CloudSat Project

A NASA Earth System Science Pathfinder Mission

CloudSat OCO2CLD-LIDAR-AUX Auxiliary Data Process Description and Interface Control Document

Product Version: P_R05

Document Revision: 0

Date: 20 November 2018

Questions concerning the document and proposed changes shall be addressed to:

Mark Richardson
Jet Propulsion Laboratory, California Institute of Technology
markr@jpl.caltech.edu
(818) 354-4270

Or

Heather Cronk
Cooperative Institute for Research in the Atmosphere, Colorado State University
Heather.cronk@colostate.edu
(970) 491-8575

Or

Philip Partain
Cooperative Institute for Research in the Atmosphere, Colorado State University
Philip.partain@colostate.edu
(970) 491-7835

Document Revision History

Date	Revision	Description	Section(s) Affected
November 2018	0	Initial Release	All

Table of Contents

1. Introduction

OCO2CLD-LIDAR-AUX combines measurements from NASA's Orbiting Carbon Observatory-2 (OCO-2) and the NASA/CNES Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite. Output is stored both on the OCO-2 native footprints and collocated with CloudSat footprints following the standard CloudSat product granule structure. It reports cloud optical depth, cloud-top pressure and cloud pressure thickness, which is geometric thickness expressed in hPa. Only liquid clouds over the ocean are retrieved. Measurements are provided over the CALIPSO and OCO-2 formation flying period from 2014-09-06 to 2017-07-14, during which time the CALIPSO lidar footprint typically fell within the OCO-2 nadir view swath with a time difference of approximately 8 minutes. Additionally, OCO-2 varies its viewing mode and only nadir view orbits are processed.

These sensors complement each other as the CALIPSO lidar precisely measures cloud top (resolution ~30 m or ~3 hPa), while many boundary layer clouds are optically thick enough to attenuate the lidar so cloud optical depth or geometric thickness cannot be retrieved. OCO-2 effectively measures photon path lengths which must then be partitioned into above-cloud and within-cloud components. CALIPSO's tight constraint on cloud top is intended to improve cloud pressure thickness retrievals by reducing uncertainty in this partitioning.

This document describes the theoretical basis for the retrieval, describes the implementation, provides validation statistics and describes the product file contents.

2. Algorithm Theoretical Basis

i. Summary

When OCO-2 is in nadir view and collocated CALIPSO data are available, a marine liquid cloudy scene is identified when:

1) the scene is over ocean,
2) the OCO-2 oxygen A-band and weak CO_2 band continua exceeds a radiance threshold, indicating a cloud within the OCO-2 footprint,
3) CALIPSO sees a single cloud layer whose cloud-top pressure, P_{top} > 680 hPa,
4) OCO-2 A-band and weak CO_2 band continua are consistent with a liquid cloud.

The CALIPSO criteria exclude some multi-layer cloud cases, ensuring a P_{top} prior that is consistent with the retrieval assumption of a single layer cloud. Non-physical results may be obtained by fitting this when, in reality, there is a multi-layer cloud system. Furthermore, OCO-2 measures reflected sunlight so only the sunlit part of any orbit is processed.

An iterative optimal estimation retrieval [*Rodgers*, 2000] using a multiple scattering radiative transfer code is performed using a micro-window of 75 A-band channels that were selected based on information content criteria [*Richardson and Stephens*, 2018]. The atmospheric profiles are provided by the European Centre for Medium Range Weather Forecasts (ECMWF) and uncertainties in measured and simulated radiances are included. Posterior values of cloud τ, P_{top} and ΔP_c are reported, along with related information such as the OCO-2 based cloud flag and phase determination.

ii. Theoretical A-band response to cloud properties

The oxygen A-band is an expression of a set of rovibrational transitions in an O_2 molecule with a typical energy change equivalent to photons of wavelength near 762 nm. Photons carry angular momentum *J* = 1 and since vibrational changes for O_2 have $\Delta J = 0$, concurrent rotational energy transitions occur to conserve angular momentum. Cases where molecular $\Delta J < 0$ are referred to as the P-branch and $\Delta J > 0$ transitions are the R-branch. Figure 1 shows an example A-band spectrum, and the doublets in each absorption peak come from the electron's internal spin angular momentum being aligned with or opposing the rotational transition.

The ocean's O_2 A-band albedo is typically near 0.02 [*Koelemeijer*, 2002], which can change somewhat with the choppiness of the surface and the illumination and viewing geometry. Since water droplets are poor A-band absorbers, more liquid cloud optical depth, τ, increases scattering and tends to brighten observed spectra. This brightening occurs most strongly in "continuum" channels in which rovibrational transitions are forbidden by quantum mechanics so there is negligible absorption by O_2.

Observed A-band spectra are strongly affected by the distribution of photon path lengths. For an absorption coefficient $k(\lambda)$ through a uniform substance, the radiance at a point $I(\lambda, z)$ is a function of the initial radiance I_0

$$I(\lambda, z) = I_0(\lambda) \exp(-k(\lambda)\Delta z) \quad (1)$$

$$\ln I(\lambda, z) = \ln I_0(\lambda) - k(\lambda)\Delta \mathrm{z} \quad (2)$$

i.e. the decrease in $\ln\ I(\lambda, z)$ from the logarithm of the initial radiance is the product of absorption coefficient and photon path length Δz. If, for example, a cloud is lower in the atmosphere, photons travel further to the cloud and back to the sensor, increasing Δz and reducing observed radiance, depending on the absorption coefficient.

This is the basic principle of A-band cloud retrievals over dark ocean surfaces: the continuum radiance I_c is near I_0 and is largely a function of cloud optical depth. Meanwhile absorption changes I/I_0 and therefore I/I_c. Figure 2 shows I and I/I_c Jacobians for the Figure 1 case, where the channels are organised by O_2 optical depth. Cloud τ changes dominate absolute radiance but for I/I_c the response to a 1 hPa change in P_{top} or ΔP_c are of the same order of magnitude as a change of 1 in cloud τ. The decrease in I/I_c with increased P_{top} or ΔP_c represents increased path length due to above-cloud and within-cloud paths respectively. The increased I/I_c for increased τ in this simulation is due to reduced mean free path of photons within the cloud due to the greater droplet number density, since in the cloud's r_{eff} and thickness are held constant for the $\partial(I/I_c)/\partial\tau$ calculation.

The ΔP_c Jacobian peaks in weaker-absorbing channels because photons that enter the cloud have travelled further than those that scatter from cloud tops. The stronger absorbing channels are almost completely attenuated over longer paths, meaning no further response to changing ΔP_c is possible.

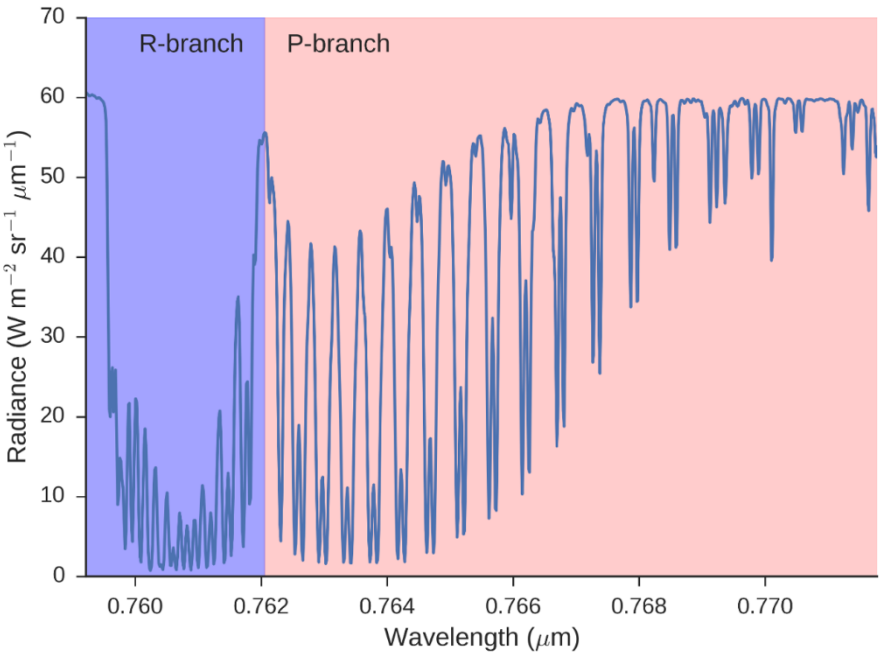


Figure 1 – Simulated OCO-2 A-band spectrum for a τ = 10 and P_{top} = 850 hPa cloud, with instrument in nadir view and a solar zenith angle of 45°. The P- and R-branches are shaded and labelled.

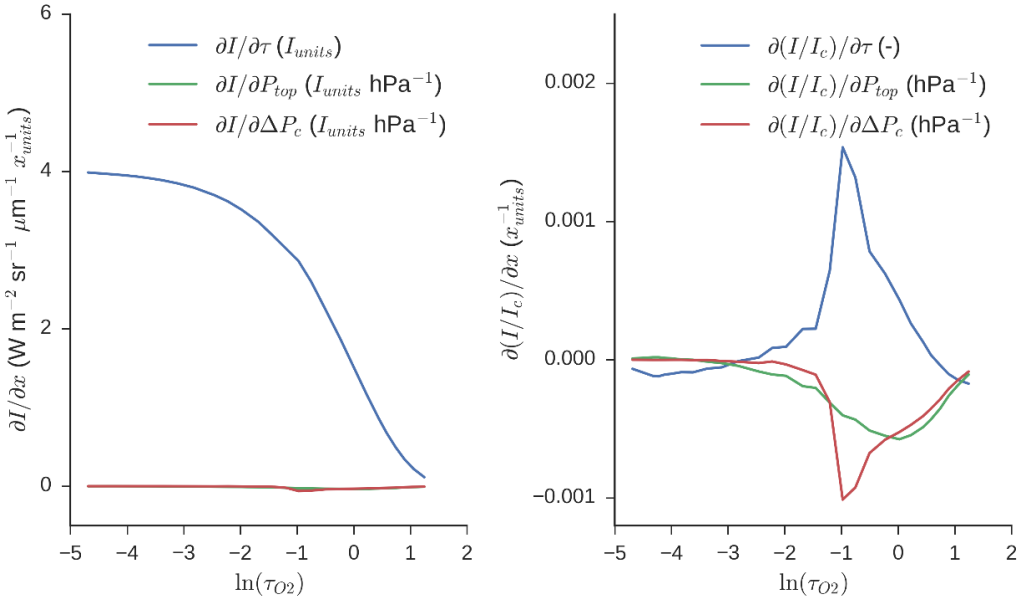


Figure 2 – Jacobians for optical depth (τ), cloud top pressure (P_{top}) and cloud pressure thickness (ΔP_c) for the Figure 1 case. Left panel is for simulated radiance, and right for radiance ratio to continuum.

iii. OCO-2 orbit and instrumentation

The OCO-2 orbit, measurement techniques and instruments are detailed in the OCO-2 Level 2 Full Physics (L2FP) Algorithm Theoretical Basis Document (ATBD, [*Boesch et al.*, 2017]) which explains its primary atmospheric column CO_2 concentration product (XCO2). Here a shorter summary is provided to aid understanding of the cloud property retrieval.

OCO-2 has two main science modes for its XCO2 product: nadir and glint. Each orbit is assigned to one of these modes, and assignment has changed through the mission with periods of alternating nadir and glint, and periods in which some orbits were set to glint only. The satellite operates in a "pushbroom" fashion, measuring 8 footprints along the swath, and the swath is not oriented perpendicular to the orbit track as the satellite rotates somewhat towards the Sun to optimise solar panel orientation. In nadir, each footprint is approximately 1.4×2.2 km^2, resulting in a total swath width near 10 km.

In the A-band, the retrieval uses a standard ocean albedo of 0.02 for the prior, with a Cox-Munk reflectance scheme[*Cox and Munk*, 1954]. In nadir view the ocean is typically so dark that the signal-to-noise ratio (SNR) is too low for reliable XCO2 retrievals, justifying the glint orbits in which the satellite looks off-nadir at the ocean glint spot. A downside of glint data is that the atmospheric path length increases so there is more absorption and a higher probability of encountering cloud or aerosol.

While the XCO2 retrieval allows for some aerosol and high-altitude cirrus, even thin layers of $\tau < 0.3$ can bias the retrievals. XCO2 retrieval development has focussed on clear sky cases and cloudy or dusty scenes are excluded when found [*Taylor et al.*, 2016]. Therefore OCO2CLD-LIDAR-AUX exploits otherwise unused OCO-2 data and adds value to the OCO-2 mission.

OCO-2 carries three co-bore sighted Fourier transform spectrometers with extremely high spectral resolution. These spectrometers cover the oxygen A-band (λ near 0.76 μm), the weak CO_2 band (λ near 1.58 μm) and the strong CO_2 band (λ near 2.06 μm). The cloud retrieval primarily uses the A-band, although the weak-CO2 band is also used to flag cloudy scenes and identify cloud phase. Table 1 lists the characteristics of the three spectrometers and Figure 3 displays a simulated cloudy scene spectrum for OCO-2 along with an approximation of the same scene measured with GOME-2-like instrument characteristics as assumed by [*Schuessler et al.*, 2014] for comparison. The 75 channel micro-window used in OCO2CLD-LIDAR-AUX is highlighted in red, and the d_s numbers in the legend are the degrees of freedom for signal when measuring τ, P_{top} and ΔP_c. This represents the number of effective pieces of information that can be obtained and a value near 3 indicates that all of those properties are well measured. For this scene the GOME-2 case only measures two properties and cannot obtain ΔP_c. Furthermore, the sub-selected 75 channels contain most of the full spectrum information.

Each of the OCO-2 spectrometers contain a nominal 1,016 channels whose output is averaged on board from multiple elements across a Focal Plane Array (FPA). The original OCO mission failed at launch and OCO-2 carries flight spares that had been placed in storage. Prior to deployment some FPA elements failed meaning that some channels are unreliable. The location of the failures depends on the across-track position, and the A-band's nominal 1,016 channels are reduced to 853 when only selecting those that are undamaged for all cross-track positions. Fortunately, the majority of the lost channels are in the continuum where there is redundancy so little cloud information is lost.

Table 1 – Instrument characteristics from Table 1 and Figure 5(a) of [*Crisp et al.*, 2016].

	O2 A-band	Weak CO2	Strong CO2
Wavelength range (μm)	0.7576—0.7726	1.5906—1.6218	2.0431—2.0834
Resolving power	17 500—18 500	19 100—20 500	19 700-19 900
Spectral sampling	2.6—3.5	2.3—3.2	2.2—3.2
Dynamic range	7.5×10^{16}—7×10^{20}	2.15×10^{16}—2.45×10^{20}	2.15×10^{16}—1.25×10^{20}
Typical SNR	200—1200	100—750	50—850

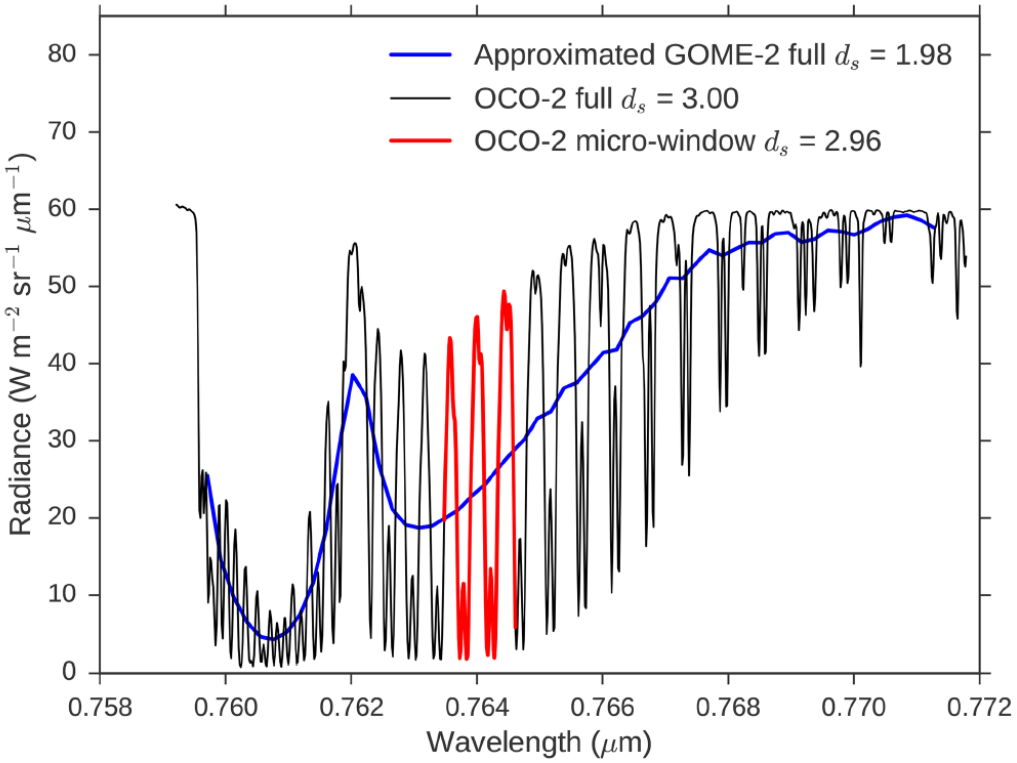


Figure 3 – Simulated A-band radiance spectra for a cloudy scene using GOME-2-like instrument characteristics (blue) and using OCO-2 instrument characteristics (black and red). The red line contains the 75 channels used in OCO2CLD-LIDAR-AUX based on an information content analysis. The degrees of freedom for signal, d_s, represent the number of properties that can be estimated from the measurements and are given in the legend for each spectrum. Source: Richardson and Stephens [2018] *, licensed under CC-BY 4.0.*

iv. Forward modelling and optimal estimation

The retrieval follows an iterative Bayesian optimal estimation procedure [*Rodgers*, 2000]. This combines prior information about the atmosphere and cloud state with observations to retrieve the posterior state. The OCO2CLD-LIDAR-AUX state vector is $\mathbf{x} = (\ln\tau, \ln P_{top}, \ln\Delta P_c)$ with logarithms to prevent non-physical negative parameter values. The output is provided as $(\tau, P_{top}, \Delta P_c)$.

It is assumed that the measured radiance spectrum vector, $\mathbf{y}$, can be reconstructed from the state vector using a forward model $F(\boldsymbol{x})$ with some additional error vector $\boldsymbol{\epsilon}$.

$$\boldsymbol{y} = F(\boldsymbol{x}) + \boldsymbol{\epsilon} \tag{3}$$

If it is assumed that uncertainty in $\mathbf{K}$, $\boldsymbol{y}$ and $\boldsymbol{x}$ can be described using Gaussian distributions and that $F(\boldsymbol{x})$ is near linear, then Bayes' theorem can be applied and the most likely posterior mean state $\hat{\mathbf{x}}$ and its covariance matrix $\hat{\mathbf{S}}$ are:

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_a + \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_\mathrm{a}\mathbf{K}^T + \mathbf{S}_{\boldsymbol{\epsilon}})^{-1}(\boldsymbol{y} - \mathbf{K}\boldsymbol{x}_a) \tag{4}$$

$$\hat{\mathbf{S}} = (\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K} + \mathbf{S}_a^{-1})^{-1} \tag{5}$$

Where the subscript "a" refers to a-priori values and ϵ those associated with observational noise. $\mathbf{K}$ is the Jacobian matrix whose elements are $K_{i,j} = dy_i/dx_j$. OCO2CLD-LIDAR-AUX sets the first guess equal to the prior, and then allows up to five iteration steps where the posterior of the previous step replaces the prior in the equations above. The $\hat{\boldsymbol{x}}$ with the lowest associated χ^2 is the retrieved state.

OCO2CLD-LIDAR-AUX uses the same radiative transfer code as the OCO-2 XCO2 Level 2 Full Physics algorithm. This is based on LIDORT [*Spurr et al.*, 2001; *Spurr*, 2002] with the 2OS second order of scattering code described in [*Natraj and Spurr*, 2007] to determine polarisation for low order scattering. This fundamentally solves the radiative transfer equation using the eigenvector method [*Flatau and Stephens*, 1988] and accounts for multiple scattering. High accuracy calculations are done at 5—20 wavelengths in each band with 16 quadrature angles, and the rest are reconstructed with a polarised single-scattering code that accurately reproduces the spectrum at all wavelengths [*O'Dell*, 2010]. The code accounts for changes in atmospheric path length of the main beam due to the curvature of the Earth but is otherwise horizontally homogeneous.

Cloud layers are vertically uniform and Mie Scattering components are precalculated based on a gamma distribution in droplet effective radius resulting from in-situ studies [*Miles et al.*, 2000]. OCO2CLD-LIDAR-AUX uses the distribution associated with r_{eff} = 12 μm and adds a component to observational uncertainty $\mathbf{S}_\epsilon$ to account for uncertainty introduced by assuming a constant r_{eff}.

ECMWF meteorological profiles are interpolated from 137 onto 20 vertical levels. The default clear-sky pressure levels are linearly interpolated from 1 Pa to the surface pressure. Here the default pressure level that is closest to the cloud centre is moved to the appropriate level subject to the requirement that there are at least 2 levels either side. If the cloud is too high or low, then the 3rd level from the top or bottom is taken. The level above is shifted to P_{top} and all levels above that are linearly interpolated down to 1 Pa. The level below the cloud centroid pressure is shifted to the cloud bottom pressure (i.e. $P_{top} + \Delta P_c$) and below that all levels are linearly interpolated to the surface pressure.

An extinction coefficient is applied to the cloud centroid pressure level, and the code linearly interpolates this extinction, producing two uniform layers between the top, centre and bottom assigned cloud levels. The extinction coefficient α is provided in units of Pa^{-1} and therefore $\tau = \alpha\Delta P_c$.

The Jacobian matrix $\mathbf{K}$ is calculated via finite differencing with increments of 0.01 in τ and 1 hPa in P_{top} or ΔP_c. For the optical depth Jacobian, the extinction coefficient α is scaled while for the P_{top} Jacobian the pressure level of all cloud layers is shifted. For ΔP_c the cloud bottom is shifted by one increment and the centroid by half of that increment while the extinction coefficient is to ensure constant τ.

The forward model implementation has been validated [*Richardson et al.*, 2017]. A comparison of simulated and observed continuum radiances for marine liquid cloud scenes during November 2015 are shown in Figure 4, where cloud properties from collocated MODIS (MYD061KM) and CALIPSO (01kmCLay) products are input to the simulation [*Taylor et al.*, 2016]. Cloud field homogeneity is estimated using the standard deviation of continuum A-band radiances in neighbouring footprints divided by the cloudy footprint's continuum. The 10 % of cases with the lowest value of this parameter are considered "smooth", and the 10 % highest are considered "broken". The centre and right panels of Figure 4 show that the radiative transfer model performs well for smooth clouds and poorly for broken clouds.

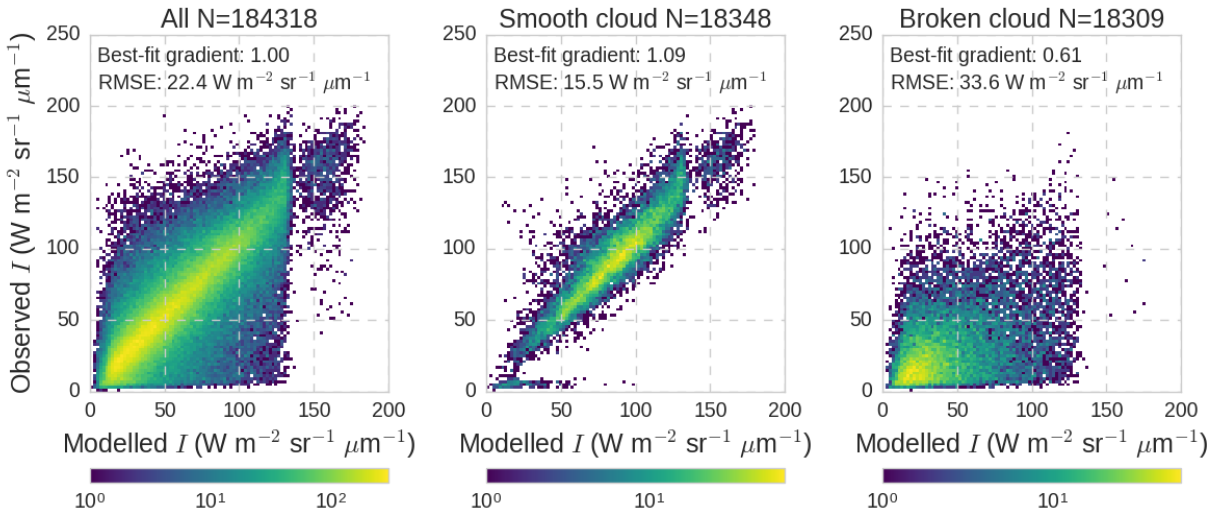


Figure 4 - 2d histograms of simulated and observed radiances across 10 continuum channels in the OCO-2 O2 A-band. Cases shown are marine liquid clouds seen by both CALIPSO & MODIS in collocated OCO-2 footprints, and simulations use the MODIS effective radius and optical depth with the CALIPSO cloud-top pressure. The left panel shows all cases during November 2015, the central panel shows the 10 % of most horizontally homogeneous cases and the right panel the 10 % most heterogeneous cases. Numbers differ slightly from Richardson et al. [2017] *Figure 5, due to selection of ranked channels 5—15 instead of 1—10.*

Smooth cloud fields do not only reduce cloud 3D effects, but also reduce the size of CALIPSO-MODIS-OCO-2 geolocation errors, and simulated A-band continuum radiance error is ±18 % for this sample [*Richardson et al.*, 2017]. This validates the radiative transfer implementation and shows the importance of cloud heterogeneity, so OCO2CLD-LIDAR-AUX also reports A-band spatial heterogeneity statistics.

In summary, the retrieval takes a prior state along with ECMWF meteorological profiles and simulates radiances with the modified LIDORT radiative transfer code that is used in the OCO-2 XCO2 retrieval. The same code also calculates the Jacobian matrix which is used with Equation (5) to iterate until the cost function χ^2 is minimised, at which point the posterior state is reported as the retrieval.

3. Algorithm Implementation

i. Matchup with CloudSat granules

Matchup distance is determined from the latitude-longitude distance between the product ground footprints as described in [*Taylor et al.*, 2016], and no account is made of how OCO-2 passes overhead

approximately 7 minutes before CloudSat. The standard output is the retrieved properties collocated with CloudSat footprints where the OCO-2 to CloudSat matchup distance is <1.25 km, about half the CloudSat along-track footprint size. This maintains consistency with other CloudSat products, although matchup characteristics and the full swath OCO-2 retrievals are also provided in datasets identified with "full_swath_*" names, such that users can perform matchups with different distance requirements. Note that retrievals are restricted to the sunlit portions of orbits for which OCO-2 was in nadir view and was formation flying with CALIPSO.

ii. Cloud detection

The standard OCO-2 level 2 preprocessors flag cloudy footprints but were not optimised for glint view over the ocean [*Taylor et al.*, 2016]. We use a simple brightness threshold instead. The mean radiances from 10 continuum channels for both the A-band and weak CO2 band are taken and divided by $\mu_0 = \cos(SZA)$. A cloud is flagged when the A-band average exceeds 6×10^{-19} ph s^{-1} m^{-2} sr^{-1} μm^{-1} and the weak CO2 band exceeds 2×10^{-19} ph s^{-1} m^{-2} sr^{-1} μm^{-1}.

The selected thresholds were based on training against collocated 1 km MODIS cloud mask values from November 2015 and the cloud mask values are also used for comparison. The threshold flag agreement is 84.86 % relative to MODIS confident only cases, or 84.39 % relative to MODIS confident or probably cloudy cases. Table 2q shows performance statistics versus 200 orbits randomly selected from outside of the training period, while Figure 55 shows the performance at different thresholds. The results show that our flag is informative about the presence of clouds.

OCO2CLD-LIDAR-AUX is only run when the CALIPSO footprint within the OCO-2 footprint's swath also contains a cloud with P_{top} > 680 hPa. The first 500 OCO2CLD-LIDAR-AUX orbits contain N = 3,624,645 cloud retrievals. Of these 95.3 % are reported as confidently cloudy by MODIS and 1.5 % as probably cloudy. The remaining 3.2 % are confidently or probably clear.

On-orbit icing of the instrument occurs, resulting in a time-dependent decay of measured radiance that is largely removed by each instrument decontamination cycle. Version 1 of OCO2CLD-LIDAR-AUX uses OCO-2 Version 7 Level 1b Science (L1bSc) radiances, which does not properly account for this icing. The worst icing led to periodic reductions of up to 18 % of the mission-start radiometric response [*Crisp et al.*, 2016]. This means that cloud flag performance is currently likely to be time dependent. Future versions will use OCO-2 L1bSc version 8 or later, which better accounts for the instrumental icing in radiance calibrations.

Table 2 – Agreement matrices between OCO2CLD-LIDAR-AUX cloud identification flag and MODIS Cloud_Mask_1km values, for MODIS confident cloudy only (top) and MODIS confident or probably cloudy (bottom).

MODIS confident only		
	OCO-2 cloudy	**OCO-2 clear**
MODIS cloudy	60.36	6.07
MODIS clear	9.07	24.50
MODIS confident or probable		
	OCO-2 cloudy	**OCO-2 clear**
MODIS cloudy	61.89	8.07
MODIS clear	7.55	22.50

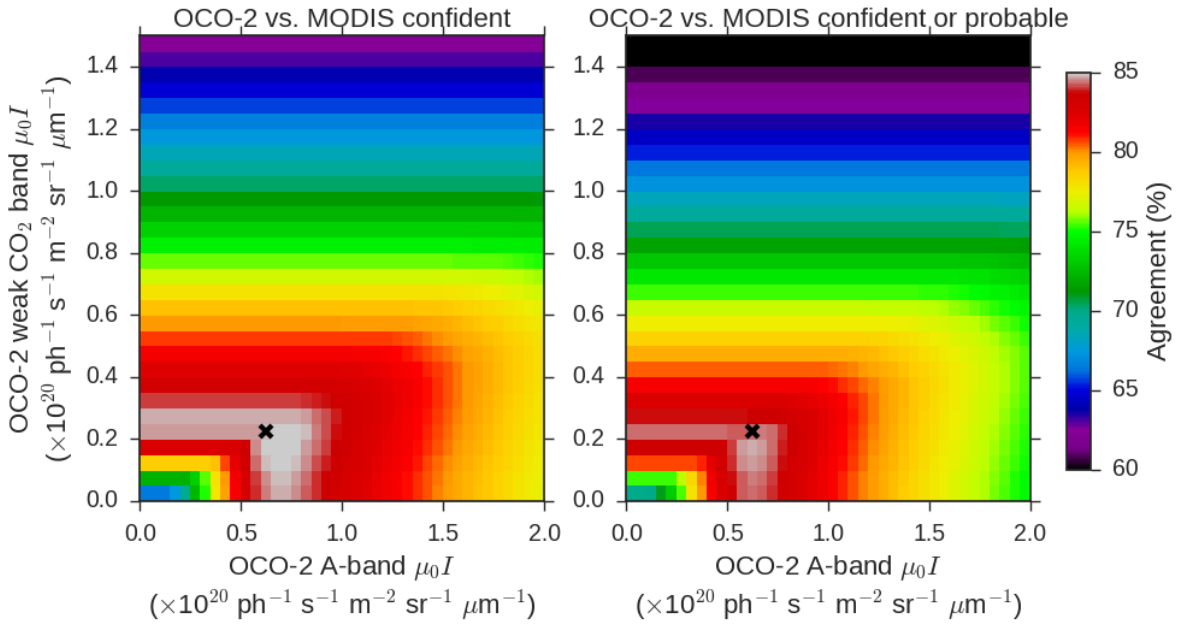


Figure 5 – Percentage of collocated soundings in which OCO2CLD-LIDAR-AUX flag agrees with MODIS Cloud_Mask_1km for different minimum radiance thresholds in the O_2 A-band (x axis) and weak CO_2 band (y axis), equivalent to the sum of the diagonal components from Table 2. The black cross shows the operational threshold.

iii. Cloud phase determination

Cloud phase is estimated from a Nakajima-King-like approach that exploits the stronger absorption of light by ice in the weak CO2 band relative to the O2 A-band. Figure 6 shows example simulated A-band and weak-CO_2 band radiances for water and ice clouds of varying τ and $r_{eff,}$ using the same continuum channels as in the cloud flag identification. A lookup table provides a threshold weak-band radiance for a given A-band radiance: if the weak-band is brighter than the threshold then cloud phase is assigned as liquid, otherwise it is assigned as ice. The thresholds were optimised based on sampling from November 2015. Table 3 shows that for the 3,919 orbits for which collocated MODIS properties are available, 90.6 % of OCO2CLD-LIDAR-AUX liquid marine clouds are also identified as such by MODIS MYD061KM.

Table 3 – MODIS classification of OCO2CLD-LIDAR-AUX identified liquid marine cloud scenes.

MODIS Classification of OCO-2 liquid cloud	Fraction (%)
Unclassified cloud flag	3.1
Clear	3.7
Liquid	90.6
Ice	1.3
Unclassified phase	0.5

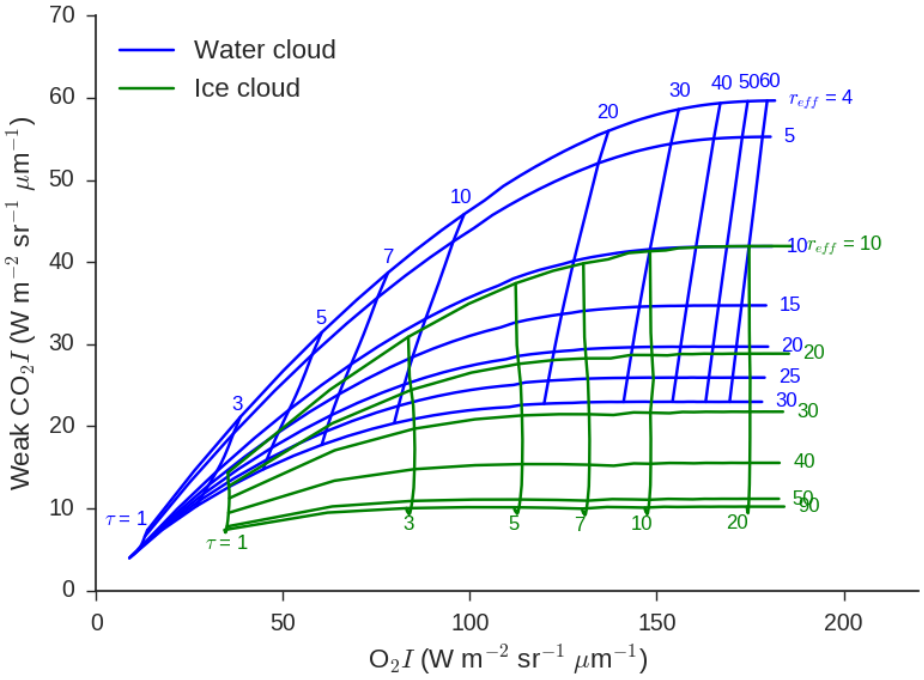


Figure 6 – Nakajima-King-like diagram for simulated cloud cases as seen by OCO-2. The x-axis contains the 10-channel mean continuum radiance from the O2 A-band and the y-axis the 10-channel mean continuum radiance in the weak CO2 band. The lines represent constant values of τ (top-to-bottom) or r_{eff} (left-to-right) as labelled. The ice cloud cases (green) are generally darker in the weak CO2 band than the water cloud cases (blue) due to ice's absorption at longer wavelengths.

iv. Radiative transfer input and prior cloud state

Atmospheric profiles are taken from ECMWF forecasts, including temperature, humidity and wind speed on 137 vertical levels interpolated down to 20 for the radiative transfer. A Cox-Munk surface with an albedo of 0.10 is used, and simulated nadir-view radiance in cloudy scenes was not found to be sensitive to this value up to <0.15.

Prior cloud τ uses the same μ_0-scaled A-band continuum radiance used in the cloud flag and phase determination input to the lookup table from [*Richardson et al.*, 2017]. An uncertainty of 20 % is assumed based on the radiative transfer validation for homogeneous cloud scenes in Figure 4.

Prior P_{top} is taken from the nearest CALIPSO 01kmCLay matchup with an uncertainty of ±5 hPa, approximately twice the vertical resolution of the CALIOP lidar products.

Prior cloud pressure thickness in metres is estimated from a subadiabatic cloud model for cloud geometric thickness [*Brenguier et al.*, 2000; *Bennartz*, 2007; *Bennartz and Rausch*, 2017]:

$$H = \sqrt{\frac{5Q_{ext}\rho_w \tau r_{eff}}{9C_w}} \quad (6)$$

where ρ_w is the density of water, τ the cloud optical depth and r_{eff} the cloud droplet effective radius, taken to be 12 μm. Q_{ext} is the scattering efficiency, approximately 2 for marine cloud droplets in the A-band. C_w is an adiabatic condensation coefficient which varies with temperature, but we take to be 1.9×10^{-3} g m^{-4} from typical values reported in [*Brenguier et al.*, 2000]. The assumed uncertainty is ±25 %,

slightly larger than estimated by [*Bennartz*, 2007]. The thickness is converted to pressure using a National Weather Service equation[1] and the state covariance matrix is assumed to be diagonal.

v. Uncertainties in simulated radiances

The observation covariance matrix $\mathbf{S}_\epsilon$ represents uncertainty in the simulated radiative transfer including instrumental uncertainty plus errors introduced by the incorrect representation of the scene. For example, atmospheric temperature or humidity profiles will not be completely accurate and this will introduce discrepancies between simulated and observed radiances. In addition, OCO2CLD-LIDAR-AUX assumes a constant droplet effective radius, whereas this varies in reality. We account for this by adding a component to the posterior error covariance.

The covariance matrix calculation is detailed in [*Richardson and Stephens*, 2018] and assumes independent contributions from instrumental uncertainty, $\mathbf{S}_I$, atmospheric temperature $\mathbf{S}_{dT}$, humidity $\mathbf{S}_{dq}$ and cloud r_{eff}, $\mathbf{S}_{dreff}$.

$$\mathbf{S}_\epsilon = \mathbf{S}_I + \mathbf{S}_{dT} + \mathbf{S}_{dq} + \mathbf{S}_{dreff} \tag{7}$$

The instrumental uncertainty is diagonal and its elements come from the OCO-2 L1bSc calibration information. Continuum signal-to-noise ratio is typically 400—800. The other components are estimated from a Monte Carlo like approach with 2,000 perturbations to temperature, humidity or r_{eff} for a variety of cloudy scenes. Temperature and humidity perturbations are based on AIRS validation [*Divakarla et al.*, 2006], being ±1.5 K in temperature and an altitude-dependent ±20 % to ±50 % in specific humidity. A lognormal distribution of r_{eff} was fit to the MODIS-observed distribution of liquid marine cloud values.

Figure 8 shows the square root of the diagonal components of the temperature, humidity and r_{eff} covariance matrices for a single case. The droplet size tends to have the largest effect, which is a uniform fraction across the band while the temperature component is important in the absorption bands. Figure 7 shows the full components of the $\mathbf{S}_\epsilon$ matrix for this case.

Matrices were determined for average meteorological profiles in “tropical” (±20° from the Equator), “mid-latitude” (±20—50°) and “high-latitude” (±50—90°) regions. It was found that linear scaling based on solar zenith angle (SZA) and τ can reliably construct one covariance matrix based on another. A single matrix for τ = 10 hPa, SZA = 45° and P_{top} = 850 hPa is stored and corrected based on each sounding’s SZA and prior τ estimate.

vi. Retrieval iteration procedure

The prior cloud state and meteorology are input to the forward model and assigned to 20 vertical pressure levels as described in Section 2iv. This also requires spacecraft instrument, geolocation and pointing information plus environmental factors such as illumination geometry which are taken from the OCO-2 L1bSc file.

For all identified liquid marine cloud soundings up to six iteration steps are allowed. For each step the previous step’s posterior state is input to the forward model. Loose constraints are applied to all steps: if the posterior state is outside $0.00001 < \tau < 150$ then the next step is taken as the closest edge of that range. The same occurs for thickness outside $0.1 < \Delta P_c < 500$ hPa, or for any cloud pressure level outside 380 hPa < P_i < Psurf. In this case the entire cloud is shifted such that it remains in that range.

[1] https://www.weather.gov/media/epz/wxcalc/pressureAltitude.pdf - last accessed 2018-03-13

The iteration with the lowest χ^2 is reported, subject to some requirements including $0.1 < \Delta P_c/\tau < 30$. If no realistic values are found, or if an exception is thrown by the retrieval or radiative transfer code in all steps, then the prior state is reported and the case is flagged.

The full OCO-2 swath retrievals are in the product as full_swath_<propertyname>, e.g. full_swath_Cloud_Top_Pressure. Standard output such as Cloud_Top_Pressure is reported for the CloudSat footprint and only contains values if the CloudSat-OCO-2 footprint separation is <1.25 km. Footprint indices and matchup distance in km between CloudSat and OCO-2 are also reported. This, along with contextual information such as the mean OCO-2 band continuum radiances and spatial statistics for a given footprint allow users to easily compare with CloudSat while exploiting the extra data available from the OCO-2 swath.

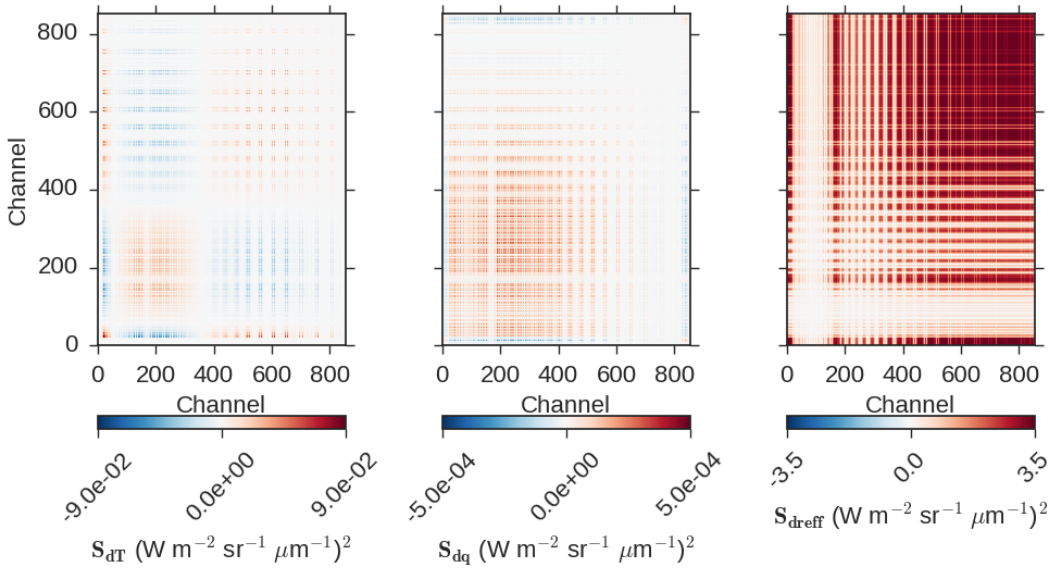


Figure 7 – Example contributions to the covariance matrix from uncertainty in atmospheric temperature (left), atmospheric humidity (centre) and cloud droplet effective radius (right). Values shown for a single τ=10 cloud case. Note the change in scale of units, and x- and y- refer to the OCO-2 channel position counting upwards in wavelength. Source: Richardson and Stephens [2018]*, licensed under CC-BY 4.0.*

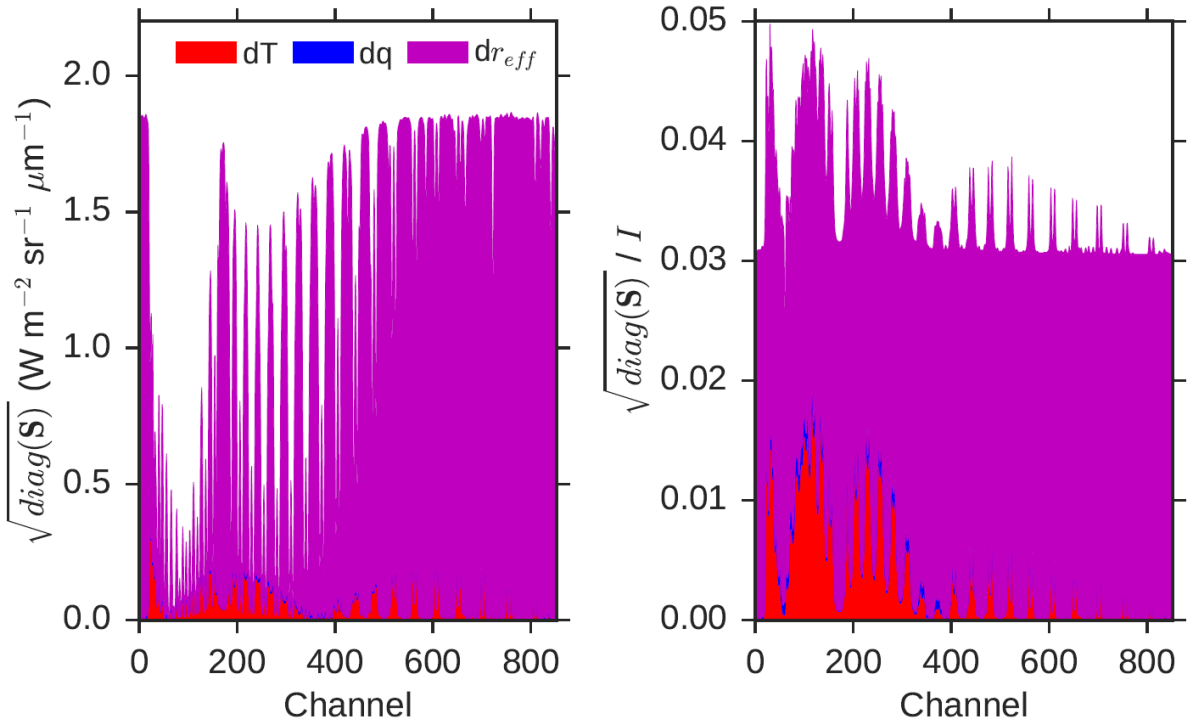
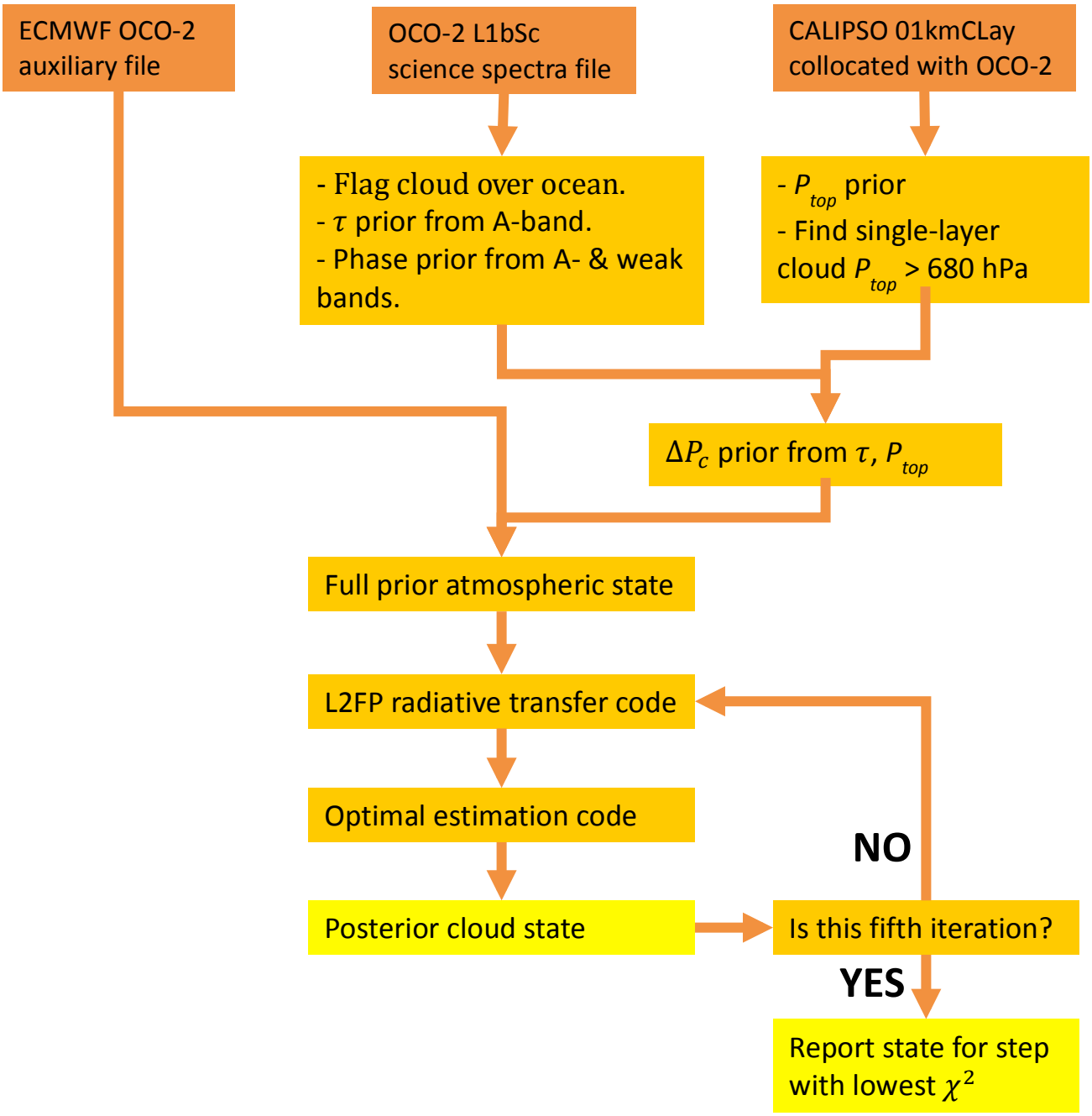


Figure 8 – Square root of diagonal elements of covariance matrices from Figure 7 stacked to show independent contribution to uncertainty from each component. Left shows the value in units of radiance, and right as a fraction of the channel radiance. This demonstrates that uncertainty is dominated by the effective radius uncertainty, with substantial contributions from temperature inside the absorption bands. Source: Richardson and Stephens [2018] *, licensed under CC-BY 4.0.*

4. Algorithm Design Summary

```
Load configuration file
Read 01kmCLay
Read ECMWF-AUX
Read L1bSc

Loop over all OCO-2 soundings
      If (sounding == over ocean
          & 1 layer whose Ptop > 680 hPa seen by CALIPSO
          & scene is bright enough to indicate cloud) then
            Estimate phase
            If (phase == liquid) then
                  Estimate prior state
                  Call retrieval algorithm
                  If (retrieval != failed) then
                        Report cloud state with lowest χ²
                  Else
                        Flag as failed
                  End
            End
      End
End
```



5. Retrieval Performance and Statistics

i. Algorithm throughput

Table 4 shows throughput statistics for the 4,055 OCO2CLD-LIDAR-AUX orbits. 13.34 % of OCO-2 footprints are identified as liquid clouds over ocean and 94.56 % of attempted retrievals give valid values. Radiative transfer or optimal estimation code failures account for 1.69 %, and non-allowed retrieval states, such as cloud bottoms below the surface, account for 3.75 % of attempts.

The histograms in Figure 9 show that failures are most common for high solar zenith angle or with very thin or thick prior optical depths. The high SZA and high prior-τ failures may be related since prior τ uses $\cos^{-1}(SZA)$-scaled radiances, resulting in large outliers for very large solar zenith angles.

Meanwhile, other small prior τ cases have other failures, some of these may be partially cloudy scenes where the retrieval, which assumes a uniform cloudy scene, is not able to produce a reliable fit to the real scene conditions.

Table 4 – For the first 1,790 processed OCO-2 orbits, total number of soundings, number for which retrievals were attempted and number of failures.

-	N	Percent of soundings	Percent of done
Soundings	258,783,160	100.00	NaN
Retrievals	34,517,811	13.34	100.00
Successful	32,641,255	12.61	94.56
Bad values	1,293,178	0.50	3.75
Failure	583,378	0.23	1.69

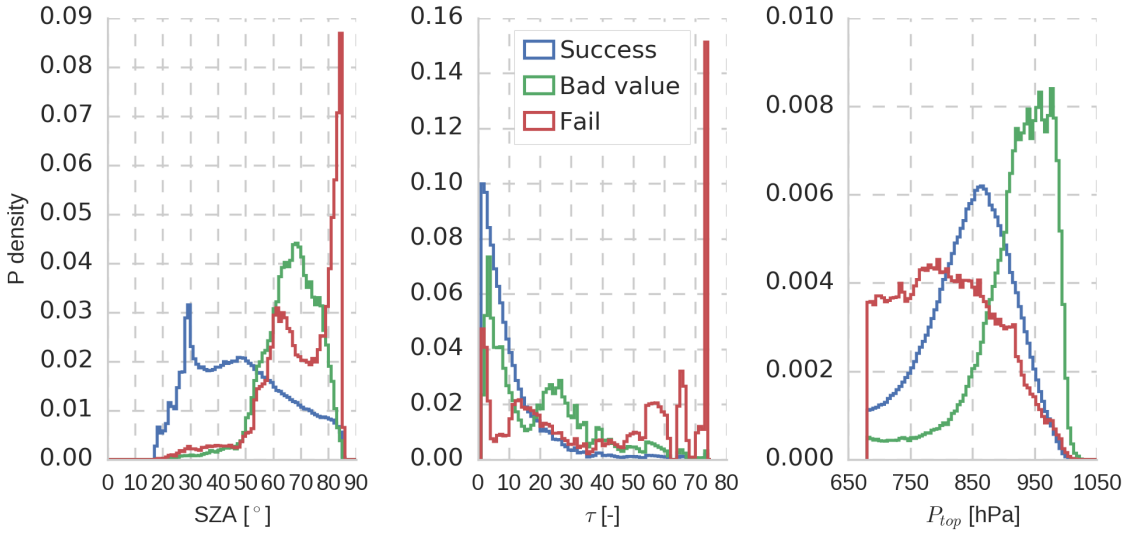


Figure 9 – Normalised histograms of solar zenith angle (left), prior cloud optical depth (centre) and prior cloud top pressure (right) for the subsets of successful retrievals (blue), cases where the radiative transfer code reported a failure (green), where another part of the retrieval failed (red) or where the posterior cloud bottom falls below the surface (purple). Sample is that from Table 4.

ii. OCO2CLD-LIDAR-AUX versus MODIS MYD061KM and CALIPSO 01kmCLay

Figure 10 shows histograms of OCO2CLD-LIDAR-AUX minus MODIS τ and OCO2CLD-LIDAR-AUX posterior minus CALIPSO prior P_{top}. The data are split for SZA below or greater than 45° and the median and 14—86 % ranges are reported. For SZA < 45° the agreement with MODIS is better and the iteration away from the prior P_{top} tends to be smaller, although the high SZA P_{top} cases appear to be bimodal. The differences relative to MODIS are partly due to collocation error and footprint size differences, meaning that this distribution places an upper limit on the OCO2CLD-LIDAR-AUX retrieval uncertainty.

The OCO-2 retrieval generally favours decreases in the retrieved cloud-top pressure, with an exception for the right hand peak in the bimodal high SZA distribution. Spatial heterogeneity and assumptions related to the cloud profile will lead to discrepancies between the retrieval and reality, the following factors are proposed as contributors:

1) Overlying aerosol or cirrus may still be present, either due to low τ or because the smaller CALIPSO footprint didn't encounter it. This subset should have lower P_{top} since scattering from the higher layer reduces mean photon path length.
2) Some mixed phase or ice clouds will still be misclassified and processed, and will have different scattering properties.
3) The assumed droplet size of 12 μm might introduce droplet-size-dependent biases in retrieved path length.
4) The assumed homogeneous cloud may result in within-cloud path biases that are compensated by cloud-top biases.

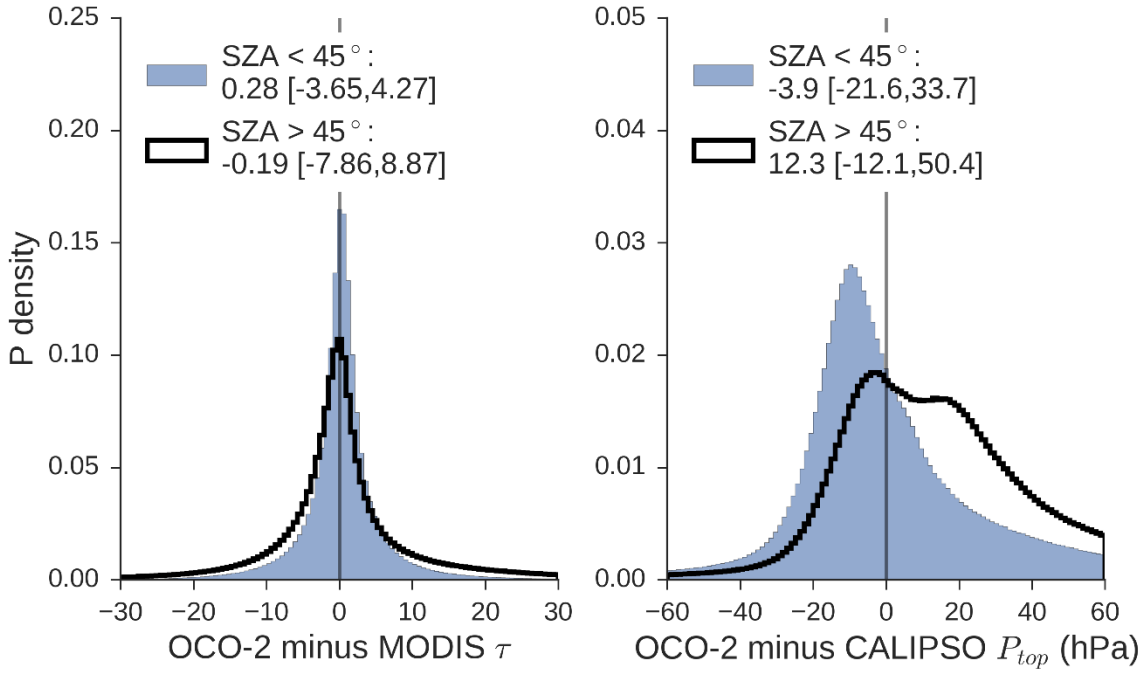


Figure 10 – Histograms of OCO2CLD-LIDAR-AUX retrieval properties minus MODIS τ (left) and CALIPSO 01kmCLay cloud-top pressure (right). In each case the solid shaded histogram is for SZA < 45° and the black line histogram for SZA > 45°.

Figure 11 shows that the subset of cases flagged as multi-layer by MODIS matches these expectations, with a larger negative P_{top} shift. Multi-layer clouds also affect the measured ratio of the weak CO_2 band continuum to that of the A-band. Lower ratios also suggest either ice clouds or larger r_{eff}. Bins of P_{top} changes by MODIS r_{eff} or the OCO-2 weak-CO_2 to O_2 A-band ratios are shown in Figure 12. Given that low I_{wk}/I_{O2} coincides with large droplet sizes, ice or multi-layer clouds, this ratio is used to flag cases where the retrieval assumptions are potentially violated, resulting in poor quality retrievals. A threshold of I_{wk}/I_{O2} = 0.28 was identified and this information is included in a Quality_flag output. Approximately 10 % of retrievals are flagged and Tables 5—7 show retrieval comparison statistics that indicate using the ratio warning flag reduces discrepancies between OCO-2 and either MODIS, CALIPSO or the prior state. These sample sizes are smaller than reported in Table 4 because only footprints with valid collocated MODIS retrievals are considered.

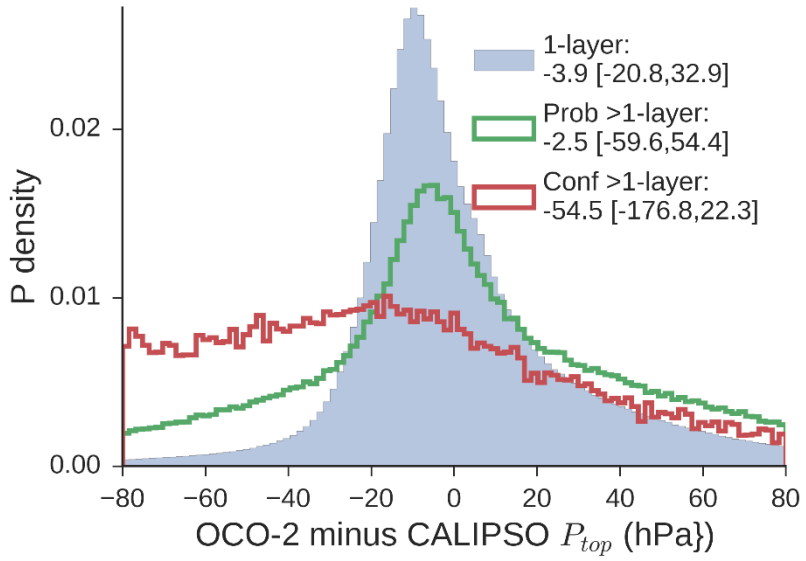


Figure 11 – Histograms of posterior OCO-2 P_{top} minus the CALIPSO prior P_{top} used in the retrieval subset by the MYD061KM Cloud_Multi_Layer_Flag, values only shown where SZA < 45°. Values of 2—5 inclusive are taken as "probably >1 layer" and values greater than 5 are assumed to be "confident >1 layer". The bias in multi-layer cloud cases is as expected, with the median (14—86 % range) shown in the legend. However, this is a minor contributor as can be seen from the shape of the 1-layer histogram. This is because multi-layer cases represent only 3.5 % of the sample.

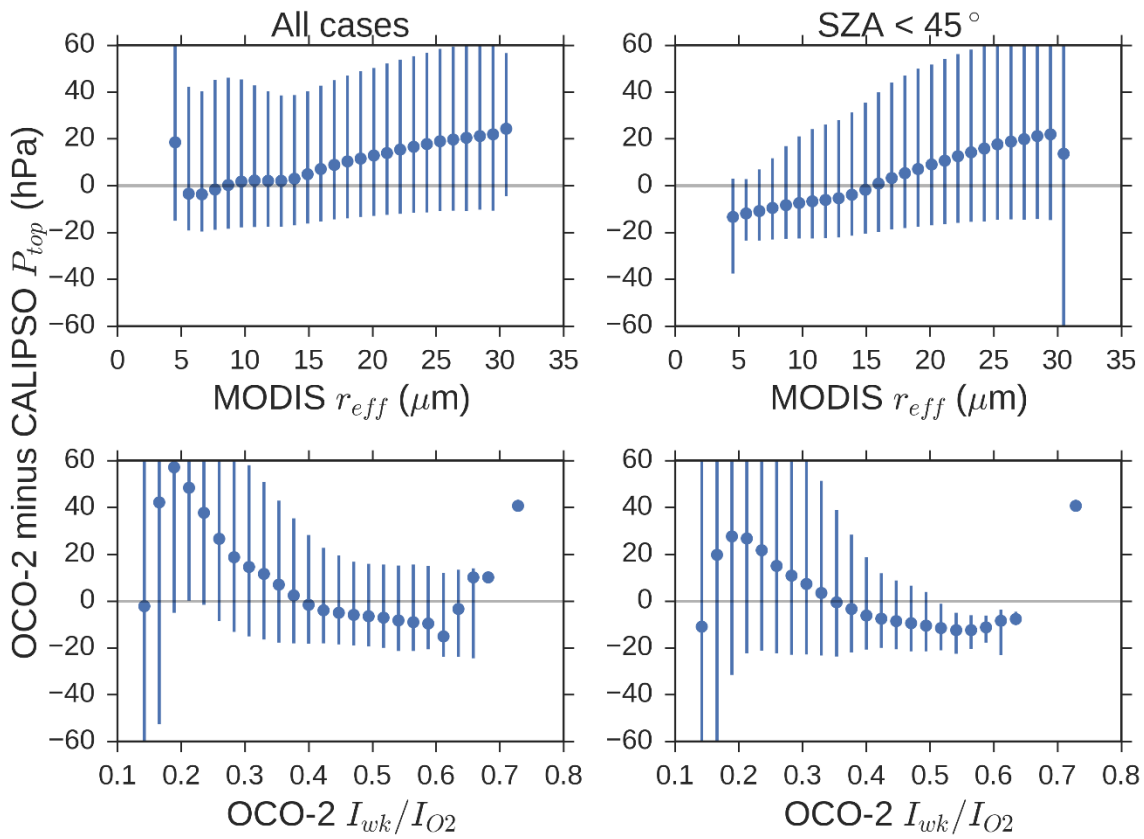


Figure 12 – Difference between OCO2CLD-LIDAR-AUX and CALIPSO P_{top} statistics binned by MODIS effective radius (top) and by the ratio of OCO-2's weak CO_2 band continuum to the oxygen A-band continuum (bottom). Left column is all cases, and right column is for SZA < 45°. Points are the distribution median and the lines span the 14—86th percentiles of the samples.

Table 5 – Statistics of MODIS minus OCO-2 samples, subset as labelled. When ratio_flag =1, this indicates cases where the oco2_band_ratio_warning_flag = 1, i.e. the ratio of I_{wk}/I_{O2} is low enough to indicate that retrieval assumptions are violated.

Case	N	Median	14—86 %	5—95 %	μ	σ
All	21697702	0.07	[-5.80,6.27]	[-14.42,16.93]	-0.38	15.28
SZA < 45°	9472916	0.28	[-3.65,4.27]	[-8.50,9.55]	0.43	7.09
SZA > 45°	12224786	-0.19	[-7.86,8.87]	[-20.58,22.40]	-1.00	19.35
I_{wk}/I_{O2} < 0.28	2380314	11.25	[-7.69,30.64]	[-30.75,44.52]	9.26	27.89
I_{wk}/I_{O2} > 0.28	19317388	-0.14	[-5.69,3.95]	[-13.34,8.59]	-1.56	12.39
SZA < 45°, I_{wk}/I_{O2} < 0.28	614478	6.41	[-5.47,21.77]	[-13.89,35.56]	7.67	17.88
SZA > 45°, I_{wk}/I_{O2} < 0.28	1765836	13.50	[-9.24,32.77]	[-41.24,46.53]	9.81	30.59
SZA < 45°, I_{wk}/I_{O2} > 0.28	8858438	0.19	[-3.58,3.51]	[-8.17,7.01]	-0.07	5.27
SZA > 45°, I_{wk}/I_{O2} > 0.28	10458950	-0.58	[-7.76,4.45]	[-18.65,10.03]	-2.83	16.01

Table 6 – As Table 5 but referring to OCO2CLD-LIDAR-AUX P_{top} minus CALIPSO P_{top}, which is used as the prior.

	hPa					
Case	**N**	**Median**	**14—86 %**	**5—95 %**	**μ**	**σ**
All	21697702	4.5	[-17.10,44.51]	[-34.83,75.58]	9.1	40.5
SZA < 45°	9472916	-3.9	[-21.56,33.67]	[-45.15,67.40]	0.5	41.4
SZA > 45°	12224786	12.4	[-12.13,50.42]	[-26.46,80.16]	15.8	38.5
I_{wk}/I_{O2} < 0.28	2380314	30.8	[-7.54,78.49]	[-29.42,108.92]	32.6	48.0
I_{wk}/I_{O2} > 0.28	19317388	2.4	[-17.63,38.17]	[-35.49,67.58]	6.2	38.5
SZA < 45°, I_{wk}/I_{O2} < 0.28	614478	17.7	[-22.35,72.79]	[-48.80,105.53]	20.2	54.0
SZA > 45°, I_{wk}/I_{O2} < 0.28	1765836	34.2	[-1.06,80.08]	[-19.33,109.91]	36.9	44.9
SZA < 45°, I_{wk}/I_{O2} > 0.28	8858438	-4.5	[-21.53,29.81]	[-44.88,62.34]	-0.8	40.0
SZA > 45°, I_{wk}/I_{O2} > 0.28	10458950	9.7	[-13.03,42.87]	[-27.34,70.93]	12.2	36.1

Table 7 – As Table 5 but referring to OCO2CLD-LIDAR-AUX posterior cloud pressure thickness minus the prior, which is based on a subadiabatic cloud assumption.

	hPa					
Case	**N**	**Median**	**14—86 %**	**5—95 %**	**μ**	**σ**
All	21697702	0.25	[-3.88,10.70]	[-7.78,25.87]	3.56	15.3
SZA < 45°	9472916	-0.36	[-5.15,4.57]	[-9.03,13.12]	0.61	12.59
SZA > 45°	12224786	1.14	[-2.68,15.77]	[-6.31,32.25]	5.84	16.75
I_{wk}/I_{O2} < 0.28	2380314	5.07	[-4.13,25.53]	[-14.40,42.93]	9.03	21.03
I_{wk}/I_{O2} > 0.28	19317388	0.11	[-3.87,8.56]	[-7.44,22.15]	2.88	14.29
SZA < 45°, I_{wk}/I_{O2} < 0.28	614478	0.58	[-8.21,13.61]	[-23.67,27.61]	2.32	19.79
SZA > 45°, I_{wk}/I_{O2} < 0.28	1765836	7.13	[-2.43,28.74]	[-11.06,46.25]	11.36	20.94
SZA < 45°, I_{wk}/I_{O2} > 0.28	8858438	-0.39	[-5.03,4.06]	[-8.64,11.71]	0.49	11.92
SZA > 45°, I_{wk}/I_{O2} > 0.28	10458950	0.75	[-2.70,12.88]	[-5.95,28.37]	4.91	15.74

6. Output File Contents

i. Summary and variable lists

Output is provided in an HDF-EOS structure with geolocation data, ancillary properties and metadata. Table 8 lists the Geolocation Fields entries and Table 9 the Data Fields entries. CloudSat has no swath and *nray* footprints along track. The OCO-2 convention is that the along track position is a "frame" and across track is a "sounding". The footprints are stored in an array of shape *nframe x nsounding* where *nsounding* = 8.

Retrieved properties are provided both for OCO-2 and CloudSat footprints, with the CloudSat footprints containing fill_value if there is no OCO-2 observed cloud within 1.25 km. Matchup indices and matchup distance are also provided for all footprints within 50 km. This allows users to easily compare results across different distance scales.

Table 8 Geolocation Fields

	Source	**Name**
(G1)	CloudSat 1B-CPR	Latitude
(G2)	CloudSat 1B-CPR	Longitude
(G3)	CloudSat 1B-CPR	Profile_time
(G4)	CloudSat 1B-CPR	UTC_start
(G5)	CloudSat 1B-CPR	TAI_start
(G6)	CloudSat 1B-CPR	DEM_elevation
(G7)	OCO-2 L1bSc	full_swath_sounding_id
(G8)	OCO-2 L1bSc	full_swath_sounding_latitude
(G9)	OCO-2 L1bSc	full_swath_sounding_longitude

Table 9 Data Fields

	Source	**Name**
(1)		CALIPSO_OCO2_matchup_distance_km
(2)		Cloud_Optical_Depth
(3)		Cloud_Pressure_Thickness
(4)		Cloud_Top_Pressure
(5)		CloudSat_OCO2_matchup_distance_km
(6)		CloudSat_OCO2_pixel_index_along_track
(7)		CloudSat_OCO2_pixel_index_cross_track
(8)	OCO2CLD-LIDAR-AUX	Quality_flag
(9)	OCO-2 L1bSc	sounding_solar_zenith
(10)	CALIPSO 01kmCLay	full_swath_CALIPSO_number_of_layers
(11)		full_swath_CALIPSO_OCO2_matchup_distance_km
(12)	OCO2CLD-LIDAR-AUX	full_swath_chi_squared
(13)	OCO2CLD-LIDAR-AUX	full_swath_cloud_flag
(14)	OCO2CLD-LIDAR-AUX	full_swath_Cloud_Optical_Depth
(15)	OCO2CLD-LIDAR-AUX	full_swath_Cloud_Pressure_Thickness
(16)	OCO2CLD-LIDAR-AUX	full_swath_Cloud_Top_Pressure
(17)		full_swath_CloudSat_OCO2_matchup_distance_km
(18)	OCO2CLD-LIDAR-AUX	full_swath_o2_local_avg

(19)	OCO2CLD-LIDAR-AUX	full_swath_o2_local_std
(20)	OCO2CLD-LIDAR-AUX	full_swath_phase_prior
(21)	OCO2CLD-LIDAR-AUX	full_swath_Quality_flag
(22)	OCO2CLD-LIDAR-AUX	full_swath_radiance_o2_continuum
(23)	OCO2CLD-LIDAR-AUX	full_swath_radiance_wk_continuum
(24)	OCO-2 L1bSc	full_swath_sounding_id
(25)	OCO-2 L1bSc	full_swath_sounding_latitude
(26)	OCO-2 L1bSc	full_swath_sounding_longitude
(27)	OCO-2 L1bSc	full_swath_sounding_solar_zenith
(28)	OCO-2 ECMWF	full_swath_surface_pressure_ecmwf
(29)	OCO-2 ECMWF	full_swath_two_meter_temperature_ecmwf
(30)	OCO-2 ECMWF	full_swath_2m_minus_700hPa_temperature_ecmwf
(31)		CALIPSO_file
(32)		L1bSc_file
(33)		ecmwf_file

ii. Geolocation Fields

(G1) Latitude (nray) [float]
CloudSat spacecraft geodetic latitude.

(G2) Longitude (nray) [float]
CloudSat spacecraft geodetic longitude.

(G3) Profile_time (nray) [float]
Seconds since the start of the granule for each profile. The first profile is 0.

(G4) UTC_start (nray) [float]
The UTC seconds since 00:00 Z of the first profile in the data file.

(G5) TAI_start (nray) [float]
The TAI timestamp in seconds for the first profile in the data file. TAI is International Atomic Time: seconds since 00:00:00 Jan 1 1993.

(G6) DEM_elevation (nray) [float]
Elevation in meters above Mean Sea Level. A value of -9999 indicates ocean. A value of 9999 indicates an error in calculation of the elevation.

(G7) full_swath_id (nframe,nsounding) [int]
A unique 16-digit code associated with a single OCO-2 sounding.

(G8) full_swath_latitude (nframe,nsounding) [float]
Geodetic latitude of OCO-2 sounding.

(G9) full_swath_longitude (nframe,nsounding) [float]
Geodetic longitude of OCO-2 sounding.

iii. Data Fields

(1) CALIPSO_OCO2_matchup_distance_km (nray) [float]
Distance in km between estimated CALIPSO and OCO-2 footprint locations at surface.

(2) Cloud_Optical_Depth (nray) [float]

Posterior cloud optical depth from OCO-2 if footprint within 1.25 km of CloudSat's.

(3) Cloud_Pressure_Thickness (nray) [float]

Posterior cloud pressure thickness in hPa if OCO-2 footprint within 1.25 km of CloudSat's.

(4) Cloud_Top_Pressure (nray) [float]

Posterior cloud top pressure in hPa if OCO-2 footprint within 1.25 km of CloudSat's.

(5) CloudSat_OCO2_matchup_distance_km (nray) [float]

Distance in km between estimated CloudSat and OCO-2 footprint locations at surface.

(6) CloudSat_OCO2_pixel_index_along_track (nray) [int]

For each nray location, this is the closest along-track index for the OCO-2 data. In OCO-2 nomenclature, it is the closest "frame".

(7) CloudSat_OCO2_pixel_index_cross_track (nray) [int]

For each nray location, this is the closest cross-track index for the OCO-2 data. In OCO-2 nomenclature, it is the closest "sounding" within the frame.

(8) Quality_flag (nray) [int]

For each valid OCO2CLD-AUX retrieval within 1.25 km of a CloudSat footprint, reports an integer whose bits correspond to the following conditions, which are related to retrieval quality. Quality_flag = 0 represents the best quality data, and more positive values represent more severe warnings. All triggered quality control flag values from the table below are summed for the final value, such that e.g. 11 represents SZA > 45°, a warning due to the I_{wk}/I_{O2} ratio plus a retrieved state outside the recommended range.

Quality_flag value	Meaning
-999999	No retrieval attempted
0	Retrieval successful with no warnings
1	SZA > 45°
2	Low I_{wk}/I_{O2} ratio, risk of poor retrieval
4	Cosmic ray strike on detector
8	Retrieved state outside recommended range
16	NOT USED
32	Code failure, example causes include invalid inputs to radiative transfer such as negative surface pressure, or no valid solution to optimal estimation iteration.

The recommended ranges are:

$0.3 <$ Cloud_Optical_Depth < 150

$680 <$ Cloud_Top_Pressure $< P_{surf}$

$0.1 <$ Cloud_Pressure_Thickness / Cloud_Optical_Depth < 30

Therefore the warning flag highlights very optically thick or thin clouds, clouds with tops that are likely to be ice or are retrieved below the surface, and clouds whose retrieved thickness leads to unrealistically dense or sparse liquid water content profiles.

(9) sounding_solar_zenith (nray) [float]

Solar zenith angle in degrees for OCO-2 footprint, when within 1.25 km of CloudSat footprint.

(10) full_swath_CALIPSO_number_of_layers (nframe,nsounding) [int]

Number of cloud layers in nearest CALIPSO footprint for all OCO-2 footprints.

(11) full_swath_CALIPSO_OCO2_matchup_distance_km (nframe,nsounding) [float]

Distance in km between CALIPSO footprint used for prior and OCO-2 footprint.

(12) full_swath_chi_squared (nframe,nsounding) [float]
χ^2 value calculated from posterior state for all OCO-2 footprints.

(13) full_swath_cloud_flag (nframe,nsounding) [int]
Set to 1 when OCO-2 radiance thresholds indicate presence of a cloud, 0 otherwise. For all OCO-2 footprints

(14) full_swath_Cloud_Optical_Depth (nframe,nsounding) [float]
Posterior cloud optical depth for all OCO-2 footprints

(15) full_swath_Cloud_Top_Pressure (nframe,nsounding) [float]
Posterior cloud top pressure in hPa for all OCO-2 footprints.

(16) full_swath_Cloud_Pressure_Thickness (nframe,nsounding) [float]
Posterior cloud pressure thickness in hPa for all OCO-2 footprints.

(17) full_swath_CloudSat_OCO2_matchup_distance_km (nframe,nsounding) [float]
Distance in km between nearest CloudSat and OCO-2 footprints, for all OCO-2 footprints.

(18) full_swath_o2_local_avg (nframe,nsounding) [float]
Mean radiance across all neighbouring footprints for 10 OCO-2 A-band continuum channels, in default OCO-2 L1bSc units: photons s^{-1} m^{-2} sr^{-1} μm^{-1}. Neighbouring footprints are those within ±1 along *or* cross track, including diagonally. Footprints on the edge of the OCO-2 swath therefore have five neighbours, and others have eight.

(19) full_swath_o2_local_std (nframe,nsounding) [float]
Standard deviation of neighbouring footprints' mean continuum radiances in default OCO-2 L1bSc units: photons s^{-1} m^{-2} sr^{-1} μm^{-1}. Neighbouring footprints are those within ±1 along *or* cross track, including diagonally. Footprints on the edge of the OCO-2 swath therefore have five neighbours, and others have eight.

(20) full_swath_phase_prior (nframe,nsounding) [int]
Estimate of cloud phase with numbering matching MODIS MYD061KM product: 2 = liquid and 3 = ice. Reported for all footprints, including land and non-cloudy scenes.

(21) full_swath_Quality_flag (nframe,nsounding) [int]
See entry Quality_flag, value provided for all OCO-2 footprints.

(22) full_swath_radiance_o2_continuum (nframe,nsounding) [float]
Radiance averaged across 10 O2 A-band continuum channels within footprint in default OCO-2 L1bSc units: photons s^{-1} m^{-2} sr^{-1} μm^{-1}. For all OCO-2 footprints.

(23) full_swath_radiance_wk_continuum (nframe,nsounding) [float]
Radiance averaged across 10 weak CO_2 band continuum channels within footprint in default OCO-2 L1bSc units: photons s^{-1} m^{-2} sr^{-1} μm^{-1}. For all OCO-2 footprints.

(24) full_swath_sounding_id (nframe,nsounding) [int]
Unique identifier as provided in all OCO-2 standard output files to facilitate comparison with other OCO-2 products if desired. For all OCO-2 footprints.

(25) full_swath_sounding_latitude (nframe,nsounding) [float]
Latitude in degrees of OCO-2 footprint, for all footprints.

(26) full_swath_sounding_longitude (nframe,nsounding) [float]
Longitude in degrees of OCO-2 footprint, for all footprints.

(27) full_swath_sounding_solar_zenith (nframe,nsounding) [float]
Solar zenith angle in degrees for all OCO-2 footprints.

(28) full_swath_surface_pressure_ecmwf (nframe,nsounding) [float]
Surface pressure in hPa collocated with each OCO-2 footprint as provided by ECMWF.

(29) full_swath_two_meter_temperature_ecmwf (nframe,nsounding) [float]
Temperature in Kelvin collocated with each OCO-2 footprint as provided by ECMWF.

(30) full_swath_2m_minus_700hPa_temperature_ecmwf (nframe,nsounding) [float]
Difference between 2 m and 700 hPa temperature in Kelvin estimated from ECMWF. 700 hPa value calculated from linear interpolation. Reported for all OCO-2 footprints.

(31) CALIPSO_file (1) [string]
Name of 01kmCLay matchup file used as input.

(32) L1bSc_file (1) [string]
Name of OCO-2 L1bSc file used as input.

(33) ecmwf_file (1) [string]
Name of OCO-2 ECMWF auxiliary file used as input.

Acknowledgement: The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Bennartz, R. (2007), Global assessment of marine boundary layer cloud droplet number concentration from satellite, *J. Geophys. Res.*, *112*(D2), D02201, doi:10.1029/2006JD007547.

Bennartz, R., and J. Rausch (2017), Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations, *Atmos. Chem. Phys.*, *17*(16), 9815–9836, doi:10.5194/acp-17-9815-2017.

Boesch, H. et al. (2017), *Orbiting Carbon Observatory (OCO)-2 Level 2 Full Physics Algorithm Theoretical Basis Document*, Pasadena, CA.

Brenguier, J., H. Pawlowska, L. Schüller, R. Preusker, J. Fischer, and Y. Fouquart (2000), Radiative Properties of Boundary Layer Clouds: Droplet Effective Radius versus Number Concentration, *J. Atmos. Sci.*, *57*(6), 803–821, doi:10.1175/1520-0469(2000)057<0803:RPOBLC>2.0.CO;2.

Cox, C., and W. Munk (1954), Statistics of the sea surface derived from sun glitter, *J. Mar. Res.*, *13*, 198–227.

Crisp, D. et al. (2016), The On-Orbit Performance of the Orbiting Carbon Observatory-2 (OCO-2) Instrument and its Radiometrically Calibrated Products, *Atmos. Meas. Tech. Discuss.*, 1–45, doi:10.5194/amt-2016-281.

Divakarla, M. G., C. D. Barnet, M. D. Goldberg, L. M. McMillin, E. Maddy, W. Wolf, L. Zhou, and X. Liu (2006), Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts, *J. Geophys. Res.*, *111*(D9), D09S15, doi:10.1029/2005JD006116.

Flatau, P. J., and G. L. Stephens (1988), On the fundamental solution of the radiative transfer equation, *J. Geophys. Res.*, *93*(D9), 11037, doi:10.1029/JD093iD09p11037.

Koelemeijer, R. B. A. (2002), Global distributions of effective cloud fraction and cloud top pressure derived from oxygen A band spectra measured by the Global Ozone Monitoring Experiment: Comparison to ISCCP data, *J. Geophys. Res.*, *107*(D12), 4151, doi:10.1029/2001JD000840.

Miles, N. L., J. Verlinde, and E. E. Clothiaux (2000), Cloud Droplet Size Distributions in Low-Level Stratiform Clouds, *J. Atmos. Sci.*, *57*(2), 295–311, doi:10.1175/1520-0469(2000)057<0295:CDSDIL>2.0.CO;2.

Natraj, V., and R. J. D. Spurr (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. Transf.*, *107*(2), 263–293, doi:10.1016/j.jqsrt.2007.02.011.

O'Dell, C. W. (2010), Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation, *J. Geophys. Res.*, *115*(D10), D10206, doi:10.1029/2009JD012803.

Richardson, M., and G. L. Stephens (2018), Information content of OCO-2 oxygen A-band channels for retrieving marine liquid cloud properties, *Atmos. Meas. Tech.*, *11*(3), 1515–1528, doi:10.5194/amt-11-1515-2018.

Richardson, M., J. McDuffie, G. L. Stephens, H. Q. Cronk, and T. E. Taylor (2017), The OCO-2 oxygen A-band response to liquid marine cloud properties from CALIPSO and MODIS, *J. Geophys. Res. Atmos.*, doi:10.1002/2017JD026561.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding Theory and Practice*, World Scientific, Singapore.

Schuessler, O., D. G. Loyola Rodriguez, A. Doicu, and R. Spurr (2014), Information Content in the Oxygen <inline-formula> <tex-math notation="TeX">A</tex-math></inline-formula>-Band for the Retrieval of Macrophysical Cloud Parameters, *IEEE Trans. Geosci. Remote Sens.*, *52*(6), 3246–3255, doi:10.1109/TGRS.2013.2271986.

Spurr, R. J. D. (2002), Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, *J. Quant. Spectrosc. Radiat. Transf.*, *75*(2), 129–175, doi:10.1016/S0022-4073(01)00245-X.

Spurr, R. J. D., T. P. Kurosu, and K. V. Chance (2001), A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval, *J. Quant. Spectrosc. Radiat. Transf.*, *68*(6), 689–735, doi:10.1016/S0022-4073(00)00055-8.

Taylor, T. E. et al. (2016), Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: validation against collocated MODIS and CALIOP data, *Atmos. Meas. Tech.*, *9*(3), 973–989, doi:10.5194/amt-9-973-2016.

FAQ

Q: Why are only some orbits processed?

A: OCO-2 commits each orbit to either nadir or glint viewing. Only nadir view orbits are processed. Other issues remove other orbits or parts thereof, such as when OCO-2 performs target measurements for validation, or is undergoing decontamination.

Q: Why do only parts of each CloudSat granule ever have data?

A: OCO-2 only retrieves in the sunlit portion of each orbit.

Q: What's the difference between full_swath_Cloud_Optical_Depth and Cloud_Optical_Depth?

A: OCO-2 takes 8 across track footprints, output from these is stored in the full_swath_* datasets in the same along-track and across-track shape as the OCO-2 products. The standard output such as Cloud_Optical_Depth is the OCO-2 retrieval within a swath that is nearest to the CloudSat footprint.

Q: Why are only some months processed?

A: OCO2CLD-LIDAR-AUX is only available for the CALIPSO-OCO-2 formation flying period.

Q: How do I find which input data versions were used?

A: The filenames stored in CALIPSO_file and L1bSc_file and ecmwf_file contain the version information.