ESA Climate Change Initiative Plus (CCI+)

Algorithm Theoretical Basis Document version 3 (ATBDv3) for the University of Leicester Full-Physics XCO_2 TanSat data product (CO2_TAN_OCFP v1.2)

-

The University of Leicester Full-Physics Retrieval Algorithm for the retrieval of XCO_2 from TanSat

for the Essential Climate Variable (ECV)

Greenhouse Gases (GHG)

Version 3, revision 1

14th February 2022

Written by
Simon Preval, Dongxu Yang, and Hartmut Boesch
Department of Physics and Astronomy, National Centre for Earth Observation, and Space Park Leicester
University of Leicester
Leicester
United Kingdom

Document history:

Version	Revision	Date	Description/Comments
1D	0	03/08/2019	Initial draft version of ATBD based on existing GHG-CCI ATBD.
1	1	22/08/2019	First version of ATBD document for CO2_TAN_OCFP v1.0
2	0	23/07/2021	Second version of ATBD document for CO2_TAN_OCFP v1.0. References and authors updated.
3	0	08/11/2021	First version of ATBD document for CO2_TAN_OCFP v1.1.
3	1	14/02/2022	Updated ATBD document correcting errors in previous CO2_TAN_OCFP v1.1 product, and creation of new product CO2_TAN_OCFP v1.2.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Purpose

This algorithm theoretical basis document (ATBD) presents the University of Leicester algorithm used to retrieve column averaged dry air mole fraction of carbon dioxide (XCO_2) and other parameters included in the Level 2 Product generated from TanSat spectra. This document details various input data required for retrievals, physical theory and mathematical background underlying retrieval assumptions, and outlines retrieval implementation and limitations of the approach used. This document applies to the latest release of the CO2_TAN_OCFP v1.2 data product.

1.2 Scope

Covering the algorithm theoretical basis for the parameters of the Level 2 Products that are routinely retrieved at the University of Leicester, section 1 provides the purpose and scope of the document. Section 2 explains the mission and instrument. Section 3 describes the processing and algorithm and section 4 gives references for cited publications.

2 TanSat Overview

2.1 The TanSat Mission

Human activities including fossil fuel combustion and land use change have led to increases in atmospheric CO_2 concentrations from a pre-industrial level of 280 parts per million (ppm) to more than 395 ppm today. Networks of surface in-situ greenhouse gas sensors provide accurate measurements of the global CO_2 concentrations and temporal, seasonal and latitudinal variations. In-situ sensors provide globally sparse observations, preventing calculations of sub-continental carbon budgets, and indirectly limiting future climate change forecast accuracy. If acquired with high accuracy and precision, satellite observations have the potential to overcome such limitations by providing globally dense datasets of column CO_2.

The Chinese Global Carbon Dioxide Monitoring Scientific Experimental Satellite (TanSat) is the first Chinese CO_2 monitoring satellite, launched on 22 December 2016. TanSat provides global measurements of total column CO_2 from its shortwave infrared (SWIR) bands combined with a near infrared (NIR) band. The mission aims are to monitor the column density of CO_2 precisely and frequently worldwide, to study the absorption and emission levels on a regional scale over time, and to develop and establish advanced technologies that are essential for precise CO_2 observations [Liu et al., 2018].

2.2 Measurement Approach

TanSat was established by the National High Technology Research and Development Program of China and supported by the Ministry of Science and Technology of China, the Chinese Academy of Sciences, and the China Meteorological Administration. The main objective of TanSat is to monitor the dry air column-averaged mole fraction of CO_2 (XCO_2) and CO_2 flux at the regional to global scale and to provide the environmental administration of China with an assessment of forest carbon balances, regional emissions and absorptions. Research using TanSat may provide an enhanced understanding of the global distribution and temporal variations of CO_2. This could advance our knowledge of the carbon cycle on a global scale and its influence on climate, which is essential for the prediction of future climate change and its possible impacts. Additionally, it aims at leading to new developments in both Earth observation satellite technologies and the approach of CO_2 measurements.

In December 2016, TanSat was launched successfully, and on-orbit tests and calibration commenced. TanSat is an agile satellite platform deployed in a sun-synchronous orbit, which operates in three observation modes, namely, nadir, sun-glint, and target [Chen et al., 2013]. There are two scientific instruments onboard TanSat, namely, a hyperspectral grating spectrometer (Atmospheric Carbon dioxide Grating Spectrometer, ACGS) and a moderate-resolution imaging polarization spectroradiometer (Cloud and Aerosol Polarization Imager, CAPI).

2.3 TanSat Instrumentation

As the primary instrument onboard TanSat, ACGS is designed to measure NIR/SWIR backscattered sunlight in the molecular oxygen (O_2) A band (0.76 μm) and two CO_2 bands (1.61 and 2.06 μm). Total column CO_2 is mainly determined from measurements of its absorption lines in the weak band (1.61 μm). Sunlight is significantly scattered and absorbed by air molecules and suspended particles (e.g., clouds and aerosols), which would result in serious errors in CO_2 retrievals. Consequently, more information from cloud and aerosol measurements is required for the CO_2 retrieval to correct the light path. This is acquired by the O_2A band that provides information on altitude and total amount (optical depth) of aerosols and clouds due to almost constant and stable O_2 concentrations in the atmosphere. In comparison, the interference from water Vapour absorption is relatively weak. However, the CO_2 weak band is spectrally far away from the O_2A band, and aerosol and cloud optical properties depend on wavelength. Thus, it is also necessary to constrain this spectral variation which is one of the purposes of the strong CO_2 band. The strong CO_2 band also provides information on water vapour and temperature, which reduces impacts from uncertainties in these parameters.

The design of the optical layout of ACGS and the specifications of instrument optical parameters can be found in a previous study [Lin et al., 2017]. The footprint on the ground is 2 km × 2 km in the nadir mode with 9 footprints in each swath and a total width of the field of view (FOV) of ~18 km at nadir.

Although the O_2A band and the strong CO_2 band are used to constrain the aerosol loading and wavelength dependence of aerosol effects, the information on aerosol and cloud scattering that can be obtained from the spectra themselves is limited. Therefore, the auxiliary instrument CAPI was designed and is located onboard TanSat. It observes the reflected sunlight in five bands from UV to NIR. To achieve more information on aerosol size, which significantly affects the wavelength dependence of the optical properties of aerosols, CAPI includes two polarization channels that can measure the first three elements (*I, Q*, *U*) of the Stokes vector at 0.67 μm and 1.64 μm. In addition, a channel centered at 1.375 μm is employed to improve screening for cirrus clouds.

2.3.1 Observing Modes

The nadir mode is the most common one used over land surfaces, in which the instrument records data along the satellite ground track. The ocean surface has low surface reflectance and thus the nadir mode cannot yield high-precision measurements due to the low signal-to-noise ratio (SNR). To deal with this problem, the satellite tracks the sun glint spot, where sunlight is reflected specularly from the ocean, with the instrument boresight pointed within five degrees of the principal plane. TanSat also has a target mode that observes a stationary surface target as the satellite flies overhead. The main purpose of this mode is to support the validation of measurements with ground-based observations, but it can also record multi-angle observations over one surface target to investigate emissions from hot spots. These measurements can also be compared to ground-based observations to validate the quality of the satellite CO_2 measurements.

2.3.2 Data Product Delivery

TanSat L1B data is provide by International Reanalysis Cooperation on Carbon Satellites Data (IRCSD) funded by International Partnership Program of Chinese Academy of Sciences (GJHZ201903), and FENGYU Satellite Data Center of the National Satellite Meteorological Center. Each observation is then processed further to estimate the XCO_2 (Level 2), along with other parameters such as surface pressure, temperature, water Vapour, albedo, aerosol profiles, cirrus profile, CO_2 column averaging kernels, and various other quality control products.

3 Input and Auxiliary Data

3.1 Profile of Carbon Dioxide Concentration

The CO_2 a priori profile is obtained from LMDZ Monitoring Atmospheric Composition and Climate - Interim Implementation (MACC-II) CO_2 model fields. The CO_2 profile is interpolated with latitude, longitude, and time to the specific location desired. If the model is not available for latter years, the CO_2 concentration is incremented according to yearly global CO_2 increases reported by NOAA. Currently the MACC-II CO_2 model data is available up to December 2015.

3.2 Surface Pressure

The European Centre for Medium-Range Weather Forecasts (ECMWF) is an assimilation model that uses observations from surface buoy and satellite measurements [ECMWF Technical Notes, 2008]. ECMWF (ERA interim) provide atmospheric profiles of pressure, temperature, and specific humidity on a 0.75º x 0.75º global grid with 91 levels. Given the latitude, longitude, and altitude of a site of interest the surface pressure can be determined from these profiles. ECMWF provides potential data for the lowest level of the same grid, which can be used to find the geopotential height of each grid point level as:

$$\text{Height} = \text{Geopotential}/\text{g}, \ (3.1)$$

where gravitational acceleration, *g*, 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:

$$P = P_0 e^{\left(-\frac{Z}{Z_0}\right)}, \ (3.2)$$

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:

$$Z_0 = \frac{RT}{Mg}, \ (3.3)$$

where *R* is the ideal gas constant, *T* is the average temperature across the differential, *M* is the Molar mass of wet air. The Molar mass of wet air can be calculated by:

$$M = \rho_d(1 - SH) + \rho_w SH, \ (3.4)$$

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 can be obtained from a global digital elevation model with a horizontal grid spacing of 30 arc seconds named GTOPO30 provided by the U.S. Geological Survey. 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 computation, 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 3-1 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). Figure 3-1 illustrates that over a mountainous topography the surface pressure can still be resolved to within 1 hPa. Hence, the surface pressure a priori can be constrained to less than 0.1%, although we use a 1-sigma (1σ) error of 4 hPa on the a priori surface pressure estimate to allow for more difficult topographies.

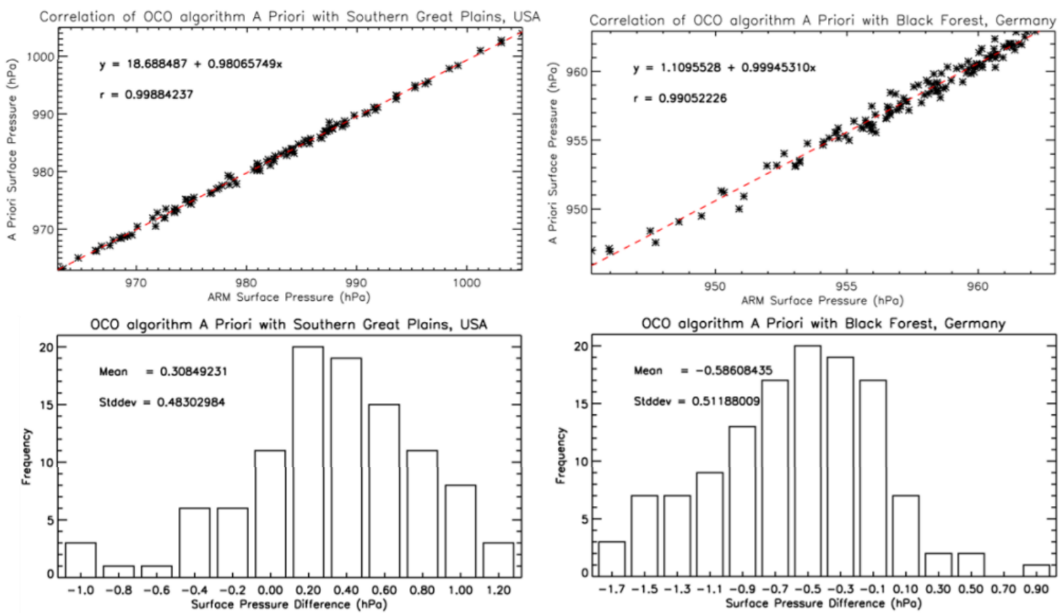

Figure 3-1. Validation of a priori surface pressure with in-situ ground-based measurements from ARM instruments in Southern Great Plains/USA and Black Forest/Germany.

3.3 Temperature and Water Vapour

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

$$\mathrm{H_2O}_{VMR} = 10^6\left(\frac{SH}{(R_d/R_w)-\left(SH\left((R_d/R_w)-1\right)\right)}\right), \quad (3.5)$$

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 $\mathrm{H_2O}_{VMR}$, are interpolated with latitude, longitude and time to the specific observation. A constant O_2 Volume Mixing Ratio (VMR) profile of 0.2095 is also used as a priori. In the retrieval, we assume that the temperature and H_2O profile shapes are sufficiently accurate, and we allow the retrieval to scale them only (additive in the case of temperature and multiplicative in the case of water vapour).

3.4 Aerosol and cloud properties

For aerosol type and corresponding optical properties, we use a dynamic (scene-dependent) setup informed by the Copernicus Atmosphere Monitoring Service (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, SU are treated as hydrophilic and DU as hydrophobic only. Furthermore, SS and DU are differentiated at three size bins. The optical properties of all particles (Table 3-1) were calculated by closely following the original MACC scheme as outlined in Morcrette et al. (2009) and Reddy (2005). Given the assumed size distributions, the optical properties of the MACC/CAMS aerosols are calculated using Mie theory and the code provided in Mishchenko et al. (2002).

For aerosol concentration and vertical distribution, we use CAMS model data as a climatology which is created from the CAMS data for the years 2014-2016. For each aerosol species we take 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 CAMS.

Table 3-1: Basic aerosol properties for CAMS 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 log-normal 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 distribution 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
Sulphate	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

The retrieval uses two aerosol retrieval types to describe aerosol scattering, which represent small and large particles. The aggregation of individual aerosol particles to aerosol type 1 (small) and aerosol type 2 (large) is given in Table 3-2. To mitigate errors from assuming a spherical shape for dust particles as in the CAMS, the dust particles of CAMS at three size bins (DU1, DU2, DU3) are replaced by the medium and coarse dust particles used in the MISR retrievals (labelled as P19 and P20). 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 CAMS. 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.

Table 3-2: Assignment of MACC aerosol types to small and large aerosol for UoL-FP retrieval. For DU, also the respective replacement with non-spherical particles P19 and P21 is given. Note that DU1 is assigned to both aerosol types at a certain percentage.

	BC_{phil}	BC_{phob}	OM_{phil}	OM_{phob}	SU	SS1	SS2	SS3	DU1	DU2	DU3
small aerosol	x	x	x	x	x	x			25% P19		
large aerosol							x	x	75% P19,P21[a]	P21	P21

[a] DU1 is distributed to 20% P19 and 80% P21

CAMS aerosol profiles are interpolated to the individual time and location of TanSat observations and downscaled to 20 levels used in the retrieval. Covariance matrices are calculated for each individual scenario depending on the assigned a priori total AOD. Expected greater variability of aerosol properties are 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.

In retrieval, the serious cloud contamination sounding from thick cloud, such as Stratus and Cumulus, has been screened out. The very thin cloud like Cirrus also impact the retrieval accuracy and cannot be ignored. A habit of roughened, aggregated solid columns (ASC) with a mean diameter 40 μm is adopted in retrieval. This habit is used very commonly in ice clouds that approached from Photochemistry of Ozone Loss in the Arctic Region in Summer (POLARIS) measurements for all seasons and locations. The total column vertical optical depth of cloud is set to a fixed value of 0.01. A Gaussian profile is applied to describe the a priori of vertical distribution of cloud extinction coefficient with a latitude dependent height centre and a fixed width (1 km), and then this profile is optimized as one of state vectors in retrieval.

3.5 Surface Properties

The albedo is calculated from the spectral continuum from the TanSat L1B data using the reflectivity:

$$\mathrm{Albedo} = \frac{\pi I_{TanSat}}{F_{Solar}\cos(\mathrm{SZA})}, \quad (3.6)$$

where SZA is the solar zenith angle, and the solar irradiance and observed TanSat radiance are given by F_{Solar} and I_{TanSat} 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 very loosely constrained (with standard deviation of 1) and the slope a priori error is set such that the band edges can vary by 50%.

3.6 Instrument Properties

3.6.1 Dispersion

The dispersion is given in the TanSat L1B data as a polynomial with the spectral channel orders *j*,

$$\lambda_j = \sum_{i=0}^{5} c_i \cdot j^i, \quad (3.7)$$

This dispersion is characterized during preflight and calibrated during in-flight tests, but small adjustments are still needed and these parameters are retrieved.

3.6.2 Noise

The noise estimates for each NIR/SWIR band can be calculated from the measured spectrum by using the noise parameters α_1 and α_2 given in the TanSat L1B data according to:

$$\mathrm{SNR} = \frac{I_{TanSat}}{\sqrt{\alpha_1^2 \cdot I_{TanSat} + \alpha_2^2}}, \ (3.8)$$

3.6.3 Instrument Line Shape Function (ILS)

To describe the response of the instrument to spectral light an Instrument Line Shape function (ILS) is used, which is given in the TanSat L1B data for each footprint and spectral channel. The ILS is provided for 200 wavelength points relative to the centre of each channel over a spectral range of ±0.2, ±0.925 and ±0.7 nm for the three different TanSat bands (Figure 3-2).

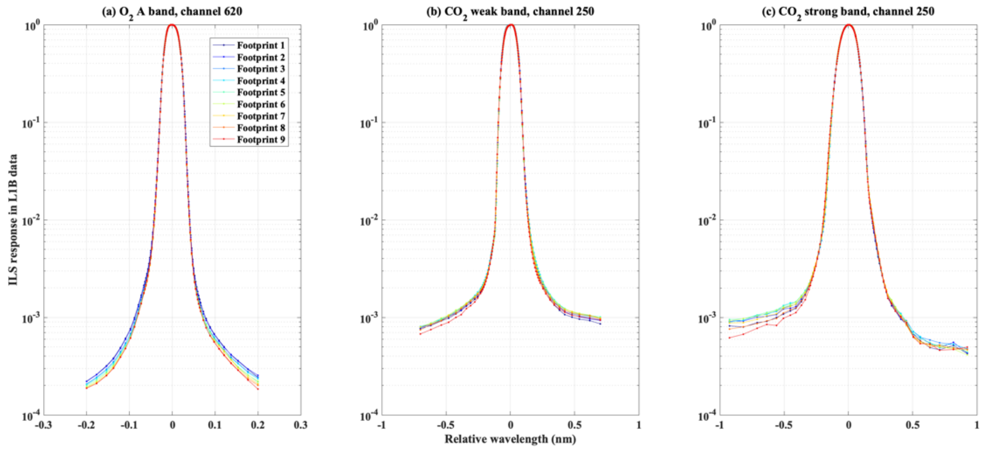

Figure 3-2. The instrument line shape (ILS) of each band of ACGS/TanSat. Only channels 620, 250 and 250 of the O_2A, CO_2 weak and strong bands are shown in this figure with footprints indicated by different colours. Note that the ILS response is provided with a peak unity in L1B data and needs to be normalized in practice.

3.7 Absorption cross-section

The retrieval algorithm uses pre-calculated absorption cross-sections for O_2, CO_2 and H_2O. These are tabulated for 70 pressure levels and 17 temperature values on a spectral grid of 0.01 cm^{-1}. These cross-section tables are based on the ABSCO version 5.0 from NASA JPL.

4 Algorithm Description

4.1 Algorithm 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_2A Band spectrum at 0.76 µm and the CO_2 bands at 1.61 and 2.06 µm as measured by the OCO (OCO-2) instrument. Whilst the algorithm was developed to retrieve XCO_2 from OCO (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 from observations from SCIAMACHY, GOSAT and OCO-2 [Boesch et al., 2006, 2013; Parker et al., 2011]. The retrieval has been modified for CO_2 retrievals from TanSat which includes the use of L1B data from TanSat and the TanSat instrument specific information. In addition, new features have been added described in section 4.2.3.6.

The UoL-FP 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 $\boldsymbol{x}$, from measured, calibrated spectral radiances.

The forward model describes the physics of the measurement process and relates measured radiances to the state vector $\boldsymbol{x}$. It consists of a radiative transfer (RT) model (RTM) 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. Alternatively, the code can use a method based on principal component analysis (PCA) [Somkuti et al., 2018].

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

$$\hat{\mathbf{S}} = \left(\mathbf{K}^{\mathbf{T}}\mathbf{S}_{\boldsymbol{\varepsilon}}{}^{-\mathbf{1}}\mathbf{K} + \mathbf{S}_{\mathbf{a}}{}^{-\mathbf{1}}\right)^{-\mathbf{1}}, \quad (4.1)$$

and the averaging kernel matrix $\boldsymbol{A}$,

$$\mathbf{A} = {}^{\partial\hat{\mathbf{x}}}/_{\partial\mathbf{x}} = \hat{\mathbf{S}}\mathbf{K}^{\mathbf{T}}\mathbf{S}_{\boldsymbol{\varepsilon}}{}^{-\mathbf{1}}\mathbf{K}, \quad (4.2)$$

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

$$\mathrm{XCO_2} = \mathbf{h}^{\mathbf{T}}\hat{\mathbf{x}}. \quad (4.3)$$

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

$$(\mathbf{a}_{\mathrm{CO2}})_{\mathrm{j}} = \frac{\partial \mathrm{XCO_2}}{\partial \mathbf{u}_{\mathrm{j}}}\frac{1}{\mathbf{h}_{\mathrm{j}}} = \left(\mathbf{h}^{\mathbf{T}}\mathbf{A}\right)_{\mathrm{j}}\frac{1}{\mathbf{h}_{\mathrm{j}}}, \quad (4.4)$$

and the variance of XCO_2 by

$$\sigma_{XCO2} = \mathbf{h}^{\mathrm{T}}\hat{\mathbf{S}}\mathbf{h}. \quad (4.5)$$

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 $\boldsymbol{a}_{CO2}$.

4.2 Forward Model

4.2.1 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 (https://mark4sun.jpl.nasa.gov/toon/solar/solar_spectrum.html). 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.

4.2.2 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 TOA Stokes parameters *I*, *Q*, and *U* on a fine spectral resolution of 0.01 cm^{-1} (0.00058, 0.0026 and 0.0042 nm in the O_2A, CO_2 weak and strong band, we use wavenumber unit with cm^{-1} here in after). 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 reflectance, to obtain 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. We therefore adopt an approximate approach called “Low Streams Interpolation” (LSI), described fully in [O’Dell, 2010] and references therein.

Rather than performing full-accuracy calculations with a large number of angular streams at all monochromatic wavelengths, calculations are only performed at a few tens of wavelengths. Very fast, low accuracy calculations are performed at all the monochromatic wavelengths; these are combined with the small number of high accuracy calculations to provide an 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 LIDORT scalar multiple-scattering model [Spurr et al., 2001], and a second-order-of-scattering polarization model called Two Orders of Scattering (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 less than a tenth of a percent [O’Dell, 2010].

4.2.2.1 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 RTM that generates radiances and Jacobians (derivatives of the radiance with respect to atmospheric and surface properties) simultaneously. The Jacobians (or known as weighting function) 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 considered to be locally plane parallel.

4.2.2.2 2OS (Two Orders of Scattering) 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 Kawabata and Ueno (1988) scalar model to compute the first two orders of scattering 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.

4.2.2.3 Two Stream Model

Making use of LSI allows a significant speed-up of RT calculations when used in combination with a dedicated RT model for the low-stream line-by-line calculations. We utilize the two-stream model TWOSTR [Spurr et al., 2011] which decreases the computation time even further compared to LIDORT using N=2 streams.

4.2.3 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.

4.2.3.1 Pixel-Wavelength Mapping

The dispersion gives the pixel-wavelength mapping and consists of 6 parameters for each band as a polynomial (see the eq. 3.7), which has been described in detail in section 3.6.1.

4.2.3.2 Convolution with Instrument Line Shape Function (ILS)

The monochromatic radiance and Jacobians that are simulated by the forward model need to be convolved with the ILS to simulate the measurements. The detailed TanSat ILS model and parameters are provided in the L1B data for each channel of each band and footprint, as described in section 3.6.3.

4.2.3.3 Zero-level Offset

An additive zero-level offset parameter with wavelength dependence as described by a linear slope is retrieved to correct for an intensity offset socket in the O_2A band and the CO_2 weak band.

4.2.3.4 Polarization

Since TanSat only measures one direction of polarization through a linear polarizer, the simulated Stokes parameters {*I, Q, U*} need to be combined to the incident signal I_{TanSat} by,

$$I_{TanSat} = p_1 \cdot I + p_2 \cdot Q + p_3 \cdot U, \quad (4.6)$$

in which p_1 (=0.5), p_2 and p_3 are the Stokes parameter coefficient for *I*, *Q* and *U* respectively, and p_2 and p_3 defined by the angle (polarization angle, σ) between the local meridian plane and scattering plane,

$$p_2 = 0.5 \cdot \cos 2\sigma, \ (4.7)$$

$$p_3 = 0.5 \cdot \sin 2\sigma. \ (4.8)$$

The detailed computation for the polarization angle can be found in the literature [Liou, 2002; Mishchenko et al., 2002] and publication [Yang et al., 2020]. TanSat L1B data also provide the polarization angle for each individual measurement.

4.2.3.5 Solar Induced Fluorescence (SIF) and radiance zero offset

The Solar Induced Chlorophyll Fluorescence (SIF) signal over land behaves similar to an additive offset of the radiance signal. For the TanSat retrieval, we retrieve a zero level offset in the O_2A band and its wavelength dependence which will compensate the SIF impact on the O_2A band surface pressure retrieval. Therefore, the TanSat XCO_2 retrieval does not include the physical SIF retrieval but uses only the O_2A band zero offset which has a strong correlation with SIF.

4.2.3.6 Additional correction of the continuum

A stable spectral pattern has been found in the O_2A band range in both the analysis of solar calibration measurement and the surface pressure retrievals. This pattern has a significant impact on the retrieval accuracy, and hence needs to be eliminated in the retrieval. We found that a radiometric calibration and stray light-like model can be used for correcting this pattern [Yang et al., 2020], and accordingly an eight orders Fourier series is applied as a scale factor on the simulated continuum as a correction to compensate the measurement pattern during retrieval. The corrected simulated radiance L can be represented using the original simulation L_{TanSat},

$$L = \left(\gamma + \sum_{i=1}^{8}(\alpha_i \cdot \cos(i \cdot \omega \cdot \Delta\lambda) + \beta_i \cdot \sin(i \cdot \omega \cdot \Delta\lambda))\right) \cdot L_{TanSat}, \ (4.9)$$

Where α_i and β_i are coefficients of i^{th} order of Fourier series of cosine and sine components with γ the zero-order coefficient and ω the scale coefficient of frequency. $\Delta\lambda$ is the wavelength difference of channels to the reference of 750 nm. In practice, γ is fixed as 1 to avoid the conflict with surface reflectance (albedo), and all other 17 coefficients in this equation (4.9) are retrieved. Estimates of a priori values are inferred from solar calibration measurement fitting and analysis. The same correction is also applied to the CO_2 weak band.

4.3 Retrieval Setup

4.3.1 Description of CO_2 Retrievals

The retrievals for CO_2 use a state vector that includes parameters for the atmosphere, surface and instrument. The retrieval only uses the O_2A band and CO_2 weak band (two-bands retrieval). The state vector of our retrieval consists of a 20-level profile (19 layers with 20 surfaces) for the CO_2 VMRs, 20-level profiles of the extinction coefficients for two aerosol types and cirrus, scaling factors for H_2O VMR and temperature profiles offset, surface albedo and a spectral shift/stretch. A wavelength slope is also retrieved for the surface albedo of each band to allow for its spectral dependence. In addition, a zero offset with its wavelength dependence slope and continuum correction coefficients is retrieved. In total, our retrieval state vector presently consists of 138 elements, as shown in Table 4-1.

Table 4-1. State Vector for CO_2 retrievals.

Description	Parameters	Number of Elements
Aerosols	2 x 20 levels	40
Cirrus	20 levels	20
CO_2	20 levels	20
Albedo	2 x 2 bands	4
Dispersion	6 x 2 bands	12
Zero-Level Offset	2 x 2 bands	4
Continuum coefficients	17 x 2 bands	34
Surface Pressure	Scalar	1
Temperature profile scaling	Scalar	1
Water Vapour profile scaling	Scalar	1
CH_4 profile scaling	Scalar	1
Total		**138**

4.4 Screening

4.4.1 Pre-Processing Screen

TanSat L1B nadir measurement mode soundings over land (land fraction > 99%) are screened prior to the retrieval to remove observations with a bad measurement flag given in the L1B data. Observations taken at solar zenith angles of > 70° are removed to eliminate low SNR spectra taken for very long light paths through the atmosphere.

4.4.2 Cloud Screen

Thick clouds within the instrument's instantaneous field of view (FOV) reflect incoming solar radiation from a height represented by the cloud top pressure which is (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.

A screen has been derived through performing a fast O_2A band retrieval where the apparent surface pressure is retrieved without considering aerosol and cloud scattering. Differences of greater than 20 hPa between retrieved O_2A band and a priori surface pressure values are used to indicate the presence of thick clouds. Soundings where the retrieved reduced chi square > 30 are also removed.

4.4.3 Post-Processing Screen

4.4.3.1 Carbon Dioxide Screen

Post-screening of the XCO_2 retrieval is based on the results from a Genetic Algorithm applied to TanSat retrievals in target mode over sites of the Total Carbon Column Observing Network (TCCON). The XCO_2 error is calculated by the difference between the TanSat retrievals and the mean values from TCCON within 500 km around but only in ±3º latitude/longitude box spatially and within ±1 hour of the overpass.

The selected filtering criteria are: the retrieved changes of layer CO_2 gradient between 700 hPa and the surface, the retrieved changes of the surface pressure difference to the a priori, the continuum correction coefficients of the cos(x) term for the O_2A band, the zero offset wavelength dependence slope of the CO_2 weak band, and the surface albedo of CO_2 weak band. For each criterium an upper and lower boundary threshold is used to filter the 'bad' retrieval [Yang et al., 2020] with ~18.3% retrievals passing the post-screening compared to the original measurement.

4.4.3.2 CO_2 bias correction

A multiple linear regression model is used for bias correction with the parameter P_i being the same as used in post-screening (see the section 4.4.3.1):

$$\Delta_{XCO2} = \sum_{i=1}^{5} A_i \cdot P_i + B. \quad (4.10)$$

Those parameters are selected by the rank of the XCO_2 error correlations (correlation coefficient). The coefficient A_i and offset *B* are optimized in the regression and used for bias correction of each individual retrieval.

5 References

[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.

[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.

[Chen et al., 2012] Chen., W., Zhang, Y., Yin, Z., Zheng, Y., Yan, C., Yang, Z., and Liu, Y.: The TanSat Mission: Global CO2 Observation and Monitoring, Proceedings of the 63rd IAC (International Astronautical Congress), Naples, Italy, Oct, IAC-12-B4.4.12, 1-5, 2012.

[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.

[ECMWF Technical Notes], MARS User Guide, User Support, Operations Department, 2008.

[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.

[Lin et al., 2017] Lin C, Li C, Wang L, et al. Preflight spectral calibration of hyperspectral carbon dioxide spectrometer of TanSat (in Chinese). Opt Precision Eng, 25: 2064-2075, 2017.

[**Liou 2002**], Liou, K.N., An Introduction to Atmospheric Radiation, Elsevier, 2002.

[Liu et al., 2018] Liu, Y., Wang, J., Yao, L., et al., The TanSat mission: preliminary global observations, *Science Bulletin,* 63, 1200-1207, 2018.

[**Mishchenko et al., 2002**] Mishchenko, M. I., L. D. Travis, and A. A. Lacis, Scattering, absorption, and emission of light by small particles, Cambridge University Press, Cambridge; New York, 2002.

[Morcrette et al. 2009] Morcrette, J.-J., O. Boucher, L. Jones, D. Salmond, P. Bechtold, A. Beljaars, A. Benedetti, A. Bonet, J. W. Kaiser, M. Razinger, M. Schulz, S. Serrar, A. J. Simmons, M. Sofiev, M. Suttie, A. M. Tompkins, and A. Untch, Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward modeling, Journal of Geophysical Research, 114(D6), doi:10.1029/2008JD011235, 2009.

[Natraj and Spurr, 2007] Natraj, V. and Spurr, R. J. D.: 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 XCO2 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.

[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.

[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'Dowd et al., 1997] O'Dowd, C. D., J. A. Lowe, M. H. Smith, B. Davison, C. N. Hewitt, and R. M. Harrison, 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, 1997.

[Parker et al., 2011] Parker,R., Boesch,H, Cogan,A., Fraser, A., Feng, L, Palmer, P., Messerschmidt, J., Deutscher, N., Griffth, D, Notholt, J., Wennberg, P. and Wunch, D., Methane Observations from the Greenhouse gases Observing SATellite: Comparison to ground-based TCCON data and Model Calculations, Geophys. Res. Lett., doi:10.1029/2011GL047871, 2011.

[Reddy 2005] Reddy, M. S., Estimates of global multicomponent aerosol optical depth and direct radiative perturbation in the Laboratoire de Mtorologie Dynamique general circulation model, Journal of Geophysical Research, 110(D10), doi:10.1029/2004JD004757, 2005.

[Rodgers 2000] Rodgers, C. D.: "Inverse Methods for Atmospheric Sounding: Theory and Practice", World Scientific, Singapore, 2000.

[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.

[Somkuti et al., 2017] Somkuti, P., Boesch, H., Natraj, V. Kopparla, P., Application of a PCA-Based Fast Radiative Transfer Model to XCO_2 Retrievals in the Shortwave Infrared, JGR, 2017.

[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 et al., 2011] Spurr, R., & Natraj, V., A linearized two-stream radiative transfer code for fast approximation of multiple-scatter fields. Journal of Quantitative Spectroscopy and Radiative Transfer, 112(16), 2630-2637, 2011.

[Yang et al., 2020] Yang, D., Boesch, H., Liu, Y., Somkuti, P., et al., (2020), Toward High Precision XCO2 Retrievals From TanSat Observations: Retrieval Improvement and Validation Against TCCON Measurements, Journal of Geophysical Research: Atmospheres, 125, e2020JD032794.

6 APPENDICES

6.1 Appendix A Acronyms

ACGS	Atmospheric Carbon dioxide Grating Spectrometer, ACGS
ARM	Atmospheric Radiation Measurement program
AOD	Aerosol Optical Depth
ATMOS	Atmospheric Trace Molecule Spectroscopy
CAMS	Copernicus Atmosphere Monitoring Service
CAPI	Cloud and Aerosol Polarization Imager
CO_2	Carbon Dioxide
ECMWF	European Centre for Medium Range Weather Forecasting
FTS	Fourier Transform Spectrometer
GOSAT	Greenhouse gases Observing SATellite
GTOPO30	Global 30 Arc-Second Elevation
H_2O	Water
IRCSD	International Reanalysis Cooperation on Carbon Satellites Data
iFOV	Instantaneous Field of View
ILS	Instrument Line Shape
km	Kilometer
LMDZ	Laboratoire de Météorologie Dynamique
LSI	Low Stream Interpolator
MACC	Monitoring Atmospheric Composition and Climate
NIR	Near infrared
O_2	Oxygen
OCO	Orbiting Carbon Observatory
PCA	Principal component analysis
ppm	Parts per million
RT	Radiative transfer
RTM	Radiative transfer model
SNR	Signal to noise ratio
SIF	Solar Induced Chlorophyll Fluorescence
SWIR	Shortwave infrared
TanSat	Chinese Global Carbon Dioxide Monitoring Satellite
TOA	Top of atmosphere

VMR	Volume Mixing Ratio
XCO_2	Dry air column-averaged mole fraction of CO_2