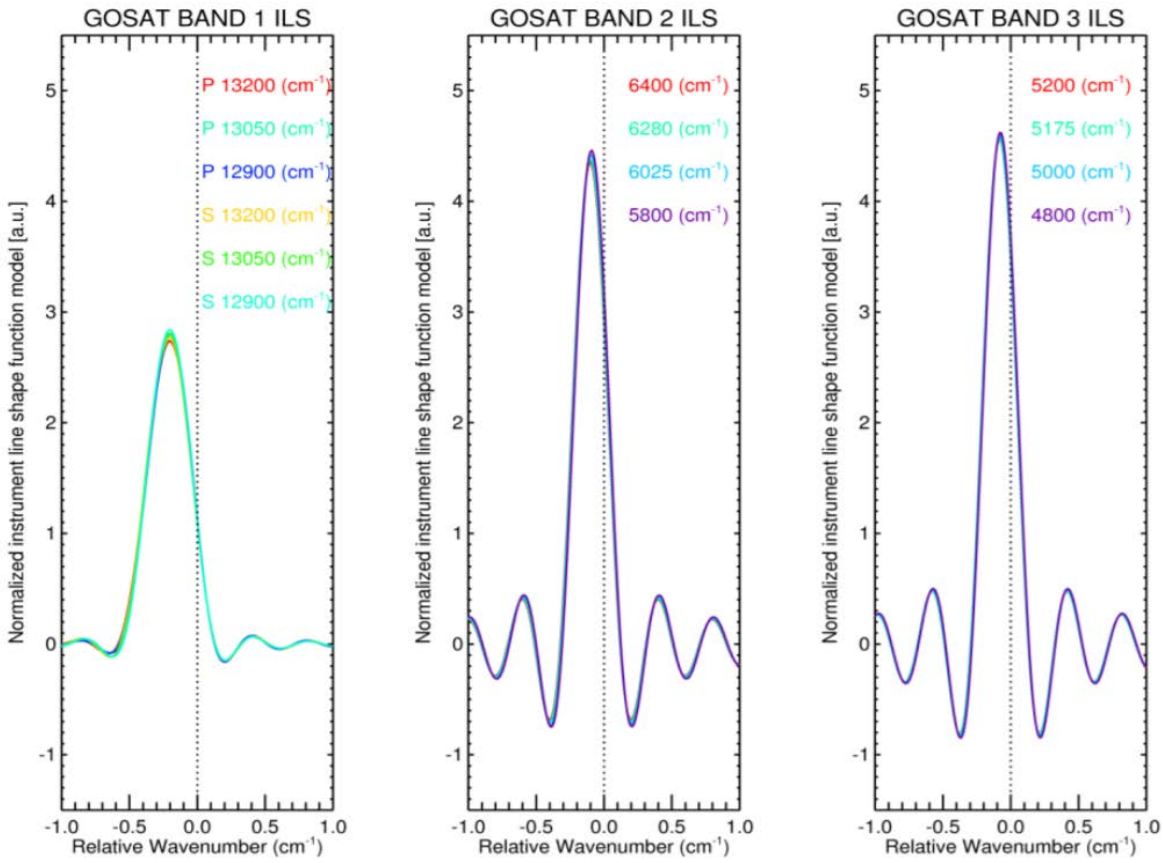
$$v_i = W_k + iD_k. \quad (13)$$

3.2.6.2 Instrument Line Shape Function (ILS)

To describe the response of the instrument to light with wavelength an instrument line shape function (ILS) is used, which is obtained from the JAXA ILSF model. This model is evaluated with the parameter adjustment by using the GOSAT pre-flight test results. Band 1 has the ILS provided for both polarizations due to band 1 being sensitive to displacement of the optical axis with its shorter wavelengths, whereas bands 2 and 3 have accurate alignment. The ILS model gives the ILS for three different wavelengths in band 1 and four different wavelengths in bands 2 and 3, which are interpolated between. The latest version of the ILS as provided by JAXA/NIES is used. The ILS model provided by JAXA can be seen in Figure 4, which gives the ILS function for 50 wavenumbers. However,

to optimise the retrieval we implement only a ±30 cm^{-1} spectral window, with little reduction of accuracy.

Figure 4: ILS used in the retrievals for each GOSAT spectral band with different wavenumbers given.



3.2.6.3 Polarization

The polarization option is turned off for GOSAT retrievals, but the OCO Full Physics Algorithm has the option to enable this in the future if necessary.

3.2.6.4 Solar Induced Fluorescence (SIF)

The Full Physics retrieval considers sun-induced chlorophyll fluorescence (SIF) over land as an additional radiance contribution in the oxygen A Band. Based on the method described in Frankenberg et al., 2011, SIF is first retrieved separately for two narrow micro windows at 775 nm and 772 nm, whereas the SIF value at the 772 nm window is multiplied by a factor of 1.7. Only the S polarised spectra are utilised.

Detector non-linearities are compensated using the same procedure as in Frankenberg et al., 2011, in which the values are calibrated as a function of the average radiance in the entire band. The reference areas, which are assumed to be devoid of any fluorescing sources, are chosen by matching each GOSAT sounding with the ESA-CCI 300 m Land Cover Product Version 1.6. Those GOSAT footprints which cover at least 85% bare, snow-covered or urban areas are used for the calibration procedure, which is performed for every month individually.

The entire SIF record is averaged spatio-temporally using a weighted mean. For each retrieved value, all GOSAT soundings within a given spatial cut-off *D* and a temporal cut-off *T* are identified and applied as weights to calculate the new smoothed value. The weights are calculated in the following manner:

$$w_d(d) = \exp\left(-\frac{d^2}{2(D/4)^2}\right) \qquad (14)$$

$$w_t(t) = \exp\left(-\frac{t^2}{2(T/4)^2}\right) \qquad (15)$$

where *d* and *t* are the distances in spatial and temporal dimension from the GOSAT measurement that is to be smoothed. *D* and *T* were chosen to be 250 km and 30 days, respectively. The weights are then multiplied to obtain the total weights, and the new, smoothed SIF value is then:

$$F_s^{avg} = \frac{\sum F_S(d,t) \cdot w(d,t)}{\sum w(d,t)} \qquad (16)$$

Where $F_s(d,t)$ is the retrieved and corrected SIF value at distances *d* and *t*. The sum runs over all GOSAT measurements within the chosen cut-offs, including the measurement at the origin (*d,t*) = (0,0). If there is no other valid GOSAT measurement within the specified cut-offs, the expression simply reduces to $F_s(0,0)$ and the SIF value prior smoothing is retained.

Within the Forward model of the Full Physics retrieval, the SIF value at the surface is propagated to the top of the atmosphere using the Beer-Lambert law, not taking into account atmospheric scattering. The inclusion of SIF in the retrieval is necessary in order to avoid systematic biases in surface properties (albedo, surface pressure) as well as atmospheric scattering due to aerosols, Frankenberg et al., 2011. In terms of the retrieved XCO_2, there is a clear trend of higher fluorescence values reducing the XCO_2 values. Since SIF has a seasonal component, the inclusion of it in the retrievals also affects the seasonality of XCO_2.

3.3 Retrieval setup

3.3.1 Description of CO_2 retrievals (CO2_GOS_OCFP)

The retrievals for Carbon Dioxide use a state structure that represents the atmosphere, surface and instrument. The state vector of our retrieval consists of a 20-level profile for the CO_2 VMRs, 20-level profiles of the extinction coefficients for two aerosol types and for cirrus, scaling factors for H_2O VMR and temperature profiles, surface albedo and a spectral shift/stretch. A 2nd order polynomial is also retrieved for the albedo to allow for a spectrally dependent albedo. In addition, over land SIF is also retrieved. In total, our retrieval state vector presently consists of 97 elements for glint cases and 98 elements for land cases, as shown in Table 2. The parameters listed in that table are discussed in sections 2.1 and 2.2.

Table 2: State vector for CO_2 retrievals.

Description	Parameters	Number of elements
Aerosol extinction vertical profile	3 x 20 levels	60
CO_2 vertical profile (VMR)	20 levels	20
Albedo (mean and gradient)	2 x 3 bands	6
Dispersion	2 x 3 bands	6
Zero-level Offset (O_2-A band only)	1	1
SIF (land only)	1	1
Surface Pressure	Scalar	1
Temperature (additive)	Scalar	1
Water vapour (multiplicative)	Scalar	1
CH_4	Scalar	1
Total		97 (glint) – 98 (land)

3.3.2 Description of Full-Physics CH_4 retrievals (CH4_GOS_OCFP)

The Full-Physics Methane retrieval is similar to the CO_2 retrieval described in Section 3.3.1 with the target retrieval species changed from CO_2 to CH_4. Methane is only retrieved in Band 2 between 5986.0 and 6136.6 cm^{-1}, using the state vector of the full physics CO_2 retrieval as a priori. The parameters used in the retrieval are taken from the same sources defined in section 3.3.1. The state vector is described in Table 3.

Table 3: State vector for Full-Physics CH_4 retrievals.

Description	Parameters	Number of elements
CH_4 vertical profile (VMR)	20 levels	20
Albedo (mean and gradient)	2 x 3 bands	6
Dispersion	2	2
Total		28

3.3.3 Description of Proxy CH_4 retrievals (CH4_GOS_OCPR)

For this CH_4 retrieval, we adopt the CO_2 Proxy method of Frankenberg et al., 2006. CO_2 is known to vary in the atmosphere much less than CH_4 and as the CO_2 absorption band is spectrally close to that of CH_4 we can use the CO_2 as a proxy for the light path to minimize common spectral artefacts due to aerosol scattering and instrumental effects (Frankenberg et al., 2008; Butz et al., 2010). CH_4 and CO_2 retrievals are carried out sequentially with channels at 1.65 µm and 1.61 µm respectively. In order to obtain a volume mixing ratio (VMR) of CH_4, it is necessary to multiply the XCH_4/XCO_2 ratio by a model XCO_2. We obtain the CO_2 VMRs from the median of a CO_2 model ensemble that comprises of data from GEOS-Chem (University of Edinburgh), NOAA CarbonTracker and LMDZ/MACC-II (Chevallier et al. 2010), convolved with scene-dependent instrument averaging kernels obtained from the GOSAT CO_2 retrieval.

The state vector of our retrieval consists of a 20-level profile for CH_4 and CO_2 VMRs, scaling factors for H_2O VMR and temperature profiles, surface albedo and a spectral shift/stretch. A 3rd order polynomial is also retrieved for the albedo to allow for a spectrally dependent albedo. Other parameters in the state vector are taken from the same sources defined in section 3.3.1. The state vector for the proxy CH_4 retrieval is described in Table 4.

Table 4: State vector for Proxy CH_4 retrievals.

Description	Parameters	Number of elements
CH_4 vertical profile (VMR)	20 levels	20
Albedo	4	4
Dispersion	2	2
Temperature (additive)	Scalar	1
Water Vapour (multiplicative)	Scalar	1
CO_2 vertical profile (VMR)	Scalar	1
Total		29

3.4 Data screening and bias correction

3.4.1 Pre-processing screen

GOSAT L1B soundings are screened prior to retrieval to remove observations with a signal to noise (SNR) of < 20, or if a sounding contains any invalid L1B ancillary data. Observations taken at solar zenith angles of > 75° are removed to eliminate a further source of low SNR spectra taking very long light paths through the atmosphere in local summer time. Soundings over Antarctica, whose spectra are known to cause retrieval failure, owing to the region's high albedo caused by clouds and snow / ice, are also filtered with a latitude screen of < -60°. In addition to this, only soundings where the ratio of the retrieved XCO_2 from the strong and weak CO_2 bands are kept for further analysis.

3.4.2 Cloud filtering

A-priori surface pressure values are used in the UoL Full-Physics retrievals, with the optimal estimation method attempting to fit surface pressure in a range around this prescribed value. Thick clouds within the instrument's instantaneous field of view reflect incoming solar radiation from the cloud's top at a pressure (typically) much lower than the (well constrained) Earth surface pressure. This renders the a priori surface pressure value incorrect for radiative transfer model light path calculations, necessitating a filter for exposures affected by thick cloud in this manner.

A screen can be quickly derived through performing a fast O_2-A band retrieval where the surface pressure is retrieved over the range: 12950.0-13200.6 cm^{-1}. Differences greater than 30 hPa between retrieved O_2 A band and a priori surface pressure values are used to indicate the presence of thick clouds.

3.4.3 Post-processing screen

3.4.3.1 Full-Physics screen

CO_2 is post-screened based on algorithm convergence criteria, as well as on the normalised χ^2 value for each spectral band. UoL-OCFP retrievals are also subject to spectral interference from factors such as aerosol and clouds, therefore we apply a number of extra screens to reduce these effects, which are detailed in the PUGS document.

It should also be noted that the a posteriori error calculated from the CO_2 retrieval is underestimated. Through comparison with TCCON ground-based validation data we have determined that the calculated error should be increased by a factor, which is also detailed in the PUGS document [D2]. The CH_4 data retrieved by the UoL-FP algorithm is also subject to the same screening as the CO_2; that is, the CH_4 OCFP dataset contains retrievals from soundings that have also cleared the CO_2 screening.

3.4.3.2 Proxy screen

CH_4 retrieved using the Proxy retrieval is post-screened based on the algorithm convergence criteria as above (i.e. outcome =1, cost_div = F, conv_flag = T) as well as on the normalised χ^2 for both the CO_2 and CH_4. Soundings below -60° latitude are also removed.

3.4.3.3 Bias correction

Following filtering a small amount of residual bias remains in the dataset when compared against all co-located TCCON observations available. To correct for this, a bias correction is derived based on a number of parameters. The parameter coefficients *m* are given in the PUGS document [D2]. For XCO_2 corrected with a set of *n* parameter coefficients, the correction takes the form:

$$\text{correction} = c + m_0 x_0 + m_1 x_1 + \ldots + m_{n-1} x_{n-1} \quad (17)$$

$$\text{XCO}_{2\,\text{final}} = \text{XCO}_2 - \text{correction} \quad (18)$$

4. Output data

The University of Leicester provides several Level 2 products for XCO_2 and XCH_4 based on GOSAT spectral data:

- CO2_GOS_OCFP
- CH4_GOS_OCFP
- CH4_GOS_OCPR

These products store XCO_2 and XCH_4 data retrieved by the UoL "full-physics" (OCFP) and "proxy" (OCPR) retrieval algorithms, which are discussed in Section 3. Quality assessment is performed on the entire dataset, where problematic retrievals are flagged based on retrieval statistics. Additional flagging is also performed based on direct comparisons with collocated ground-based measurements from the Total Carbon Column Observing Network (TCCON), along with post-retrieval bias correction using retrieval parameters that strongly correlate with observed GOSAT-TCCON biases.

From the retrieval output, the original and bias corrected XCO_2/XCH_4 is published as daily netCDF files. Additional a priori state vector information and other useful data such as quality flags, averaging kernels, and geolocation information are also recorded in these files.

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].

References

Boesch et al., 2006: Boesch, H., Toon, G. C., Sen, B., Washenfelder, R. A., Wennberg, P. O., Buchwitz, M., de Beek, R., Burrows, J. P., Crisp, D., Christi, M., Connor, B. J., Natraj, V., and Yung, Y. L.: Space-based near-infrared CO_2 measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, J. Geophys. Res., 111, D23302, doi:10.1029/2006JD007080, 2006.

Boesch et al., 2011: Boesch, H., D. Baker, B. Connor, D. Crisp, and C. Miller, Global characterization of CO_2 column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 mission, Remote Sensing, 3 (2), 270-304, 2011.

Butz et al., 2010: Butz, A., O. P. Hasekamp, C. Frankenberg, J. Vidot, and I. Aben (2010), 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.

Chevallier et al., 2010: Chevallier, F., Feng, L., Boesch, H. Palmer, P., and Rayner, P., On the impact of transport model errors for the estimation of CO_2 surface fluxes from GOSAT observations, Geophys. Res. Let., 37, L21803, 2010.

Connor et al., 2008: Connor, B. J., H. Boesch, G. Toon, B. Sen, C. Miller, and D. Crisp, Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res., 113, D05305, doi:10.1029/2006JD008336, 2008.

Crisp et al., 2012: D. Crisp, B. M. Fisher, C. O'Dell, C. Frankenberg, R. Basilio, H. Bösch, L. R. Brown, R. Castano, B. Connor, N. M. Deutscher, A. Eldering, D. Griffith, M. Gunson, A. Kuze, L. Mandrake, J. McDuffie, J. Messerschmidt, C. E. Miller, I. Morino, V. Natraj, J. Notholt, D. M. O'Brien, F. Oyafuso, I. Polonsky, J. Robinson, R. Salawitch, V. Sherlock, M. Smyth, H. Suto, T. E. Taylor, D. R. Thompson, P. O. Wennberg, D. Wunch, and Y. L. Yung, Atmos. Meas. Tech., 5, 687-707, 2012.

Dee et al., 2011: Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., et al.: The ERA‐Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553-597, 2011

Frankenberg et al., 2006: Frankenberg, C., J. F. Meirink, P. Bergamaschi, A. P. H. Goede, M. Heimann, S. Körner, U. Platt, M. van Weele, and T. Wagner (2006), Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: Analysis of the years 2003 and 2004, J. Geophys. Res., 111, D07303, doi:10.1029/2005JD006235.

Frankenberg et al., 2008: Frankenberg, C., P. Bergamaschi, A. Butz, S. Houweling, J. F. Meirink, J. Notholt, A. K. Petersen, H. Schrijver, T. Warneke, and I. Aben (2008), Tropical methane emissions: A revised view from SCIAMACHY onboard ENVISAT, Geophys. Res. Lett., 35, L15811, doi:10.1029/2008GL034300.

Frankenberg et al., 2011: Frankenberg, C., Aben, I., Bergamaschi, P., et al., Global column-averaged methane mixing ratios from 2003 to 2009 as derived from SCIAMACHY: Trends and variability, *J. Geophys. Res*., doi:10.1029/2010JD014849, 2011.

Hamazaki, 2006: Hamazaki, T., GOSAT Mission Overview, 3rd IWGGMS Meeting, May 30 2006, Tsukuba, Japan, 2006.

Hersbach et al., 2020: Hersbach, H., Bell, W., Berrisford, P., Hirahara, S., Horányi, A. J., et al.: The ERA5 global reanalysis, Q. J. Roy. Meteorol. Soc., 146, 1999-2049, 2020

Kahn et al., 2001: Kahn, R.; Banerjee, P.; McDonald, D. Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean. J. Geophys. Res., 106, D16., 2001.

Kawabata and Ueno, 1988: Kawabata, K., and S. Ueno, The first three orders of scattering in vertically inhomogeneous scattering–absorbing media, Astrophys. Space Sci., 150(2), 327–344, 1988

Kuze et al., 2009: Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: 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., 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.

Mishchenko et al., 2002: Mishchenko, M.I., L.D. Travis, and A.A. Lacis, 2002: Scattering, Absorption, and Emission of Light by Small Particles. Cambridge University Press.

Morcrette et al., 2009: Morcrette, J.-J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J. W., Razinger, M., Schulz, M., Serrar, S., Simmons, A. J., Sofiev, M., Suttie, M., Tompkins, A. M., and Untch, A.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward modelling. J. Geophys. Res.-Atmos., 114, D06206, doi: 10.1029/2008JD011235, 2009.

Nakajima and Tanaka, 1988: Nakajima, T. and Tanaka, M.: Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation. J. Quant. Spectrosc. Radiat. Transfer, 40, 51–69, doi:10.1016/0022-4073(88)90031-3, 1988.

Natraj and Spurr, 2007: Natraj, V., and Spurr, R. J. D. (2007). A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering–absorbing media. J. Quant. Spectrosc. Radiat. Transfer, 107, 263-293, doi: 10.1016/j.jqsrt.2007.02.011, 2007

Natraj et al., 2008: Natraj, V., Boesch, H., Spurr, R. J. D., and Yung, Y. L.: Retrieval of XCO_2 from simulated Orbiting Carbon Observatory measurements using the fast linearized R-2OS radiative transfer model, J. Geophys. Res., 113, D11 212, doi:10.1029/2007JD009017, 2008.

O'Dell, 2010: O'Dell, C.W.: Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation, J. Geophys. Res., 115, D10 206, doi:10.1029/2009JD012803, 2010.

O'Dell et al., 2012: O'Dell, C. W., B. Connor, H. Boesch, D. O'Brien, C. Frankenberg, R. Castano, M. Christi, D. Crisp, A. Eldering, B. Fisher, M. Gunson, J. McDuffie, C. E. Miller, V. Natraj, F. Oyafuso, I. Polonsky, M. Smyth, T. Taylor, G. C. Toon, P. O. Wennberg, and D.Wunch, The ACOS CO_2 retrieval algorithm – Part 1: Description and validation against synthetic observations, Atmos. Meas. Tech., 5, 99–121, 2012.

O'Dowd et al., 1997: O'Dowd, C. D., J. A. Lowe, M. H. Smith, B. Davison, C. N. Hewitt, and R. M. Harrison (1997), Biogenic sulphur emissions and inferred non-sea-salt-sulphate cloud condensation nuclei in and around Antarctica, Journal of Geophysical Research, 102(D11), 12,839, doi:10.1029/96JD02749.

Parker et al., 2011: Parker, R., Boesch, H., Cogan, A., et al., Methane Observations from the Greenhouse gases Observing SATellite: Comparison to ground-based TCCON data and Model Calculations, Geophys. Res. Lett., 38, L15807, doi:10.1029/2011GL047871, 2011.

Parker et al., 2020: Parker, R., Webb, A., Boesch, H., et al., A decade of GOSAT Proxy satellite CH4 observations, Earth Syst. Sci. Data, 3383–3412, doi: 10.5194/essd-12-3383-2020, 2020.

Reddy et al., 2005: Reddy, M. S., O. Boucher, N. Bellouin, M. Schulz, Y. Balkanski, J.-L. Dufresne, and Pham, M.: Estimates of global multicomponent aerosol optical depth and direct radiative perturbation in the Laboratoire de Météorologie Dynamique general circulation model, J. Geophys. Res., 110, D10S16, doi: 10.1029/2004JD004757, 2005.

Rodgers, 2000: Rodgers C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing, 2000.

Somkuti et al., 2017: Somkuti, P., Boesch, H., Natraj, V., Kopparla, P., Application of a PCA‐Based Fast Radiative Transfer Model to XCO2 Retrievals in the Shortwave Infrared, Journal of Geophysical Research: Atmospheres, 122(19), doi: 10.1002/2017JD027013, 2017.

Spurr et al., 2001: Spurr, R. J. D., Kurosu, T. P., and Chance, K. V.: A linearized discrete ordinate radiative transfer model 940 for atmospheric remote-sensing retrieval, J. Quant. Spectrosc. Radiat. Transfer, 68, 689–735, doi:10.1016/S0022-4073(00)00055-8, 2001.

Spurr, 2002: Spurr, R. J. D., Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, J. Quant. Spectrosc. Rad. Tran., 75(2), 129–175, 2002.

Spurr and Natraj, 2011: Spurr, R., and Natraj, V. (2011). A linearized two-stream radiative transfer code for fast approximation of multiple-scatter fields. J. Quant. Spectrosc. Radiat. Transfer, 112(16), 2630-2637, doi: 10.1016/j.jqsrt.2011.06.014, 2011.

Yang et al., 2020: Yang, D., Boesch, H., Liu, Y., et al. (2021). Towards high precision XCO2 retrievals from TanSat observations: Retrieval improvement and validation against TCCON measurements. J. Geophys. Res.-Atmos., 125, e2020JD032794, doi: 10.1029/2020JD032794, 2020.

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: https://support.ecmwf.int