OCO D-55207

Orbiting Carbon Observatory-2 (OCO-2)

Level 2 Full Physics Retrieval Algorithm Theoretical Basis

Version 3.0 Rev 0
December 7, 2017
Data Release 8

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ORBITING CARBON OBSERVATORY (OCO) - 2

LEVEL 2 FULL PHYSICS ALGORITHM

Theoretical Basis Document

Hartmut Boesch	University of Leicester
Linda Brown	Jet Propulsion Laboratory
Rebecca Castano	Jet Propulsion Laboratory
Mathew Christi	Colorado State University
Brian Connor	BC Consulting Limited
David Crisp	Jet Propulsion Laboratory
Annmarie Eldering	Jet Propulsion Laboratory
Brendan Fisher	Jet Propulsion Laboratory
Christian Frankenberg	Jet Propulsion Laboratory
Michael Gunson	Jet Propulsion Laboratory
Robert Granat	Jet Propulsion Laboratory
James McDuffie	Jet Propulsion Laboratory
Charles Miller	Jet Propulsion Laboratory
Vijay Natraj	Jet Propulsion Laboratory
Denis O'Brien	Colorado State University
Chris O'Dell	Colorado State University
Gregory Osterman	Jet Propulsion Laboratory
Fabiano Oyafuso	Jet Propulsion Laboratory
Vivienne Payne	Jet Propulsion Laboratory
Igor Polonski	Colorado State University
Mike Smyth	Jet Propulsion Laboratory
Rob Spurr	RT Solutions Inc.
David Thompson	Jet Propulsion Laboratory
Geoff Toon	Jet Propulsion Laboratory

Approved by:

_____

Version 3.0
December 7, 2017

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Document History:

Version	Revision	Date	Description/Comments
1.0	1		Initial version of OCO ATBD
1.0	2	4/18/2009	Final version of OCO ATBD
1.0	3	10/15/2010	First version of OCO-2 ATBD
1.0	4	11/10/2010	Final pre-external review version, minor fixes
2.0	0	4/17/2014	Prelaunch version for OCO-2
2.0	1	2/21/2015	Draft postlaunch version for OCO-2
2.0	2	3/30/2015	Accompanying V6 and V6R public data release
3.0	0	12/7/2017	Updated for Version 8 public data release

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

TABLE OF CONTENTS

1 Introduction 1
1.1 Purpose 1
1.2 Scope 1
2 Experiment Overview 2
2.1 Objectives 2
2.2 Measurement Approach 3
2.3 Instrument Characteristics 5
2.3.1 Observing Modes 7
2.3.2 Changes to Glint Mode Operations 9
2.3.3 Data Product Delivery 9
3 Algorithm Description 10
3.1 Algorithm Overview 10
3.2 Forward Model 15
3.2.1 Optical Properties 17
3.2.2 Solar Spectrum Model 32
3.2.3 Radiative Transfer 33
3.2.4 Instrument Model 37
3.3 State Vector 41
3.3.1 Basic Description of State Vector 41
3.3.2 The *a Priori* State Vector and Covariance Matrix 42
3.4 Jacobians 51
3.5 Inverse Method 52
3.5.1 Formulation and Implementation 52
3.5.2 Goodness of Spectral Fit 54
3.6 X_{CO2}, Characterization, and Error Analysis 55
3.6.1 Pressure Weighting Function 55
3.6.2 X_{CO2} 56
3.6.3 The Column Averaging Kernel a_{CO2} 56
3.6.4 Smoothing and Interference due to the State Vector $\tilde{a}_c$ 57
3.6.5 Correlation of X_{CO2} with non-CO_2 State Vector Elements $\tilde{\mathbf{s}}_1$ 58
3.6.6 Components of X_{CO2} Variance 58
4 Algorithm Specification 60
4.1 Data System Context 60
4.2 Level 2 Preprocessing 60
4.2.1 ABO2 Preprocessor 60
4.2.2 IMAP Preprocessor (and SIF Retrievals) 61
4.3 Level 2 Full Physics Inputs 61
4.3.1 Requirements on Inputs 62
4.3.2 Attributes of Input Products 62
4.4 Level 2 Full Physics Outputs 63

4.5 Description of Full Physics Major Code Sections 64
4.6 Lua Configuration 65
4.7 Level 2 Processing PGEs 66
4.7.1 ECMWF Resampler 66
4.7.2 L2 A-Band Preprocessing PGE 67
4.7.3 L2 IMAP-DOAS Preprocessing PGE 67
4.7.4 Sounding Selection PGE 67
4.7.5 L2 FP Algorithm 67
4.7.6 FP Diagnostic Product Generator 68
4.7.7 FP Standard Product Generator 68
4.7.8 Level 2 PGE Multiplicity Summary 68
4.8 Error Handling 69
4.9 Commercial off-the-Shelf Components 69
4.10 Quality Assessment and Recording 69
5 References 70
6 Acronyms 79

LIST OF FIGURES

Figure 2-1. The OCO-2 instrument showing the major optical components and optical path (right) and images of spectra recorded by the FPA in the three spectral channels (left). 5

Figure 2-2. (a) The illumination and readout scheme used for the OCO-2 FPAs. (b) Spatial layout of eight cross track footprints for nadir observations over Washington, D.C. 6

Figure 2-3. Nadir, glint, and target observations. (a) Nadir observations are acquired over the sunlit hemisphere at latitudes where the surface solar zenith angle is <85°. On all orbits except downlink orbits, as the Observatory passes over the northern terminator, it pitches up to point the instrument aperture at the Sun for solar radiometric calibrations. (b) Glint observations are made at latitudes on the sunlit hemisphere where the solar zenith angle of the glint spot is less than 75°. (c) For target observations, the spacecraft points the instrument at a stationary surface target as it flies over. A small-amplitude sinusoidal oscillation in the pitch axis is superimposed on the nominal pointing to scan the spectrometer slit across the target. 7

Figure 2-4. (a) The principal plane is a plane defined by the Sun, surface footprint, and instrument aperture. (b) The spacecraft azimuth changes during the orbit to maintain the alignment of the long axis of the spectrometer slits (which are roughly parallel with the axis of the solar panels) perpendicular to the principal plane [Crisp et al., 2007]. 9

Figure 3-1. Viewing geometry and a few possible optical paths traversed by solar photons that are recorded by the instrument. 11

Figure 3-2. Spectral dependence of the absorption coefficients of CO_2 at wavelengths within the strong CO_2 band near 2.08 μm are shown for pressures near 0.1, 100, and 600 hPa (mbar). 13

Figure 3-3. Flow diagram for the L2 retrieval algorithm. 15

Figure 3-4. Example spectral residuals from GOSAT retrievals in the O_2 A band, showing persistent systematic features. The red line shows residuals using the HITRAN 2008 positions and intensities, while the blue line shows the residuals using the Long et al. [2010] positions and intensities. The residuals are very similar, although the new parameters did bring the retrieved surface pressures more in line with ECMWF values. 20

Figure 3-5. Residuals for a 10-airmass atmospheric spectrum at 1.6 μm. The spectrum is a ground-based sounding recorded at 0.014 cm^{-1} resolution with the TCCON Park Falls FTS station, 22 December, 2004. The residuals in red are from the line mixing software and line list from Lamouroux et al. [2010], while the residuals in blue are for the present software and database from Benner et al. [2011], which combines a speed-dependent profile and tridiagonal line mixing. Figure from Thompson et al. [2012]. 21

Figure 3-6. Residuals for a 10-airmass atmospheric spectrum at 2.06 μm. See text associated with Figure 3-5. Figure from Thompson et al. [2012]. 22

Figure 3-7. Optical depth contributions of different molecules in the O_2 A band for a US standard atmosphere. 23

Figure 3-8. Optical depth contributions for different gases in the 1.6 μm CO_2 band for a tropical standard atmosphere. 23

Figure 3-9. Optical depth contributions from different molecules in the 2.06 μm CO_2 band for a tropical standard atmosphere. 24

Figure 3-10. Absorption coefficient calculation processing and sources. 25

Figure 3-11. Independent phase matrix elements for the cloud ice particles used in the L2 retrieval algorithm. Shown are both the exact phase functions, as well as the M=5000 delta-M. 28

Figure 3-12. Aerosol and cloud optical properties assumed in the L2 code. Left: Extinction Efficiency relative to that at 755 nm. Middle: Single Scattering Albedo. Right: Asymmetry Parameter. DU: Dust, SS: Sea Salt, BC: Black Carbon, OC: Organic Carbon, SO: Sulfate, WC: Water Cloud, IC: Ice Cloud. The three OCO-2 bands are demarcated by the dashed vertical lines. 31

Figure 3-13. Percent errors in OCO-simulated TOA radiances due to the LSI method for four different cases. Each case includes one or multiple clouds. The three OCO bands are shown in blue (O_2 A), green (weak CO_2), and red (strong CO_2). Taken from O'Dell [2010], Figure 8..... 36

Figure 3-14. Illustration of non-uniform sampling approach. Top: Training. From a full monochromatic spectrum (black diamonds) certain spectral points are selected for RT calculations (blue squares). The remaining points are calculated from linear interpolation of the selected points. Bottom: Interpolation errors in the non-selected points for different interpolation schemes. Linear interpolation is used because its errors are comparable to higher-order interpolation schemes..... 37

Figure 3-15. Geometry of the reference plane for polarization, i.e., the local meridian plane. This plane is formed by the local normal unit vector and the ray from the target FOV to the satellite. The direction of propagation of measured light is the latter. Two example polarizations are shown depicting light polarized both parallel (||) and perpendicular (⊥) to this plane. The polarization angle of the instrument is defined as the angle between the axis of accepted polarization and the reference plane for polarization. 40

Figure 3-16. *A priori* CO_2 correlation matrix (in colors), together with the 1-sigma errors for each pressure level. 43

Figure 3-17. Prior aerosol profiles uses the L2FP retrieval. 45

Figure 3-18. Global distribution of aerosol type on July. (Panel A) dominant type; (Panel B) second-most dominant type. 46

Figure 3-19. Global distribution of aerosol optical depth on July. (Panel A) dominant type; (Panel B) second-most dominant type; (Panel C) fractional contribution of the two dominant types to total AODs..... 46

Figure 3-20. Example of first two leading eigenvectors for current GOSAT retrievals in the O_2 A band (left) and strong CO_2 band (right). 50

Figure 4-1. Data flow through the L2 FP software components..... 65

Figure 4-2. L2 Processing overview. 66

LIST OF TABLES

Table 3-1. Inputs and outputs of the forward model. 17
Table 3-2. ABSCO v. 4.2 parameter sources. 25
Table 3-3. Density weightings and RH values chosen for the L2 algorithm. 29
Table 3-4. Description of state vector elements. 42
Table 3-5. Aerosol prior profile quantities and *a priori* uncertainties. 45
Table 3-6. Description of *outcome* values. 53
Table 3-7. Inverse method products recorded with each sounding. 56
Table 4-1. L1B Science product attributes. 62
Table 4-2. ECMWF product attributes. 62
Table 4-3. L2 Single Sounding product attributes. 63
Table 4-4. L2Dia product attributes. 63
Table 4-5. L2Std product attributes. 64
Table 4-6. Daily L2 PGE instances. 68

1 Introduction

1.1 Purpose

This algorithm theoretical basis document (ATBD) describes the algorithm used to retrieve the column-averaged CO_2 dry air mole fraction X_{CO2} and other quantities included in the Level 2 (L2) Product from the spectra collected by the Orbiting Carbon Observatory-2 (OCO-2). In particular, this document identifies sources of input data, which are required for retrievals; describes the physical theory and mathematical background underlying the use of this information in the retrievals; includes implementation details; and summarizes the assumptions and limitations of the adopted approach.

1.2 Scope

This document covers the algorithm theoretical basis for the parameters of the Full Physics (FP) Product that are to be routinely retrieved at the OCO-2 Science Data Operations System. Specialized products or parameters are not discussed. This document describes data release Version 8 (v8). Section 1 describes the purpose and scope of the document. Section 2 provides a brief overview of the mission and instrument. The processing concept and algorithm description are presented in Section 3. References for publications cited in the text are given in Section 5. Section 6 provides a list of acronyms and their definitions.

2 Experiment Overview

2.1 Objectives

The Orbiting Carbon Observatory is the first National Aeronautics and Space Administration (NASA) mission designed to collect space-based measurements of atmospheric carbon dioxide (CO_2) with the precision, resolution, and coverage needed to characterize the processes controlling its buildup in the atmosphere [Crisp et al., 2004; 2007]. After a launch mishap, which prevented the original OCO mission from reaching orbit, the Orbiting Carbon Observatory-2 (OCO-2) mission was formulated to meet the original OCO objectives.

Fossil fuel combustion and other human activities are now emitting almost 40 billion tons of CO_2 into the atmosphere every year. Atmospheric CO_2 measurements currently being collected by a global network of surface stations indicate that less than half of the CO_2 is accumulating in the atmosphere. The remainder is apparently being absorbed by CO_2 sinks in the ocean and the terrestrial biosphere [see Le Quéré et al., 2009; 2013]. While the existing surface greenhouse gas monitoring network has expanded continuously over the past 50 years and now provides the accuracy and coverage needed to quantify the abundance of this gas on global scales, it still lacks the spatial and temporal resolution and coverage needed to identify and quantify CO_2 sources and sinks on regional scales or to quantify emissions from discrete point sources.

One way to improve the coverage and resolution of these measurements is to collect spatially-resolved, global measurements of the column-averaged CO_2 dry air mole fraction X_{CO2} from space [Rayner and O'Brien, 2001]. Although natural processes and human emissions can change the atmospheric CO_2 mixing ratio by as much as 8% near the surface (>30 ppm out of the ~400 ppm background), the amplitude of these variations decreases rapidly with altitude, such that X_{CO2} variations rarely exceed 2% (8 ppm) on regional to global scales. East-west variations are typically no larger than 0.3 to 0.5% [Miller et al., 2007]. Because of this, modeling studies show that space-based measurements of X_{CO2} can substantially improve our understanding of surface fluxes only if they have the accuracy, precision, coverage, spatial resolution, and temporal sampling needed to describe X_{CO2} variations with amplitudes no larger than 0.3 to 0.5% (1 to 2 ppm) on scales ranging from <100 km over continents to ~1000 km over the ocean [Rayner and O'Brien, 2001].

Systematic biases with amplitudes larger than 0.3% on spatial scales of 100 to 1000 km will introduce spurious X_{CO2} gradients that would be indistinguishable from those produced by true CO_2 sources or sinks. Absolute X_{CO2} accuracies better than 0.3% on these scales are therefore essential for retrieving CO_2 fluxes. Truly global biases are less of a concern because they will not introduce spurious X_{CO2} gradients. However, such biases could compromise validation of the space-based measurements against other standards for atmospheric CO_2, such as the World Meteorological Organization (WMO) [see Zhao et al., 1997].

Space-based measurements of X_{CO2} are likely to make their most significant contributions to our understanding of the carbon cycle over the ocean and over tropical land masses because these regions are poorly sampled by the existing ground-based network. X_{CO2} estimates over the ocean are needed to quantify their large natural CO_2 sources and sinks and to facilitate the tracking of CO_2 emissions transported over the ocean by the prevailing winds. X_{CO2} measurements must also be collected over nearly the full range of latitudes on the sunlit hemisphere to avoid uncertainties introduced by the transport of air in and out of the field of regard.

To resolve CO_2 fluxes on spatial scales ranging from <100 to ~1000 km, data must be collected at higher resolution to discriminate natural sinks from nearby sources. A small sampling footprint also helps to ensure that some cloud-free soundings can be obtained even in partially cloudy regions, since the probability of measuring a cloud-free scene is inversely proportional to footprint size. A small sounding footprint is also needed to quantify CO_2 emissions from discrete point sources, such as individual power plants or cities, because the minimum detection limit (measured in kg of CO_2) associated with a given concentration change (e.g., a 1-ppm variation in X_{CO2}) is inversely proportional to the area of the footprint.

The natural processes responsible for the uptake and release of CO_2 are driven primarily by photosynthesis and respiration on land and by the solubility of CO_2 in the ocean. The efficiency of these natural processes varies on diurnal, seasonal, and interannual time scales. CO_2 emissions from human activities also vary on these time scales. Existing ground-based measurements indicate that while diurnal CO_2 variations in the vicinity of local sources and sinks can be large (>10 ppm), these variations are confined near the surface and rarely contribute to X_{CO2} variations larger than 0.3% [see Miller et al., 2007]. While these small differences would be difficult to detect from space, X_{CO2} should be estimated from measurements acquired at the same, fixed time of day everywhere on Earth to avoid introducing a spatially-varying diurnal bias. Global measurements are needed at semimonthly intervals over a complete annual cycle to identify changes in the natural and human contributions to atmospheric CO_2 over the seasonal cycle. More than one seasonal cycle must be observed to resolve the relative contributions of seasonal and interannual variability to the atmospheric CO_2 buildup.

2.2 Measurement Approach

OCO-2 carries and points a single instrument that incorporates three high-resolution imaging grating spectrometers that make coincident measurements of reflected sunlight in the short-wave-infrared (SWIR) CO_2 bands centered near 1.61 and 2.06 µm and in the near-infrared (NIR) molecular oxygen (O_2) A band at 0.765 µm [Crisp et al., 2007]. Simultaneous, co-boresighted measurements from these three spectral regions are combined to define a single "sounding." Each sounding is analyzed with remote-sensing retrieval algorithms to produce an estimate of X_{CO2} for the atmospheric path between the Sun, the reflecting surface, and the OCO-2 instrument. Because the dry air mole fraction of O_2 is well known and essentially constant, measurements of O_2 A band absorption provide direct constraints on the surface pressure and uncertainties in the atmospheric optical path length introduced by cloud and aerosol scattering and pointing errors. Measurements of absorption by the weak and strong CO_2 absorption bands near 1.61 and 2.06 µm, respectively, provide information about both the CO_2 column abundance and the wavelength-dependent scattering by aerosols along the same optical path.

The Observatory flies in formation with the 705-km Afternoon Constellation (A-Train). This 705-km altitude, 98.8-minute, Sun-synchronous orbit has a 98.2 degree inclination, a 1:36:30 PM mean ascending equator crossing time, and a 16-day (233 orbit) ground track repeat cycle. OCO-2 follows a ground track that crosses the equator 217.3 km east of the World Reference System-2 ground track that is followed by the NASA Aqua platform, such that it overflies the CloudSat radar and CALIPSO lidar ground footprints. The orbit's early afternoon mean local time is well-suited for acquiring observations of the absorption of reflected sunlight by CO_2 and O_2 because the Sun is high, maximizing the available signal. It also facilitates coordinated calibration and validation campaigns with other A-Train instruments, and synergistic use of OCO-2 data with that from other A-Train platforms.

The OCO-2 instrument must measure CO_2 and O_2 absorption with adequate precision to yield X_{CO2} estimates with a precision better than 0.3% on spatial scales smaller than 100 km over continents and 1000 km over the ocean across more than 90% of latitudes on the sunlit hemisphere of the Earth. To meet these requirements, the instrument must have a high sensitivity and a high signal-to-noise ratio (SNR) over a wide dynamic range [Crisp et al., 2007; Crisp, 2008]. A high spectral resolving power ($\lambda/\delta\lambda$ >20,000) is needed to resolve the CO_2 and O_2 lines from the adjacent continuum to maximize the sensitivity to small (<0.3%) variations in X_{CO2}. Measurements across the entire O_2 and CO_2 band are needed at high SNR because a 0.3% variation in X_{CO2} must be inferred from substantially smaller variations in O_2 and CO_2 absorption strength. The retrieval algorithm must then perform an optimal estimation fit to dozens of lines within each band to yield X_{CO2} retrievals with precisions near 0.3%. A wide dynamic range is needed because the contrast between line cores and the adjacent continuum can exceed 100:1, and because the signal level depends on the intensity of the sunlight reflected from the surface, which decreases with increasing solar zenith angle (latitude) and with decreasing surface reflectance.

The OCO-2 instrument collects eight soundings over its 0.8-degree-wide swath every 0.333 seconds, yielding surface footprints with downtrack dimensions <2.25 km and cross-track dimensions that vary from 0.1 to 1.3 km at nadir [Crisp et al., 2007]. The high spatial resolution facilitates the discrimination of natural sinks from nearby sources and enhance the coverage by increasing the probability of collecting some cloud-free soundings, even in partially cloudy conditions. The high sampling rate provides dozens to hundreds of samples over downtrack distances that are small compared to those that characterize the spatial variability of X_{CO2} over continents (<100 km) and ocean (1000 km), even when less than 10% of samples are sufficiently cloud-free to yield full-column X_{CO2} measurements. X_{CO2} estimates from these soundings can therefore be averaged to improve the precision. While the rapid downtrack sampling yields high spatial resolution along the orbit tracks; the east-west (left/right of ground track) resolution is largely determined by the distance between orbit tracks. The 98.8-minute orbit period yields ~14.56 orbits each day that are separated by ~24.7° of longitude. The orbit track spacing decreases to ~13° after two days, and to 1.5°after a full 16-day repeat cycle.

To ensure their accuracy, the space-based X_{CO2} estimates are validated through comparisons with near-simultaneous measurements of X_{CO2} acquired by ground-based Fourier Transform Spectrometers (FTS) in the Total Carbon Column Observing Network (TCCON) [Wunch et al., 2011]. This network currently includes over 20 stations distributed over a range of latitudes from Lauder, New Zealand (45S) to Eureka, Canada (80N), and it is continuing to add new facilities. To relate TCCON measurements to the WMO CO_2 standard, aircraft profiles have been measured over several stations, using the same in situ CO_2 measurement approaches used to define that standard [Wunch et al., 2010; Messerschmidt et al., 2011; Washenfelder et al., 2006; Deutscher et al., 2010]. OCO-2 targets a TCCON site as often as once each day, acquiring thousands of measurements as it flies overhead. These measurements are analyzed to identify biases smaller than 0.2% (1 ppm) at these sites. The space-based X_{CO2} estimates are further validated through comparisons with in situ CO_2 and surface pressure measurements from ground-based sites with the aid of data assimilation models to provide a more complete global assessment of measurement accuracy.

2.3 Instrument Characteristics

The three OCO-2 spectrometers use similar optical designs and are integrated into a common structure to improve system rigidity and thermal stability. They share a common housing and a common F/1.8 Cassegrain telescope. The light path is illustrated in Figure 2-1. Light entering the telescope is focused at a field stop and then re-collimated before entering a relay optics assembly. From there it is directed to one of the three spectrometers by a dichroic beam splitter and transmitted through a narrowband pre-disperser filter. The pre-disperser filter for each spectral range transmits light with wavelengths within ~±1% of the central wavelength of the CO_2 or O_2 band of interest and rejects the rest. The light is then refocused on the spectrometer slits by a reverse Newtonian telescope. Each spectrometer slit is ~3 mm long and ~25 μm wide. These long, narrow slits are aligned to produce co-boresighted fields of view that are ~0.0001 radians wide by ~0.0146 radians long. Because the diffraction gratings efficiently disperse light that is polarized in the direction parallel to the slit, a polarizer is included in front of the slit to reject the unwanted polarization before it enters the spectrometer, where it could contribute to the scattered light background.

Once the light enters a spectrometer slit, it is collimated by a two-element refractive collimator, dispersed by a reflective planar holographic diffraction grating, and then focused by a two-element camera lens onto a two-dimensional focal plane array (FPA) after traversing a second narrowband filter. The narrowband filter just above the FPA is cooled to ~180 K to reject thermal emission from the spectrometer optical bench.

The spectral range of each channel is wide enough to include the complete molecular absorption band as well as some nearby continuum to provide constraints on the optical properties of the surface and aerosols, as well as absorbing gases. To meet these requirements, the O_2 A band channel covers 0.758 to 0.772 μm with a resolving power of >17,000, while the 1.61 and 2.06 μm CO_2 channels cover 1.594 to 1.619 μm and 2.042 to 2.082 μm, respectively, with a resolving power of ~20,000.

The gratings disperse the light onto the FPAs in the direction orthogonal to the long dimension of the slits (Figure 2-2). The fields of view are resolved spatially along the slits. The FPAs are 1024 x 1024 pixel arrays with 18 μm x 18 μm pixels that have a 100% fill factor (i.e., there are no spatial or spectral gaps between the pixels). The slit is imaged on the FPA, which samples its full width at half-maximum with two to three pixels. The quantum efficiency of the FPAs is between 75 and 90%, and the read noise is <30 electrons/pixel/exposure. The FPA temperatures are maintained at <120 K by a pulse-tube cryocooler that is thermally coupled to an

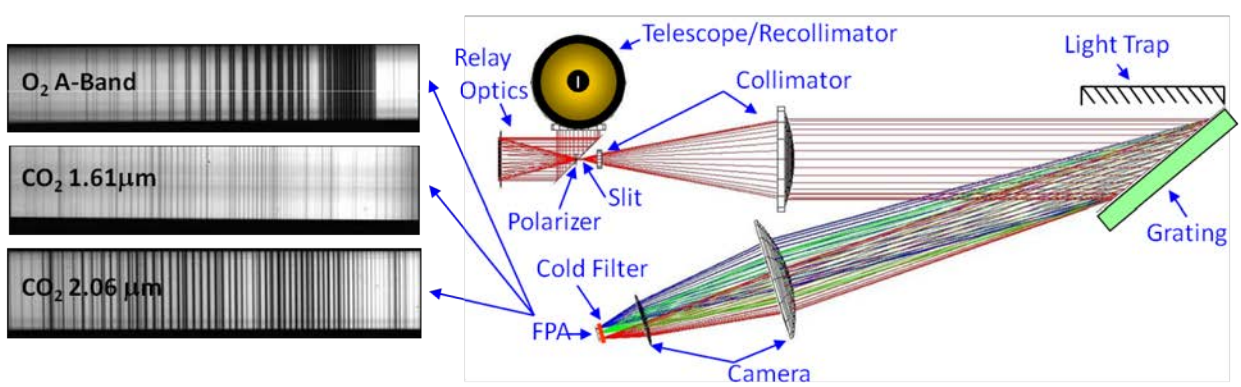

Figure 2-1. The OCO-2 instrument showing the major optical components and optical path (**right**) and images of spectra recorded by the FPA in the three spectral channels (**left**).

external radiator though variable conductance heat pipes. At this temperature, thermal noise from the FPAs is negligible during the short exposure time (0.333 seconds). The optical bench is maintained at 268 K by a thermal radiative shroud that is coupled to an external radiator by variable conductance heat pipes.

The spectrum produced by each channel is dispersed to illuminate all 1016 pixels in spectral dimension on each FPA. The length of the slit limits spatial field of view to only ~190 pixels in the spatial dimension (Figure 2-2a). OCO-2 soundings use an along-slit field of view as defined by ~160 of these 190 pixels. In normal science operations, the FPAs are continuously read out at 3 Hz using a "rolling readout" scheme that precludes gaps between adjacent frames. To reduce the downlink data rate and increase the SNR, 20 adjacent pixels in the FPA dimension parallel to the slit (i.e., the spatial direction in Figure 2-2a) are summed on board to produce up to eight spatially-averaged spectra (Figure 2-2b). The along-slit angular field of view of each of these spatially-averaged "super-pixels" is ~1.8 mrad (0.1° or ~1.3 km at nadir from a 705-km orbit). The angular width of the narrow dimension of the slit is only 0.14 mrad, but the telescope focus is purposely blurred to increase the effective full width at half-maximum of each slit to ~0.6 mrad. This blurring largely eliminates illumination variations across the slit, which could affect the instrument line shape, and reduces the impact of small errors in the boresight alignment among the three spectrometer slits.

In addition to the eight spatially-binned, 1016-element spectra, each spectrometer also returns four to 20 spectral samples without on-board spatial binning to provide the full along-slit spatial resolution. Each of these full-resolution "color stripes" covers a 220-pixel-wide region of the FPA. This includes all of the pixels illuminated by the slit (190 pixels) as well as a few pixels beyond the ends of the slit (Figure 2-2). These full-spatial-resolution color stripes are used to characterize the spatial variability within each of the spatially-summed super-pixels and to monitor the thermal emission and scattered light within the instrument.

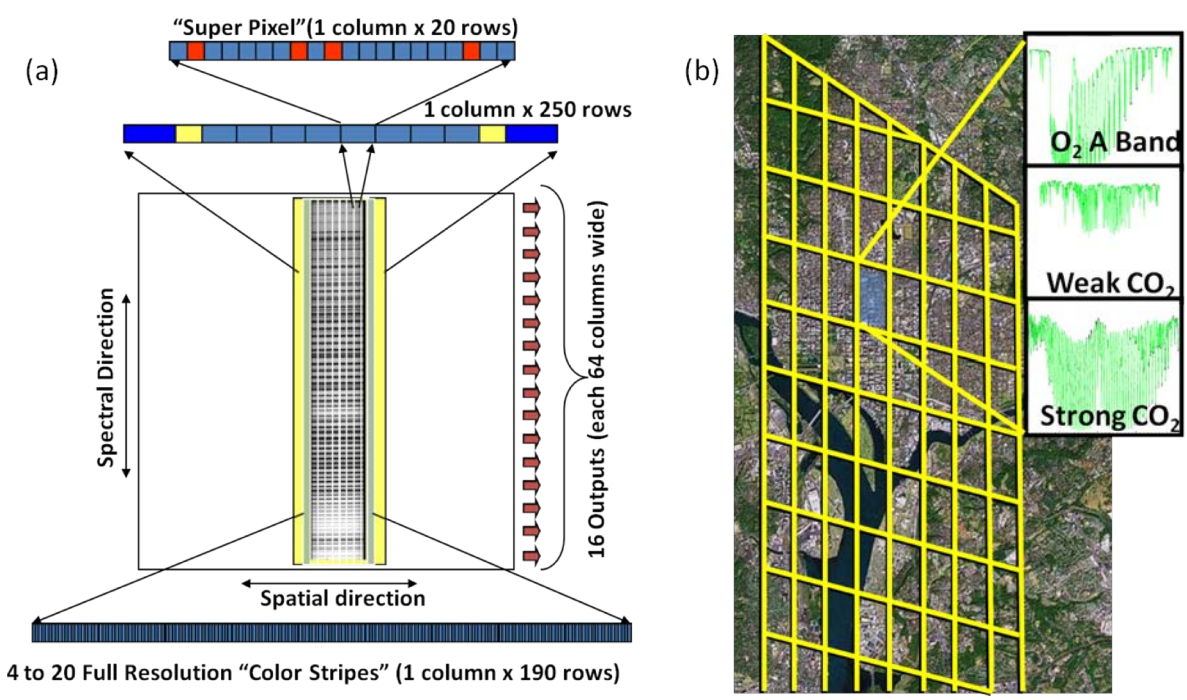

Figure 2-2. (**a**) The illumination and readout scheme used for the OCO-2 FPAs. (**b**) Spatial layout of eight cross track footprints for nadir observations over Washington, D.C.

2.3.1 Observing Modes

The spacecraft bus orients the instrument to collect science observations in nadir, glint, and target modes (Figure 2-3) [Crisp et al., 2007; Crisp 2008]. In nadir mode, the attitude control system (ACS) points the instrument aperture to the local nadir, so that data can be collected along the ground track just below the spacecraft. In glint mode, the ACS is programmed to point the instrument aperture between the local nadir the bright "glint" spot, where sunlight is specularly reflected from the Earth's surface.. In target mode, the ACS points the instrument aperture at specific stationary surface targets as the satellite flies overhead. To ensure that the target is not missed, the ACS superimposes a small-amplitude sinusoidal oscillation (±0.23° about the spacecraft y-axis) in the direction perpendicular to the long dimension of the spectrometer slit to scan the slits over a region centered on the nominal target location (Figure 2-3c). This target scan, combined with the instrument's 0.8°-wide field of view, creates a 0.46° x 0.8° viewing box around the target. The period of this sinusoidal oscillation will be less than 24 seconds, such that the slit is scanned over the target >20 times in a nine-minute target observation.

For nadir and glint observations, the ACS is required to point the instrument's field of view to within 0.15° of its intended target (1.9 km at nadir). For target observations, a pointing accuracy of 0.22° is required. The nominal plan is to switch from nadir to glint observations on alternate 16-day global ground-track repeat cycles so that the entire OCO-2 ground track is mapped in each mode every 32 days. Comparisons between nadir and glint observations will provide opportunities to identify and correct for biases introduced by the viewing geometry. This plan may be altered later in the mission to maximize the number of usable soundings. Target observations will be acquired over an OCO-2 validation site roughly once each day.

The same rate of data sampling is used for nadir, glint, and target observations. The instrument collects eight adjacent, spatially resolved samples every 0.333 seconds (24 samples per second). At this data collection rate, the Observatory collects ~400 soundings per degree of

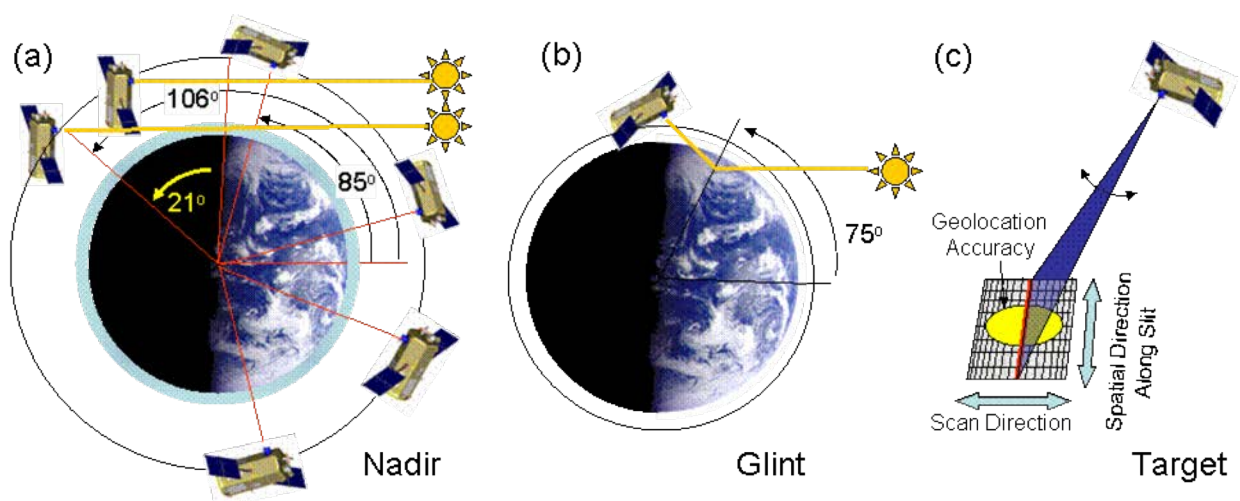

Figure 2-3. Nadir, glint, and target observations. (**a**) Nadir observations are acquired over the sunlit hemisphere at latitudes where the surface solar zenith angle is <85°. On all orbits except downlink orbits, as the Observatory passes over the northern terminator, it pitches up to point the instrument aperture at the Sun for solar radiometric calibrations. (**b**) Glint observations are made at latitudes on the sunlit hemisphere where the solar zenith angle of the glint spot is less than 75°. (**c**) For target observations, the spacecraft points the instrument at a stationary surface target as it flies over. A small-amplitude sinusoidal oscillation in the pitch axis is superimposed on the nominal pointing to scan the spectrometer slit across the target.

latitude as it travels from pole to pole, or ~14 million soundings over the sunlit hemisphere every 16-day ground repeat cycle. Clouds, aerosols, and other factors will reduce the number of soundings available for X_{CO2} retrievals, but the small sounding footprint ensures that some data will be sufficiently cloud free on regional scales at monthly intervals.

Nadir observations will be collected at all locations where the surface solar zenith angle is less than 85°. This mode provides the highest spatial resolution and is expected to return more usable soundings in regions that are partially cloudy or have significant surface topography. However, nadir observations are expected to have limited SNR over dark ocean or ice-covered surfaces at high latitudes. Glint observations are expected to provide much greater SNR over these surfaces. Glint soundings will be collected at all latitudes where the surface solar zenith angle is less than 75°. Target observations will be conducted over OCO-2 validation sites that are within 61° of the local spacecraft nadir along the orbit track and spacecraft viewing angles between 30° west of the ground track and 5° east of the ground track. When the target is near the ground track, a single pass can last up to nine minutes, providing 12000 to 24000 soundings in the vicinity of the target. This large number of soundings reduces the impact of random errors and provides opportunities to identify spatial variability in the X_{CO2} field near the target.

While the sunlight incident at the top of the Earth's atmosphere is unpolarized, both reflection from the surface and scattering by the atmosphere can polarize the radiation field measured by the OCO-2 instrument. These processes act primarily to reduce the intensity of the radiation that is polarized in the direction parallel to the principal plane, defined by the Sun, the surface target and the instrument aperture.. Polarization has a much smaller effect on the intensity polarized in the direction perpendicular to the principal plane. As noted above, the OCO-2 instrument is only sensitive to light polarized in the direction parallel to the orientation of the long axis of the spectrometer slits. The nadir and glint observing strategies have therefore been designed such that the long axis of the spectrometer slits (which are roughly parallel to which are roughly parallel to the axis of rotation of the solar arrays) remains oriented perpendicular to principal plane to maximize signal and minimize polarization errors (Figure 2-4a). As the Observatory ascends over the southern terminator, its long is pointed north-northwest along the orbit track and the spectrometer slits are oriented almost perpendicular to the orbit track (Figure 2-4b). In this orientation, the instrument collects data in a conventional push-broom fashion, where the sounding surface footprint is determined by the cross-track instantaneous field of view (0.1°) and the integration time (0.333 seconds). For nadir observations, this yields eight cross-track footprints along the spectrometer slit with dimensions of 1.29 km x 2.25 km.

As the Observatory proceeds northward along its orbit, it rotates counterclockwise about its nadir vector, such that the long axis of the spacecraft points westward and the long axis of the spectrometer slits are aligned with the orbit track just north of the sub-solar latitude. At this point, each spatially-resolved surface footprint is determined by the projected width of the slit (<0.03°) and the exposure time. For nadir observations at the sub-solar latitude, each of the footprints is ~0.4 km x 2.25 km and there is spatial overlap between footprints acquired in successive exposures by the spatial elements along the slit. The counter-clockwise azimuth rotation continues as the Observatory approaches the northern terminator, where the long axis of the spacecraft is pointing southwest, along the orbit track, and the spectrometer slit is once again oriented almost perpendicular to the orbit track.

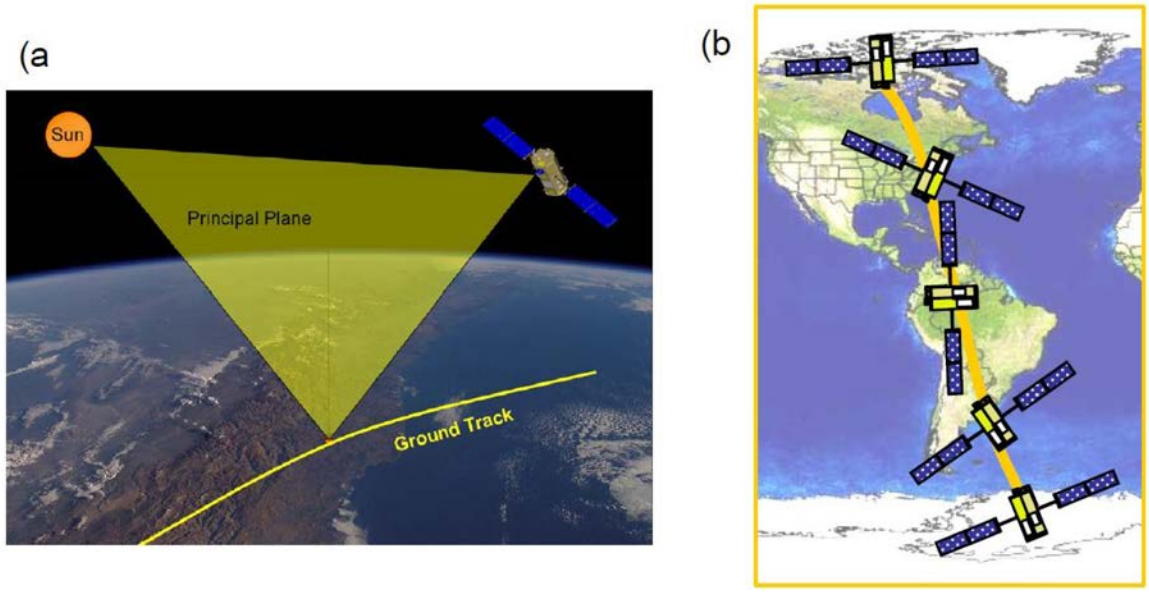

Figure 2-4. (**a**) The principal plane is a plane defined by the Sun, surface footprint, and instrument aperture. (**b**) The spacecraft azimuth changes during the orbit to maintain the alignment of the long axis of the spectrometer slits (which are roughly parallel with the axis of the solar panels) perpendicular to the principal plane [Crisp et al., 2007].

2.3.2 Changes to Glint Mode Operations

As designed, the intended orientation of the axis of accepted polarization was to be horizontal to the ground (i.e., perpendicular to the local meridian plane). However, a problem with the actual polarization plane of the instrument was discovered on orbit. A mitigation strategy was employed to yaw the spacecraft by 30° in glint mode, but not in nadir mode (as of the time of this writing). In target mode, because of the complicated slewing of the spacecraft, the polarization angle can vary widely. A more detailed description of the treatment of polarization and the changes to glint mode operations can be found in the OCO-2 L1B ATBD.

2.3.3 Data Product Delivery

Science and housekeeping data are transmitted to a NASA Near Earth Network station in Alaska once each day. The data are then transferred to the Earth Science Mission Operations center at the NASA Goddard Space Flight Center (GSFC), where the raw telemetry is converted to time-ordered raw radiance spectra (Level 0 Products). This product is then delivered to the OCO-2 Science Data Operations System at the NASA Jet Propulsion Laboratory (JPL), where full orbits are first processed to yield radiometrically-calibrated geolocated spectral radiances within the O_2 and CO_2 bands (Level 1 Products). The boresighted spectra for each coincident CO_2/O_2 sounding are then processed to estimate the column-averaged CO_2 dry air mole fraction, X_{CO2} (L2 Products). Other L2 data products retrieved from each sounding include the surface pressure, surface-weighted estimates of the column-averaged water vapor and atmospheric temperature, the vertical distribution and optical depth of optically-thin clouds and aerosols, the CO_2 column averaging kernels, and a number of diagnostic products.

3 Algorithm Description

3.1 Algorithm Overview

The primary purpose of the retrieval algorithm is to derive estimates of the column-averaged atmospheric CO_2 dry air mole fraction, X_{CO2}, and other L2 data products from the spectra returned by the OCO-2 mission. X_{CO2} is defined as the ratio of the column abundances of CO_2 to the column abundance of dry air:

$$X_{CO2} = \frac{\int_0^\infty N_{CO2}(z)dz}{\int_0^\infty N_{air}(z)dz} \qquad (3\text{-}1)$$

Here, $N_{CO2}(z)$ is the altitude (z)-dependent number density of CO_2 (e.g., number of CO_2 molecules per cubic meter) and $N_{air}(z)$ is the altitude-dependent number density of dry air. Because O_2 constitutes 0.20935 N_{air}, X_{CO2} can also be expressed as:

$$X_{CO2} = \frac{0.20935 \cdot \int_0^\infty N_{CO2}(z)dz}{\int_0^\infty N_{O2}(z)dz} \qquad (3\text{-}2)$$

The number densities of CO_2 and O_2 can be inferred from precise spectroscopic observations of reflected sunlight because the measured intensity of the sunlight at wavelengths where these gases absorb is proportional to the total number of molecules along the optical path S:

$$I(\lambda, \theta, \theta_0, \varphi - \varphi_0) = F_0(\lambda)\cos\theta_0 \cdot R(\lambda, \theta, \theta_0, \varphi - \varphi_0)\left\langle \exp\left\{-\int_0^s \sum_{m=1}^{M} [\sigma_m(\lambda, s)N_m(s)]ds\right\}\right\rangle \qquad (3\text{-}3)$$

Here, $I(\lambda, \theta, \theta_o, \varphi\text{-}\varphi_o)$ is the observed intensity at wavelength λ, θ and φ are the observation zenith and azimuth angles, and θ_o and φ_o are the corresponding solar zenith and azimuth angles. $F_o(\lambda)$ is the solar flux at the top of the atmosphere, $R(\lambda, \theta, \theta_o, \varphi\text{-}\varphi_o)$ is the reflectance of the surface, $\sigma_m(\lambda, S)$ and $N_m(S)$ are the absorption cross sections of the mth absorbing constituent (i.e., CO_2, O_2, or other absorber), and the integration is performed along an optical path S, which extends from the top of the atmosphere to the reflecting surface and back to the spacecraft. The brackets < > indicate an average over the range of possible optical paths that the photons could travel. A few of the possible optical paths are illustrated in Figure 3-1.

Because the argument of the exponential in Equation (3-3) depends linearly on the optical cross section per molecule, σ, and length of the optical path, S, as well as the number density of absorbing molecules, errors or uncertainties in either σ or S can introduce compensating errors and uncertainties in the retrieved number densities. For example, consider a sounding footprint that is partially obscured by a cloud, which scatters a fraction of the radiation back to the sensor before it has traversed the complete path from the top of the atmosphere to the surface and back to the spacecraft. If the scattering by this cloud is neglected in the retrieval algorithm, the reduced optical path could be misinterpreted as evidence for a reduced number density of absorbing molecules. Similarly, if multiple scattering between the surface and an airborne aerosol layer produces an optical path that is longer than the direct path between the Sun, surface, and spacecraft, and this multiple scattering is neglected, the enhanced optical path could

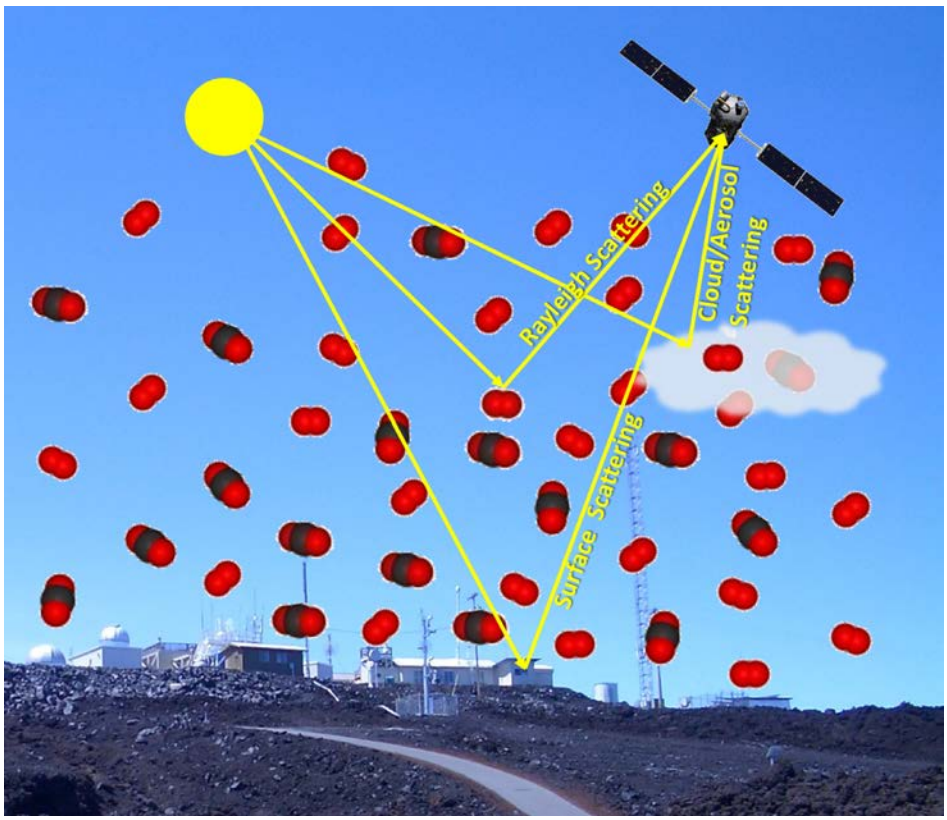

Figure 3-1. Viewing geometry and a few possible optical paths traversed by solar photons that are recorded by the instrument.

be misinterpreted as evidence for an increased number density of absorbing molecules. Accurately estimating the atmospheric optical path is therefore a primary focus of the retrieval algorithm.

If we divide the atmosphere into a series of discrete layers where the vertical distributions of the absorbing constituents are uniform or at least spatially-uncorrelated, we can rewrite the quantity:

$$\int_0^s \sum_{m=1}^{M} [\sigma_m(\lambda, s) N_m(s)] ds = \sum_{n=1}^{N} \sum_{m=1}^{M} \int_{s(n)}^{s(n+1)} [\sigma_m(\lambda, s) N_m(s)] ds \tag{3-4}$$

If we consider a linear path through a plane-parallel layer *n* at a local zenith angle of θ, the optical path between point *s(n)* and *s(n+1)* can be approximated as $ds = dz/\cos\theta$. We can then express the quantity included in the integrand in terms of the vertical optical depth in each layer:

$$\tau_{m,n}(\lambda, z) = \int_{s(n)}^{s(n+1)} [\sigma_m(\lambda, z) N_m(z)] dz \tag{3-5}$$

The total column-integrated vertical optical depth can then be given as:

$$\tau(\lambda) = \sum_{n=1}^{N} \sum_{m=1}^{M} \int_{s(n)}^{s(n+1)} [\sigma_m(\lambda, z) N_m(z)] dz \tag{3-6}$$

If we assume that the atmosphere is plane-parallel and ignore scattering by gases and airborne particles, the intensity observed by the spacecraft for a solar zenith angle θ_0 and observation angle θ is given by:

$$\begin{aligned} &I(\lambda, \theta, \theta_0, \varphi - \varphi_0) \\ &\quad = F_0(\lambda) \cos\theta_0 \cdot R(\lambda, \theta, \theta_0, \varphi - \varphi_0) \times \exp\left\{-\tau(\lambda) \frac{\cos\theta_0 + \cos\theta}{\cos\theta_0 \cdot \cos\theta}\right\} \end{aligned} \tag{3-7}$$

In the above discussion, the complexity associated with scattering by gases, airborne particles, and the surface was avoided by introducing the optical path averaging operator < > or was simply ignored (as in Equation (3-6)). More generally, the observed intensity will depend on both absorption and scattering, which can alter the optical path traversed by the solar photons. Fortunately, all of these optical processes can be accurately simulated by solving the equation of radiative transfer:

$$\mu \frac{d\,\mathbf{I}(\tau, \theta, \phi)}{d\tau} = \mathbf{I}(\tau, \theta, \phi) - \mathbf{J}(\tau, \theta, \phi) \tag{3-8}$$

where

$$\mathbf{J}(\tau, \theta, \phi) = \frac{\omega(\tau)}{4\pi} \int_{-1}^{1} \int_{0}^{2\pi} \mathbf{\Pi}(\tau, \theta, \theta', \phi - \phi')\, \mathbf{I}(\tau, \theta', \phi') d\phi' d\theta' + \mathbf{Q}(\tau, \mu, \phi) \tag{3-9}$$

Here, ω is the single scattering albedo, $\mathbf{\Pi}$ is the phase matrix for scattering, and $\mu = \cos\theta$. The first term in Equation (3-9) represents multiple scattering contributions. If we ignore thermal emission (an accurate assumption over the range of wavelengths observed by OCO-2), the inhomogeneous source term $\mathbf{Q}(\tau, \mu, \varphi_o)$ describing single scattering of the (attenuated) solar beam can be expressed as:

$$\mathbf{Q}(\tau, \mu, \varphi - \varphi_0) = \frac{\omega(\tau)}{4\pi} \mathbf{\Pi}(\tau, \mu - \mu_0, \phi - \phi_0)\, \mathbf{I}_0 \exp[-\tau / \mu_O] \tag{3-10}$$

Within the context of an atmospheric remote-sensing retrieval algorithm, the equation of radiative transfer is a component of the forward model.

Monitoring small changes in the abundances of gases such as O_2 and CO_2 using NIR absorption poses special challenges at NIR wavelengths because their absorption cross sections vary rapidly with wavelength (Figure 3-2).

The strongest absorption occurs near the centers of the narrow vibration-rotation lines. At other wavelengths, the absorption can be orders of magnitude weaker. Hence, for small number densities and/or optical path lengths, changes in these properties will produce their largest variations in the radiance field at wavelengths near the line centers. Eventually, however, as the

number densities or optical path lengths increase, virtually all of the sunlight is absorbed at these wavelengths (e.g., the line cores become saturated), and the sensitivity to further changes decreases dramatically. The region of peak sensitivity then moves further down the line wings.

Even with its relatively high spectral resolving power ($\lambda/\delta\lambda$ ~20,000), the OCO-2 instrument does not completely resolve the narrow cores of spectral lines. In spite of this, the spectral variability must be resolved by the forward model to accurately quantify the absorption within the spectral range sampled by the instrument line shape (ILS) function of each individual detector pixel. The most straightforward way to do this is to divide the spectral range of interest into a series of discrete spectral intervals that are narrow enough to completely resolve the spectral features contributed by atmospheric gases, the incident solar radiation field, airborne particles, or the reflecting surface. The equation of radiative transfer must then be evaluated within each spectral interval. Finally, the spectrum observed by each detector pixel is the product of the full-resolution simulated spectrum and the ILS function for the pixel.

The primary advantages of this direct spectrum-resolving approach are its conceptual simplicity and intrinsic accuracy. It can fully accommodate all of the physics of the atmospheric and surface processes that contribute to the absorption and scattering of solar radiation. Its primary drawback is its computational expense, because the equation of radiative transfer must be evaluated thousands of times to resolve the spectral structure within an NIR O_2 or CO_2 band. A number of methods (correlated-k, spectral mapping, etc.) have been developed to reduce the number of spectral intervals needed, but all of these methods employ simplifications and approximations that can introduce unacceptable errors in the computed radiance spectra.

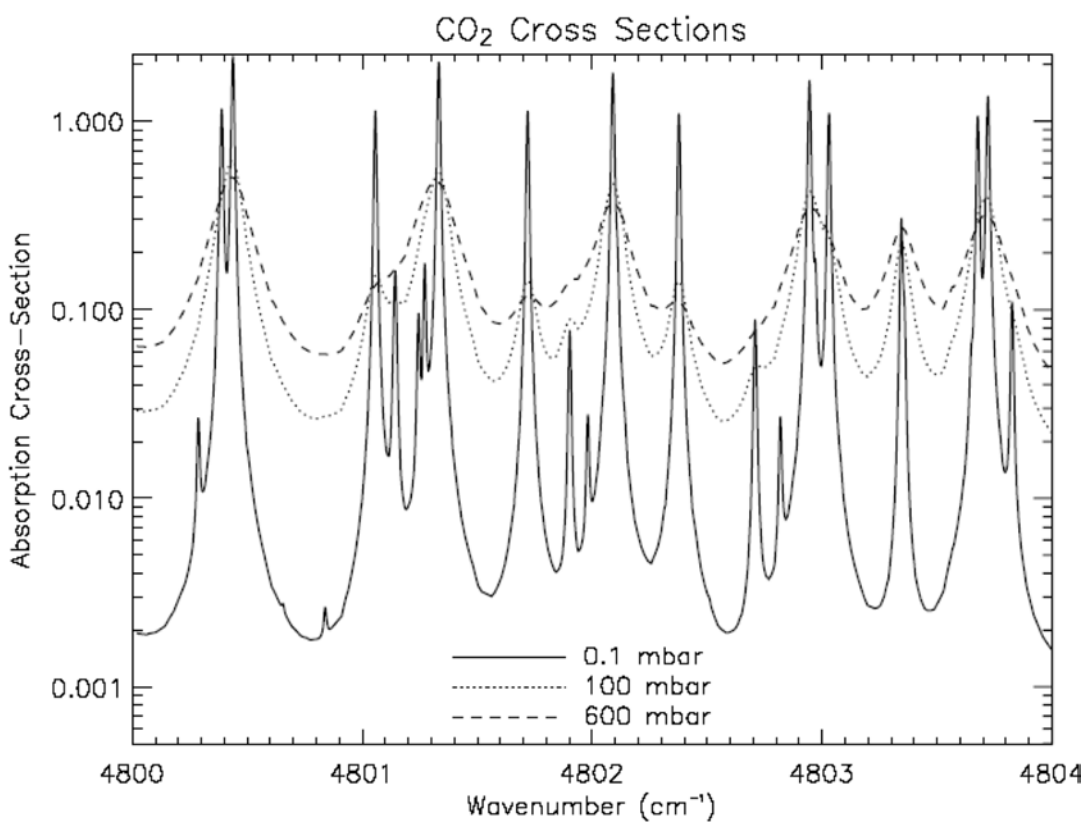

Figure 3-2. Spectral dependence of the absorption coefficients of CO_2 at wavelengths within the strong CO_2 band near 2.08 μm are shown for pressures near 0.1, 100, and 600 hPa (mbar).

To retrieve estimates of $N_{CO2}(z)$ or $N_{O2}(z)$ (or $\tau_{m,n}$) from observations of reflected solar radiation, the equation of transfer must be inverted. In general, this cannot be done analytically because of the complexity and nonlinearity of the equation of radiative transfer. However, these quantities can be derived using conventional nonlinear least-squares fitting techniques, where these quantities are treated as unknown coefficients and the equation of radiative transfer is the fitting function. In addition to the fitting function, these techniques require the first derivatives of the intensities with respect to any specific component of the state vector $x_{m,n}$:

$$\mathrm{K_{m,n}} = \frac{\partial \mathrm{I}}{\partial \mathrm{x_{m,n}}} \tag{3-11}$$

These radiance Jacobians are also generated by the forward model. Because the equation of radiative transfer is highly nonlinear with respect to most elements of the state vector, radiances and Jacobians must be derived for a surface/atmospheric state that is close to the actual state or the retrieval algorithm may not converge.

The primary components of the OCO-2 X_{CO2} retrieval process are summarized in Figure 3-3. Given an initial guess for the atmospheric and surface states and the observing geometry for a specific sounding, the forward model generates polarized synthetic radiance spectra and radiance Jacobians in the O_2 A band and in the two CO_2 bands observed by the OCO-2 instrument. This model first generates these synthetic spectra on a spectral grid that fully resolves the solar spectrum at the top of the atmosphere, the absorption and scattering cross sections for each atmospheric gas and airborne cloud and aerosol particle type included in the state vector, and the reflecting surface. The instrument model then convolves this full-resolution synthetic radiance spectrum using the ILS function and corrects for instrument polarization to simulate the spectrum recorded by the instrument. The forward model is described in Section 3.2.

The state structure includes atmospheric and surface properties that affect the spectrally-dependent radiances observed by the satellite. The surface-atmosphere state vector includes those components of the state structure that will be optimized by the retrieval algorithm to improve the fit between the observed and simulated radiances. The state vector is described in Section 3.3. Some of aspects of the instrument throughput (e.g., dispersion, instrument line shape) can also be included in the state vector. The methods used to generate radiance Jacobians are described in Section 3.4.

The inverse method is based on a Rodgers [2000]-type optimal estimation approach and has been described in Connor et al. [2008] and Boesch et al. [2006]. This model modifies the initial state vector to minimize differences between the observed and simulated spectra from each sounding. The inverse method is described in Section 3.5.

Once the atmospheric state yielding the best match to the observed spectrum has been found, the algorithm then determines X_{CO2}, errors in X_{CO2} from different sources (such as vertical smoothing, measurement noise, etc.), and the X_{CO2} column averaging kernel. This step is necessary because X_{CO2} is not itself an element of the state vector. Rather, it is determined from the profile of CO_2, which is part of the state vector. It is formally given by the total number of CO_2 molecules in the column divided by the total number of dry air molecules in the column. This step is labeled “Error Analysis” in Figure 3-3 and is described in Section 3.6.

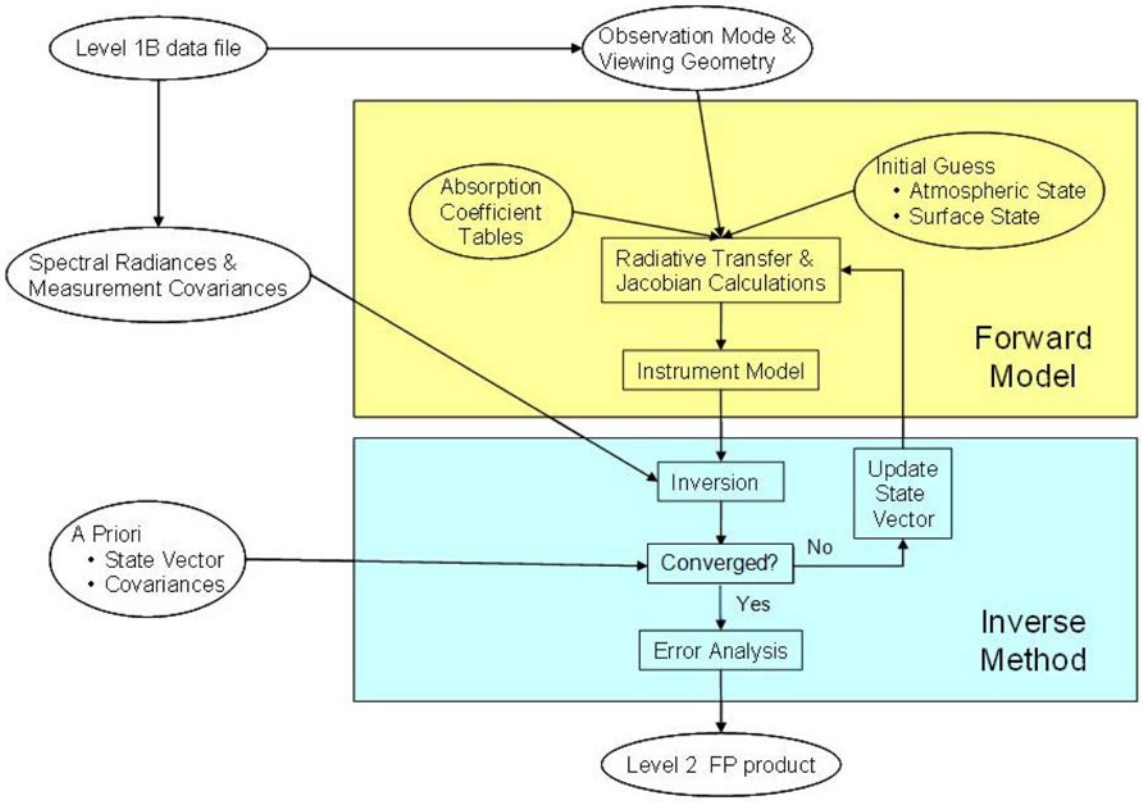

Figure 3-3. Flow diagram for the L2 retrieval algorithm.

Finally, given an estimate of X_{CO2}, a series of screening tests are performed to evaluate the quality of the retrieval. The criteria for a successful retrieval are currently under investigation and are subject to change prior to the final delivery of the data. Examples of success criteria that have been used in the past are:

- Measurement relative χ^2 in each band < threshold
- *a posteriori* X_{CO2} error < threshold
- Retrieved aerosol optical depth < threshold

A new method to assign warn levels to the data has been developed, as described in Mandrake et al. [2013]. This approach uses an objective method to bin the data into 20 groups and assign a warn level, with the lowest numbers indicating the best data quality, and increasing warn levels as data quality decreases. These warn levels will be included with the data as an alternative to the empirical data screening described above.

3.2 Forward Model

The forward model simulates solar spectra and radiance Jacobians for analysis of observations acquired by the OCO-2 spacecraft in nadir, glint, and target modes. A simple cartoon illustrating the atmospheric and surface optical processes that contribute to the absorption and scattering of solar radiation as it traverses the atmosphere and is reflected by the surface is shown in Figure 3-1. The forward model can also generate synthetic spectra and Jacobians for the analysis of

direct observations of the solar disk collected by the OCO-2 instrument during thermal vacuum testing or those collected by the fourier transform spectrometers in the TCCON.

The software components of the forward model include:

- Spectrally-dependent surface and atmosphere optical properties
  - Cloud and aerosol single scattering optical properties
  - Gas absorption and scattering cross-sections
  - Surface reflectance
- Spectrum effects
  - Solar model
  - Fluorescence
  - Instrument Doppler shift
- Atmosphere/surface radiative transfer model
- Instrument model
  - Spectral dispersion
  - ILS
  - Polarization response
  - Zero-level offset (optionally)

The methods used to generate the spectrally dependent gas and particle optical properties are described in Section 3.2.1. The solar spectrum at the top of the atmosphere is generated by a model described in Section 3.2.2. Synthetic spectra are generated by a radiative transfer algorithm based on a spectrum-resolving, multi-stream multiple scattering radiative transfer model that incorporates the linear discrete ordinate radiative transfer model (LIDORT) [Spurr et al., 2001; Spurr, 2002], a polarization correction based on a fast two-orders-of-scattering model [Natraj and Spurr, 2007], and a series of other techniques to improve its accuracy, speed, and range of validity. This model is described in Section 3.2.3. The instrument model simulates the spectral dispersion, ILS function, polarization dependence, and any other necessary corrections, such as zero-level offset. The instrument model is described in Section 3.2.4. The inputs and outputs of the forward model are enumerated in **Table 3-1**.

Several assumptions have been made for analyzing the OCO-2 data. First, it is assumed that measured radiances have been radiometrically calibrated. Second, it is assumed that thermal emission from the surface and atmosphere is negligible compared with reflected sunlight. This is certainly true in the O_2 A and weak CO_2 bands, and is a reasonably good assumption in the strong CO_2 band at 2.06 microns. Third, soundings that contain optically-thick clouds or aerosols can be rejected because they contain no useful information about CO_2 in the lower troposphere. It is assumed that inelastic scattering processes (e.g., Raman) are negligible at the wavelengths of interest to OCO-2. Further, the effects of airglow and absorption by the Chappuis bands of ozone in the O_2 A band are also currently ignored.

Table 3-1. Inputs and outputs of the forward model.

Inputs	Outputs
State structure: • Atmospheric state • Surface state • Instrument (ILS, dispersion)	Radiance spectrum
Observation geometry: zenith angles, altitude, etc.	Partial derivatives (Jacobians) of the radiance spectrum with respect to each of the state vector elements
Solar absorption and continuum tables	
Gas absorption and scattering cross sections	
Aerosol optical properties	
Surface bidirectional reflectance distribution function parameters	

3.2.1 Optical Properties

This section describes the atmospheric absorption and scattering calculation in the forward model. The radiative transfer model uses these values to predict the solar energy absorbed or scattered at each wavelength within an atmospheric layer at a given temperature and pressure. To improve computational efficiency, lookup tables for the absorption cross-sections for each absorbing gas are computed in advance. The OCO-2 forward model interpolates between table entries at runtime. For absorbing gases, the table generation process records representative cross sections k_ν (cm^2/molecule) at *every* applicable wavelength for the full range of applicable pressures (Pa), temperatures (K), frequencies (cm^{-1}), and broadeners such as water vapor. The resulting absorption coefficients (ABSCO) table is a four-dimensional lookup table for each absorbing gas in the forward model. For airborne cloud and aerosol particles, the wavelength single-scattering optical cross sections and scattering phase functions are tabulated for specific particle modes. The wavelength-dependent Rayleigh scattering cross sections and phase functions are calculated for each iteration for each retrieval.

We begin with a broad overview of the gas absorption calculations and physical models used to produce the cross sections, followed by a brief review of the laboratory experiments and spectroscopic parameters used to inform these calculations. We then describe software that implements the algorithm.

3.2.1.1 Gas Absorption Cross Section Calculation

The OCO-2 L2 algorithm uses precalculated lookup tables of absorption coefficients (ABSCO tables) for the calculation of gas absorption cross sections. The ABSCO tables contain molecular absorption cross sections over the range of relevant wavelengths, temperatures, pressures, and broadeners in units of cm^2/molecule. The forward model in the L2 algorithm computes atmospheric absorption at each relevant temperature, pressure, and wavelength using linear interpolation. New laboratory results and theoretical models have been incorporated into successive versions of ABSCO tables in order to obtain more accurate absorption coefficients. ABSCO tables are currently available for O_2, CO_2, H_2O, and CH_4, and are released intermittently.

For a single molecule, the absorption associated with a single transition k between two states i and j centered at wavenumber ν_k $(=\nu_{ij})$ can be written as the product of the corresponding line intensity $S_k(= S_{ij})$ and a line shape function $f(\nu, \nu_k)$. The monochromatic absorption coefficient at frequency ν_k of a given species is given by a summation over all relevant lines:

$$k_\nu^{\ x} = \sum_\nu S_k f(\nu, \nu_k) \tag{3-12}$$

The line shape function most commonly used within codes for the calculation of atmospheric absorption is the Voigt line shape. This shape is a convolution between the Lorentz (or pressure-broadening) function, which is applicable to higher pressures where the shape of the spectral lines is dominated by collisions between molecules, and the Doppler line shape, which describes a low-pressure regime in which the line shape is determined by the distribution of molecular velocities.

The Voigt line shape offers a number of advantages in terms of the availability of fast algorithms and the availability of line parameters. For example, the High Resolution Transmission (HITRAN) 2012 database [Rothman et al., 2013] provides Voigt line shape parameters for 47 different molecules, including over 470,000 lines for CO_2 alone. However, a number of studies have shown that the Voigt line shape is not sufficient for accurate retrievals of CO_2 and O_2 from near-IR solar spectra at the level required to meet OCO-2 science objectives (e.g., Long and Hodges [2012], Tran and Hartmann [2008], Predoi-Cross et al. [2008a], and Hartmann et al. [2009]). The OCO-2 retrievals require consideration of more subtle phenomena such as quantum-mechanical line mixing, speed dependence, and collision-induced absorption.

Line mixing occurs when the transition energies of a given molecular species overlap such that rotationally-inelastic collisions can transfer intensity from one line to another. Line mixing can change the population of states (i.e., the distribution of line strengths within a band) as well as the shape of individual spectral lines, overestimating the absorption in the wings of the band and underestimating the absorption in the peaks of the P and R branches [Tran et al., 2006]. Speed dependence of molecular collisions leads to deviation from the basic Voigt line shape (where all collisions are assumed to occur at the peak of the velocity distribution). Collision-induced absorption (CIA) occurs at high pressures, where interactions between molecules are strong enough and/or operate over long enough periods to violate the quantum rules that govern the "allowed" energy states for transitions of individual, isolated molecules, thereby introducing (or "turning on") additional transitions.

Line mixing, speed dependence, and CIA are all currently considered within the OCO-2 ABSCO model. However, line shape is still an active area of research and there is still significant room for improvement in the modeling of absorption coefficients in the OCO-2 bands. Further details on specific molecules are provided below.

3.2.1.2 Spectroscopy for OCO-2 Bands

In order to achieve improvements to the near-IR spectroscopy of CO_2 and O_2 in the three spectral channels at 0.76, 1.61, and 2.06 μm, members of the OCO-2 Science Team have formed international collaborations with leading experts to obtain high-quality laboratory measurements and theoretical modeling of high-resolution laboratory spectra of these two constituents.

In the last ten years, the state of the art has greatly advanced in measuring the spectral line positions, line intensities, and pressure-broadened line shapes and pressure shifts of CO_2 and O_2 absorption features (e.g., Miller et al. [2004a, b; 2005], Toth et al. [2006a, b; 2007a, b; 2008a, b],

Devi et al., [2007a, b], Predoi-Cross et al. [2009], and Sung et al. [2009] for CO_2, and Robichaud et al. [2008a, b], Robichaud et al. [2009], Predoi-Cross [2008a, b], and Tran and Hartmann [2008] for O_2). Improvements came from careful control of the experimental process applied for the laboratory measurements, use of reliable spectrometers (the McMath-Pierce Fourier transform spectrometer at Kitt Peak National Observatory, a Bruker 125 FTS at JPL, a Cavity ring-down spectrometer [CRDS] at National Institute of Standards and Technology [NIST] in Maryland, a Bomem FTS at the National Research Council in Canada, a Bruker 120 FTS in Giessen, Germany, and a Bruker FTS 66V at Laboratoire Interuniversitaire des Systèmes Atmospheriques in France), and sophisticated retrieval methods that evoked non-Voigt line shapes in a multispectrum fitting procedure (e.g., Devi et al. [2007a, b]).

3.2.1.2.1 Oxygen A Band

To include the effects of line mixing and CIA, we have adopted versions of the codes developed by Tran et al. [2006] (with updates described in Tran and Hartmann [2008]) for the main isotopologue of O_2. The original database supplied by Tran and Hartmann has been modified to utilize intensity and position parameters from CRDS measurements [Long et al., 2010]. Note that the Long et al. [2010] air-broadened half widths were *not* incorporated. The reason is that we found these parameters to be highly coupled with our line mixing strategy, so that any change from the values provided in Tran and Hartmann [2008] degraded performance. For consistency, we retain a Voigt line shape with widths and line mixing modelled exactly the same as in Tran and Hartmann [2008]. Those widths are based on a parameterization first found in Yang et al. [2005]. Pressure shifts are from Robichaud et al. [2008b] and Predoi-Cross et al. [2008] found in HITRAN 2008, with modifications to the R-branch by Brown et al. [2009]. Minor transitions and isotopologues are computed using measurements from Long et al. [2011]. (Isotopologue abundances from HITRAN 2008 are used.) The Galatry line shape [Galatry, 1961] was used for these minor lines, since narrowing parameters are available for lines in the Long et al. [2011] list. We supplemented them with a few additional Voigt-profile lines from the HITRAN 2008 database to cover all remaining weak O_2 lines in the entire 12850 – 12900 cm^{-1} window.

Broadening of O_2 by H_2O was measured in the 60 GHz Q-branch (50 – 119 GHz) and in the pure rotational S-branch (424 – 1850 GHz) by Drouin et al. [2014]. We assume no vibrational dependence and apply the Drouin et al. [2014]. H_2O broadening coefficients to the O_2 A band also.

Tables generated using the parameters and software detailed above have been validated using spectra from independent laboratory measurements and ground-based FTS measurements from the TCCON network and the Greenhouse Gases Observing Satellite (GOSAT). These comparisons show systematic residuals above the level of instrument noise, indicating significant room for improvement in modeling of O_2 absorption. Example residuals are shown in Figure 3-4.

The L2 algorithm contains the capability to apply empirical scaling factors to the ABSCO tables. For ACOS v3.4 GOSAT retrievals, an empirical scaling of 1.0125 had been applied to the ABSCO v4.2 O_2 tables in order to obtain agreement between the retrieved surface pressure and the European Centre for Medium-Range Weather Forecasting (ECMWF) surface pressure fields. (The ECMWF surface pressures should be a good estimate on average, particularly in data-rich regions.) In the current version of the OCO-2 algorithm, no empirical scaling is applied to the the ABSCO v4.2 O_2 tables.

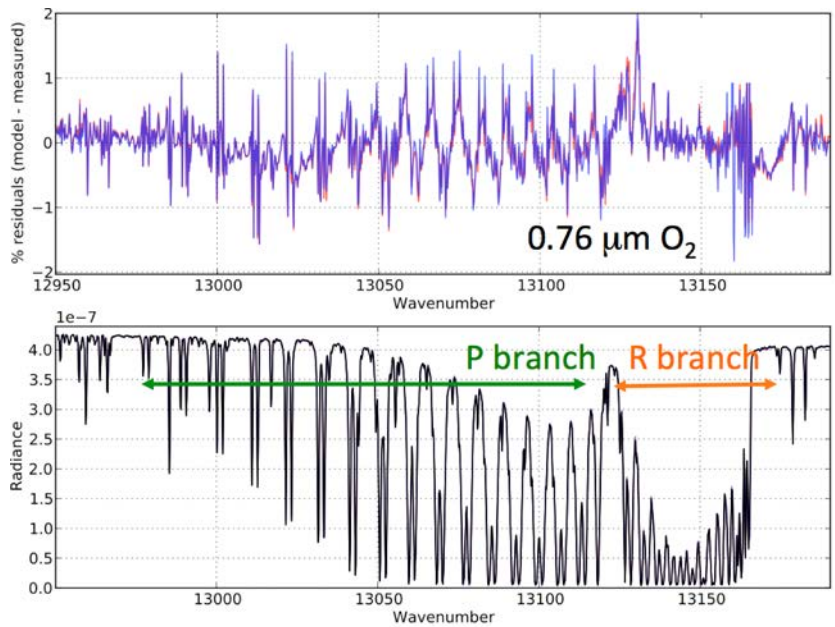

Figure 3-4. Example spectral residuals from GOSAT retrievals in the O_2 A band, showing persistent systematic features. The red line shows residuals using the HITRAN 2008 positions and intensities, while the blue line shows the residuals using the Long et al. [2010] positions and intensities. The residuals are very similar, although the new parameters did bring the retrieved surface pressures more in line with ECMWF values.

Work is underway to incorporate new laboratory measurements into a multispectrum fit for the O_2 A band. This approach will allow us to capitalize on the strengths of a range of different complementary laboratory measurement techniques. High-quality CRDS measurements (from NIST) offer extremely high precision and the potential of better constraints on the line shape (including line mixing and CIA effects). FTS measurements can ensure consistency of measurement across wide spectral ranges. New FTS measurements (taken at JPL) offer a wealth of potential new information on line shape, line mixing, and temperature dependence. Photoacoustic spectroscopy measurements taken at Caltech can offer information at high pressure, providing new insights into CIA.

3.2.1.2.2 Carbon Dioxide Bands

The CO_2 bands use line parameters and mixing models derived from studies by Benner et al. [2011] and Devi et al. [2007a]. Additional details on validation experiments for these databases, which are used in all ABSCO versions after v4.0, are found in Thompson et al. [2012]. The computation incorporates a speed-dependent line profile with tridiagonal line mixing. These databases were found to result in improved spectral residuals for atmospheric measurements compared to the databases generated using first-order line mixing, computed using software from Niro et al. [2005a, b]. Parameters for foreign broadening of CO_2 by H_2O come from studies by Sung et al. [2009]. This Sung et al. study treated the 4.3 μm band only. However, it is likely that these H_2O broadening parameters can be safely extrapolated to the other bands as well, and we favor this approach until additional direct laboratory data is available.

We modify the HITRAN 2008 isotopologue abundances slightly in the CO_2 band. These abundances are typical for measurements taken at sea level, but higher in the atmosphere CO_2 fractionation effects can modify the apparent abundances. Specifically, the 628 isotopologue abundance is increased by a factor of 1.04 to account for these apparent differences, and for the

intrinsic uncertainty in band strength as measured in the laboratory. This factor was derived empirically from tests using the TCCON network of ground-based, upward-looking FTS stations.

The possibility of a contribution of CO_2 CIA underlying the allowed ro-vibrational transitions in the 2.06 μm CO_2 band remains open. As discussed in Thompson et al. [2012], we found it necessary to incorporate an additional "continuum" absorption in the 2.06 μm CO_2 band, which takes the shape of two Gaussian distributions centered at 4853.5 and 4789 wavenumbers, with standard deviations of 10 and 8 wavenumbers and intensity scalings of $2.1*10^{-24}$ and $4.2*10^{-25}$, respectively (this is the intensity at 1 Atm, but absorption at other levels scales in proportion to pressure). These parameters were set to minimize error in retrievals with TCCON up-looking FTS spectra. Ideally, in the future we would hope that an improved line shape model and/or theoretical calculation of CIA would negate the need for empirically-based spectral features in the ABSCO tables.

The L2 algorithm provides the capability to apply scaling factors to the ABSCO tables. For ABSCO v. 4.2, scaling factors of 1.0038 and 0.9946 are applied to the 1.6 and 2.06 μm CO_2 bands respectively, in order to bring the retrievals to faster convergence and to provide improved agreement with validation measurements.

Example residuals in the OCO-2 CO_2 bands are shown in Figure 3-5 and Figure 3-6. The current ABSCO tables offer improvements over previous versions, but the resulting residuals do show evidence of persistent systematic residuals, particularly in the 2.06 μm band. Future work will involve utilization of new low temperature laboratory FTS measurements (from JPL) and CRDS measurements (from NIST and Caltech) in the multispectrum fits for CO_2, in order to provide improved constraints on temperature dependence and line shape.

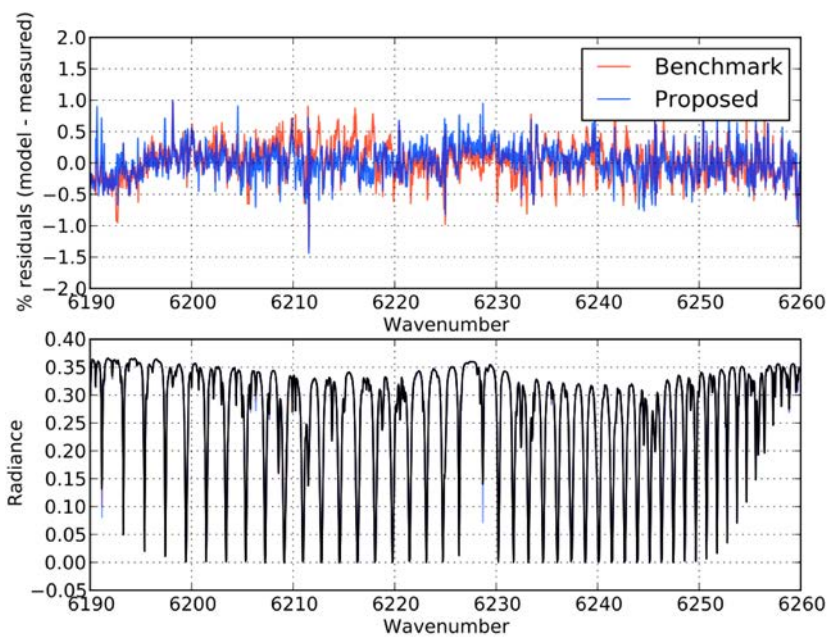

Figure 3-5. Residuals for a 10-airmass atmospheric spectrum at 1.6 μm. The spectrum is a ground-based sounding recorded at 0.014 cm^{-1} resolution with the TCCON Park Falls FTS station, 22 December, 2004. The residuals in red are from the line mixing software and line list from Lamouroux et al. [2010], while the residuals in blue are for the present software and database from Benner et al. [2011], which combines a speed-dependent profile and tridiagonal line mixing. Figure from Thompson et al. [2012].

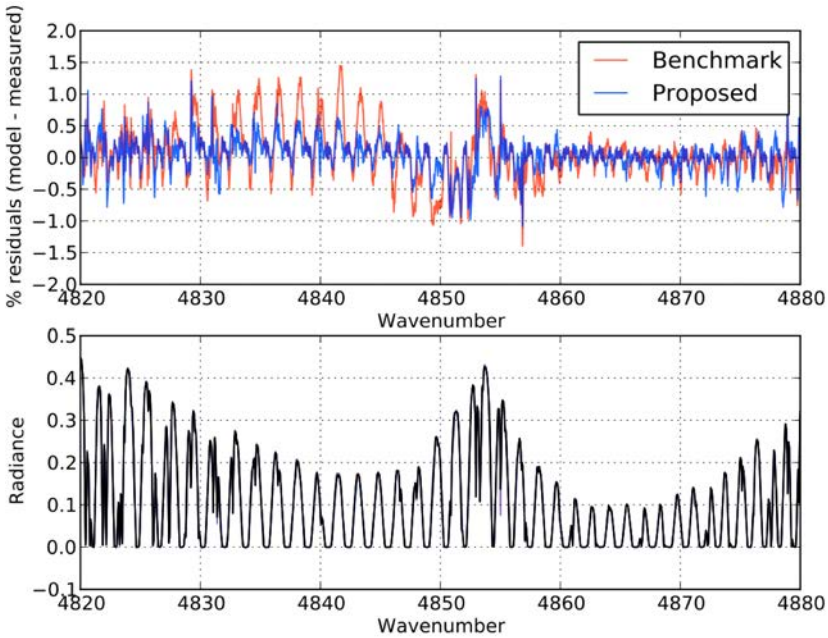

Figure 3-6. Residuals for a 10-airmass atmospheric spectrum at 2.06 µm. See text associated with Figure 3-5. Figure from Thompson et al. [2012].

3.2.1.2.3 Other Absorbing Gases

Several gases besides CO_2 and O_2 absorb within the spectral ranges used by OCO-2. Optical depth contributions for these gases (calculated using the Line by Line Radiative Transfer Model available at http://rtweb.aer.com) are shown in Figure 3-7, Figure 3-8, and Figure 3-9.

The most important gas is water vapor (H_2O), which produces measureable absorption in all three bands. Line parameters for H_2O transitions in the OCO-2 channels are currently taken from a custom list supplied by Iouli Gordon (Smithsonian Astrophysical Observatory). This custom list is similar to the list that went into the HITRAN 2012 compilation [Rothman et al., 2013]. The H_2O tables will be updated in future ABSCO versions to maintain consistency with the latest HITRAN compilation. The H_2O continuum model used in the OCO-2 ABSCO tables is a custom version of the MT CKD model [Mlawer et al., 2012], with updates as described in a 2014 OCO-2 Science Team Meeting presentation by Mlawer et al.

The second most important interferent is methane (CH_4). Absorption coefficients for CH_4 have been calculated using the HITRAN 2008 database [Rothman et al., 2009]. Again, CH_4 tables will be updated in future ABSCO versions to maintain consistency with the latest HITRAN compilation. To date, the CH_4 absorption coefficients have not been utilized within the L2 algorithm, but the bandpasses selected avoid the stronger methane lines. The ozone (O_3) Chappuis-Wulf bands produce weak, quasi-continuum absorption within the O_2 A band (2 to $4*10^{-22}$ cm^2/molecule). O_3 is not currently included in the OCO-2 forward model. The absorption by the other molecules shown in Figure 3-7, Figure 3-8, and Figure 3-9 is too small to produce measureable effects in the X_{CO2} retrieval, and these molecules are currently omitted from the algorithm.

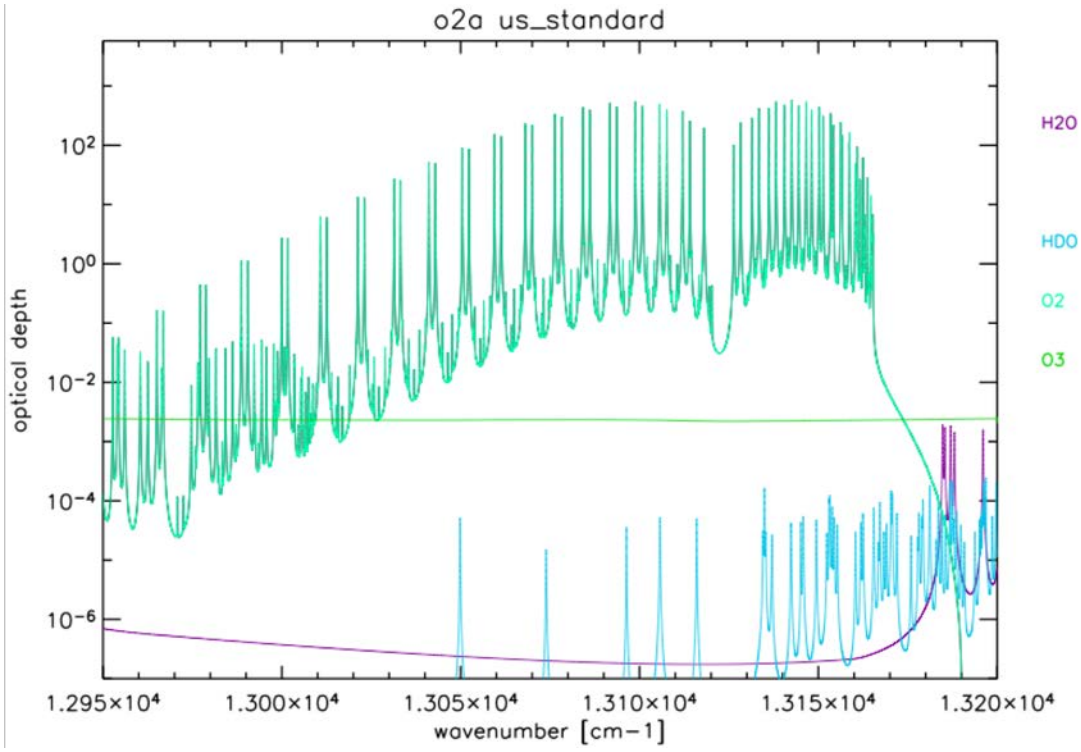

Figure 3-7. Optical depth contributions of different molecules in the O_2 A band for a US standard atmosphere.

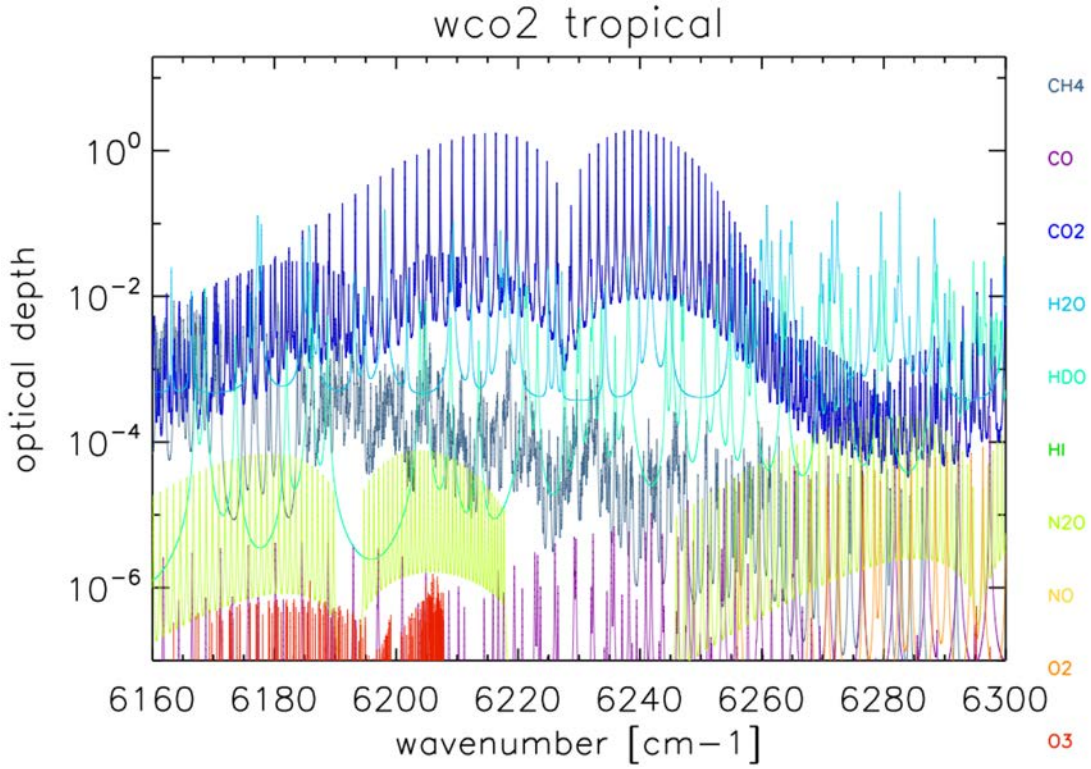

Figure 3-8. Optical depth contributions for different gases in the 1.6 μm CO_2 band for a tropical standard atmosphere.

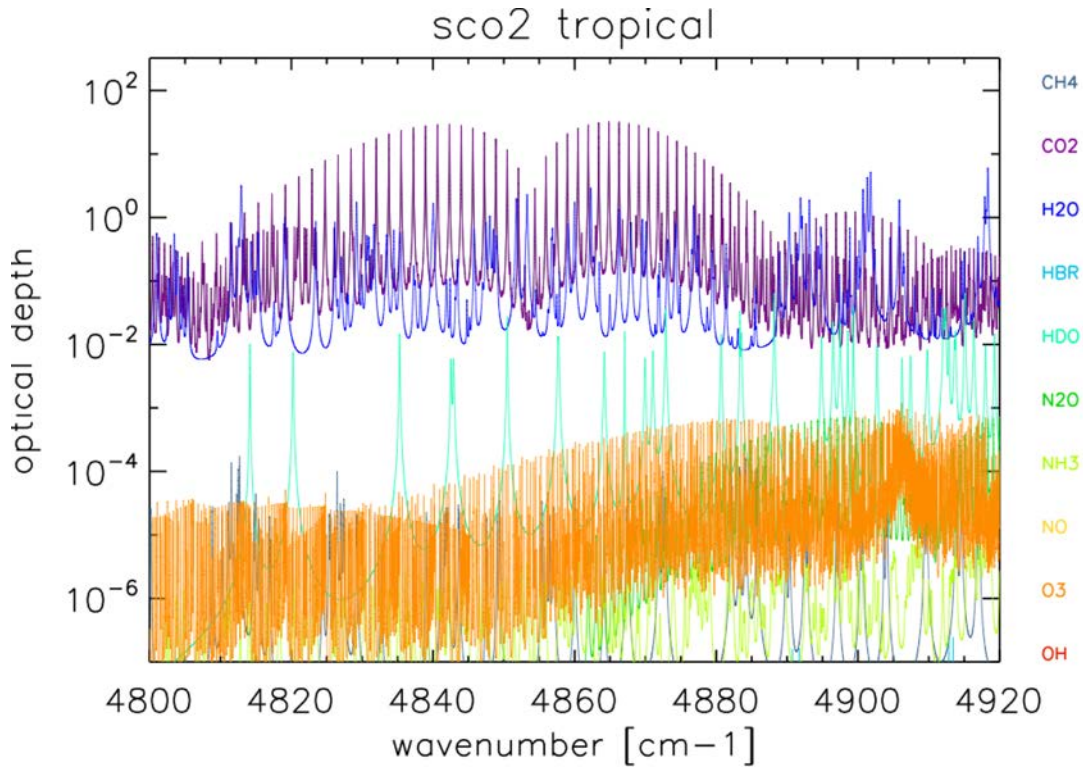

Figure 3-9. Optical depth contributions from different molecules in the 2.06 µm CO_2 band for a tropical standard atmosphere.

3.2.1.3 Software Implementation

Due to the modular nature of the ABSCO calculation, different bands rely on different codebases to compute the absorption cross sections. In summary, the LABFIT codebase from Benner et al. [2011] provides CO_2 cross sections. This implements databases from Benner et al. [2011] and Devi et al. [2007a] for 2.0 µm and 1.6 µm CO_2, respectively. The procedure and modeling assumptions of this approach are detailed at greater length in Thompson et al. [2012]. The line mixing codebase from Tran and Hartmann [2008] computes the major isotope of O_2. We use the ABSCOI codebase, a routine adapted from the GFIT software package in use by TCCON [Wunch et al., 2011], for the minor isotopes of O_2 and all CH_4 and H_2O lines. A flowchart showing the principal line list sources and the computation strategies for each line appears below in Figure 3-10.

The table format is a four-dimensional (4D) array, allowing cross sections to be parameterized by temperature, pressure, and the volume mixing ratio of one other atmospheric gas (H_2O). All tables are provided in the 4D format, although the calculations only incorporate broadening by H_2O of O_2 and CO_2, and not for CH_4.

Table 3-2 below gives more detailed information on several key parameter sources.

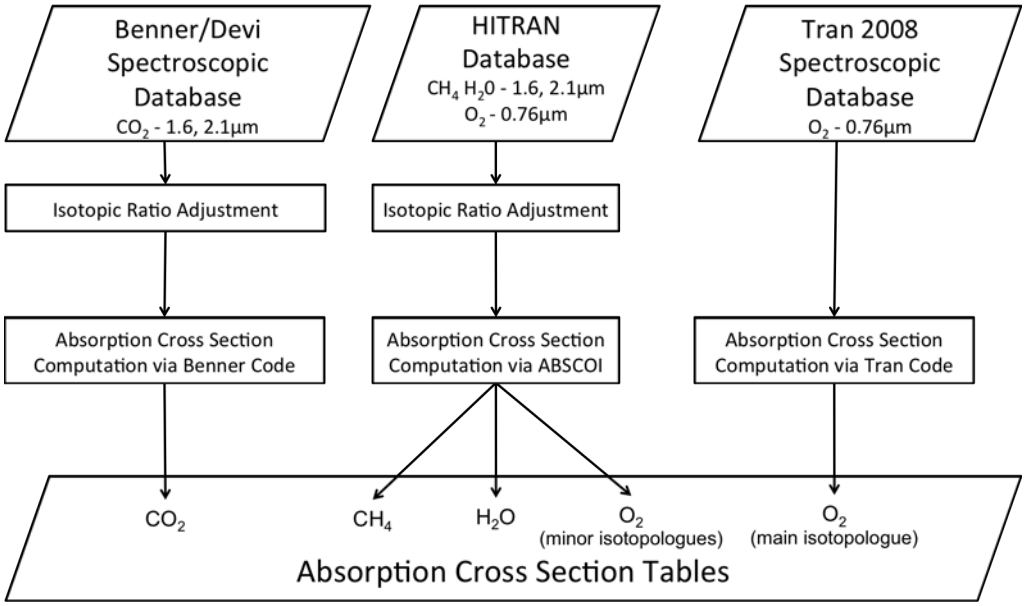

Figure 3-10. Absorption coefficient calculation processing and sources.

Table 3-2. ABSCO v. 4.2 parameter sources.

	0.76 µm O_2	**1.61 µm CO_2**	**2.06 µm CO_2**	**H_2O**
Spectral range	12745–13245 cm^{-1}	4700–6500 cm^{-1}	4700–6500 cm^{-1}	12745–13245 cm^{-1} 4700–6500 cm^{-1}
Spectral resolution	0.01 cm^{-1} or 0.002 cm^{-1}	0.01 cm^{-1} or 0.002 cm^{-1}	0.01 cm^{-1} or 0.002 cm^{-1}	0.01 cm^{-1} or 0.002 cm^{-1}
Position	Long (2010) Long (2011)	Devi (2007)	Benner/Devi (2011)	Gordon (2012) Rothman (2010)
Intensities	Long (2010) Long (2011)	Devi (2007)	Benner/Devi (2011)	Gordon (2012) Rothman (2010)
Air-widths	Tran (2008)	Predoi-Cross (2009)	Benner/Devi (2011)	Gordon (2012) Rothman (2010)
Air-shifts	Brown (2009) Robichaud (2008a) Predoi-Cross (2008)	Devi (2007b)	Benner/Devi (2011)	Gordon (2012) Rothman (2010)
Temp. dep.	Brown (2000)	Predoi-Cross (2009)	Benner/Devi (2011)	Gordon (2012) Rothman (2010)
Line shapes	Voigt/Galatry	Speed-dependent Voigt	Speed-dependent Voigt	Voigt
Isotopologue abundance	Rothman (2009)	Rothman (2009)	Rothman (2009)*	Rothman (2009)
H_2O broadening	Drouin (2014)	Sung (2009)	Sung (2009)	N/A
Air-line mixing	Tran (2008)	Devi (2007)	Benner/Devi (2011)	N/A
Temp. dep.	Tran (2008)	N/A	N/A	N/A
Speed dep.	N/A	Devi (2007)	Benner/Devi (2011)	N/A
Continuum	CIA via Tran (2008)	N/A	Empirical, gaussian*	Mlawer (2012)*

**See text for further details.*

3.2.1.4 Calculation of Gas Absorption Optical Depths in each Atmospheric Layer

Once we have the lookup tables of cross section on pressure P, temperature T, and wavelength λ, we must use the actual profiles of P, T, and gas concentration to calculate optical depth for each layer and gaseous species.

Let us suppose we have a thick layer that goes from a lower pressure P_{bot} to an upper pressure P_{top} and we wish to calculate the optical depth of this layer for a certain species of gas. Let the dry-air volume-mixing ratio of this gas be q_{gas}, which is defined as the number of moles of the gas per mole of dry air. Further, let us call the absorption cross section of the gas (in $m^2\ mol^{-1}$) $k(P, T)$ at a particular wavenumber. Because the temperature T is an intrinsic function of the pressure P for the given profile, we write the gas absorption coefficient for this specific layer as $k(P)$.

In an infinitesimal layer, it can be shown that the number of molecules of *dry air* (against which the volume mixing ratio of any gas is defined) can be written:

$$N_{dry}\, dz = -\frac{dP}{gM_{dry}}(1-q) \tag{3-13}$$

where dZ is the layer width in meters, N_{dry} is the number of moles of dry air per cubic meter, g is the local gravitation acceleration, q is the local specific humidity, and M_{dry} is the molar mass of dry air in $kg\ mol^{-1}$. Thus, the total optical depth of a thick layer can be written:

$$\tau_{gas} = \int_{P_{top}}^{P_{bot}} q_{gas} \frac{k(P)[1-q(P)]}{g(P)M_{dry}} dP \tag{3-14}$$

Note that we have expressly allowed g, q, q_{gas}, and k to vary as a function of pressure across this layer. Unfortunately, we only know each quantity at the boundary of the layer in question. Therefore, we must make assumptions about how each varies across the layer in order to evaluate the integral. Initially, each quantity was simply evaluated at the center of the layer assuming a linear variation; however, this approach can cause small biases in the total optical depth of a given gas and can lead to corresponding biases in retrieved X_{CO2}. This is mostly due to layers at the top of the atmosphere that span close to two orders of magnitude in pressure (say from 1 to 100 mbar)—the variation across this layer in k is so great (and quite nonlinear) that we cannot simply ignore it.

A straightforward approach that largely eliminates this bias is to divide each layer into a number of sublayers. Then the optical depth for each sublayer is evaluated according to the approximation:

$$\tau_{sub} = \frac{q_{gas,cen}(1-q_{cen})}{g_{cen}M_{dry}} \cdot K(P_{cen}, T_{cen}) \cdot \Delta P \tag{3-15}$$

where the *cen* subscript indicates a quantity to be evaluated at the center of the sublayer; all such quantities are evaluated using linear interpolation in pressure.

For each sublayer, $k(P_{cen}, T_{cen})$ is evaluated using 2D linear interpolation in both pressure and temperature. The optical depth of the full layer is then simply the sum of the optical depths of its component sublayers.

The L2 FP code subdivides each model layer into 10 sublayers of equal pressure width, plus adding additional sublayers at each of the ECMWF pressure levels. The 10 sublayers captures

the variation of the ABSCO table with pressure, while using the ECMWF pressure levels captures the full temperature profile (which is specified on the ECMWF pressure levels). Once the sublayer optical depths are evaluated, they can be summed according to Simpson's rule to obtain the total gas optical depth of each layer (and for each species of gas). This algorithm for integration was compared to a full (and much slower) integration using the Gnu Scientific Library, giving a maximum difference of 0.6%.

Currently, the L2 code makes this calculation for O_2 in the O_2 A band and for both CO_2 and H_2O in the weak and strong CO_2 bands. Abundances of the minor isotopologues of each are a fixed fraction of the total dry-air volume-mixing ratio, set inside the spectroscopic lookup tables.

3.2.1.5 Rayleigh Scattering Cross Sections

Rayleigh scattering is accounted for in all OCO bands, though it is most important in the O_2 A band. The Rayleigh scattering cross section is calculated using the following classic model:

$$\sigma_R = \frac{24\pi^3}{N_s^2\lambda^4}\frac{(n_s^2-1)^2}{(n_s^2+2)^2}\frac{6+3\rho}{6-7\rho} \tag{3-16}$$

where λ is the wavelength, ρ is the depolarization factor, N_s is the number density of air at some pressure and temperature, and n_s is the index of refraction of standard air *at the same pressure and temperature,* and ρ is the depolarization factor, which describes the effect of molecular anisotropy. According to Lorentz-Lorenz theory, the product of the two terms involving these two variables is independent of pressure and temperature. We take $N_s = 2.687 \bullet 10^{19}$ molecules/cm^3 at 273.15 K and 1013.25 mbar. The prefactor $24\pi^3/N_s^2$ may be calculated to be $1.031 \bullet 10^{-20}$ m^2 µm^4 per molecule, such that specifying λ in µm will result in the cross section having units of m^2/molecule.

The index of refraction for air at standard pressure and 273.15 K is taken from Allen [1964], using the following parameterization:

$$n_s = 1 + a * (1 + b\lambda^2) \tag{3-17}$$

where λ is in microns, a = $2.871*10^{-4}$, and b = $5.67*10^{-3}$. The depolarization factor is taken to be a constant over all the OCO bands, with a value of 0.0279 as suggested by Young [1980]. The Rayleigh phase matrix including polarization is a standard formulation [e.g., Hansen and Travis, 1974].

In order to calculate the Rayleigh optical depth τ_R for a given layer, it is necessary to multiply the cross section by the total number of air molecules per m^2 in the layer. For a given layer with pressure width ΔP given in Pa and mean gravity g in m/s^2, this is given approximately by:

$$N_{tot}\Delta z = \frac{N_A \Delta P}{g M_{dry}}\left(1 + \frac{1-\varepsilon}{\varepsilon} q\right) \tag{3-18}$$

where $N_A = 6.02214*10^{23}$ is Avogadro's number, M_{dry} = 28.96e-3 is the molar mass of dry air in kg/mole, and ε is the ratio of the molar mass of dry air to that of water vapor ($\varepsilon \approx 0.622$). Note that this represents the total number of air molecules in a layer, and therefore includes both the wet and dry contributions.

3.2.1.6 Cloud and Aerosol Optical Properties

The single-scattering optical properties (cross sections, single-scattering albedos, and scattering phase matrices) of cloud and aerosol particles are precomputed and tabulated. Currently, the OCO-2 retrieval algorithm includes five different types of scattering particles for each sounding. The first two are liquid water and ice clouds, described in the next paragraph. Another two are the two dominant aerosol types, set a priori to live in the lower troposphere (see section 3.3.2.3 for details), selected from the five available in the MERRA model. The last type is a sulfate aerosol that is confined to live in the upper atmosphere (section 3.3.2.3), and uses the same optical properties as the tropospheric sulfate aerosol type.

The wavelength-dependent optical properties for liquid water clouds were calculated from Mie theory assuming a Gamma particle size distribution and an effective radius of 8 μm. The optical properties of ice clouds with a single effective diameter of 70 μm were obtained from a database of ice crystal optical properties, compiled for Moderate-Resolution Imaging Spectroradiometer (MODIS) Collection-6 [Baum et al., 2014] . This database makes use of severely roughened ice particles and has more realistic phase functions that that from Collection-5, which was used in previous ACOS retrieval versions. This database also has the advantage that it was calculated at high wavelength resolution, and therefore did not have to interpolate across important spectral features, as occurred in the previous version. Figure 3-11 shows the six independent phase matrix elements of the 70 μm severely roughened ice crystals at a wavelength of 765 nm. To save on processing time, we employ delta-M truncation [Wiscombe, 1977] with 5000 moments. As shown in Figure 3-11, there is no visible difference between the exact phase function and the truncated one.

The five aerosol types are based on the aerosol types used in the Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis system: dust, smoke, sea salt, sulfate aerosol, organic carbon, and black carbon. In that system, the sea salt, carbon types, and smoke are all hygroscopic (dependent upon relative humidity). Further, five differently-sized

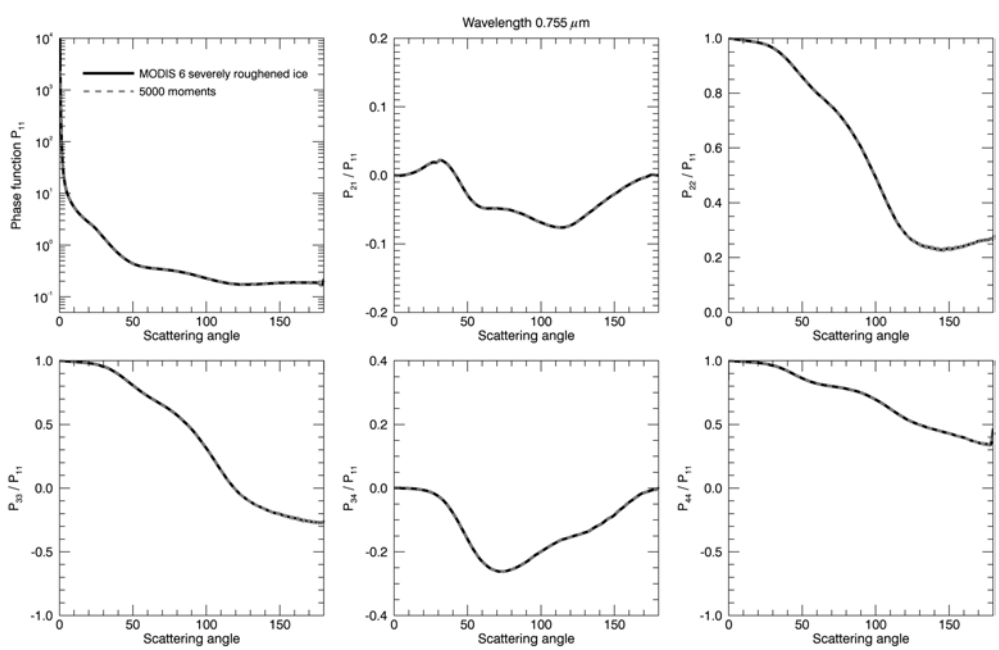

Figure 3-11. Independent phase matrix elements for the cloud ice particles used in the L2 retrieval algorithm. Shown are both the exact phase functions, as well as the M=5000 delta-M.

bins each of dust and sea salt are separately tracked in the model. These bins plus the hygroscopicity lead to a huge range of effective aerosol types and associated optical properties.

In the L2 retrieval, we wish to use the types most commonly associated with a given location and month, but representing that huge level of detail is likely unnecessary. Therefore, we find mean relative humidity values for each subtype and calculate optical properties only for those relative humidity values. Second, we composite the different dust and sea salt bins into a single dust type and a single sea salt type. The L2 algorithm has flexibility of the number of aerosol types used. The default configuration uses the two most prevalent types, with the optical depth of these two types scaled to match the optical depth of all five types in MERRA. Below, we first describe how we choose the mean relative humidity values, and then we give the mathematics that describe the compositing of the aerosol optical properties.

3.2.1.6.1 Aggregation of MERRA Aerosol Subtypes

In choosing how to combine different subtypes into an average aerosol type, we use the fact that the densities of some aerosol types or subtypes are highly correlated with one another, such that we can form a mixture of the two. The only things that remain are (1) to determine the relative fractions of each type to use (in terms of fraction of the aerosol density), and (2) to determine "typical" relative humidity (RH) values to use, as some of the types have optical properties that depend on RH.

In the following analysis, we selected 100 3-hour MERRA reanalysis files at random. This was done a couple of times to make sure the results were fairly stable. In **Table 3-3**, we give the results for the density weightings that were chosen for L2 algorithm, as well as chosen typical values of RH to use. The latter was based primarily on weighting by optical depth in the O_2 A band.

Table 3-3. Density weightings and RH values chosen for the L2 algorithm.

Aggregate Name	Subtypes	Density Fractions [%]	RH values [%]
Dust	DU Bins 1, 2, 3, 4, 5	15, 40.5, 30.5, 11.5, 2.5	N/A
Sea Salt	SS Bins 1, 2, 3, 4, 5	0.2, 2.2, 19.4, 60.6, 17.6	88 (all)
Black Carbon	BC Phobic, Philic	21, 79	54
Organic Carbon	OC Phobic, Philic	16, 84	77
Sulfate Aerosol	SU (Type 1)	100	70

3.2.1.6.2 Composite Optical Properties of the Aggregate MERRA Aerosols

The MERRA aerosol optical properties are a function of relative humidity and effective radius for a number of types, where each type has potentially a number of size bins that are kept separate. Two or more of these subtypes can be put together as a composite type. We are given the complete optical properties for the individual aerosol subtypes by colleagues at NASA GSFC. Let us first define the following symbols:

- Q: extinction efficiency
- S: physical cross section per unit volume in m^{-1}
- β: volume extinction coefficient (cross section per unit volume) in m^{-1}
- k: mass extinction coefficient in m^2/kg
- λ: wavelength
- ω: single scattering albedo
- f: fraction of density

- ρ: mass of aerosol per m^3 of air
- *g*: asymmetry parameter

For any aerosol type or mixture, we have:

$$\beta(\lambda) = k(\lambda)\rho \tag{3-19}$$

where β and *k* can be for extinction (*e*) or scattering (*s*). Let our composite aerosol type have β(λ), *k(*λ), and ρ. Let us define the fractionation between subtypes in terms of aerosol density. We do this as follows for two contributing subtypes (1 and 2), but this formalism is trivially expanded to more subtypes:

$$\rho = \rho_1 + \rho_2 = f_1\rho + f_2\rho \tag{3-20}$$

and $f_1+f_2 = 1$. The total volume extinction coefficient β is just the sum of the individual βs:

$$\begin{aligned} \beta_e &= \beta_{e,1} + \beta_{e,2} \\ &= k_{e,1}\rho_1 + k_{e,2}\rho_2 \\ &= k_{e,1}f_1\rho + k_{e,2}f_2\rho \\ &= (k_{e,1}f_1 + k_{e,2}f_2)\rho \end{aligned} \tag{3-21}$$

From Equation (3-19) and Equation (3-21), we see that the composite mass extinction coefficient is given by:

$$k_e = k_{e,1}f_1 + k_{e,2}f_2 \tag{3-22}$$

So, this implies that we can easily calculate mass extinction coefficient k_e for the composite type. The mass scattering coefficient k_s is calculated similarly. Finally, the composite single scattering albedo ω is just the ratio of the composite scattering coefficient to the composite extinction coefficient:

$$\omega = \frac{k_s}{k_e} = \frac{k_{s,1}f_1 + k_{s,2}f_2}{k_{e,1}f_1 + k_{e,2}f_2} = \frac{k_{s,1}f_1 + k_{s,2}f_2}{k_e} \tag{3-23}$$

Finally, we want to determine phase function quantities, such as *g*, and phase matrix expansion coefficients. All phase function quantities will composite exactly like *g*. Therefore, it is sufficient to show how to find the composite *g*, and this can be applied to all phase function objects (expansion coefficients, etc.). It is straightforward to show that:

$$g = \frac{k_{s,1}g_1f_1 + k_{s,2}g_1f_2}{k_{s,1}f_1 + k_{s,2}f_2} = \frac{k_{s,1}g_1f_1 + k_{s,2}g_1f_2}{k_s} \tag{3-24}$$

Note that we have not calculated any *Q*-type quantities yet (extinction efficiency, etc.). Formally, these are never needed in the L2 FP algorithm, as the wavelength-dependence of *Q* (*e* or *s*) is exactly the same as for *k* (*e* or *s*) for any given type or subtype. However, they are the quantities we typically specify in our optical properties files, and so we discuss them here.

Physically, Q is the extinction cross section per unit volume divided by the physical cross section per unit volume:

$$Q = \frac{\beta}{s} \tag{3-25}$$

It can be shown that the effective Q over a full size distribution of spherical particles is given by:

$$Q = \frac{4}{3} k \cdot R_{eff} \cdot \rho_{bulk} \tag{3-26}$$

where ρ_{bulk} is the bulk density of the aerosol material, and in general may be a function of the relative humidity (but not the wavelength!). GSFC gives Q, β, and R_{eff} for each aerosol type and as a function of relative humidity. Therefore, the bulk density they use is easily calculated for each of their five main types as a function of RH. For dust, they use a bulk density of 2650 kg/m^3, except for the first size bin where they seem to have used a bulk density of 2500 kg/m^3. Using the equation for Q and knowing that S of a composite type is simply the sum of the Ss of the contributing subtypes, it is easily shown that the composite Q_e can be calculated from the individual Q_es according to:

$$Q_e = \frac{k_e}{\frac{k_{e,1} f_1}{Q_{e,1}} + \frac{k_{e,2} f_2}{Q_{e,2}}} \tag{3-27}$$

and similarly for Q_s.

The extinction efficiency, single scattering albedo, and asymmetry parameters of the five aerosol types and two cloud types are shown in Figure 3-12. The asymmetry parameter represents the degree of forward scattering in the intensity phase function. All scattering types include a fully polarized treatment of the scattering phase function using a standard formalism [de Rooij and van der Stap, 1984]. The scattering phase matrix for ice was shown in Figure 3-11 above.

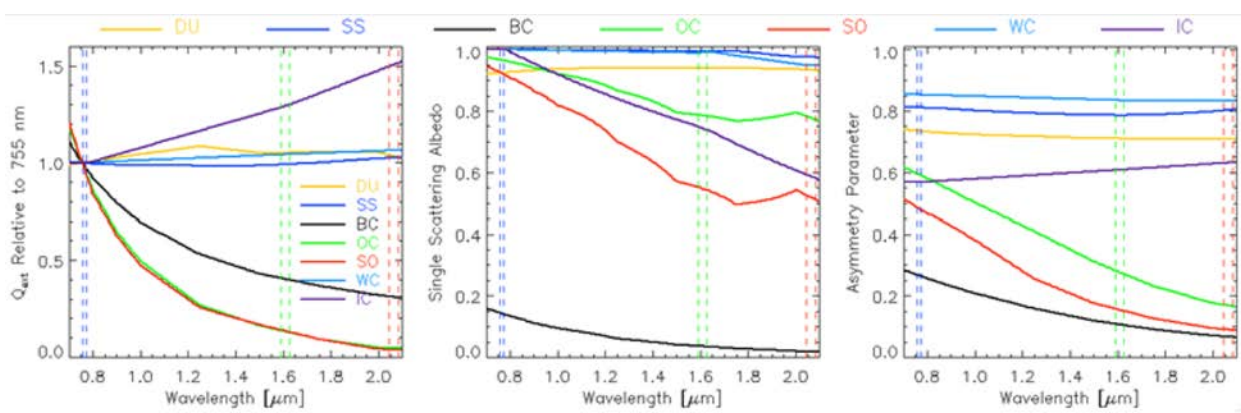

Figure 3-12. Aerosol and cloud optical properties assumed in the L2 code. **Left:** Extinction Efficiency relative to that at 755 nm. **Middle:** Single Scattering Albedo. **Right:** Asymmetry Parameter. DU: Dust, SS: Sea Salt, BC: Black Carbon, OC: Organic Carbon, SO: Sulfate, WC: Water Cloud, IC: Ice Cloud. The three OCO-2 bands are demarcated by the dashed vertical lines.

3.2.2 Solar Spectrum Model

The solar spectrum used in the OCO-2 FP forward model is comprised of two parts: a pseudo-transmittance spectrum and a solar continuum spectrum. The product of these two pieces results in the high-resolution, absolutely-calibrated input solar spectrum for the forward model. Doppler shifts due to variations in Sun-Earth distance and Earth rotation are individually calculated for each sounding.

3.2.2.1 Pseudo-Transmittance Spectrum

The pseudo-transmittance spectrum is derived from an empirical solar line list together with a subroutine that contains a simple solar line shape function, both of which were developed by G. Toon and colleagues [Toon et al., 1999]. By moving to a line list and line shape solar model, they have been able to combine data from several sources to broaden the spectral coverage and remove the instrumental and atmospheric effects of the individual measurements and provide a selectable-resolution, exo-atmospheric solar pseudo-transmittance spectrum.

The line list contains more than 35,000 lines and covers the 600–25,000 cm^{-1} region. For the near-infrared (NIR) region covered by OCO-2, the line list is based primarily on ground-based Kitt Peak spectra, with the exception of the O2 A-band regions which are based primarily on Denver University balloon spectra. Additionally, some ground-based TCCON spectra from Park Falls were used. The spectrum has been validated against Izana TCCON spectra in the NIR as well as GOSAT spectra at lower resolution.

The empirical solar line shape function gives a Doppler shape near the center with exponentially decaying wings. The line shape does not have a physical basis but has been found to give a reasonable representation.

Readers should contact Geoffrey Toon directly for more detailed information regarding the derivation and implementation of the line list and line shape function.

The use of the line list and line shape model based pseudo-transmittance spectrum offers advantages for this application over solar spectra generated by other methods. Since the line list and solar line shape are predefined, the pseudo-transmittance spectrum can be pre-calculated at high resolution and stored for quick table lookup during the retrievals. Additionally, since both instrumental and atmospheric effects are removed during the line list fitting process, they may be ignored in the OCO-2 cod

3.2.2.2 Solar Continuum Model

The solar continuum model calculates the solar Planck function, which is then multiplied with the solar pseudo-transmittance spectra to obtain the calibrated solar spectrum. The solar continuum model is based on a 9th order polynomial fit to the NIR part of the low-resolution extra-terrestrial solar spectrum acquired by the Solar Spectrum (SOLSPEC) instrument [Thuillier et al., 2003]. The polynomial continuum model is multiplied by the pseudo-transmittance spectrum which has had its resolution reduced to match the SOLSPEC resolution. Only the OCO-2 spectral bands are weighted in the fit, although a single polynomial is fit across all three bands.

The uncertainty of the SOLSPEC data is estimated to be 1–3%. We assume that the solar continuum is invariant over time. The sensitivity of the OCO-2 retrieval to the errors in the continuum had not been directly evaluated at the time of writing.

The SOLSPEC spectra are given for a Sun-Earth distance of 1 Astronomical Unit (AU; 1 AU = 149,597,871 km), so that it is necessary to scale the spectrum with the square of the seasonally-varying Earth-Sun distance, which has been taken from http://eclipse.gsfc.nasa.gov.

3.2.2.3 Solar Doppler stretch

The relative velocity of the Earth and the Sun caused by the Earth's rotation and the ellipticity of the Earth's orbit about the Sun produces daily and seasonal Doppler shifts in the solar lines with respect to the telluric lines. The Doppler shift $\Delta\lambda$ for a wavelength λ is calculated assuming a small (non-relativistic) relative velocity v_{rel}:

$$\frac{\Delta\lambda}{\lambda} = \frac{-v_{rel}}{c} \tag{3-28}$$

where c is the speed of light.

The relative velocity is given by the sum of the Earth-Sun radial velocity and the Earth rotational velocity component. The Earth-Sun velocity is calculated from the derivative of a 6th order polynomial fit of solar distance derived from NASA Reference Publication 1349—Twelve Year Planetary Ephemeris: 1995 – 2006. Note that this correction does not include the Doppler shift due to the spacecraft's ~7 km/sec velocity around the Earth, which affects solar and telluric lines equally. Such a correction is made within the instrument model.

3.2.3 Radiative Transfer

The radiative transfer (RT) module encapsulates all the physics associated with the modification of the solar radiation during its passage through the atmosphere and reflection by the surface. This is greatly complicated by the fact that some of the radiation is reflected back to space (by cloud/aerosol) before hitting the ground, causing the average photon path to be wavelength-dependent, as well as by the fact that reflected radiation is strongly polarized, especially in glint mode over ocean.

Because OCO-2 is polarization-sensitive and the radiation can be highly polarized, it is essential that the forward model calculations take this into account and calculate the fully-polarized radiation. We describe the radiance at a given wavelength incident upon the instrument by a three-element Stokes vector $\{I, Q, U\}$. We neglect the circular polarization component because it is generally negligible and OCO-2 is not sensitive to it. Traditionally, a fully-polarized RT code such as vector LIDORT (or VLIDORT) would be used for the calculation, but this is too computationally expensive. Instead, we resort to a piecemeal approach. First, we approximate the Stokes vector as:

$$I = I_1 + I_{ms} + I_{corr} \tag{3-29}$$

$$Q = Q_1 + Q_2 \tag{3-30}$$

$$U = U_1 + U_2 \tag{3-31}$$

where the subscript *1* denotes first order of scattering (and includes the direct beam reflected from the surface), *2* denotes second order of scattering, *ms* denotes all orders of scattering above the first, and I_{corr} represents a polarization correction to I_{ms}. The various terms are then calculated as follows:

1. I_1, Q_1, and U_1 are calculated by a fast, fully-polarized single order of scattering module
2. I_{ms} is calculated by LIDORT, a multiple-scattering scalar RT code
3. Q_2, U_2, and I_{corr} are calculated by a fully-polarized, second order of scattering technique called 2OS.

Our goal is a spectrum of *I*, *Q*, and *U* at each of the ~40,000 high-resolution spectral points in order to convolve with the instrument response. However, even with the savings of not using a vectorized, all-orders-of-scattering code, a direct brute-force process is still too expensive. Therefore, a method dubbed the "Low Streams Interpolator" is employed to drastically speed up the RT process and avoid the direct, brute-force approach.

3.2.3.1 LIDORT

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

3.2.3.2 Polarization Correction: The Two Orders of Scattering Model

The OCO-2 FP code uses LIDORT as the scalar RT model. However, to fully describe the state of light observed by the detector, polarization has to be taken into account. Calculations of multiple scattering with full treatment of polarization are computationally very expensive. Since multiple scattering is depolarizing, it is reasonable to expect that the polarization could be accounted for by a low-order scattering approximation. Natraj and Spurr [2007] extended the scalar Kawabata and Ueno [1988] model to compute the first two orders of scattering (2OS) for vertically inhomogeneous scattering media with polarization included. To enable accurate computations for the range of solar viewing angles encountered by OCO-2, atmospheric transmittances for the incoming solar beam are treated for a curved spherical-shell atmosphere [Spurr, 2002]. For the 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 vectorized multiple-scattering computation.

3.2.3.2.1 Theory

The formulation of the 2OS model is based on the invariant imbedding theory of RT. The contribution from the pth order of scattering is related to that from the $p\text{-}1^{th}$ order of scattering by an integro-differential equation. Starting from direct transmission, the expressions for subsequent orders of scattering can be derived. This is particularly straightforward for the first two orders of scattering. A Fourier series expansion is done to solve the problem efficiently. The expressions for the Stokes parameters *I*, *Q,* and *U* are as follows:

$$I_{OCO} = I_{sc} + I_{corr} \tag{3-32}$$

$$Q_{OCO} = Q_1 + Q_2 \tag{3-33}$$

$$U_{OCO} = U_1 + U_2 \tag{3-34}$$

where the subscripts *OCO*, *sc*, *1*, and *2* are, respectively, the OCO-2 measurement, scalar computation (from LIDORT), and first and second order computations (from the 2OS model). I_{corr} is the intensity correction computed by the 2OS model. Equation (3-32) implies that the LIDORT is employed in multiple-scattering mode (single scattering is computed using the 2OS model). The reasons for this will become apparent in Section 3.2.3.3.

3.2.3.2.2 Evaluation

The OCO-2 FP RT model has been tested against the full multiple scattering vector model VLIDORT [Spurr, 2006] for the OCO-2 nadir, glint, and target modes of operation. Computations were performed for different solar zenith angles, surface types, and aerosol loadings [Natraj et al., 2008]. The radiance errors using the hybrid LIDORT-2OS model are an order of magnitude (or more) smaller than the errors arising from the use of LIDORT alone. In addition, a linear error analysis study showed that the errors in the retrieved X_{CO2} using the hybrid model are much lower than the measurement noise and smoothing errors. On the other hand, the scalar model alone induces X_{CO2} errors that could dominate the retrieval error budget.

3.2.3.3 Acceleration of the RT Calculations

3.2.3.3.1 Low-Streams Interpolation

The scalar RT code used to calculate the multiple-scattering component of the total intensity *(I_{ms}* from Equation (3-29)*)*, LIDORT, relies on a quadrature-based approach wherein the code effectively only treats radiation coming from specific discrete zenith angles or "streams," and interpolates between these angles to find the radiation at an arbitrary angle. In standard nomenclature, a code is said to be run with *N* streams total, where *N/2* is the number of streams in either the upper or lower hemisphere. The computational burden in theory goes as N^3, because the codes largely rely on multiplication and inversion of *N* x *N* matrices. However, often there is other significant computational overhead, which lessens the extreme dependence on *N* to be more like N^2, especially for lower numbers of streams.

The number of streams required to achieve a given accuracy depends on the type of atmosphere in question. To obtain sufficient accuracy for most scenes and viewing geometries, eight to 32 streams are required. However, doing this in a line-by-line approach is cost-prohibitive. To speed this process, we employ the method of low-streams interpolation (LSI), fully explained in O'Dell [2010]. In a nutshell, this process allows one to perform just a handful (five to 20 per band) of high-accuracy calculations, alongside a low-accuracy RT calculation for every monochromatic wavelength. The monochromatic high-accuracy calculations can then be reconstructed with very low errors using the LSI method. In addition, because the technique is so simple (involving just a couple of very simple equations), it was very straightforward to linearize such that it can handle Jacobians as well. In our particular setup, we have:

- Low accuracy (monochromatic) calculations: two-stream scalar RT plus polarized single-scattering code
- High accuracy calculations: 16-stream scalar RT plus polarized 2OS calculations (Section 3.2.3.2)

The errors from a similar setup are shown in Figure 3-13 (taken from O'Dell, 2010). Internal tests have shown that the errors in X_{CO2} from this approach tend to be random and typically less than 0.3 ppm.

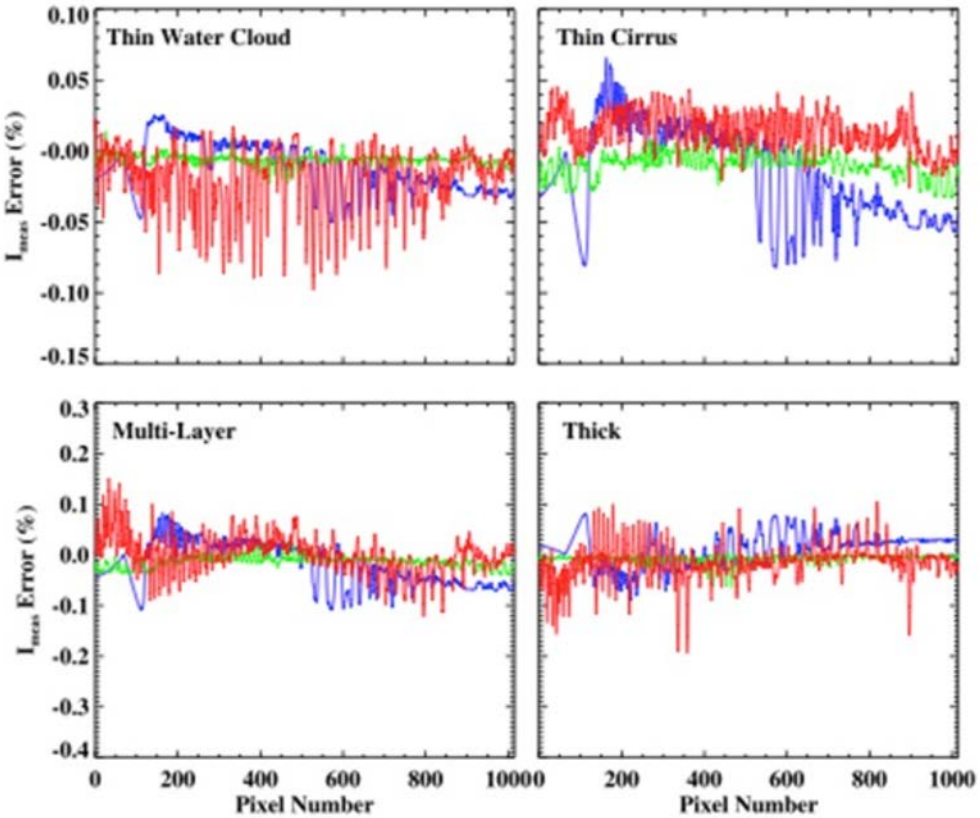

Figure 3-13. Percent errors in OCO-simulated TOA radiances due to the LSI method for four different cases. Each case includes one or multiple clouds. The three OCO bands are shown in blue (O_2 A), green (weak CO_2), and red (strong CO_2). Taken from O'Dell [2010], Figure 8.

3.2.3.3.2 Non-Uniform Spectral Sampling

Although the errors in the LSI are acceptable, the resulting RT is still slower than necessary. The low-accuracy monochromatic RT calculations typically dominate the overall calculation time. Therefore, we found it acceptable to "skip" some of these wavelengths and simply interpolate from neighboring spectral points to obtain these values. Recall that we perform our RT calculations on a 0.01 cm^{-1} spaced grid for all three NIR bands. It turns out this this spacing most over-samples the O_2 A band, though there is some oversampling in each band.

We use a simple approach to determine which spectral points can be skipped. It is a semi-automated process that requires monochromatic radiance errors from interpolation to be less than some threshold, typically 0.1%. We have investigated different interpolation techniques and generally find that linear interpolation is comparable in accuracy to higher order interpolation schemes, as shown in Figure 3-14 below.

Implementing "non-uniform sampling" leads to speed savings of about 50% in the low-accuracy, monochromatic RT calculations. The savings is highest in Band 1 and is almost negligible in Band 3, which is the least oversampled. The L2 retrieval realizes an overall speed savings of ~20%, though this depends on the details of the low- and high-accuracy RT calculation setup. The errors in retrieved X_{CO2} due to non-uniform spectral are virtually always less than 0.3 ppm, and are typically less than 0.1 ppm.

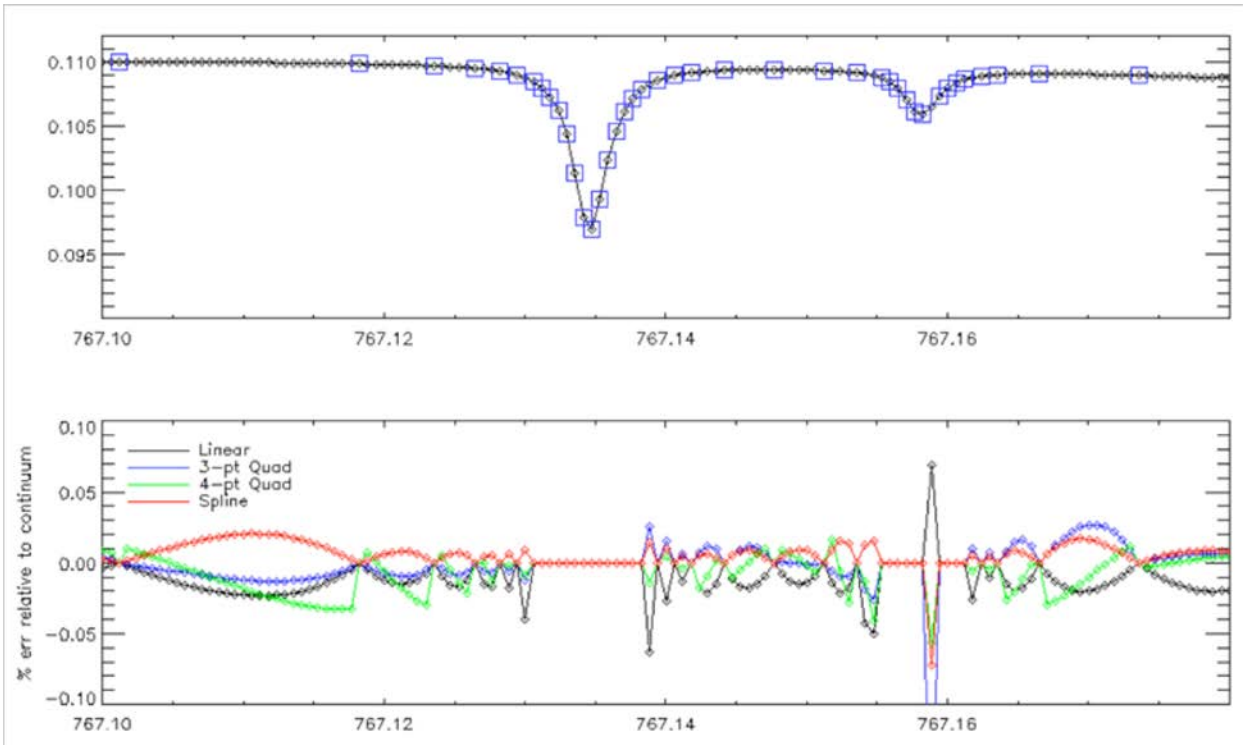

Figure 3-14. Illustration of non-uniform sampling approach. **Top:** Training. From a full monochromatic spectrum (black diamonds) certain spectral points are selected for RT calculations (blue squares). The remaining points are calculated from linear interpolation of the selected points. **Bottom:** Interpolation errors in the non-selected points for different interpolation schemes. Linear interpolation is used because its errors are comparable to higher-order interpolation schemes.

3.2.3.4 Fluorescence Radiative Transfer

The fluorescence state vector elements are discussed in Section 3.3.2.4.3. This shows to how to calculate the solar-induced fluorescence (SIF) signal at the surface, which is taken to be isotropic and unpolarized. As discussed in that section, the emitted radiance due to fluorescence is modeled in the O_2 A band as linear in wavenumber. This would break down over a large spectral range, but is appropriate for the narrow O_2 A band from 759 – 772 nm. To propagate the SIF signal at the surface, SIF_{gnd}, to the top of atmosphere (TOA), we only take into account absorption by molecular oxygen:

$$SIF_{TOA} = SIF_{gnd} \cdot exp\left(\frac{-\tau_{oxygen}}{\mu}\right) \tag{3-35}$$

where τ_{oxygen} is the total zenith optical depth due to O_2, and μ is the cosine of the sensor zenith angle; the equation is implicitly a function of (monochromatic) wavelength. SIF_{TOA} is then simply added to the upwelling solar radiance at TOA. For simplicity, we have ignored scattering and extinction due to clouds and aerosols, as we typically only retrieve X_{CO2} in reasonably clear scenes. In reality, the extinction sink and scattering source of radiance due to clouds and aerosols often partially cancel, so this approximation is typically adequate for our purposes.

3.2.4 Instrument Model

The RT model produces high-resolution (~0.01 cm^{-1}) synthetic spectra of I, Q, and U for a specified observing geometry and surface/atmospheric state. Before these synthetic spectra can

be compared to the calibrated, geolocated spectral radiances, the synthetic *I*, *Q*, and *U* must then be corrected for instrument polarization, corrected for Doppler shift, and convolved with the instrument line shape (ILS) function. Each of these steps is defined below. The instrument model performs these steps.

Pixel-wavelength mapping: the wavelength λ for detector pixel *i* and spectrometer *k* is given by:

$$\lambda_i = \sum_{n=1}^{N} d_k^n \times i^{n-1} \tag{3-36}$$

where d are the N dispersion coefficients for spectrometer. For OCO-2, i ranges from 1 to 1016 for each spectral band and there are six coefficients for each band. These relationships also vary for each of the eight footprints within a band. The dispersion coefficients are stored in each Level-1B file. In the V8 product, the retrospective L1B dispersion is modified to include corrected coefficients for the first two terms (d^0 & d^1). These corrections are derived from the L2 retrieval algorithm and account for Doppler and thermo-mechanical shifts observed, in flight, in the dispersion (refer to the Level 1B Algorithm Theoretical Basis document).

3.2.4.1 Instrument Line Shape Function

The ILS function describes the response of an individual OCO-2 detector pixel to light with a wavelength λ. Mathematically, this can be described as the product of the monochromatic intensity *I* with the ILS:

$$I(\lambda) = \int_{-\infty}^{+\infty} I(\lambda') \times ILS(\lambda, \lambda')\, d\lambda' \tag{3-37}$$

The ILS is a varying function of spectral color (pixel) within each footprint and band. Therefore, OCO-2 has 1016 x 8 x 3 = 24384 different line shape functions. Note that for GOSAT data based on an FTS, this step requires a convolution of the ILS with the monochromatic spectrum.

3.2.4.2 Radiance Scaling

The calculated spectrum can be scaled by a detector pixel-dependent scale factor. For downlooking spectra, radiance scaling is not used. For uplooking spectra, where there is no analog to the scaling that occurs through ground model, this radiance scaling can be used to match the absolute values of the measured and modeled spectrum:

$$C(\lambda) = \sum_{n=1}^{N} c_k^n \times (\lambda - \lambda_c)^{n-1} \tag{3-38}$$

where *c* are *N* radiance scaling coefficients for spectrometer *k*, and λ_c is the center wavelength of the spectral window.

3.2.4.3 Instrument Doppler Correction

The wavelength scale is modified based on the relative velocity of the spacecraft and the Earth target to account for the Doppler shift. This relative velocity is calculated from the spacecraft velocity and the Earth's rotation. The equation for the Doppler correction is:

$$\lambda_{adj} = \lambda_{org}\left(1 + \frac{v_{rel}}{c}\right) \tag{3-39}$$

where c is the speed of light. The L2 retrieval is then performed on this adjusted wavelength grid.

3.2.4.4 Zero-Level Offset

As discussed in Butz et al. [2011] and Crisp et al. [2012], Band 1 of the TANSO-FTS instrument on GOSAT is known to have a nonlinear response to intensity. This leads to systematic features in the Band 1 spectra, the size of which are intensity-dependent. To account for this, we include a wavelength-independent zero-level-offset, originally shown by Butz et al. [2011] to partially mitigate this instrument problem. This is modeled simply as a constant radiance added to the spectrum at TOA. It is given a prior value of zero with a very loose prior constraint.

For OCO-2, we do not expect to need to retrieve a zero-level-offset for any band, though the capability is there if it proves to be necessary.

3.2.4.5 Treatment of Polarization

OCO-2 fundamentally measures a single linear polarization of the incoming light. The axis of polarization accepted by the instrument depends on the yaw angle of the satellite and other instrument parameters. For convenience, this is given in the L1B file via two sets of parameters. The first is the polarization angle of the axis of the accepted polarization, and the second is a set of "Stokes Coefficients" $\{m_1, m_2, m_3, m_4\}$, which define the response of the instrument to Stokes vector of the incoming light:

$$I_{meas} = m_1 I + m_2 Q + m_3 U + m_4 V \tag{3-40}$$

The polarization angle ϕ_{pol} is defined such that $m_2 = ½ \cos(2\ \phi_{pol})$ and $m_3 = ½ \sin(2\ \phi_{pol})$, and physically represents the angle between the axis of accepted polarization of the instrument and the reference plane for polarization. Stokes parameters Q and U also depend upon the chosen reference plane for polarization. For OCO-2, we choose the local merdian plane to be this reference plane for polarization, which is the plane containing the local normal unit vector and the vector pointing from the target field of view (FOV) to the satellite (Figure 3-15). Stokes parameter Q is then defined as:

$$Q = I_{||} - I_{\perp} \tag{3-41}$$

i.e., Q is the difference in intensities of the light polarized parallel and perpendicular to this plane. U is similar but requires a rotation by 45°:

$$U = I_{45} - I_{135} \tag{3-42}$$

and further the definition of Stokes parameter U requires a definition of the "sense of rotation". Excerpting from Schutgens et al. [2004]:

> "The sense of rotation is usually defined by demanding that the three unit vectors ⊥, ∥, and prop form a right-hand coordinate frame, see Figure 2. Chandrasekhar [1960], van de Hulst [1981], and Hovenier and de Haan [1985] assumed ⊥ × ∥ = prop, while Azzam and Bashara [1987] and Slijkhuis [2000] assumed ∥ × ⊥ = prop. The difference in Stokes vector is the sign of U (and hence χ 180° $-\chi$). Note that POLDER data [Bréon and

CNES Project Team, 1997] uses the latter convention, while SCIAMACHY data [Slijkhuis, 1998] currently uses the first."

where "prop" denotes a unit vector pointing in the direction of propagation. For OCO-2, we use the former convention of Chandrasekhar [1960], van de Hulst [1981], and Hovenier and de Haan [1985], also adopted by the SCIAMACHY team [Slijkhuis, 1998], i.e., we assume that $\perp \times \parallel = \text{prop}$. This convention is used both in the definition of the polarization angle and Stokes coefficients provided in the OCO-2 L1B files, as well as the L2 radiative transfer solvers.

It is critical to note that even in nadir mode, the target FOV never sits directly at the subsatellite point. It is always offset slightly, and there is a different geometry for each of the eight OCO-2 footprints in a given frame. Therefore the polarization reference plane, defined above, can be significantly different for each of the eight footprints in a given frame, and thereforesubsequently m_Q and m_U may vary significantly from one footprint to the next in a given frame.

In previous documents, we implied that for OCO-2 the polarization reference plane would be defined as the principal plane, which is formed by the ray from the target to the Sun and the target to the satellite. This frame, while simpler in the sense that the definitions of Q and U would not vary much in nadir mode from one footprint to the next, is typically not used in the RT community and therefore we have avoided it in favor of the local meridian plane. In glint mode, the principal and local meridian planes are nearly identical, but in nadir and target modes they can differ significantly.

Further, the orientation of the axis of accepted polarization is not always horizontal to the ground (i.e., perpendicular to the local meridian plane). While this was indeed the intent for OCO-2, a problem with the actual polarization plane of the instrument was discovered on orbit. A mitigation strategy was employed to yaw the spacecraft by 30° in glint mode, but not in nadir mode (as of the time of this writing). In target mode, because of the complicated slewing of the spacecraft, the polarization angle can vary widely.

In the special case of glint measurements in the principal plane when $\phi_{pol} = -30°$, the measured intensity reduces to:

$$I'_{||} = 0.5I + 0.25Q - 0.433U \quad (3\text{-}43)$$

However, in nadir and target modes, the polarization angle may vary with the footprint index from near 0° to near 180°.

Finally, we note that in the L1B software, identical target locations are assumed for all three spectral bands, and for each channel in a given band. Therefore, the polarization angle and Stokes coefficients take a single set of values for a given sounding.

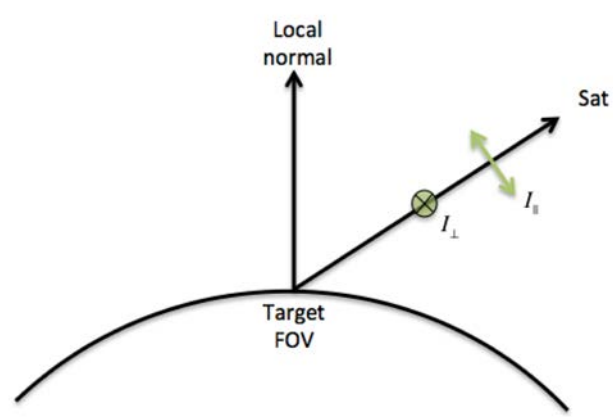

Figure 3-15. Geometry of the reference plane for polarization, i.e., the local meridian plane. This plane is formed by the local normal unit vector and the ray from the target FOV to the satellite. The direction of propagation of measured light is the latter. Two example

3.2.4.6 Application of Stokes Vector

As shown in the previous section, for OCO-2 the only parameter necessary to calculate the Stokes coefficients is the angle of the

slit with respect to the principal plane. These are calculated and stored in the OCO-2 L1B files. Within the Ll 2 RT code, we simply calculate *I*, *Q*, and *U* at the top of the atmosphere, and construct the measured radiance as in Equation. Most generally, we write this as:

$$I_{meas} = m_1 I + m_2 Q + m_3 U \tag{3-44}$$

where the m_i are the Stokes coefficients, and V is not calculated by the RT software (it is assumed to be negligible). Equation (3-46) is performed for each channel and band on OCO-2. The same equation is used for GOSAT, but the construction of the GOSAT Stokes parameters is significantly more complicated and is described in detail in O'Brien et al. [2013].

3.3 State Vector

3.3.1 Basic Description of State Vector

The section below describes the state vector and setup used in the at-launch version of the L2 algorithm (B4.0). A previous version, B2.9, was described fully in the literature [O'Dell et al., 2012]. Below we primarily describe updates to that algorithm version.

A state vector specifies those aspects of the state of the atmosphere being measured and of the instrument measuring it, which together determine the value of the resulting measurement. We will use the term *state structure* to indicate the complete set of parameters required by the forward model to simulate a measurement to the necessary accuracy, and the term state vector to indicate the set of parameters being retrieved. This may be a simple subset of the state parameters that describe changes to the state structure (such as an offset to a temperature profile).

The state structure contains all the information about the atmosphere, the surface, the instrument, and the Sun needed by the forward model, as described in Section 3.2. These are captured in six substructures:

- Atmosphere: temperature profile, water vapor profile, and surface pressure
- Absorber: gaseous volume mix ratio profiles (CO_2, H_2O, O_2, [O_3, and CH_4—not currently used])
- Aerosol: aerosol shape parameters (aerosol optical depth [AOD] at 755 nm, peak height, and width)
- Ground: Bidirectional reflectance distribution function parameters—either BRDF amplitude (meand and spectral slope), ocean BRDF parameters, including a lambertian mean and spectral slope, as well as the surface wind speed.
- Instrument: spectral dispersion and ILS
- Solar: parameters of solar continuum model
- Empirical orthogonal function (EOF) amplitudes for residual fitting

The state structure is completely flexible and can include an arbitrary number of absorber and aerosol extinction profiles with an arbitrary number of vertical levels. For operational L2 retrievals, 20 levels are used in the state structure for CO_2, but 91 levels are used for temperature and water vapor. These are discussed below in Section 3.3.2.

The state vector currently consists of 60–61 elements (OCO-2) or 52 – 53 elements (GOSAT) as shown in **Table 3-4**. The exact number depends on whether the retrieval is over land or ocean, and is slightly different for OCO-2 versus GOSAT.

Table 3-4. Description of state vector elements.

Description	Parameters	Number of Elements
Aerosols	5 x 3 parameters	15
Temperature offset	Scalar	1
Water vapor multiplier	Scalar	1
Surface pressure	Scalar	1
BRDF amplitude, mean and spectral slope (land) Lambertian albedo mean & spectral slope (ocean)	2 x 3 bands	6
Surface wind speed (ocean only)	Scalar	1
CO_2	20 levels	20
Spectral dispersion offset	1 shift per band	3
Spectral dispersion slope (OCO-2 only)	1 slope per band	3
O_2 A band radiance offset (GOSAT only)	Scalar	1
Chlorophyll fluorescence (O_2 A band) (land only)	Mean + slope	2
Residual EOF amplitudes	1 per band (GOSAT) 3 per band (OCO-2)	3 (GOSAT) 9 (OCO-2)
OCO-2 total land (ocean)		**61 (60)**
GOSAT total land (ocean)		**53 (52)**

In the next section, we describe the initial value and associated uncertainties of each state vector quantity.

3.3.2 The *a Priori* State Vector and Covariance Matrix

One of the most important factors in any optimal estimation retrieval algorithm is the setup of the *a priori* values and their corresponding covariance matrices. They represent our best guess of the values of the parameters given our prior knowledge, along with an estimate of their 1-sigma uncertainties (assuming Gaussian statistics). When the *a prio*ri constraint on a parameter is relatively loose (large *a priori* covariance), the retrieval relies more on the measurement in order to determine the parameter; conversely, when the *a priori* constraint is tight (small *a priori* covariance), the measurements have less effect on the retrieved parameter value. The overall *a priori* covariance matrix is the matrix of error variances and covariances of the *a priori* values of all state vector parameters; in principle, the *a priori* errors of any two parameters may be correlated.

The retrieval algorithm relies in part on constraining solutions so that they are statistically reasonable. In order to do this, we must have a realistic *a priori* state vector and a covariance matrix which constrains poorly measured components of the state. Note that, for convenience, the *a priori* state vector is also used for the first guess in our retrieval, as it ideally represents our best guess on the state of the atmosphere and surface before taking the OCO-2 observations into account.

The *a priori* vector is a simple vector of parameters with their *a priori* values. We take the corresponding *a priori* covariance matrix to be diagonal, with the exception of the carbon dioxide profile block. That is, the *a priori* uncertainty of every parameter is assumed to be uncorrelated with that of every other parameter, except within the *a priori* CO_2 profile, in which uncertainties at one vertical level may be correlated with uncertainties at other levels. We now describe the *a priori* values and assumed uncertainties for each parameter in turn.

3.3.2.1 Profile of Carbon Dioxide Concentration

The *a priori* profile of CO_2 is taken from the same algorithm used by TCCON [Wunch et al., 2011]. As stated in that work:

> "The CO_2 *a priori* profiles are from a climatology based on the GLOBALVIEW dataset, and change based on the time of year and the latitude of the site. Stratospheric CO_2 profiles are generated from the age of air relationship derived by Andrews et al. [53] (2001)."

A key parameter is the height of the tropopause, which is determined from the GEOS-FP first guess temperature profile (see Section 3.3.2.2). We use the methodology of Reichler et. al. (2003) to find a tropopause pressure/altitude consistent with the WMO standard lapse-rate criterion (WMO, 1957), limited to be within the pressure range of 75-450 hPa. To mitigate the influence of vertical temperature perturbations found in the high resolution input data that are not related to the tropopause, we smooth the input temperature profile to a vertical resolution that more closely matches our forward model (20 levels).

The CO_2 covariance submatrix naturally has the most impact on retrieval of X_{CO2}. Currently, a single CO_2 covariance matrix is used for all retrievals. This covariance has been constructed by assuming a root-mean-square variability of X_{CO2} of 12 ppm, which is an estimate of global variability [Dufour and Bréon, 2003]. Variability as a function of height is assumed to decrease rapidly, from ~ 10% at the surface to ~1% in the stratosphere. The covariance among altitudes in the troposphere was estimated based on the Laboratoire de Météorologie Dynamique general circulation model, but the correlation coefficients were reduced arbitrarily to ensure numerical stability in taking its inverse. The total variability embodied in this covariance is unrealistically large for most of the world (all relatively clean-air sites). It is intended to be a minimal constraint on the retrieved X_{CO2}; the use of a single covariance everywhere at all times eliminates the covariance matrix as a source for variation in retrieval characteristics. Figure 3-16 shows a rendering of the CO_2 covariance matrix (taken from O'Dell et al. [2012]).

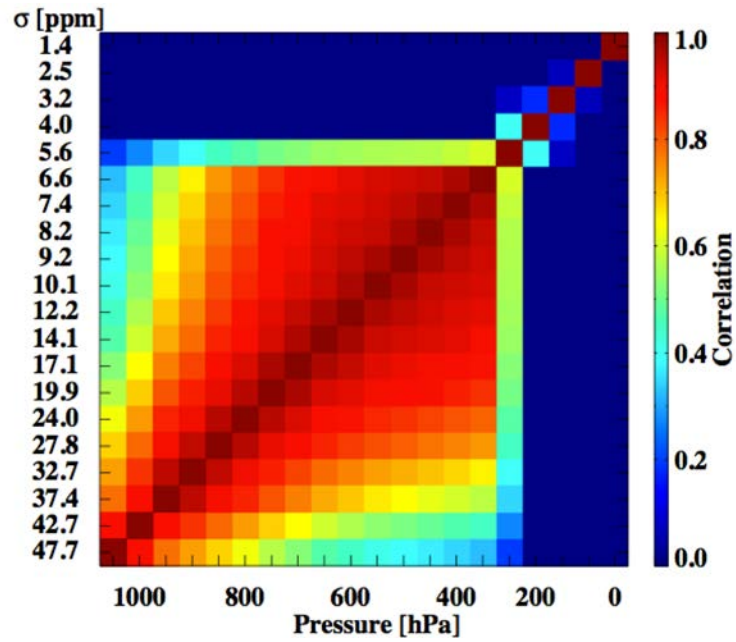

Figure 3-16. *A priori* CO_2 correlation matrix (in colors), together with the 1-sigma errors for each pressure level.

3.3.2.2 Meteorological Variables: Surface Pressure, Temperature, and Water Vapor

In OCO-2 data versions prior to v8, prior and first-guess meteorological variables are taken from a short-term forecast from the ECMWF model. The forecast length is between zero and nine hours. Details of the ECMWF model can be found at http://www.ecmwf.int/. The model is on 137 vertical levels and has a horizontal resolution of 0.25° in latitude and longitude. ECMWF forecast files are downloaded daily.

Select fields are interpolated from the ECMWF files to the locations and times of the OCO-2 soundings, including surface pressure and profiles of temperature and specific humidity. Because the elevation of the model grid box is often slightly different from the actual elevation digital elevation map [Zong, 2008], a further correction is made based on the hydrostatic equation (often called a hypsometric adjustment) to account for this elevation difference.

The ECMWF surface pressures are expected to have errors of approximately 1 – 2 hPa [Salstein et al., 2008]. It is unknown what the true error will be when over complex topography at scales of 1 – 2 km. Currently, the code assumes a 1-sigma error of 4 hPa on the *a priori* surface pressure estimate. This may be tightened in the future once the errors from OCO-2 are better understood. In contrast, for ACOS-GOSAT retrievals, a smaller prior error of 1 hPa is assumed.

As stated previously, the L2 FP code does not try to retrieve the entire temperature profile in the state vector. Instead, it is assumed that the ECMWF forecast temperature profile can be corrected by a single small offset. The *a priori* 1-sigma uncertainty in this offset is taken to be 5 K. It is likely that there is more information on temperature than just a single number, but thus far retrieval experiments indicate that this assumption for a temperature correction is sufficient.

Similarly to temperature, the water vapor profile from the ECMWF forecast is assumed to be correct up to an overall scale factor. This is consistent with the fact that there is not much water vapor information in the OCO-2 spectra, and most of it comes from the strong CO_2 band. Tests thus far indicate that this assumption is justified. We assume a 1-sigma error of 0.5 on the water vapor scale factor, which emphasizes the measurement in the retrieval.

In v8, we transitioned to GEOS FP-IT data as the source for the meteorological data (https://gmao.gsfc.nasa.gov/weather_prediction/). The team evaluated the impact on the L2 data when the source of meteorological data was changed from ECMWF to GEOS FP-IT, with a focus on the surface pressure difference from the prior (dP), and XCO_2. There were no strong spatial patterns of difference, and the standard deviation of XCO_2 differences was 0.4 ppm. The standard deviation of dP for ECMWF was 2.4 hPa and for GEOS FP-IT it was 2.7 hPa. The ECMWF code was modified to be used with GEOS FP-IT, changing for example the pressure grid.

The GEOS FP-IT analysis is performed at a horizontal resolution of 0.625-degree longitude by 0.5-degree latitude and at 72 levels, extending to 0.01 hPa. All products are generated at the native resolution of the horizontal grid. Three-hourly intervals are used for three-dimensional products.

3.3.2.3 Aerosol Profiles

As stated previously, the FP model attempts to retrieve profiles of five scatterers: two cloud types, two tropospheric aerosol types, and one stratospheric aerosol type. The cloud types are a water cloud type and an ice cloud type. The two tropospheric aerosol types are drawn from dust, smoke, sulfate aerosol, organic carbon, and black carbon. In Version 8 of the full-physics retrieval, a stratospheric aerosol was added to the retrieval state vector, which, for convenience, uses the same optical properties as the tropospheric sulfate aerosol type. The optical properties of these types were explained in Section 3.2.1.6 above. For each type, we retrieve three quantities, which characterize the amount and vertical profile of that scatterer. The vertical profiles are assumed to be Gaussian in the relative pressure $x = P/P_s$, where P is the local pressure and P_s is the surface pressure. The shape of the Gaussian is defined by the location of

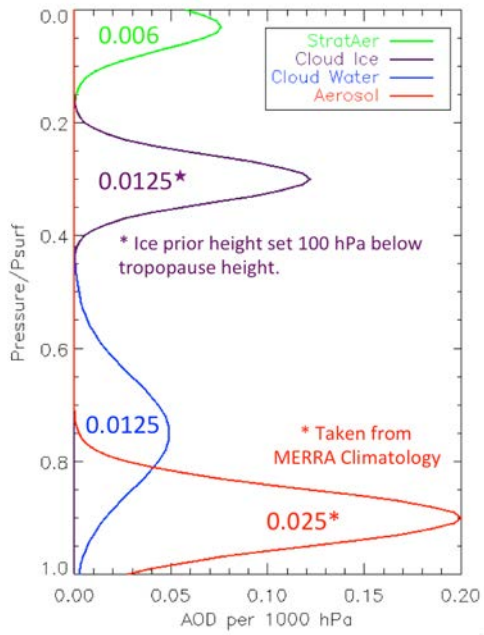

Figure 3-17. Prior aerosol profiles uses the L2FP retrieval.

the peak height and the 1-sigma width (both are defined in units of $x = P/P_s$). The optical depth per unit pressure is given by:

$$\frac{d\tau_\lambda}{dx} = \frac{Q_\lambda}{Q_{755}} \frac{AOD_{755} S(x)}{\int_0^1 S(x)dx} \tag{3-45}$$

where Q_λ is the extinction efficiency at a wavelength λ, AOD_{755} is the desired AOD at 755 nm (our reference wavelength), and $S(x)$ is the profile shape given by:

$$S(x) = exp\left\{-\frac{(x - x_a)^2}{2\sigma_h^2}\right\} \tag{3-46}$$

where x_a is the peak height of aerosol type a, and σ_a is the 1σ Gaussian width. The *a priori* values and uncertainties are given in **Table 3-5** below. Currently, the aerosol *a priori* error covariance matrix is strictly diagonal. Prior aerosol profiles are shown in Figure 3-17, below.

Table 3-5. Aerosol prior profile quantities and *a priori* uncertainties.

Type	AOD_{755}	Ln(AOD) ± 1σ	Height ± 1σ	Width ± 1σ
Water cloud	0.0125	−4.382 ± 1.8	0.75 ± 0.4	0.1 ± 0.01
Ice cloud	0.0125	−4.382 ± 1.8	variable† ± 0.2	0.04 ± 0.01
Aerosol 1	Variable*	variable* ± 2.0	0.9 ± 0.2	0.05 ± 0.01
Aerosol 2	Variable*	variable* ± 2.0	0.9 ± 0.2	0.05 ± 0.01
Stratospheric Aerosol	0.006	−5.116 ± 1.8	0.03 ± 1e-4	0.04 ± 0.01

**AOD priors for aerosol types are chosen from the MERRA climatologies, but a maximum of AOD_{755} 0.2 is set.*
† The ice cloud height is set to be 100 hPa below the tropopause pressure (see Section 3.3.2.1)

The two aerosol types are chosen from a two-year MERRA climatology (2009 – 2010). They are the types that form the highest and second highest fraction of the AOD at 755 nm for a given

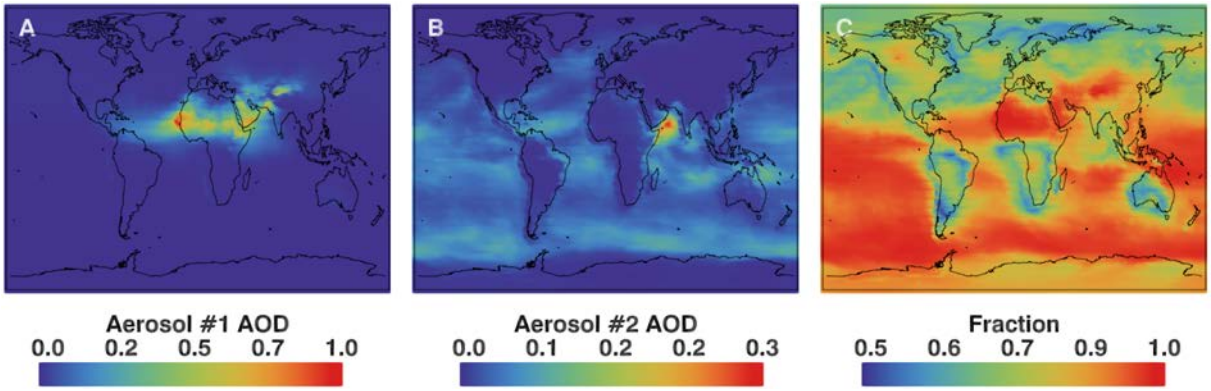

Figure 3-19. Global distribution of aerosol optical depth on July. (**Panel A**) dominant type; (**Panel B**) second-most dominant type; (**Panel C**) fractional contribution of the two dominant types to total AODs.

MERRA grid cell when averaged over a month. Figure 3-18 shows maps of the dominant and second-most dominant type for July. Figure 3-19 shows the corresponding AODs for each of these types. Rather than use the peak height from MERRA, we have found it more numerically stable to set the peak heights and widths to fixed values as given in **Table 3-5**.

However, this may be revisited if we can change from monthly climatologies to actual MERRA coincidences. For all scatterers, the peak width prior uncertainty is set so low (0.01) that it is effectively not retrieved. This is because in practice our spectra have very little sensitivity to the width of a thin scattering layer.

3.3.2.4 Surface Properties

3.3.2.4.1 Land Surface a Priori

Definition of the BRDF model:

A general BRDF model can made up of two kernels, for the unpolarized and polarized components, respectively, of the reflection. For azimuthally symmetric BRDF models, this can be expressed as:

$$\rho(\theta_i, \theta_r, \Delta\varphi) = r\left[r_s F_s(\theta_i, \theta_r, \varphi) + r_p F_p(\theta_i, \theta_r, \varphi)\right] \tag{3-47}$$

where:

ρ is the BRDF

r is the overall amplitude

r_s is the amplitude of the scalar (RPV) model

F_s is the shape of the scalar (RPV) model

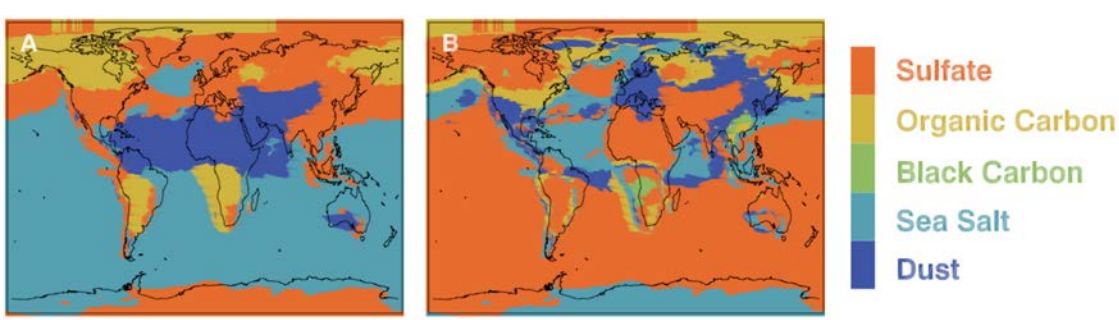

Figure 3-18. Global distribution of aerosol type on July. (**Panel A**) dominant type; (**Panel B**) second-most dominant type.

r_p is the amplitude of the polarized (Breon) model
F_p is the shape of the polarized (Breon) model
θ_i is the incident (solar) zenith angle
θ_r is the reflection (observer) zenith angle
φ is the relative azimuth angle

The BRDF models are functions of three angles, under the assumption that the surface is azimuthally symmetric. Thus, only the relative azimuth angle (difference between solar and observer azimuth) is needed.

For the OCO-2 forward model, we assume that $r_p = 0$ and $r_s = 1.0$, making the BRDF unpolarized, and only dependent on the RPV model. Dropping the polarized kernel, we can write the BRDF as:

$$\rho(\theta_i, \theta_r, \varphi) = r r_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \tag{3-48}$$

where now the BRDF shape is described by certain kernel parameters: ρ_0, Θ, k. The OCO-2 forward model takes fixed values for these parameters, which define the actual shape of the RPV (Rahman – Pinty – Verstraete) kernel [Rahman et al. 1993].

Further, r can be defined as follows, to introduce a linear relationship of the BRDF amplitude with wavenumber.

$$r = w + s(\nu - \nu_0) \tag{3-49}$$

where w is the "BRDF weight" and s is the "BRDF weight slope". For a Lambertian model, these would be the albedo and albedo slope, respectively. ν and ν_0 are the wavenumber and reference wavenumber (at which the surface properties are retrieved), respectively.

We can therefore rewrite the BRDF in the following manner:

$$\rho(\theta_i, \theta_r, \varphi) = \rho_w(\theta_i, \theta_r, \varphi) + \rho_s(\theta_i, \theta_r, \varphi)(\nu - \nu_0) \tag{3-50}$$

where:

$$\rho_w(\theta_i, \theta_r, \varphi) = w r_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \tag{3-51}$$

$$\rho_s(\theta_i, \theta_r, \varphi) = s r_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \tag{3-52}$$

The OCO-2 L2 full physics retrieval labels the variables as follows: first the two retrieved state variables:

w : brdf_weight
s : brdf_weight_slope

and then the 5 fixed state parameters:

r_s : brdf_rahman_factor =	1.0
r_p : brdf_breon_factor =	1e-20 (effectively zero)
ρ_0 : brdf_hotspot_parameter =	0.05
Θ : brdf_asymmetry_parameter =	–0.1
k : brdf_anisotropy_parameter =	0.75

These values for the fixed BRDF parameters were used as the initial guess for an inversion algorithm using multi-angle polarimetric satellite data (Table 1 and 6 in Dubovik et al. [2011]).

Note that the two state variables are retrieved independently within each band, but all bands use the same values of the fixed parameters.

Output Variables:

The two retrieved state variables (w and s, the brdf_weight and brdf_weight_slope) are converted to more physically meaningful surface reflectance values (brdf_reflectance and brdf_reflectance_slope) for the L2 retrieval output. The reflectance and reflectance slope are computed by evaluating the weighted BRDF kernel model at the primary observation geometry (solar incident angle, reflection angle to the sensor). The result is a reflectance factor, equal to the ratio between the surface reflected radiance and the radiance from a purely reflecting Lambertian surface at the same geometry [Schaepman-Strub et al. 2006]. For conditions with no atmospheric scattering, this should be equivalent to a Lambertian albedo. Since the Lambertian albedo was used in the OCO-2 retrieval in Version 7 and earlier, the effective albedo computed in this way should be approximately consistent.

Evaluating the RPV kernel at the primary angle can be expressed as follows, where the subscript 0 values denote the solar and observer angles:

$$\rho_w\left(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0\right) = wr_s F_s\left(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0; \rho_0, \Theta, k\right) \quad (3\text{-}53)$$

$$\rho_s\left(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0\right) = sr_s F_s\left(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0; \rho_0, \Theta, k\right) \quad (3\text{-}54)$$

3.3.2.4.2 Ocean Surface a Priori in Glint Mode

Over water surfaces (defined as land fraction less than 20%), the surface is assumed to be described by a combination of two surface types: a Cox-Munk distribution of planar facets [Cox and Munk, 1954] plus a Lambertian surface. The latter is necessary because while the Cox-Munk surface can usually describe our observations quite well, it cannot perfectly fit the continua in all three spectral bands with a single free parameter (the surface wind speed).

There are a number of options that could be used to fit the surface. One could (unphysically) fit a separate wind speed for each band, but this still leaves minor problems with the slope of the continuum in a given band being inconsistently fit. Alternatively, one could fit the water index of refraction (and its slope) in each spectral band. We choose to fit a low-reflectivity Lambertian surface on top of the Cox-Munk surface. As over land, the water Lambertian surface has a mean albedo and albedo slope separately specified for each band. The *a priori* mean values are set to 0.02 in the O_2 A, weak CO_2, and strong CO_2 bands. We constrain the strong CO_2 mean albedo so it cannot take a value below 0.02, and we require non-negative Lambertian albedos in each band. The mean albedos in the other two bands have *a priori* uncertainties of 0.2. The slopes have a mean of zero and a 1-sigma prior uncertainty as over land.

For the Cox-Munk surface, the refractive index is taken to be constant within a band, with values of 1.331, 1.318, and 1.303 in the O_2 A, weak CO_2, and strong CO_2 bands, respectively. The *a priori* wind speed is taken from the ECMWF forecast, with a prior 1-sigma uncertainty of 6.325 m/s.

3.3.2.4.3 Chlorophyll Fluorescence

Over land surfaces, the state vector also contains the value of chlorophyll fluorescence at the surface at λ_{ref} = 755 nm, plus its slope with respect to wavelength. Our approach was introduced

in Frankenberg et al. [2012], where it was shown that ignoring chlorophyll fluorescence in X_{CO2} retrievals that use the O_2 A band would generally induced biases in the retrieved X_{CO2}, regardless of whether the O_2 A band was only used to constrain aerosols, surface pressure, or both. Specifically we assume that the SIF signal at the surface, SIF_{surf}, can be adequate parameterized across the O_2 A band as:

$$SIF_{surf}(\lambda) = \left[f_{755} + \gamma \left(\frac{10^7}{\lambda_{ref}} - 13245\, cm^{-1} \right) \right] I_{TOA}(\lambda_{ref}) \tag{3-55}$$

where λ is the wavelength in nm. The two state vector parameters are f_{755}, the value of fluorescence at the surface relative to the continuum at the reference wavelength, and the slope term γ. The *a priori* values and 1-sigma errors are set as:

$$\begin{aligned} f_{755} &= 0 \pm 0.008 \\ \gamma &= 0.0018 \pm 0.0007\, per\ cm^{-1} \end{aligned} \tag{3-56}$$

The fluorescence signal is modeled as isotropic and unpolarized. The RT model propagates the surface signal to TOA, as discussed in Section 3.2.3.4. It should be noted that the primary purpose of the fluorescence implementation here is to minimize biases in X_{CO2} and not provide the best possible fluorescence product. For fluorescence data users, fits based purely on Fraunhofer lines will be more robust and will be distributed within the IMAP pre-processor.

3.3.2.5 Wavelength Grid Offset

Based on our experience with GOSAT retrievals, there could be a wavelength offset. We measured the wavelength grid of OCO-2 on the ground. This is a wavelength (in μm) for each of the 1016 channels for each footprint and band. On orbit this may be slightly displaced due to instrumental and Doppler effects. This offset is fit separately for each spectral band, but a good first guess is necessary. If the offset is off by more than about one spectral sample, the fit will often fail to converge.

We will initially try using a simple instrument Doppler correction. If this correction is not enough, we will try and estimate the grid offset initial guess. We estimate this offset based on observing the actual position of a strong solar line near 12985.16326 cm^{-1} and assuming that the Doppler multiplier is the same for each band. The first guess is typically good to approximately 0.02 cm^{-1}, sufficient for a good first guess. The *a priori* uncertainty of this value is set relatively large, so in reality the L2 fit completely constrains this parameter (for each footprint and band).

3.3.2.6 Residual EOF Parameters

Note: This subsection describes a set of state vector elements that was introduced in Atmospheric CO_2 Observations from Space (ACOS) algorithm Version B3.3. It is not described in either O'Dell et al. [2012] or Crisp et al. [2012]. It replaces the older "empirical noise" formulation used in previous versions of the algorithm.

3.3.2.6.1 Using Empirical Orthogonal Functions to Minimize Fit Residuals

Owing to remaining inaccuracies of the spectroscopic input data and also to instrumental effects, we almost never obtain residuals consistent with the detector noise estimates, especially at higher SNR. If systematic residuals persist, the reduced χ^2 will thus scale with SNR and reach 1 only at very low SNR values. χ^2 values would thus not truly represent the goodness of the fit anymore and make filtering based on these values difficult.

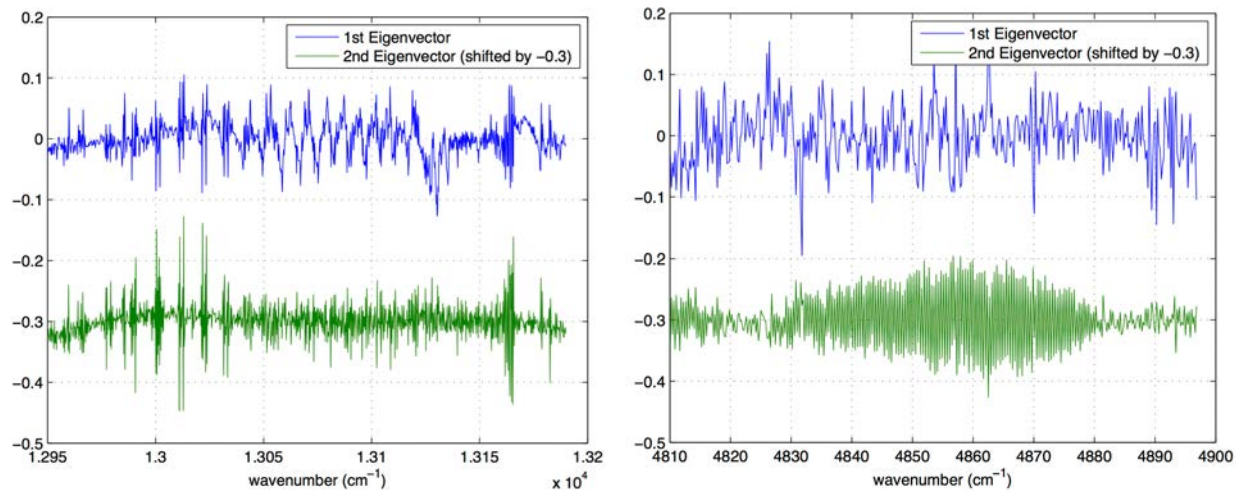

Figure 3-20. Example of first two leading eigenvectors for current GOSAT retrievals in the O_2 A band (**left**) and strong CO_2 band (**right**).

Based on experience with GOSAT retrievals, we adopted an alternative approach using a singular value decomposition of many fit residuals. The method works as follows: For a diverse training set of good fits (without compensating for systematic residuals), we compose three matrices ***M*** constituted by the spectral residuals of the spectral fits within each spectral band (typically >10000 soundings are used). Each matrix ***M*** is then decomposed into its eigenvectors using singular value decomposition:

$$\mathbf{M} = \mathbf{U}\ \mathbf{S}\ \mathbf{V}^{\mathbf{T}} \tag{3-57}$$

with the columns of ***U*** spanning orthonormal basis vector of the systematic residuals observed in the training dataset. By convention, the first eigenvector explains the largest fraction of the total variance, as indicated by descending order of singular values (the diagonal elements of ***S***). In most cases encountered here, the first singular vector will be very similar to the mean residual. However, the analysis of higher order singular vector can reveal other systematic features (be they instrumental or spectroscopic) and can sometimes contribute to the overall observed variance in residuals.

Examples of typical eigenvectors encountered in fit residuals with GOSAT data can be seen in Figure 3-20. In the case of the O_2 A band, these features are mostly related to errors in the oxygen spectroscopy, especially in terms of its line mixing behavior as well as collision-induced absorption and specific line shape model. The second eigenvector appears to be related to a change in line width, which might be an instrumental effect in this band (as the instrument line shape might vary if the field of view is not homogenously illuminated). For the strong CO_2 band, the first eigenvector is again mostly related to spectroscopy and potentially also the broad-band structure of the solar irradiance. The second eigenvector is related to an instrumental artifact that has been eliminated in the latest GOSAT L1B version. These examples show what can be learned from the systematic analysis of spectral residuals, and we expect the leading eigenvector to be similar for OCO-2, unless major changes in the spectroscopy are achieved by that time.

3.3.2.6.2 Retrieval Implementation

The procedure for including singular vectors in the L2 FP retrievals works as follows. We know that the systematic features can be explained by a linear combination of eigenvectors as illustrated above. We can then modify our original forward modeled radiance values $I_b(\boldsymbol{x})$, which represent a vector of radiances in a given spectral band b, by adding a linear combination of all eigenvector $\boldsymbol{u}_{i,b}$, multiplied by a scalar a_i. Here i represents the eigenvector number. The factors $a_{i,b}$ (for each band) are additional parts of the state vector to be retrieved. The corresponding Jacobians are simply the vectors $\boldsymbol{u}_i$. The modification to the modeled radiance is then simply:

$$\vec{I}_b'(\boldsymbol{x}) = \vec{I}_b(\boldsymbol{x}) + \sum_i a_i \vec{u}_i \tag{3-58}$$

For GOSAT, we used one eigenvector per band, and have achieved reduced χ^2 values close to 1 irrespective of signal level. For OCO-2, the number of eigenvectors used in the retrieval will depend on the fractional contribution to the overall variance, which will only be known once it is in orbit and in initial retrievals are made.

3.3.2.6.3 Advantages and Risks

One of the advantages of this method is that the goodness of the fit will be largely improved with better convergence. Theoretically, the inclusion of residual-pattern eigenvectors in the fit should be mostly cosmetic, but it can affect the weighting of different spectral regions and thereby also the retrieval. Especially broad-band systematic residuals that deviate from zero will be penalized strongly by the least-squares fit routine. The inclusion of the eigenvectors of the fit residuals ensures that most of the noise is weighted properly with the mean residual being close to flat.

The risk is that systematic problems may be hidden by the eigenvectors. Ideally, these features should not occur in the first place and our goal should be to minimize the need for eigenvectors as much as possible. Also, the generation of eigenvectors is somewhat circular. For every code or spectroscopy change, a large dataset has to be run without the EOF fit turned on to generate a training dataset for the computation of the updated eigenvectors.

3.4 Jacobians

In the inverse method we use, the entire K-matrix of the state vector is required, which represents the partial derivative of each spectral channel radiance with respect to each state vector element. Mathematically, this is expressed as:

$$K_{ij} = \frac{\partial F_i(\mathbf{x})}{\partial x_j} \tag{3-59}$$

where x_j is the state vector element j and F_i is the forward modeled radiance for channel i. The typical retrieval setup for OCO-2 has m = 3048 channels and n = 45 state vector elements, meaning that K is a (3048, 45) matrix.

There are two components to the calculation of the Jacobian. The LIDORT code we use is linearized: it can calculate the Jacobian of its output. This Jacobian is in terms of the inputs and outputs of LIDORT. The inputs to LIDORT are ***not*** the same as our state vector elements, although they are calculated from them. Likewise, the outputs from LIDORT are not the same as our forward model, although the forward model is calculated using LIDORT and the state vector elements.

We perform the mapping from state vector elements to the LIDORT input and the LIDORT output to the forward model automatically using the standard calculus chain rules implemented by automatic differentiation.

Automatic Differentiation

We make use of a technique called Rall numbers, as described in *Scientific and Engineering C++* by Barton and Hackman [1997] (named for the original book by L.B. Rall *Automatic Differentiation*, Springer-Verlag, Berlin, New York, 1981). In our code, instead of using a standard C *double* we use a new C++ class *AutoDerivative* that keeps track of both a value and the gradient of the value with respect to the state vector elements. All the normal mathematical functions are overloaded to use the chain rule to propagate the gradient through the calculation.

One of the advantages of this particular technique is that it works well with code that calculates Jacobians in other ways. This is one of the primary reason we selected this technique over some of the other automatic derivative techniques. For example, LIDORT is already calculating Jacobians—we have no interest in replacing anything in LIDORT. Instead, we calculate the Jacobian of the parameter up to LIDORT, use LIDORT itself to propagate through the RT calculation, and then take the results of LIDORT, wrap them up as *AutoDerivative* objects, and continue using this to propagate through the remaining forward model calculation.

Although the use of an *AutoDerivative* is computationally fast, it is not as fast as hand-coded derivative calculations. For locations in the code that profiling show as bottlenecks, we can replace the *AutoDerivative* calculation with hand-created code that gives the same results.

3.5 Inverse Method

3.5.1 Formulation and Implementation

As mentioned in Section 3.1, the notation and concepts of Rodgers [2000] are employed for the inverse method. This section builds on Section 2.1 of Connor et al. [2008]. The spectrum, or measurement vector $\boldsymbol{y}$, is expressed symbolically as $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{\varepsilon}$ where $\boldsymbol{x}$ is the state vector, $\boldsymbol{F}$ is the forward model, and $\boldsymbol{\varepsilon}$ is the vector of spectral errors due to the measurement and forward model.

The solution of the OCO-2 inverse method is the state vector $\hat{\boldsymbol{x}}$ with maximum *a posteriori* probability, given the measurement $\boldsymbol{y}$. Our inverse method employs the Levenberg-Marquardt modification of the Gauss-Newton method.

The operational inverse method consists of a set of routines, which are essentially mathematical and independent of the physics embodied by the measurement and state vectors. This implies that the structure of both vectors may be varied, so the routines are readily applied to other experiments, such as the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography or ground-based FTS [Boesch et al., 2006]. The ability to retrieve X_{CO2} from space-based and ground-based measurements using the same algorithm is critical for detecting and removing biases from the space-based data and forms a critical component of the OCO-2 validation strategy [Crisp et al., 2004].

The result of the inverse method is classified according to the parameter *outcome* as described in **Table 3-6**.

Table 3-6. Description of *outcome* values.

outcome	Description
1	Convergence was reached and the spectral fit was consistent with assumed uncertainties
2	Convergence was reached but the spectral fit was poor relative to assumed uncertainties
3	Convergence was not reached in the maximum allowed number of iterations
4	The maximum number of diverging steps (described below) was exceeded

To find the state vector that produces the maximum *a posteriori* probability, we minimize the following standard cost function *c*:

$$c = (\mathbf{y} - F(\mathbf{x}))^T \mathbf{S}_{\boldsymbol{\varepsilon}}^{-1} (\mathbf{y} - F(\mathbf{x})) + (\mathbf{x}_a - \mathbf{x})^T \mathbf{S}_a^{-1} (\mathbf{x}_a - \mathbf{x}) \tag{3-60}$$

where the weighting function matrix (Jacobian) is:

$$\mathbf{K} = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \tag{3-61}$$

and $\boldsymbol{x_a}$ is the *a priori* state vector, $\boldsymbol{S_a}$ is the *a priori* covariance matrix, and $\boldsymbol{S_\varepsilon}$ is the measurement covariance matrix. On each iteration *i*, we solve for the state vector update $\boldsymbol{dx_{i+1}}$ using a slightly modified form of Rodgers' Equation 5.36 to improve numerical accuracy by avoiding inversion of a large matrix:

$$((1 + \gamma)S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i)dx_{i+1} = [K_i^T S_\varepsilon^{-1} (y - F(x_i)) + S_a^{-1}(x_a - x_i)] \tag{3-62}$$

where γ is the Levenberg-Marquardt parameter. This method reduces to regular Gauss-Newton minimization when $\gamma = 0$. γ is initialized with a value of 10.0.

After each calculation of $\boldsymbol{dx_{i+1}}$, before using it to update $\boldsymbol{x_i}$, we assess the impact of forward model nonlinearity on the reduction in the cost function. To do this, we calculate the following quantity:

$$R = (c_i - c_{i+1})/(c_i - c_{i+1,FC}) \tag{3-63}$$

Here, c_i and c_{i+1} are the cost function from Equation (3-60) after iterations *i* and *i+1*, and $c_{i+1,FC}$ is calculated from Equation (3-60) assuming that $F(x+dx_{i+1}) = F(x) + Kdx_{i+1}$

This is simply the ratio of the actual reduction in the cost function *c* (between iterations *i* and *i+1*) to the forecast reduction in the cost function based on an assumption of linearity. For a perfectly linear forward model, the cost function reduction will equal the forecast cost function reduction, and *R* will equal unity. A step in which the cost function actually increases will yield $R < 0$. Cases for which $0 < R < 1$ are cases for which the cost function decreased, but not by as much as expected.

The use of *R* to assess $\boldsymbol{dx_{i+1}}$ and to change γ is an area of active research and may be modified in future versions of the algorithm. In the algorithm immediately post launch, the detailed treatment is as described in the following paragraph.

Cases with $R \leq 0.0001$ are considered divergent. γ is increased by a factor 10 and a new value of $\mathbf{dx_{i+1}}$ is calculated. If the total number of diverging steps passes a certain threshold, the entire

retrieval is abandoned, and *outcome* is set to 4. In cases with 0.0001< R, $\boldsymbol{dx_{i+1}}$ is used to update $\boldsymbol{x_i}$. If 0.0001< R <0.25, γ is increased by a factor 10 for the next iteration. For steps with 0.25< R <0.75, γ is left unchanged. Those cases with R >0.75 are deemed quite linear, and in these cases the Levenberg-Marquardt parameter γ is reduced by a factor of two.

After each successful, non-divergent iteration, we again test for convergence. To facilitate that, we compute the error variance derivative:

$$d\sigma_i^2 = \mathbf{dx_{i+1}}^T \hat{\mathbf{S}}^{-1} \mathbf{dx_{i+1}} \tag{3-64}$$

where $\hat{\boldsymbol{S}}$ denotes the covariance of the error in the retrieved state using the relation:

$$\hat{\mathbf{S}}^{-1} \mathbf{dx_{i+1}} \cong [\mathbf{K_i}^T \mathbf{S_\varepsilon}^{-1} (\mathbf{y} - F(\mathbf{x_i})) + \mathbf{S_a}^{-1}(\mathbf{x_i} - \mathbf{x_a})] \tag{3-65}$$

and $\boldsymbol{dx_{i+1}}$ is the state vector update assuming γ = 0 (this is generally larger than the actual step size $\boldsymbol{dx_{i+1}}$ computed for nonzero γ). $d\sigma_i^2$ is effectively the square of the state vector update in units of the solution variance.

If $d\sigma_i^2 < fn$ (where n is the number of state vector elements, and f ~1), convergence is reached. We then update the state vector a final time.

Lastly, we compute the retrieval covariance matrix $\hat{\boldsymbol{S}}$ and the averaging kernel matrix $\boldsymbol{A}$. $\hat{\boldsymbol{S}}$ is given by:

$$\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \tag{3-66}$$

and the averaging kernel matrix $\boldsymbol{A}$ is given by:

$$\mathbf{A} = \hat{\mathbf{S}}\, \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} \tag{3-67}$$

Finally, the degrees of freedom for signal are given by the trace of the matrix $\boldsymbol{A}$; the degrees of freedom for the CO_2 profile are the trace of the CO_2-only submatrix.

We note that the OCO-2 retrieval problem as described is underdetermined. There are typically ≤ 2 degrees of freedom for the CO_2 profile (compared to 20 profile levels) and ~20 total degrees of freedom (compared to ~50 elements). However, the use of an *a priori* constraint guarantees that the problem is well posed and well conditioned. Although there are no more than two degrees of freedom for the CO_2 profile (and thus there is little ability to discriminate between altitudes), the profile, rather than just the CO_2 column, is formally retrieved. This is to allow for the large variations of CO_2 in the boundary layer; such variations can produce bias in the results if a fixed profile shape is used in the retrieval.

3.5.2 Goodness of Spectral Fit

The goodness of fit is estimated from χ^2 for each spectral band:

$$\chi_i^2 = \frac{\frac{1}{n}\Sigma(y - F(x))^2}{\frac{1}{n}\Sigma\varepsilon^2} \tag{3-68}$$

If $\chi^2_i < max\_chi2$ for all i, then the fit is provisionally acceptable, and *outcome* is set to 1; otherwise, *outcome* = 2.

Because of the presence of non-random spectral errors, the values of *max_chi2* must be determined empirically by inspection of a large set of retrieved results. Thus, they (and hence the assessment of fit quality) are subject to refinement. Therefore users may wish to include retrievals with both *outcome* = 1 and *outcome* = 2 in their initial analyses.

3.6 X_{CO2}, Characterization, and Error Analysis

3.6.1 Pressure Weighting Function

We define the pressure weighting function $\boldsymbol{h}$ to relate the local CO_2 mixing ratio values specified on the discrete pressure levels to the profile-weighted average such that $X_{CO2} = \boldsymbol{h}^T \hat{\mathbf{x}}$. This procedure is described in detail in O'Dell et al. [2012]. We only briefly summarize the procedure here—the interested reader is referred to that work for details. Recalling that X_{CO2} is defined as the CO_2 column divided by the dry air column, it may be seen that the vector $\boldsymbol{h}$ represents the pressure intervals assigned to the state vector levels, normalized by the surface pressure and corrected for the presence of water vapor such that $\hat{\mathbf{x}}$ is weighted by the number of moles of dry air per layer.

First, we describe our pressure-leveling scheme. This is a simple sigma coordinate system where:

$$p_i = b_i p_{surf} \tag{3-69}$$

where the b_i constants are fixed. Currently, we have N_{lev}=20 levels, and b = {0.0001, 1/19, 2/19, …, 18/19,1.0}. Thus, the lowest level is equal to the surface pressure (this is slightly easier to work with than a previous approach we used, wherein the pressure levels were fixed). For p_{surf} = 1000 hPa, the pressure width of all the layers (except the top layer) is 52.63 hPa; for the top layer it is 52.53 hPa. The typical model lid is therefore roughly 0.1 hPa.

In order to calculate $\boldsymbol{h}$ on model levels (i.e., layer boundaries), we must relate the column of any gas to its dry mixing ratio profile, and the pressure interval in each layer must be conceptually divided by assigning fractions of it to the two adjacent levels in such a way that integrating over all levels conserves both total pressure and CO_2 column.

The column of any gas X may be shown to be given by:

$$C_X = \sum_i \frac{u_i^X}{g_i\, M_{dry}\left(1 + \varepsilon\, u_i^{H2O}\right)} \Delta p_i \tag{3-70}$$

where u_i is the mixing ratio on level i, Δp_i is the pressure interval assigned to that level, g is the local acceleration due to gravity, and M_{dry} is the mean molecular weight of dry air (taken to be 0.0289644 kg/mol). It is straightforward to show that this is equivalent to:

$$C_X = \sum_i \frac{u_i^X (1 - q_i)}{g_i\, M_{dry}} \Delta p_i \tag{3-71}$$

where q_i is the specific humidity of dry air at level i. Appendix A of O'Dell et al. [2012] details the calculation of the pressure weighting function on atmospheric levels and is not repeated here.

3.6.2 X_{CO2}

Finally, X_{CO2} and its uncertainty σ_{XCO2} are calculated. The CO_2 dry air mole fraction X_{CO2} is obtained by averaging the CO_2 profile, weighted by the pressure weighting function $\boldsymbol{h}$ such that $X_{CO2} = \boldsymbol{h}^T \hat{\mathbf{x}}$. The formal error variance in the retrieved X_{CO2} is therefore given by:

$$\sigma^2_{XCO2} = \mathbf{h}^T \hat{\mathbf{S}}\, \mathbf{h} \tag{3-72}$$

A series of standard error analysis calculations are performed as part of each retrieval to characterize the retrieval results and quantify their uncertainties. The retrieval uncertainties and averaging kernels are calculated from the measurement Jacobian $\boldsymbol{K}$ of Section 3.5.1 evaluated at the retrieved state, the measurement covariance matrix, and the *a priori* covariance matrix. From these we compute the column averaging kernel $\boldsymbol{a_{CO2}}$, the uncertainty due to smoothing and interference $\tilde{\boldsymbol{a}}_c$, the correlation of non-CO_2 state vector elements to X_{CO2} $\tilde{\mathbf{s}}$, and the components of X_{CO2} variance due to measurement error, smoothing, and interference. These last six quantities are defined and discussed below. A complete list of operational retrieval products is given in **Table 3-7**.

Table 3-7. Inverse method products recorded with each sounding.

Name	Size	Comment
χ^2	3	Sum of squares of normalized residuals in each spectrometer
$\hat{\mathbf{x}}$	n	Retrieved state vector
$\hat{\boldsymbol{S}}_{ii}$	n	Diagonal elements of $\hat{S}$ (error covariance matrix)
$\hat{S}_{CO2}$	q^2	CO_2-only submatrix of $\hat{S}$
$\tilde{\mathbf{s}}_1$	n-q	Correlation of X_{CO2} with non-CO_2 elements of $\boldsymbol{x}$
$\boldsymbol{A_{CO2}}$	q^2	CO_2-only submatrix of averaging kernel ($\boldsymbol{A}$)
$\boldsymbol{a_{CO2}}$	n	Column averaging kernel
$\tilde{\boldsymbol{a}}_c$	n	Error in X_{CO2} due to smoothing and interference
X_{CO2}	1	Column-weighted CO_2 dry air mole fraction
σ^2_m	1	Variance of X_{CO2} due to measurement noise
σ^2_s	1	Variance of X_{CO2} due to smoothing
σ^2_i	1	Variance of X_{CO2} due to interference
σ^2_{XCO2}	1	Total variance of X_{CO2} ($\sigma^2_m + \sigma^2_s + \sigma^2_i$)
d_f	1	Degrees of freedom (full state vector)
d_{CO2}	1	Degrees of freedom (CO_2 profile only)

3.6.3 The Column Averaging Kernel a_{CO2}

Let $\hat{\mathbf{u}}$ be the retrieved CO_2 mixing ratio profile and $\hat{\mathbf{e}}$ be the retrieved vector of all non-CO_2 quantities:

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{e}} \end{pmatrix} \tag{3-73}$$

and let:

$$\mathbf{A}_{ij} = \frac{\partial \hat{\mathbf{x}}_i}{\partial \mathbf{x}_j} \text{ i, j = 1,n} \tag{3-74}$$

so:

$$\frac{\partial \mathrm{X}_{CO2}}{\partial \mathbf{x}_j} = (\mathbf{h}^T \mathbf{A})_j \tag{3-75}$$

Now consider only the CO_2 part of the state vector, the first q elements u_i , i=1 to q, and define the column averaging kernel:

$$(\mathbf{a}_{CO2})_j = \frac{\partial \boldsymbol{X}_{CO2}}{\partial \mathbf{u}_j} \frac{1}{\mathbf{h}_j} = (\mathbf{h}^T \mathbf{A})_j \frac{1}{\mathbf{h}_j} \tag{3-76}$$

The column averaging kernel $\boldsymbol{a}_{CO2}$ has the property that its elements all equal 1 in the ideal case, where the retrieved X_{CO2} responds to changes in $\boldsymbol{u}$ exactly as the true value of the profile-weighted mixing ratio. For a real retrieval, the elements of $\boldsymbol{a}_{CO2}$ may be more or less than 1, and will have values much less than 1 in regions where the *a priori* CO_2 profile is important.

3.6.4 Smoothing and Interference due to the State Vector $\tilde{\boldsymbol{a}}_c$

The vector $\tilde{\boldsymbol{a}}_c$ captures the smoothing and interference (or cross talk) errors in X_{CO2} due to each element of $\hat{\mathbf{x}}$. Thus, $\tilde{\boldsymbol{a}}_c$ reveals the sensitivity of the retrieved X_{CO2} value to uncertainties in elements of the state vector. It may be derived from the full averaging kernel matrix as follows.

The error in X_{CO2} is given by:

$$\Delta \boldsymbol{X}_{CO2} = \mathbf{h}^T(\hat{\mathbf{x}} - \mathbf{x}) = \mathbf{h}^T(\mathbf{A}_{uu} - \mathbf{I})(\mathbf{u} - \mathbf{u}_\mathbf{a}) + \mathbf{h}^T \mathbf{A}_{ue}(\mathbf{e} - \mathbf{e}_\mathbf{a}) + \varepsilon_u \tag{3-77}$$

Equation (3-77) is an adaptation of Equation (7) in Rodgers and Connor [2003]. The first term in Equation (3-77) represents smoothing error, the second interference error, and ε_u all other sources of error. Here, $\boldsymbol{A}_{uu}$ and $\boldsymbol{A}_{ue}$ are submatrices of $\boldsymbol{A}$, representing the CO_2-only component and the cross talk components (those which mix elements of the CO_2 profile $\boldsymbol{u}$ and the non-CO_2 elements $\boldsymbol{e}$), respectively. $\boldsymbol{I}$ is the identity matrix.

It follows that the error in X_{CO2} due to each state vector element is given by:

$$\tilde{\mathbf{a}}_{c,j} = \begin{cases} \mathbf{h}^T(\mathbf{A} - \mathbf{I})\sigma_j, j = 1, q \\ \mathbf{h}^T \mathbf{A} \sigma_j, j = q+1, n \end{cases} \tag{3-78}$$

where σ_j is the error in element j. Since $h_j \equiv 0$ for $j > q$, the full matrix $\boldsymbol{A}$ may be used in place of the submatrices in Eq. (3-77).

Alternatively, Eq. (3-78) may be written:

$$\mathbf{a}_{c,j} = \begin{cases} (\frac{\partial X_{CO2}}{\partial \mathbf{u}_j} - \mathbf{h}_j)\sigma_j = (\mathbf{a}_{CO2,j} - 1)\mathbf{h}_j \sigma_j, j = 1,q \\ \frac{\partial X_{CO2}}{\partial \mathbf{e}_j}\sigma_j, j = q+1,n \end{cases} \tag{3-79}$$

The first q elements of $\tilde{\boldsymbol{a}}_c$, corresponding to the CO_2 profile, are components of the smoothing error. The remaining elements represent the interference, or cross talk error.

3.6.5 Correlation of X_{CO2} with non-CO_2 State Vector Elements $\tilde{\mathbf{s}}_1$

We define $\tilde{\mathbf{s}}_1$, the correlation of X_{CO2} with the state vector elements, to aid the diagnosis and understanding of cross talk. Define a matrix $\boldsymbol{H}$ with dimension $n \times (n+1)$ $n \times (n+1)$ such that:

$$\mathbf{H}_{ij} = \begin{cases} \mathbf{h}_i, j = 1 \\ 1, i = j-1, j > 1 \\ 0, else \end{cases} \tag{3-80}$$

Thus:

$$\mathbf{H}^{\mathrm{T}}\hat{\mathbf{x}} = \begin{pmatrix} X_{CO2} \\ x_1 \\ x_2 \\ \bullet \\ \bullet \\ x_n \end{pmatrix} \tag{3-81}$$

and the correlation matrix of $\boldsymbol{H}^T\hat{\mathbf{x}}\ \hat{\mathbf{x}}$ is given by $\boldsymbol{H}^T\hat{\boldsymbol{C}}\,\boldsymbol{H}$, where $\hat{\boldsymbol{C}}$ is the correlation matrix corresponding to $\hat{\boldsymbol{S}}$.

$\tilde{\mathbf{s}}_1$ is the first row of $\boldsymbol{H}^T\hat{\boldsymbol{C}}\,\boldsymbol{H}$, where $\tilde{\mathbf{s}}_1 = (1\ \rho_{XCO2,j} \ldots)$ for $j = 1, n$. The second and subsequent elements of $\tilde{\mathbf{s}}_1$ are the correlation coefficients of X_{CO2} with each element of the state vector.

3.6.6 Components of X_{CO2} Variance

In Section 3.6.2 above, the variance in X_{CO2} is given in Equation (3-72) for pressure weighting function $\boldsymbol{h}$ and *a posteriori* error covariance $\hat{\boldsymbol{S}}$. We also calculate the component parts of the variance, as described next.

In Equation (3.16) of Rodgers [2000], the total error is broken up into measurement, smoothing, and forward model error (which includes both radiative transfer and the instrument). The operational L2 code has no explicit information on forward model error, though it may be implicitly included in the measurement error estimate provided as input. Formally then, we separate the error into measurement and smoothing components. Further, smoothing due to the full state vector may logically be separated into smoothing due to CO_2 itself and interference due to non-CO_2 elements, as described in Rodgers and Connor [2003]. Thus, we calculate variance components due to spectral error (measurement and implicit model error), smoothing (due to CO_2), and interference.

First, define the contribution function as:

$$\mathbf{G_y} = \hat{\mathbf{S}}\mathbf{K}^{\mathbf{T}}\mathbf{S}_{\varepsilon}^{-1} \tag{3-82}$$

then the measurement error covariance is:

$$\hat{\mathbf{S}}_{\mathbf{m}} = \mathbf{G_y}\mathbf{S}_{\varepsilon}\mathbf{G}_{\mathbf{y}}^{\mathbf{T}} \tag{3-83}$$

the smoothing error due to CO_2 is:

$$\hat{\mathbf{S}}_{\mathbf{s}} = (\mathbf{A_{CO2}} - \mathbf{I})\mathbf{S_{a,CO2}}(\mathbf{A_{CO2}} - \mathbf{I})^{\mathbf{T}} \tag{3-84}$$

and the interference error is:

$$\hat{\mathbf{S}}_{\mathbf{i}} = \mathbf{A_{ue}}\mathbf{S_{ae}}\mathbf{A}_{\mathbf{ue}}^{\mathbf{T}} \tag{3-85}$$

where subsets of the averaging kernel $\boldsymbol{A}$ and *a priori* covariance $\boldsymbol{S_a}$ are defined as follows:

$$\mathbf{A_{CO2}} = \mathbf{A(1{:}nz,\ 1{:}nz)} \text{ for nz = \# levels in } CO_2 \tag{3-86}$$

$$\mathbf{A_{ue}} = \mathbf{A(1{:}nz,\ nz{+}1{:}n)} \qquad \text{for n = full length of state vector} \tag{3-87}$$

$$\mathbf{S_{aCO2}} = \mathbf{S_a(1{:}nz,\ 1{:}nz)} \tag{3-88}$$

$$\mathbf{S_{ae}} = \mathbf{S_a(nz{+}1{:}n,\ nz{+}1{:}n)} \tag{3-89}$$

Finally, the variance components are determined. The measurement error is given by:

$$\sigma^2{}_{\mathrm{m}} = \mathbf{h}^{\mathrm{T}}\ \hat{\mathbf{S}}_{\mathbf{m}}\ \mathbf{h} \tag{3-90}$$

the smoothing due to CO_2 is given by:

$$\sigma^2{}_{\mathrm{s}} = \mathbf{h}^{\mathrm{T}}\ \hat{\mathbf{S}}_{\mathbf{S}}\ \mathbf{h} \tag{3-91}$$

and the interference is given by:

$$\sigma^2{}_{\mathrm{i}} = \mathbf{h}^{\mathrm{T}}\ \hat{\mathbf{S}}_{\mathbf{i}}\ \mathbf{h} \tag{3-92}$$

These components of variance have the property that:

$$\sigma_{XCO2}^2 = \sigma_m^2 + \sigma_s^2 + \sigma_i^2 \tag{3-93}$$

4 Algorithm Specification

The purpose of this section is to describe implementation details of the remote sensing algorithms that produce the OCO-2 Level 2 end-user Diagnostic (L2Dia) and Standard (L2Std) data products. The science basis for these algorithms is provided in Section 3 of this document. The algorithm specification provides a software description of that science as implemented in the operational ground system—the Science Data Operations System (SDOS). The intent of the algorithm specification is to capture the as-built operational implementation of the algorithm.

4.1 Data System Context

The SDOS Processing System consists of a set of Product Generation Executables (PGEs) and the infrastructure that initiates and routes data between them. Conceptually, the PGEs can be grouped into processing levels.

- Level 0 Processing prepares incoming datasets for higher-level processing.
- Level 1 Processing generates engineering data products and calibrated, geolocated science measurements.
- Level 2 Preprocessing and Sounding Selection selects the soundings for which to analyze in terms of X_{CO2}. SIF retrievals are also generated in the preprocessing step.
- Level 2 Processing generates the X_{CO2} science results.

A full description of the L2 PGEs is provided in Section 4.7 below.

4.2 Level 2 Preprocessing

As stated above, three preprocessing steps are run on the OCO-2 data, but within L2. The first of these is the ABO2 Preprocessor whose main function is to identify heavily cloudy and aerosol-laden scenes for exclusion from FP analysis, but plays other roles as well. The IMAP preprocessor serves as a complementary cloud-screening algorithm, and simultaneously makes retrievals of SIF. The Sounding Selection algorithm takes input from both of these processors, as well as L1B data and meteorological data in order to select which soundings to run through the FP algorithm. This is because the FP algorithm is computationally intensive, and it is not currently feasible to run all soundings through it, so care must be taken to only select soundings with a high likelihood of yielding useful X_{CO2} retrievals.

4.2.1 ABO2 Preprocessor

The O_2 A band cloud screening algorithm was developed at Colorado State University under the ACOS program. It employs a fast Bayesian retrieval to estimate surface pressure and surface albedo, assuming clear-sky conditions with only molecular Rayleigh scattering, from high-resolution spectra of the O_2 A band near 765 nm. The estimated surface pressure, surface albedo and the chi-squared goodness-of-fit statistic are used to flag scenes as cloudy, clear, or indeterminate [Taylor et al., 2011]. Further details of the ABO2 technique are described in the ABO2 Algorithm ATBD as well as an upcoming publication [Taylor et al., 2015*, in prep*].

The basic method is that, absent clouds or aerosols, the surface pressure of a clear scene can be determined to within 2 – 5 hPa accuracy using the O_2 A band spectrum of reflected sunlight. This is because of the strong oxygen absorption features that are present in this band. When surface pressure is higher, the absorption features are deeper for a given observation geometry.

When clouds or aerosols are present, they change the path lengths for most photons, either via shortening or lengthening, such that the retrieved surface pressure can be very different from the expected value based on a meteorological forecast, and also the O_2 A spectrum itself cannot be well-fit with a clear-sky assumption.

The key screening parameters are the retrieved minus *a priori* surface pressure, as well as the goodness of fit as indicated thorugh a chi^2 metric. Over ocean away from glint observations, the retrieved surface albedo is also used (because oceans are only bright due to the presence of clouds or aerosols away from the glint spot). These parameters interpreted internally via thresholds to provide a simple cloud flag. As the primary objective of the ABO2 preprocessor is to flag as cloudy scenes that are obviously cloudy or aerosol-contaminated, thresholds are set rather loosely so scenes that have some cloud or aerosol contamination will sometimes pass the filter. Furthermore, the key screening parameters are used as inputs to the genetic algorithm used for sounding selection [Mandrake, 2013], described in section 4.2.2.

4.2.2 IMAP Preprocessor (and SIF Retrievals)

The IMAP-DOAS pre-processor is a fast non-scattering retrieval algorithm that fits vertical column densities of CO_2 and H_2O [Frankenberg et al., 2005] as well as a Fraunhofer-line based SIF retrieval algorithm based on Frankenberg et al. [2011]. CO_2 and H_2O fits are performed independently in the weak and strong CO_2 bands, respectively. More details can be found in the IMAP Algorithm ATBD.

The key screening parameters are both the CO_2 and well as the H_2O ratios, which are the ratios of retrieved CO_2 and H_2O columns in the weak and strong band. These ratios deviate from unity if significant scattering is present in the atmosphere and can thus be used to identify scattering on top of the ABO2 cloud-screen and both have been shown to be complementary.

The SIF fits are performed independently at 757 and 771 nm and will be provided in the IMAP-DOAS L2 files.

4.3 Level 2 Full Physics Inputs

The L2 software utilizes both the L1B and ECMWF resampled products generated earlier in the processing pipeline. The L1B file provides radiometrically corrected spectra and geolocation information. The ECMWF Resampled product provides pressure, temperature and specific humidity profiles derived from ECMWF forecasted or reinterpreted products as described in Section 3.3.2.2. The ECMWF is interpolated in time and location along the OCO-2 orbit track to match the L1B geolocation. More information about these products can be found in the L1B Algorithm Specification Document. The L2 software also utilizes the additional static data sources:

- ABSCO tables—tabulated absorption coefficients providing gas optical properties information, as described in Section 3.2.1
- Solar spectrum information, as described in Section 3.2.2
- *A priori* CO_2 profiles, as described in Section 3.3.2.1
- MERRA *a priori*—aerosol optical depth data for the world interpreted on a monthly bases from MERRA products, as described in Section 3.3.2.3
- Aerosol Properties—scattering and optical properties of scattering particles, derived from MERRA data, as described in Section 3.3.2.3
- *a priori* surface properties, as described in Section 3.3.2.4
- *a priori* wavelength grid offset, as described in Section 3.3.2.5

- Residual empirical orthogonal function parameters, as described in Section 3.3.2.6
- *a priori* values for atmospheric components not coming from other products
  - Fluorescence
  - Coxmunk windspeed, Lambertian albedo
  - Zero Level Offset
- *a priori* covariance values for all quantities, as described in Section 3.3.2

4.3.1 Requirements on Inputs

The L1B and ECMWF input data products are required to be:

- Time-ordered—soundings occur in increasing time order
- HDF5—Hierarchical data format v5
- Single operating mode—only nadir, glint, or target soundings in one file
- Single orbit—only a single orbit of data
- Quality control flags set—quality flags do not have fill values
- Geolocated—geolocation has already been performed on pointing information and is included in L1B files

4.3.2 Attributes of Input Products

Attributes of the L1B Science and ECMWF input data products are given in **Table 4-1** and **Table 4-2**, respectively.

Table 4-1. L1B Science product attributes.

L1B Science Product	
Description	The spectra in the L1B product include factors that (1) correct for the curvature of the signal on the instrument image plane, (2) apply the radiometric calibration to the detector counts in the science telemetry, and (3) eliminate the effects of spatial cross talk in the spectrometer signal
Long Name	Spatially ordered, geolocated, calibrated science spectra
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	1.37 GB

Table 4-2. ECMWF product attributes.

ECMWF Product	
Description	ECMWF forecast data resampled to the location of the OCO-2 measurements
Long Name	Spatially ordered, geolocated meteorological forecasts
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	1.60 GB

4.4 Level 2 Full Physics Outputs

The L2 FP PGE produces a single L2 FP output file in HDF format for each sounding. These files are then aggregated by the L2Std PGE into products destined for distribution at the Distributed Active Archive Center.

Attributes of the L2 Single Sounding, L2Dia, and L2Std output data products are given in **Table 4-3**, **Table 4-4**, and **Table 4-5**, respectively.

Table 4-3. L2 Single Sounding product attributes.

L2 Single Sounding Product	
Description	Direct output of L2 FP process per sounding. Contains X_{CO2}, atmospheric profiles, error estimates, and the state vector
Long Name	L2 FP single sounding output
Frequency	1 per sounding processed
Format	HDF5
Average orbit volume	1.0 GB
Data groups*	Product Metadata RetrievalHeader RetrievalResults RetrievedStateVector SpectralParameters

**see Data User's Guide*

Table 4-4. L2Dia product attributes.

L2 Diagnostic Product	
Description	Geolocated retrieved state vectors with CO_2 column averaged dry air mole fraction - physical model. Geolocated retrieved CO_2 column averaged dry air mole fraction - physical model.
Long Name	Geolocated X_{CO2} and state vectors - physical model
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	2.09 GB
Data groups*	Product Metadata L1bScSoundingReference RetrievalHeader RetrievalGeometry PreprocessingResults RetrievalResults AlbedoResults DispersionResults AerosolResults RetrievedStateVector SpectralParameters

**see Data User's Guide*

Table 4-5. L2Std product attributes.

L2 Standard Product	
Description	Geolocated retrieved CO_2 column averaged dry air mole fraction - physical model
Long Name	Geolocated X_{CO2} - physical model
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	1.32 GB
Data groups*	Product Metadata L1bScSoundingReference RetrievalHeader RetrievalGeometry PreprocessingResults RetrievalResults AlbedoResults DispersionResults AerosolResults SpectralParameters

**see Data User's Guide*

4.5 Description of Full Physics Major Code Sections

The FP code carries out the following steps to process a single sounding:

- Load Lua configuration, as described in Section 4.6
- Set up inputs as described in Section 4.3.2 and class connections from configuration
- Perform radiative transfer, as described in Section 3.2.3
- Apply solar model on spectra, as described in Section 3.2.2
- Model instrument effects on spectra, as described in Section 3.2.4
- Non-linear least squares inversion to compare modeled and measured spectra, as described in Section 3.5
- Update state vector, as described in Section 3.5.1
- Push state vector updates into models, as described in Section 3.3
- Check for convergence, as described in Section 3.5.2
- If converged perform error analysis, as described in Section 3.6
- Write L2 Single Sounding Product output file, as described in Section 4.4

Figure 4-1 shows the flow of data through the various L2 FP software components. It is meant to give a more detailed picture of how the L2 code is organized. A diagram of the algorithm control flow can be seen in Figure 3-3.

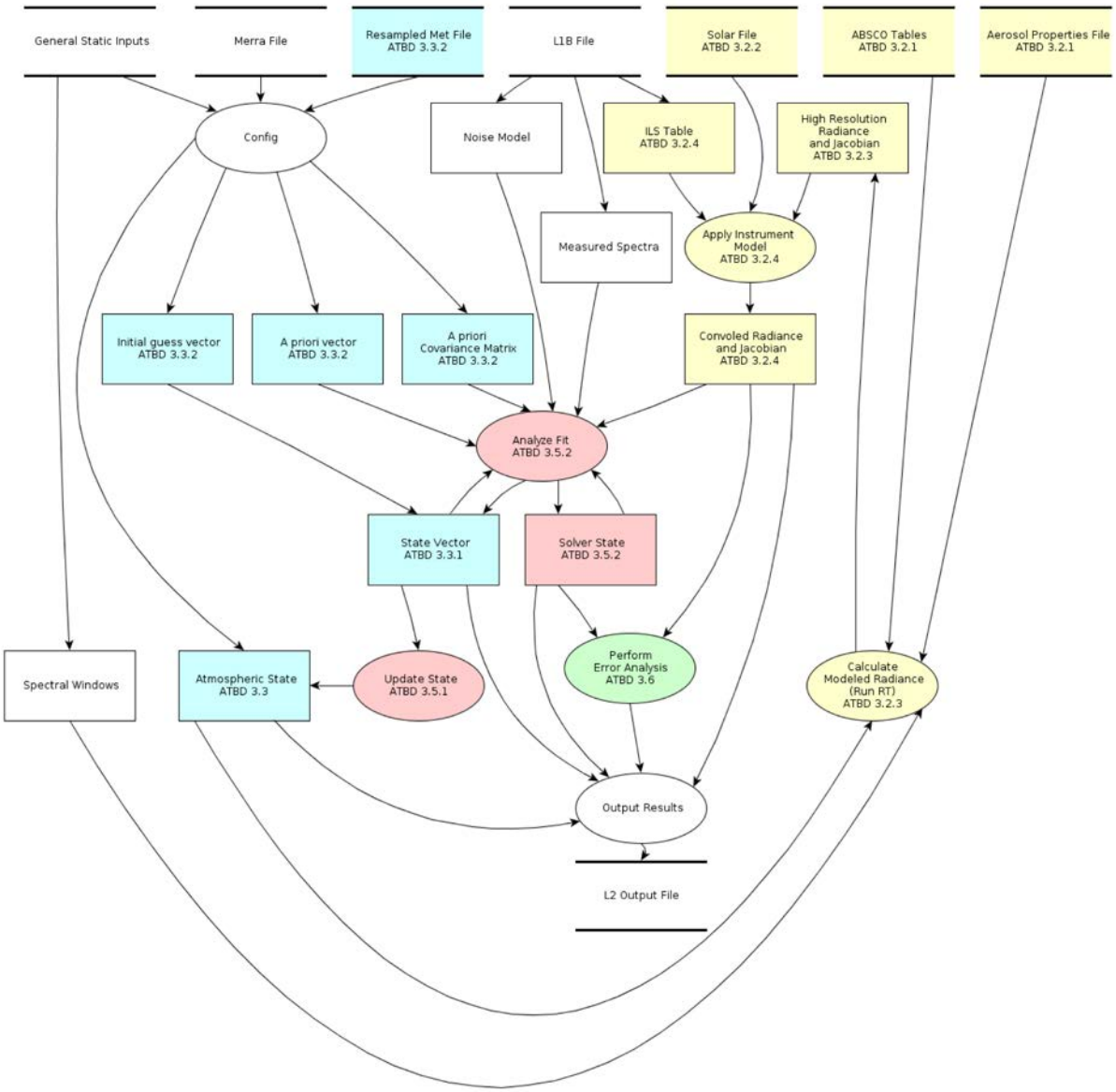

Figure 4-1. Data flow through the L2 FP software components.

4.6 Lua Configuration

Lua is a lightweight, embeddable scripting language. It was chosen to manage the L2 FP configuration over a declarative approach due to its flexibility. The L2 code itself was written in a manner such that the C++ portions are components that have minimal dependency on each other and adhere to well defined interfaces. The Lua configuration selects which components are used and makes them aware of each other. The configuration files themselves are written in layers where more specific configurations extend the general. At the bottom level are configurations that implement instrument specific arrangements of L2 FP components. These files specify which files are involved in the processing as well as algorithmic choices. At this level the Lua code looks like declarative keyword/value configuration files. This approach combines flexibility with simplicity for end users.

The configuration loads input data, building *a priori* and initial guess vectors as it processes components. Components are connected to each other and passed their input in a hierarchal manner. Dependent components are loaded first so they are initialized and can be passed to components that need them. The top most level is the OCO forward model component. It is created last and when ready called to start iterative execution as described in previous sections.

4.7 Level 2 Processing PGEs

A schematic illustrating the L2 Processing PGEs is given in Figure 4-2 below.

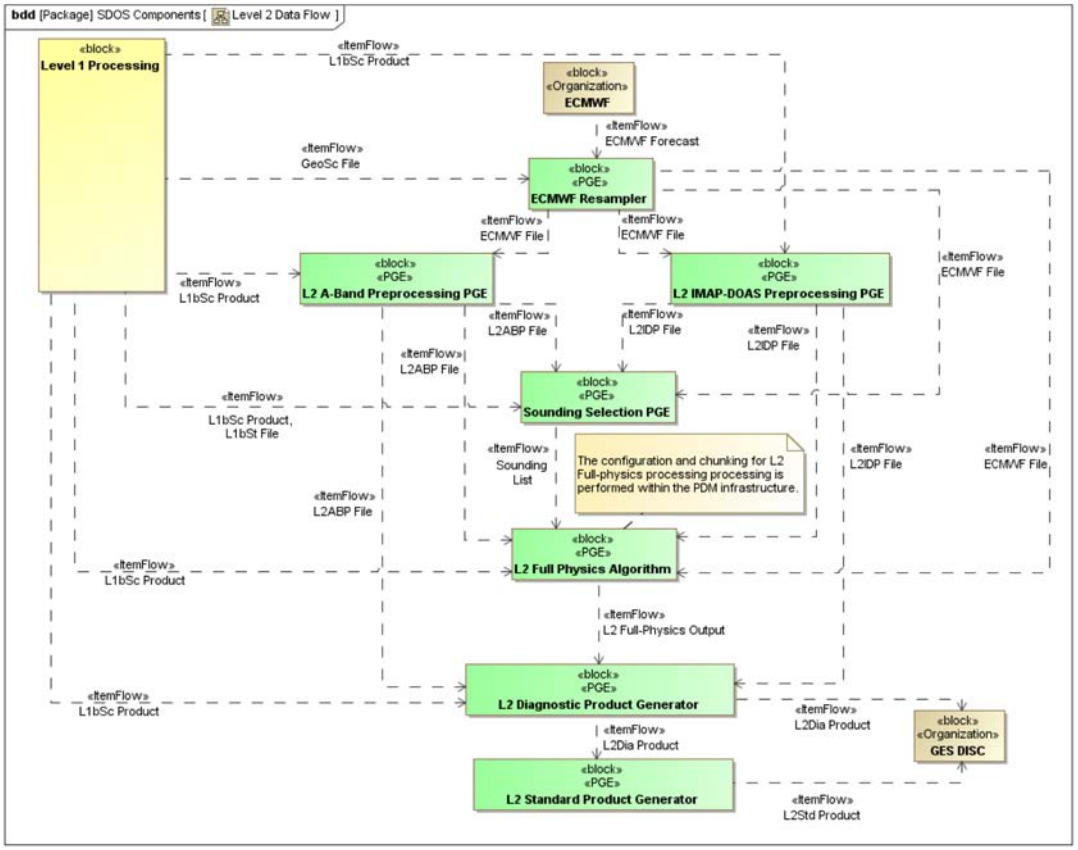

Figure 4-2. L2 Processing overview.

4.7.1 ECMWF Resampler

Purpose: The ECMWF Resampler extracts the meteorological data relevant to a single orbit from a global ECMWF dataset.

Connectivity:

- Receives an ECMWF Forecast file from ECMWF
- Receives a GeoSc file from the Geolocation PGE
- Provides an ECMWF file to the L2 A-Band Preprocessing PGE, the L2 Iterative Maximum *a Posteriori* Differential Optical Absorption Spectroscopy (IMAP-DOAS) Preprocessing PGE, the Sounding Selection PGE, and the L2 FP algorithm

Multiplicity: The Processing System runs the ECMWF Resampler once for each L1BSc product.

4.7.2 L2 A-Band Preprocessing PGE

Purpose: The L2 A-Band Preprocessing PGE derives attributes of each sounding using a retrieval on the ABO2 measurement only.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an ECMWF file from the ECMWF Resampler
- Provides an L2ABP file to the Sounding Selection PGE, the L2 FP algorithm, and the L2Dia Product Generator

Multiplicity: The Processing System runs the L2 A-Band Preprocessing PGE once for each L1BSc product.

4.7.3 L2 IMAP-DOAS Preprocessing PGE

Purpose: The L2 IMAP-DOAS Preprocessing PGE derives attributes of each sounding using an IMAP-DOAS retrieval.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an ECMWF file from the ECMWF Resampler
- Provides an L2IDP file to the Sounding Selection PGE, the L2 FP algorithm, and the L2 Diagnostic Product Generator

Multiplicity: The Processing System runs the L2 IMAP-DOAS Preprocessing PGE once for each L1BSc product.

4.7.4 Sounding Selection PGE

Purpose: The Sounding Selection PGE identifies measurements for the L2 FP Algorithm to use to generate X_{CO2} values.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an L1BSt file from the L1BStat PGE
- Receives an ECMWF file from the ECMWF Resampler
- Receives an L2ABP file from the L2 A-Band Preprocessing PGE
- Receives an L2IDP file from the L2 IMAP-DOAS Preprocessing PGE
- Provides an L2Lst file to the L2 FP algorithm
- Generates an L2Sel file that is used for analysis only

Multiplicity: The Processing System runs the Sounding Selection PGE once for each L1BSc product.

4.7.5 L2 FP Algorithm

Purpose: The L2 FP algorithm retrieves X_{CO2} for selected science measurements.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an ECMWF file from the ECMWF Resampler
- Receives an FP Configuration file from the FP Populator

- Receives an L2ABP file from the L2 A-Band Preprocessor PGE
- Receives an L2IDP file from the L2 IMAP-DOAS Preprocessor PGE
- Generates an L2 FP Output file for use by the FP Diagnostic Product Generator

Multiplicity: The Processing System runs the L2 FP algorithm once for each measurement to be processed.

4.7.6 FP Diagnostic Product Generator

Purpose: The FP Diagnostic Product Generator extracts information from The L1BSc product, the L2ABP file, the L2IDP file, and all the L2 FP Output files for an orbit and saves that information to a single file.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an L2ABP file from the L2 A-Band Preprocessor PGE
- Receives an L2IDP file from the L2 IMAP-DOAS Preprocessor PGE
- Receives a set of L2 FP Output files from the L2 FP Algorithm, one file for every measurement processed for a single orbit
- Produces an L2Dia product for use by the FP Standard Product Generator and delivery to the GES DISC

Multiplicity: The Processing System runs the FP Diagnostic Product Generator once for each L1BSc product.

4.7.7 FP Standard Product Generator

Purpose: The FP Standard Product Generator subsets the information contained in an L2Dia product to generate an L2Std product.

Connectivity:

- Receives an L2Dia product from the FP Diagnostic Product Generator
- Produces an L2Std product for delivery to the GES DISC

Multiplicity: The Processing System runs the FP Standard Product Generator once for each L1BSc product.

4.7.8 Level 2 PGE Multiplicity Summary

Table 4-6 summarizes the number of instances of L2 PGEs run during a nominal day.

Table 4-6. Daily L2 PGE instances.

PGE	Total
ECMWF Resampler	1/L1BSc product
L2 A-Band Preprocessor PGE	1/L1BSc product
L2 IMAP-DOAS Preprocessor PGE	1/L1BSc product
Sounding Selection PGE	1/L1BSc product
FP Populator	1/chunk
L2 FP algorithm	1/measurement
FP Diagnostic Product Generator	1/L1BSc product
FP Standard Product Generator	1/L1BSc product

4.8 Error Handling

Inputs for ranges and expected units checked. For example, ensuring that latitude and longitude values are within acceptable ranges.

Radiative transfer checks sanity of inputs. For example, checking that albedo is less than 1.0 or that optical depths are at acceptable levels.

Math overflow and underflow errors caught. These are caught as floating point exceptions. Additionally there are checks to flag when NaN values are created.

Exception handling catches unexpected errors using mechanisms built into C++. Many exceptions are caught within individual components and dealt with if their possible occurrences are known ahead of time. However, third party code might throw unexpected exceptions that are rare. These would be propagated to the top level of the software. The code will try to write any output data pending before existing. It also marks the output product with the *.error* extension.

4.9 Commercial off-the-Shelf Components

The Level 2 FP implementation uses a number of third party software packages. These are as follows:

- Blitz++ — http://blitz.sourceforge.net/
- Boost C++ Libraries — http://www.boost.org/
- Bzip2 — http://www.bzip.org/
- GNU Scientific Library — https://www.gnu.org/software/gsl/
- HDF5 — http://www.hdfgroup.org/HDF5/
- Lua — http://www.lua.org/
- Luabind — http://www.rasterbar.com/products/luabind.html
- Nose Python Unit Testing — https://nose.readthedocs.org/en/latest/

4.10 Quality Assessment and Recording

There are no quality assessments made in the FP Single Sounding files themselves. Instead the L2Std Product PGE assigns a quality assessment, referred to as a Warn Level, using data from the L2Dia Product. For more details on the Warn Level generation procedure, see Mandrake et al. [2012].

5 References

Allen, C.W. (1964), *Astrophysical Quantities*, Athlone Press (2nd ed.), London.

Anderson, P. W. (1949), Pressure Broadening in the Microwave and Infra-Red Regions, *Phys. Rev.*, vol. 76, Issue 5, pp. 647-661.

Andrews, A., Boering, K., Daube, B., Wofsy, S., Loewenstein, M., Jost, H., Podolske, J., Webster, C., Herman, R., Scott, D., Flesch, G. J., Moyer, E. J., Elkins, J. W., Dutton, G. S., Hurst, D. F., Moore, F. L., Ray, E. A., Romashkin, P. A., and Strahan, S. E. (2001). Mean ages of stratospheric air derived from in situ observations of CO2, CH4, and N2O, *J. Geophys. Res.-Atmos.*, 106, 32295–32314.

Barton, John J., and Lee R. Nackman (1997), *Scientific and Engineering C++: An Introduction with Advances Techniques and Examples*. Addison-Wesley.

Baum, B. A., A. J. Heymsfield, P. Yang, and S. T. Bedka (2005a), Bulk scattering properties for the remote sensing of ice clouds. Part I: Microphysical data and models, *J. Appl. Meteorol*., 44, 1885–1895.

Baum, B. A., P. Yang, A. J. Heymsfield, S. Platnick, M. D. King, Y.-X. Hu, and S. T. Bedka (2005b), Bulk scattering properties for the remote sensing of ice clouds. Part II: Narrowband models, *J. Appl. Meteorol*., 44, 1896–1911.

Baum, B. A., P. Yang, A. J. Heymsfield, A. Bansemer, B. H. Cole, A. Merrelli, C. Schmitt, and C. Wang (2014), Ice cloud single-scattering property models with the full phase matrix at wavelengths from 0.2 to 100µm. *J. Quant. Spectrosc. Radiat. Transfer*, in press.

Benner, D. C., V. M. Devi, E. Nugent, K. Sung, L. R. Brown, C. E. Miller and R. A. Toth (2011), Line parameters of Carbon Dioxide in the 4850 cm^{-1} region, poster presentation at the 22nd Colloquium on High Resolution Molecular Spectroscopy, Dijon, France.

Boesch, H., G.C. Toon, B. Sen, R.A. Washenfelder, P.O. Wennberg, M. Buchwitz, R. de Beek, J.P. Burrows, D. Crisp, M. Christi, B.J. Connor, V. Natraj, and Y.L. Yung (2006), Space-based near-infrared CO2 measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, *Journal of Geophysical Research-Atmospheres*, 111 (D23).

Brown, L. R., and C. Plymate (2000), Experimental line parameters of the Oxygen A band at 760 nm, *J. Mol. Spectrosc*. 100, 166-179.

Brown, L. (2009), Line list is available in the acos/absco repository at the following location: acos/level_2/exe/absco/input/line_mixing/o2/lm_o2_inputs_20091026/SDF.dat

Butz, A., S. Guerlet, O. Hasekamp, D. Schepers, A. Galli, I. Aben, C. Frankenberg et al. (2011), Toward accurate CO2 and CH4 observations from GOSAT. *Geophysical Research Letters* 38, no. 14.

Chandrasekhar, S., (1950), Radiative Transfer. Oxford Univ. Press, Oxford.

Chandrasekhar, S., (1960), Radiative Transfer. Dover Publications, Dover, N.Y.

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

Cox, C. and W. Munk (1954), Measurement of the roughness of the sea surface from photographs of the Sun's glitter, *J. Opt. Soc. Am*., 44, 838-850.

Crisp, D., R.M. Atlas, F.-M. Breon, L.R. Brown, J.P. Burrows, P. Ciais, B.J. Connor, S.C. Doney, I.Y. Fung, D.J. Jacob, C.E. Miller, D. O'Brien, S. Pawson, J.T. Randerson, P. Rayner, R.J. Salawitch, S.P. Sander, B. Sen, G.L. Stephens, P.P. Tans, G.C. Toon, P.O. Wennberg, S.C. Wofsy, Y.L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, D. Kenyon, S. Schroll (2004), The Orbiting Carbon Observatory (OCO) mission, *Advances in Space Research* 34, 700–709.

Crisp, D., C. E. Miller, and P. L. DeCola (2007), NASA Orbiting Carbon Observatory: Measuring the column averaged carbon dioxide mole fraction from space, *JARS*, 12/1/2007.

Crisp, D. (2008), The Orbiting Carbon Observatory: NASA's First Dedicated Carbon Dioxide Mission, *Proc. SPIE,* 7106, DOI: 10.1117/12.802194.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bo¨sch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L. (2012). The ACOS X_{CO2} retrieval algorithm, Part 2: Global X_{CO2} data characterization, *Atmos. Meas. Tech*., 5, 687-707.

Crisp, David, B. Fisher, Christopher O'Dell, Christian Frankenberg, R. Basilio, H. Bosch, L. R. Brown et al. (2012), The ACOS CO2 retrieval algorithm-Part II: Global XCO2 data characterization, 687.

de Rooij, W. A., and C. C. A. H. van der Stap (1984), Expansion of Mie scattering matrices in generalized spherical functions, *Astron. Astrophys*., 131, 237–248.

Deutscher, N. M. et al. (2010), Total column CO2 measurements at Darwin, Australia - Site description and calibration against in situ aircraft profiles, *Atmos. Meas. Tech.*, 3(4), 947–958, doi:10.5194/amt-3-947-2010.

Devi, V. M., D. C. Benner, L. R. Brown, C. E. Miller, and R. A. Toth (2007a), Line mixing and speed dependence in CO2 at 6348 cm-1: Constrained multispectrum analysis of intensities and line shapes in the 30012 ← 00001 band, *J. Mol. Spectrosc*. 242, 90-117.

Devi, V. M., D. C. Benner, L. R. Brown, C. E. Miller, and R. A. Toth (2007b), Line mixing and speed dependence in CO2 at 6227.9 cm-1: Constrained multispectrum analysis of intensities and line shapes in the 30013 ← 00001 band, *J. Mol. Spectrosc*. 245, 52-80.

Dubovik, O. et al. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. Atmos. Meas. Tech. 4, 975–1018 (2011).

Dufour, E. and F.-M. Bréon, (2003), Spaceborne Estimate of Atmospheric CO2 Column by Use of the Differential Absorption Method: Error Analysis, *Appl. Opt*. 42, 3595-3609.

Drouin, B. J., V. H. Payne, F. Oyafuso, K. Sung and E. J. Mlawer (2014), Pressure broadening of oxygen by water, *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 133, January 2014, Pages 190-198, ISSN 0022-4073.

Frankenberg, C., Platt, U., and Wagner, T. (2005), Iterative maximum *a posteriori* (IMAP-) DOAS for retrieval of strongly absorbing trace gases: Model studies for CH4 and CO2 retrieval from near- infrared spectra of SCIAMACHY onboard ENVISAT, *Atmos. Chem. Phys.*, 5, 9–22.

Frankenberg, C., Fisher, J., Worden, J., Badgley, G., Saatchi, S., Lee, J.-E., et al. (2011), New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, 38 (17), L17706.

Frankenberg, C., et al. (2012), Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO 2 retrievals. *Atmospheric Measurement Techniques* 5.8: 2081-2094.

Galatry, L. (1961), Simultaneous effect of Doppler and foreign gas broadening on spectral lines, *Phys. Rev*., 122, 1218-1223.

Hansen, J.E., and L.D. Travis (1974), Light Scattering in Planetary Atmospheres, *Space Science Reviews*, 16, 527-610.

Hartmann, J.-M., H. Tran, and G. Toon (2009), Influence of line mixing on the retrievals of atmospheric CO_2 from spectra in the 1.6 and 2.1 μm regions, *Atmos. Chem. Phys.* Disc.9,, 4873-4898.

Hovenier, J. W., and J. F. de Haan (1985), Polarized light in planetary atmospheres for perpendicular directions, *Astron. Astrophys.,*146, 185–191.

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

Lamouroux, J., H. Tran, A. L. Laraia, R. R. Gamache, L. S. Rothman, I. E. Gordon, J.-M. Hartman (2010), Updated database plus software for line-mixing in CO2 infrared spectra and their test using laboratory spectra in the 1.5–2.3 μm region, *J. Quantitative Spectroscopy & Radiative Transfer,* 111, 2321-2331.

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R. A., House, J. I., Huntingford, C., Levy, P. E., Lomas, M. R., Majkut, J., Metzl, N., Ometto, J. P., Peters, G.P., Prentice, I. C., Randerson, J. T., Running, S. W., Sarmiento, J. L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G. R., and Woodward, F. I. (2009), Trends in the sources and sinks of carbon dioxide, *Nature Geosci.*, 2, 831–836, doi:10.1038/ngeo689

Le Quéré, C., C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., van der Werf, G. R., Ahlstro¨m, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K.,

Jourdain, C., Kato, E., Keeling, R. F., Klein Goldewijk, K., Levis, S., Levy, P., Lomas, M., Poulter, B., Raupach, M. R., Schwinger, J., Sitch, S., Stocker, B. D., Viovy, N., Zaehle, S., and Zeng, N. (2013), The global carbon budget 1959–2011, *Earth Syst. Sci. Data*, 5, 165–185, doi:10.5194/essd-10 5-165-2013.

Long, D. A., D. K. Havey, M. Okumura, C. E. Miller, J. T. Hodges (2010), O 2 A-band line parameters to support atmospheric remote sensing, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111 (14), 2021-2036.

Long, D. A., D. K. Havey, S. S. Yu, M. Okumura, C. E. Miller, J. T. Hodges (2011), O 2 A-band line parameters to support atmospheric remote sensing. Part II: The rare isotopologues, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112 (16), 2527-2541.

Long, D. A., and J. T. Hodges (2012), On spectroscopic models of the O2 A-band and their impact upon atmospheric retrievals, *Journal of Geophysical Research: Atmospheres*, 117 (D12).

Messerschmidt, J. et al. (2011), Calibration of TCCON column-averaged CO2: the first aircraft campaign over European TCCON sites, *Atmos. Chem. Phys*., 11(21), 10765–10777, doi:10.5194/acp-11-10765-2011.

Miller, C.E., and L. R. Brown (2004a), Near Infrared Spectroscopy of Carbon Dioxide: 16O12C16O Line Positions, *J. Mol. Spectrosc*. 228, 329-354.

Miller, C. E., M. A. Montgomery, R. M. Onorato, C. Johnstone, T. P. McNicholas, B. Kovaric and L. R. Brown (2004b), Near Infrared Spectroscopy of Carbon Dioxide. II. $^{16}O^{13}C^{16}O$ and $^{16}O^{13}C^{18}O$ Line Positions, *J. Mol. Spectrosc*. 228, 355-374.

Miller, C. E., L. R. Brown, R. A. Toth D. C. Benner, and V. M. Devi (2005), Spectroscopic challenges for high accuracy retrievals of atmospheric CO2 and the Orbiting Carbon Observatory (OCO) experiment, *Comptes Rendus Physique* 6, 876-887.

Miller, C. E., D. Crisp, P. L. DeCola, S. C. Olsen, J. T. Randerson, A. M. Michalak,A. Alkhaled, P. Rayner, D. J. Jacob, P. Suntharalingam, D. B. A. Jones, A. S. Denning, M. E. Nicholls, S. C. Doney, S. Pawson, H. Boesch, B. J. Connor, I. Y. Fung, D. O'Brien, R. J. Salawitch, S. P. Sander, B. Sen, P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, and R. M. Law (2007), "Precision requirements for space-based X_{CO2} data", *J. Geophys. Res*. 112, D10314, [doi:10.1029/2006JD007659].

Mlawer, E. J., V. H. Payne, J.-L. Moncet, J. S. Delamere, M. A. Alvarado, D. C. Tobin (2012), Development and recent evaluation of the MT_CKD model of continuum absorption, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370 (1968), 2520-2556.

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. Rad. Tran*., 107, 263–293.

Natraj, V., H. Boesch, R. J. D. Spurr, and Y. L. Yung (2008), Retrieval of X_{CO2} from simulated Orbiting Carbon Observatory measurements using the fast linearized R-2OS radiative transfer model, *J. Geophys. Res*., 113, D11212, doi: 10.1029/2007JD009017.

Niro, F., T. von Clarmann, K. Jucks, and J.-M. Hartmann (2005a), Spectra calculations in central and wing regions of CO_2 IR bands between 10 and 20 m. III: atmospheric emission Spectra, *J. Quant. Spectrosc. Radiat. Transfer* 90, 61–76.

Niro, F., K. Jucks, and J.-M. Hartmann (2005b), Spectra calculations in central and wing regions of CO_2 IR bands. IV: software and database for the computation of atmospheric spectra, *J. Quant. Spectrosc. & Radiat. Transfer* 95, 469–481.

O'Brien, Denis M., Igor Polonsky, Chris O'Dell, Akihiko Kuze, Nobuhiro Kikuchi, Yukio Yoshida, and Vijay Natraj (2013), Testing the Polarization Model for TANSO-FTS on GOSAT Against Clear-Sky Observations of Sun Glint Over the Ocean. *IEEE Trans. on Geoscience and Remote Sensing*, 51 (12), 5199–5209.

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

O'Dell, C. W., B. Connor, H. Bösch, D. O'Brien, C. Frankenberg, R. Castano, M. Christi et al. (2012). The ACOS CO 2 retrieval algorithm–Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques* 5 (1), 99-121.

Predoi-Cross, A., K. Hambrook, R. Keller, C. Povey, I. Schofield, D. Hurtmans, H. Over, and G. Ch. Mellau (2008a), Spectroscopic lineshape study of the self-perturbed oxygen A-band, *J. Mol. Spectrosc*. 248, 85-110.

Predoi-Cross, A., C. Holladay, H. Heung, J.-P. Bouanich, G. Ch. Mellau, R. Keller, and D. R. Hurtmans (2008b), Nitrogen-broadened lineshapes in the oxygen A-band: Experimental results and theoretical calculations, *J. Mol. Spectrosc*. 251, 159–175.

Predoi-Cross, A., A.R.W. McKellar, D. C. Benner, V. M. Devi, R. R. Gamache, C. E. Miller , R. A. Toth and L. R. Brown (2009), Temperature dependences for air-broadened Lorentz half width and pressure-shift coefficients in the 30013←00001 and 30012←00001 bands of CO_2 near 1600 nm, *Can. J. Phys*. (in press).

Rahman, H., Pinty, B. & Verstraete, M. M. Coupled surface-atmosphere reflectance (CSAR) model: 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. Journal of Geophysical Research: Atmospheres 98, 20791–20801 (1993).

Rayner, P. J. and D. M. O'Brien (2001), "The utility of remotely sensed CO_2 concentration data in surface source inversions," *Geophys. Res. Lett*., 28, 175-178.

Reichler, T., Dameris, M., Sausen, R. Determining the tropopause height from gridded data, *Geo. Res. Let.* 30, 2042, doi:10.1029/2003GL018240, 2003.

Robichaud, D. J., J. T. Hodges, L. R. Brown, D. Lisak, M. Okumura, and C. E. Miller (2008a), Experimental Intensity and Line Shape Parameters of the Oxygen A-Band Using Frequency-Stabilized Cavity Ring-Down Spectroscopy, *J. Mol. Spectrosc*. 248, 1-13.

Robichaud, D. J., J. T. Hodges, P. Masłowski, L. Y. Yeung, M. Okumura, C. E. Miller, and L. R. Brown (2008b), High-Accuracy Transition Frequencies for the O_2 A-Band, *J. Mol. Spectrosc*. 251, 27-43.

Robichaud, D. J., L. Y. Yeung, D. A. Long, D. K. Havey, J. T. Hodges, M. Okumura, C. E. Miller, and L. R. Brown (2009), Experimental Line Parameters of the $b^1\Sigma g^+ \leftarrow X^3\Sigma g^-$ Band of Oxygen Isotopologues at 760 nm Using Frequency-Stabilized Cavity Ring-Down Spectroscopy, *J. Phys. Chem. A*, 2009, 113 (47), pp 13089–13099, doi: 10.1021/jp901127h.

Rodgers, C. (2000) Inverse Methods for Atmospheric Sounding: Theory and Practice. World Scientific Publishing Co Pte Ltd.

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

Rothman L. S., I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion, K. Chance, L. H. Coudert, V. Dana, V. M. Devi, S. Fally, J.-M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W. J. Lafferty, J.-Y. Mandin, S. T. Massie, S. Mikhailenko, C. E. Miller, N. Moazzen-Ahmadi, O. V. Naumenko, A. Nikitin, J. Orphal, A. Predoi-Cross, V. Perevalov, A. Perrin, C. P. Rinsland, M. Rotger, M. Šimečková, M. A. H. Smith, K. Sung, S. Tashkun, J. Tennyson, R. A. Toth, A. C. Vandaele, and J. Vander Auwera (2009), The HITRAN 2008 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer* 110,533-572.

Rothman, L.S., I.E.Gordon, Y.Babikov, A.Barbe, D.Chris Benner, P.F.Bernath, M.Birk, L.Bizzocchi, V.Boudon, L.R.Brown, A.Campargue, K. Chance, E.A.Cohen, L.H.Coudert, V.M.Devi, B.J.Drouin, A.Fayt, J.-M. Flaud, R.R.Gamache, J.J.Harrison, J.-M.Hartmann, C.Hill, J.T.Hodges, D.Jacquemart, A.Jolly, J.Lamouroux, R.J.LeRoy, G.Li, D.A.Long p, O.M.Lyulin, C.J.Mackie, S.T.Massie, S.Mikhailenko, H.S.P.Müller, O.V.Naumenko, A.V.Nikitin, J.Orphal, V.Perevalov, A.Perrin, E.R.Polovtseva, C.Richard, M.A.H.Smith v, E.Starikova, K.Sung, S. Tashkun, J.Tennyson, G.C.Toon, Vl.G.Tyuterev, G.Wagner (2013), The HITRAN 2012 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer, 130, 4.*

Salstein, D. A., Ponte, R. M., and Cady-Pereira, K. (2008), Uncertainties in atmospheric surface pressure fields from global analyses, *J. Geophys. Res.*, 113, D14107.

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S. & Martonchik, J. V. Reflectance quantities in optical remote sensing—definitions and case studies. Remote Sensing of Environment 103, 27–42 (2006).

Schutgens, A. J., Tilstra, L. G., Stammes, P. and Breon, F.-M. (2004), On the relationship between Stokes parameters *Q* and *U* of atmospheric ultraviolet/visible/near-infrared radiation, *J. Geophys. Res.*, 109 (D09205), doi:10.1029/2003JD004081.

Slijkhuis, S. (1998), SCIAMACHY level 0 to 1c processing algorithm theoretical basis document,Tech. Note ENV-ATB-DLR-SCIA-0041,Dtsch. Zent. für Luft- und Raumfahrt, Oberpfaffenhofen, Germany.

Slijkhuis, S. (2000), Calculation of polarisation from Rayleigh single scattering, Rep. ENV-TN-DLR-SCIA-0043 , Dtsch. Zent. für Luft- und Raumfahrt, Oberpfaffenhofen, Germany.

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. Rad. Tran*., 68(6), 689–735.

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. Rad. Tran*., 75(2), 129–175.

Spurr, R. J. D. (2006), VLIDORT: a linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, *J. Quant. Spectrosc. Rad. Tran*., 102(2), 316–342.

Sung, K., L. R. Brown, R. A. Toth, and T. J. Crawford (2009), FT-IR measurements of H_2O-broadened half-widths of CO_2 at 4.3 μm, *Can. J. Phys*. (in press).

Taylor, T.E., C. O'Dell, D.M. O'Brien, N. Kikuchi, T. Yokota, T. Nakajima, H. Ishida, D. Crisp, and T. Nakajima (2011), Comparison of cloud screening methods applied to GOSAT near-infrared spectra, IEEE *Trans. Geosci. Rem. Sens*., doi: 10.1109/TGRS.2011.2160270.

Taylor, T.E., C.W. O'Dell, C. Frankenberg, P. Partian, H. W. Cronk, A. Savtchenko, H. R. Pollock, D. Crisp, A. Eldering, and M. Gunson, (2015), Orbiting Carbon Observatory-2 (OCO-2) aerosol and cloud screening; validation against collocated MODIS data, manuscript in preparation.

Thompson, D. R., D. C. Benner, L. R. Brown, D. Crisp, V. M. Devi, Y. Jiang, V. Natraj, F. Oyafuso, K. Sung, D. Wunch, R. Castano, C. E. Miller (2012), Atmospheric validation of high accuracy CO 2 absorption coefficients for the OCO-2 mission, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 113 (17), 2265-2276.

Thuillier, G., M. Hersé, P. C. Simon, D. Labs, H. Mandel, D. Gillotay, and T. Foujols (2003), The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS 1-2-3 and EURECA missions, *Solar Physics*, 214(1): 1-22.

Toon, G.C., J.-F. Blavier, B. Sen, R.J. Salawitch, G.B. Osterman, J. Notholt, M. Rex, C.T. McElroy, and J.M. Russell III (1999), Ground-based observations of Arctic O3 loss during spring and summer 1997, *J. Geophys. Res*., 104(D21), 26497-26510.

Toth, R. A., L. R. Brown, C. E. Miller, V. M. Devi, and D. C. Benner (2006a), Line strengths of $^{12}C^{16}O_2$: 4550-7000 cm^{-1}, *J. Mol. Spectrosc*. 239, 221-242.

Toth, R. A., L.R. Brown, C.E. Miller, V. M. Devi, and D. C. Benner (2006b), Self-broadened widths and shifts of CO^2, *J. Mol. Spectrosc*. 239, 243-271.

Toth, R. A., C. E. Miller, L. R. Brown, V. M. Devi, and D. C. Benner (2007a), Line positions and strengths of $^{16}O^{12}C^{18}O$, $^{18}O^{12}C^{18}O$ and $^{17}O^{12}C^{18}O$ between 2200 and 7000 cm^{-1}, *J. Mol. Spectrosc*. 243, 43-61.

Toth, R. A., C. E. Miller, V. M. Devi, D. C. Benner, and L. R. Brown (2007b), Air-broadened width and pressure shift coefficients of $^{12}C^{16}O_2$: 4700 – 7000 cm^{-1}, *J. Mol. Spectrosc*. 246, 133-157.

Toth, R. A., C. E. Miller, L. R. Brown, V. M. Devi, and D. C. Benner (2008a), Line strengths of $^{16}O^{13}C^{16}O$, $^{16}O^{13}C^{18}O$, $^{16}O^{13}C^{17}O$ and $^{18}O^{13}C^{18}O$ between 2200 and 6800 cm^{-1}, *J. Mol. Spectrosc*. 251, 64-89.

Toth, R. A., L. R. Brown, C. E. Miller, V. M. Devi, and D. C. Benner (2008b), Spectroscopic database of CO_2 line parameters: 4300 – 7000 cm^{-1}, *J. Quant. Spectrosc. Radiat. Transfer* 109, 906-921.

Tran, H., C. Boulet, and J.-M. Hartmann (2006), Line mixing and collision-induced absorption by oxygen in the A band: Laboratory measurements, model, and tools for atmospheric spectra computations, *J. Geophys. Res*. 111, D15210, doi: 10.1029/ 2005JD006869.

Tran, H., and J.-M. Hartmann (2008), An improved O_2 A band absorption model and its consequences for retrievals of photon paths and surface pressures, *J. Geophys. Res*.113, D18104, doi: 10.1029/2008JD010011.

van de Hulst, H. C. (1981), Light Scattering by Small Particles , Dover, Mineola, N. Y.

Washenfelder, R. A., G. C. Toon, J.-F. Blavier, Z. Yang, N. T. Allen, P. O. Wennberg, S. A. Vay, D. M. Matross and B. C. Daube (2006), Carbon dioxide column abundances at the Wisconsin Tall Tower site, *J. Geophys. Res*. 111, D22305, doi: 10.1029/2006JD007154.

Wiscombe, W. J. (1977), The delta-M method: rapid yet accurate radiative flux calculations for strongly asymmetric phase functions, *J. Atmos. Sci.,* 34 (9), 1408-1422.

World Meteorological Organization (WMO), Meteorology A Three-Dimensional Science: Second Session of the Commission for Aerology, *WMO Bulletin IV*(4), WMO, Geneva, 134-138, 1957.

Wunch, D. , G. C. Toon, P. O. Wennberg, S. C. Wofsy, B. B. Stephens, M. L. Fischer, O. Uchino, J. B. Abshire, P. Bernath, S. C. Biraud, J.-F. L. Blavier, C. Boone, K. P. Bowman, E. V. Browell, T. Campos, B. J. Connor, B. C. Daube, N. M. Deutscher, M. Diao, J. W. Elkins, C. Gerbig, E. Gottlieb, D. W. T. Griffith, D. F. Hurst, R. Jim´enez, G. Keppel-Aleks, E. Kort, R. Macatangay, T. Machida, H. Matsueda, F. Moore, I. Morino, S. Park, J. Robinson, C. M. Roehl, Y. Sawa, V. Sherlock, C. Sweeney, T. Tanaka, and M. A. Zondlo, (2010), Calibration of the total carbon column observing network using aircraft profile data, *Atmos. Meas. Tech. Discuss*., 3, 2603–2632, doi: 10.5194/amtd-3-2603-2010.

Wunch, D., G. C. Toon, J.F.L. Blavier, R.A. Washenfelder, J. Notholt, B. J. Connor, D.W.T. Griffith, V. Sherlock, P.O. Wennberg (2011) The Total Carbon Column Observing Network, *Phil. Trans. Of the Royal Society A-Mathematical Physical and Engineering Sciences*, 369, 2087-2112, DOI: 10.1098/rsta.2010.0240.

Yang, Z., P. O. Wennberg, R. P. Cageao, T. J. Pongetti, G. C. Toon, and S. P. Sander (2005), Ground-based photon path measurements from solar absorption spectra of the O_2 A-band, *J. Quant. Spectrosc. Radiat. Transfer* 90, 309–321.

Young, A.T. (1980), Revised depolarization corrections for atmospheric extinction, *Applied Optics*, 19 (20), 3427-3428.

C. Zhao, P.P. Tans, and K.W. Thoning (1997), A high precision manometric system for absolute calibrations of CO2 in dry air, *J. Geophys. Res.* 102, 5885-5894.

Zong, J. (2008), Application of Global High-Resolution DEM in the OCO Mission, in: International Archives of the Photogrammetry, *Remote Sensing, and Spatial Information Sciences*, Vol. XXXVII, Part B4, Beijing.

6 Acronyms

2D, 4D	Two- and four-dimensional
2OS	First two orders of scattering (model)
A-Train	Earth Observing System Afternoon Constellation
ABSCO	Absorption coefficients
ACS	Attitude control system
AOD	Aerosol optical depth
ATBD	Algorithm theoretical basis document
AU	Astronomical Unit
CIA	Collision-induced absorption
CO_2	Carbon Dioxide
CRDS	Cavity ring-down spectrometer
ECMWF	European Centre for Medium-Range Weather Forecasting
EOF	Empirical orthogonal function
FOV	Field of view
FP	Full physics
FPA	Focal plane array
FTS	Fourier transform spectrometer
GOSAT	Greenhouse Gases Observing Satellite
GSFC	NASA Goddard Space Flight Center
H_2O	Water
HITRAN	High Resolution Transmission database
ILS	Instrument line shape
IMAP-DOAS	Iterative maximum *a posteriori* differential optical absorption spectroscopy
JPL	NASA Jet Propulsion Laboratory
L (0,1..)	Level 0, Level 1, etc. (data product)
LIDORT	Linearized discrete ordinate radiative transfer
LSI	Low-streams interpolation
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MODIS	Moderate-Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NIR	Near-infrared
NIST	National Institute of Standards and Technology (Maryland)
O_2	Oxygen
OCO	Orbiting Carbon Observatory
PGE	Product Generation Executable
ppm	parts per million
RH	Relative humidity
RT	Radiative transfer (model)
RTE	Radiative transfer equation
SDOS	Science Data Operations System

SIF	Solar-induced fluorescence
SOLSPEC	Solar Spectrum (instrument)
SNR	Signal-to-noise ratio
SWIR	Short-wave-infrared
TCCON	Total Carbon Column Observing Network
TOA	Top of atmosphere
VLIDORT	Vector LIDORT
WMO	World Meteorological Organization
X_{CO2}	Column-averaged CO_2 dry air mole fraction