Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm

Christopher O'Dell
Colorado State University

A Eldering
California Institute of Technology

Paul O. Wennberg
California Institute of Technology, wennberg@gps.caltech.edu

David Crisp
California Institute of Technology, david.crisp@jpl.nasa.gov

Michael Gunson
California Institute of Technology

See next page for additional authors

Publication Details
Abstract
Since September 2014, NASA’s Orbiting Carbon Observatory-2 (OCO-2) satellite has been taking measurements of reflected solar spectra and using them to infer atmospheric carbon dioxide levels. This work provides details of the OCO-2 retrieval algorithm, versions 7 and 8, used to derive the column-Averaged dry air mole fraction of atmospheric CO2 (XCO2) for the roughly 100,000 cloud-free measurements recorded by OCO-2 each day. The algorithm is based on the Atmospheric Carbon Observations from Space (ACOS) algorithm which has been applied to observations from the Greenhouse Gases Observing SATellite (GOSAT) since 2009, with modifications necessary for OCO-2. Because high accuracy, better than 0.25%, is required in order to accurately infer carbon sources and sinks from XCO2, significant errors and regional-scale biases in the measurements must be minimized. We discuss efforts to filter out poor-quality measurements, and correct the remaining good-quality measurements to minimize regional-scale biases. Updates to the radiance calibration and retrieval forward model in version 8 have improved many aspects of the retrieved data products. The version 8 data appear to have reduced regional-scale biases overall, and demonstrate a clear improvement over the version 7 data. In particular, error variance with respect to TCCON was reduced by 20% over land and 40% over ocean between versions 7 and 8, and nadir and glint observations over land are now more consistent. While this paper documents the significant improvements in the ACOS algorithm, it will continue to evolve and improve as the CO2 data record continues to expand.

Publication Details

Authors

This journal article is available at Research Online: https://ro.uow.edu.au/smhpapers1/496
Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm

Christopher W. O’Dell¹, Annmarie Eldering², Paul O. Wennberg³, David Crisp², Michael R. Gunson², Brendan Fisher², Christian Frankenberg³, Matthäus Kiel³, Hannakaisa Lindqvist⁴, Lukas Mandrake⁵, Aronne Merrelli⁶, Vijay Nattraj², Robert R. Nelson¹, Gregory B. Osterman², Vivienne H. Payne², Thomas E. Taylor¹, Debra Wunch⁶, Brian J. Drouin², Fabiano Oyafuso², Albert Chang², James McDuffie², Michael Smyth², David F. Baker⁴, Sourish Basu⁷,⁸, Frédéric Chevallier⁹, Sean M. R. Crowell¹⁰, Liang Feng¹¹,¹², Paul I. Palmer¹¹,¹², Mavendra Dubey¹³, Omaira E. García¹⁴, David W. T. Griffith¹⁵, Frank Hase¹⁶, Laura T. Iraci¹⁷, Rigel Kivi¹⁸, Isamu Morino¹⁹, Justus Notholt²⁰, Hirofumi Ohyama¹⁹, Christof Petri²⁰, Coleen M. Roehl³, Mahesh K. Sha²¹, Kimberly Strong², Ralf Sussmann²², Yao Te²³, Osamu Uchino¹⁹, and Voltaire A. Velazco¹⁵

¹Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA
²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
³Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
⁴Finnish Meteorological Institute, Helsinki, Finland
⁵SSEC, University of Wisconsin-Madison, Madison, WI, USA
⁶Department of Physics, University of Toronto, Toronto, Canada
⁷NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, CO, USA
⁸Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado, USA
⁹Laboratoire des Sciences du Climat et de l’Environnement, IPSL, CEA-CNRS-UVSQ, Gif-sur-Yvette, France
¹⁰College of Atmospheric and Geographic Sciences, University of Oklahoma, Norman, OK, USA
¹¹National Centre for Earth Observation, University of Edinburgh, UK
¹²School of GeoSciences, University of Edinburgh, UK
¹³Los Alamos National Laboratory, Los Alamos, NM, USA
¹⁴Izaña Atmospheric Research Center, Meteorological State Agency of Spain (AEMet), Santa Cruz de Tenerife, Spain
¹⁵Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, Australia
¹⁶Karlsruhe Institute of Technology, IMK-ASF, Karlsruhe, Germany
¹⁷NASA Ames Research Center, Moffett Field, CA, USA
¹⁸Finnish Meteorological Institute, Sodankylä, Finland
¹⁹National Institute for Environmental Studies (NIES), Tsukuba, Japan
²⁰Institute of Environmental Physics, University of Bremen, Bremen, Germany
²¹Royal Belgian Institute for Space Aeronomy, Brussels, Belgium
²²Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany
²³LERMA-IPSL, Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, 75005, Paris, France

Correspondence: Christopher W. O’Dell (christopher.odell@colostate.edu)

Received: 27 July 2018 – Discussion started: 16 August 2018
Revised: 12 November 2018 – Accepted: 15 November 2018 – Published: 11 December 2018

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract. Since September 2014, NASA’s Orbiting Carbon Observatory-2 (OCO-2) satellite has been taking measurements of reflected solar spectra and using them to infer atmospheric carbon dioxide levels. This work provides details of the OCO-2 retrieval algorithm, versions 7 and 8, used to derive the column-averaged dry air mole fraction of atmospheric CO$_2$ ($X_{CO_2}$) for the roughly 100,000 cloud-free measurements recorded by OCO-2 each day. The algorithm is based on the Atmospheric Carbon Observations from Space (ACOS) algorithm which has been applied to observations from the Greenhouse Gases Observing SATellite (GOSAT) since 2009, with modifications necessary for OCO-2. Because high accuracy, better than 0.25%, is required in order to accurately infer carbon sources and sinks from $X_{CO_2}$, significant errors and regional-scale biases in the measurements must be minimized. We discuss efforts to filter out poor-quality measurements, and correct the remaining good-quality measurements to minimize regional-scale biases. Updates to the radiance calibration and retrieval forward model in version 8 have improved many aspects of the retrieved data products. The version 8 data appear to have reduced regional-scale biases overall, and demonstrate a clear improvement over the version 7 data. In particular, error variance with respect to TCCON was reduced by 20% over land and 40% over ocean between versions 7 and 8, and nadir and glint observations over land are now more consistent. While this paper documents the significant improvements in the ACOS algorithm, it will continue to evolve and improve as the CO$_2$ data record continues to expand.

1 Introduction

Bias-free measurement of atmospheric CO$_2$ concentrations from space is a long-pursued goal in the carbon cycle community. Such measurements are critical for inferring sources and sinks of carbon, and how these sources and sinks change over time due to both anthropogenic and natural causes (e.g., Rayner and O’Brien, 2001; Chevallier et al., 2007; Baker et al., 2010). The first instrument capable of CO$_2$ measurements from space using the near- and short-wavelength infrared was SCIAMACHY, the SCanning Imaging Absorption spectrometer for Atmospheric CHartographY (Buchwitz et al., 2005; Reuter et al., 2011), which operated from 2002 to 2012. This was followed by the first dedicated greenhouse gas satellite, the Japanese Greenhouse gases Observing SATellite (GOSAT), which launched in January 2009 (Yokota et al., 2009). The Orbiting Carbon Observatory-2 (OCO-2) followed on 2 July 2014, with the goal of measuring the column-averaged dry air mole fraction of carbon dioxide ($X_{CO_2}$) with sufficient precision and accuracy to enable greatly enhanced understanding of the surface–atmosphere exchange of CO$_2$ on regional scales (Crisp et al., 2008; Crisp, 2015). OCO-2 was preceded by the original OCO mission, which failed due to a launch vehicle malfunction in 2009. Retrieval algorithms originally developed for OCO (Connor et al., 2008) have been continuously refined since 2009 (O’Dell et al., 2012), by application to data from GOSAT.

$X_{CO_2}$ measurements from the OCO-2 version 7 data product (Elderding et al., 2017a) have recently been used to estimate CO$_2$ fluxes from both natural (Liu et al., 2017; Chatterjee et al., 2017; Crowell et al., 2018a) and anthropogenic (Hakkarainen et al., 2016; Schwandner et al., 2017; Nassar et al., 2017) sources; see Eldering et al. (2017b) for a complete review of these findings. However, $X_{CO_2}$ measurements must be both extremely accurate and precise in order to accurately determine fluxes (Miller et al., 2007), since fluxes are determined from small (~2.5%) spatial and temporal gradients in the $X_{CO_2}$ field. Spatially coherent biases in $X_{CO_2}$ on regional scales as small as a few tenths of a part-per-million (ppm) in $X_{CO_2}$ can lead to spurious values of inferred fluxes (Chevallier et al., 2014).

The ACOS algorithm was originally developed for OCO. It was first applied to GOSAT data in 2009 and has continuously evolved and improved in the intervening years. Generally, good error statistics were shown for GOSAT observations over both land and water, with typical biases below 1 ppm based on comparisons to both ground-based (Lindqvist et al., 2015; Kulawik et al., 2016) and aircraft (Frankenberg et al., 2016) validation data. After the successful launch of OCO-2, the ACOS algorithm was further modified and tuned for application to the OCO-2 spectra. $X_{CO_2}$ error statistics are similar to those from GOSAT, with rms errors less than 1.5 ppm when compared against most ground-based Total Carbon Column Observing Network (TCCON, Wunch et al., 2010) stations (Wunch et al., 2017). However, Wunch et al. (2017) noted that important biases remain, in particular related to latitude, surface properties, and atmospheric scattering by clouds and aerosols. A particularly troubling bias evident in the Southern Hemisphere mid-latitude ocean in austral winter had amplitudes as large as several ppm. This bias was not seen in ACOS retrievals using GOSAT data, though GOSAT’s ocean glint viewing geometry was restricted and could not typically see this far south, potentially masking the problem.

The primary purpose of this paper is to describe the details of the ACOS $X_{CO_2}$ retrieval algorithm as applied to OCO-2 data (Gunson and Elderding, 2017), in particular the latest version 8 (also referred to as build 8 or B8). Because science results have already been published with version 7 (also referred to as build 7 or B7) as discussed above, we also discuss the differences between versions 7 and 8. This paper is organized as follows: Sect. 2 discusses prescreening of the data to remove cloudy and difficult-to-retrieve soundings. Section 3 lists the details of the retrieval algorithm and its evolution since O’Dell et al. (2012). Section 4 discusses the methodology and results of the post-retrieval filtering and bias correction. Section 5 provides a brief evaluation of $X_{CO_2}$ from both
Table 1. Prescreening filter criteria.

<table>
<thead>
<tr>
<th>Category</th>
<th>Land criterion</th>
<th>Ocean criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Successful measure</td>
<td>Sounding_Qual_Flag = 0</td>
<td>Same as land</td>
</tr>
<tr>
<td>A-Band preprocessor</td>
<td>Cloud_Flag = 0</td>
<td>Same as land</td>
</tr>
<tr>
<td>Solar geometry</td>
<td>SZA &lt; 85° (nadir), &lt; 80° (glint)</td>
<td>Same as land</td>
</tr>
<tr>
<td>IMAP preprocessor</td>
<td>0.985 &lt; co2_ratio &lt; 1.045</td>
<td>Same as land</td>
</tr>
<tr>
<td>Band 1 SNR</td>
<td>SNR1 ≥ 100</td>
<td>Same as land</td>
</tr>
<tr>
<td>Band 3 SNR</td>
<td>SNR3 ≥ 75</td>
<td>Same as land</td>
</tr>
<tr>
<td>Land fraction</td>
<td>f_land ≥ 80 %</td>
<td>f_land ≤ 20 %</td>
</tr>
</tbody>
</table>

Figure 1. Fraction of soundings passing the OCO-2 B8 prescreening filter in December 2015 (a, c) and June 2016 (b, d), for both nadir mode (a, b) and glint mode (c, d). Starting in November 2015, about one-third of all orbits are performed in nadir mode, and two-thirds are performed in glint mode.

2 Data and prescreening

Because only scenes with sufficient signal and nearly devoid of cloud and aerosol contamination can yield successful $X_{CO_2}$ retrievals, a prescreener is used for OCO-2 soundings before processing by the Level-2 “Full-Physics” (L2FP) $X_{CO_2}$ retrieval algorithm. Our prescreening module requires outputs from two fast algorithms, described in detail in Taylor et al. (2016). First, the “A-band Preprocessor” (ABP) performs a fast retrieval of surface pressure using the O$_2$A band only, assuming that no clouds or aerosols are present. Poor spectral fits and differences between the retrieved and a priori surface pressure greater than 25 hPa are used to identify the presence of cloud or aerosol contamination. Scenes without sufficient signal to noise in the O$_2$A band are skipped altogether. Second, the “IMAP-DOAS” preprocessor performs fast, clear-sky fits to the weak and strong CO$_2$ bands at 1.61 and 2.06 µm, respectively. While this preprocessor solves for a number of variables, the CO$_2$ and H$_2$O columns, which are fit independently of each of these two bands, are most relevant for cloud screening. From these spectral fits, the strong-to-weak ratios of the column-integrated CO$_2$ and H$_2$O are derived. The CO$_2$ ratio must be within a certain range (near unity) for the scene to be deemed sufficiently clear to warrant a Full-Physics retrieval. Other screens are used to remove soundings unlikely to yield successful $X_{CO_2}$, such as those at high solar zenith angle or for which the continuum SNR levels are too low. Unlike in version 7 of the OCO-2 algorithm, there is no explicit screen for snow- and ice-covered surfaces. However, the surface albedo in the strong CO$_2$ band is low over snow and ice, and therefore the strong CO$_2$ band SNR is...
filter will remove many of those scenes. The full prescreening criteria for OCO-2 B8 are given in Table 1.

In total, roughly 26% of land soundings pass our prescreener (28% land nadir, 25% land glint) and 27% of ocean glint soundings pass it as well. Generally these fractions are strong functions of both location and time of year. To illustrate this, the fractions of soundings passing the prescreening criteria for December 2015 and June 2016 are shown in Fig. 1. A number of features are observed. A higher fraction of soundings are passed in the tropics than at higher latitudes relative to the sub-solar latitude (∼−23° in December and +23° in June), and the passing rates tend to be higher over bright versus dark surfaces. Also, few soundings survive over the tropical rainforests in South America and Africa, which are often cloudy. A significant number of soundings survive prescreening over the Greenland and Antarctic ice sheets during their summer season (this was not the case in version 7), though it is shown later that most of these fail the post-retrieval quality screening (Sect. 4.2). About 10% of nadir soundings over ocean pass the prescreening criteria; this occurs in regions where the nadir view is relatively close to the glint geometry, typically near the sub-solar latitude. These nadir ocean soundings are currently removed by post-retrieval filtering, as their quality relative to the glint ocean observations has not yet been evaluated. A final obvious feature is that fewer soundings are available in nadir mode than in glint – this is because many orbits over the Atlantic and Pacific oceans became “full-time” glint-mode orbits beginning in November 2015 (Crisp et al., 2017). Prior to that, there were equal numbers of nadir and glint orbits, but after that change, approximately one-third of all orbits are nadir and two-thirds are glint.

3 The NASA ACOS XCO2 retrieval algorithm as applied to OCO-2

The original ACOS XCO2 retrieval algorithm over land (version 2.9) was described in O’Dell et al. (2012), with details specific to GOSAT given in Crisp et al. (2012). Details of the spectroscopy used at that time were published in Thompson et al. (2012). In this section, we give an overview of the evolution from ACOS version 2.9 to OCO-2 versions 7 and 8, including spectroscopy, aerosol treatment, and a number of other changes.

Briefly, the NASA ACOS algorithm uses optimal estimation to solve for parameters of a state vector to obtain the best match to spectra from the three GOSAT or OCO-2 near-infrared bands and consistent with a prior constraint. These bands are the O2A band at 0.76 μm (band 1), the weak CO2 band at 1.61 μm (band 2), and the strong CO2 band at 2.06 μm (band 3). The state vector parameters, listed in Table 2, include the profile of CO2 at twenty atmospheric levels along with a number of ancillary parameters to which the GOSAT and OCO-2 near-infrared spectra are sensitive. These include surface pressure, surface albedo parameters (over land only), a temperature profile offset and water vapor profile multiplier, and parameters related to the wavelength scale of the spectra (dispersion shift and stretch). The latter are relative to the preflight values of these parameters, described in Lee et al. (2017). Because telluric line positions are known with high accuracy, the retrieval solves for them with virtually no dependence on the prior. To account for scattering effects of thin cloud or aerosol, the retrieval also solves simultaneously for amounts and Gaussian vertical profiles (as described in Sect. 3.1) of five different kinds of scatterers with fixed optical properties: a water cloud type, an ice cloud type, two fixed aerosol types, and beginning in version 8, an Upper Tropospheric/Lower Stratospheric (UTLS) sulfate aerosol layer. In addition, the retrieval also fits scaling factors for three spectral patterns per band, to account for imperfections in the spectroscopy, solar model, and instrument model, and determined using singular value decomposition of our fit residuals run on clear-sky soundings (Sect. 3.3). For solar-induced fluorescence (SIF) emission from plants on land, we fit for two SIF parameters which are needed to account for this fluorescence in the L2 spectra (Sect. 3.5). These SIF parameters are not the official SIF data product; that product is derived from the IMAP prescreener through a dedicated fit (Sun et al., 2018). In total, there are typically 55 fitted parameters for land retrievals and 53 for ocean1. With the exception of CO2, the a priori covariance matrix is diagonal, with the 1σ uncertainties as given in Table 2.

The first documented algorithm version, B2.9 as described in O’Dell et al. (2012), had several deficiencies which occasionally produced large biases in the retrieved XCO2 (Wunch et al., 2011a). This early version of the algorithm also contained some cumbersome traits, such as a variable number of vertical levels from sounding to sounding, which made the output difficult to use. The observed XCO2 biases were partially related to the aerosol parameterization, demonstrated by the fact that clear-sky retrievals of clear-sky simulations did not exhibit substantial biases (O’Dell et al., 2012). Furthermore, errors in the O2 and CO2 spectroscopy were suspected to be an additional source of bias. Over the course of several years, a number of changes to the algorithm were therefore implemented to yield the present version, B8. The changes are too numerous to fully describe here, but the most important ones are listed in Table 3. The changes fall into several major categories, with spectroscopy, aerosol treatment, treatment of the ocean surface, and chlorophyll fluorescence being the most important. In B8, the meteorology used to prescribe the a priori temperature profile, water vapor profile, and surface pressure was also changed (Sect. 3.5).

Further, as listed in Table 4, some minor retrieval differences exist between the GOSAT and OCO-2 versions of the algorithm. Besides using instrument models specific to each

1This excludes parameters in our state vector with prior uncertainties close to zero, such as cloud and aerosol layer widths.
Table 2. General setup of the ACOS state vector.

<table>
<thead>
<tr>
<th>Element</th>
<th>No. of elements</th>
<th>Prior value</th>
<th>Prior uncertainty (1σ)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂ values</td>
<td>20</td>
<td>Same as TCCON</td>
<td>Same as ACOS B2.9</td>
<td>Defined on sigma pressure levels</td>
</tr>
<tr>
<td>Temperature offset</td>
<td>1</td>
<td>0 K</td>
<td>5 K</td>
<td>Rel. to prior profile</td>
</tr>
<tr>
<td>Surface pressure</td>
<td>1</td>
<td>From prior meteorology</td>
<td>4 hPa</td>
<td>Prior unc. 1 hPa for B3.5</td>
</tr>
<tr>
<td>H₂O scale factor</td>
<td>1</td>
<td>1.0</td>
<td>0.5</td>
<td>Multiplier on prior profile</td>
</tr>
<tr>
<td>Aerosol type 1,2 OD₇₅₅</td>
<td>2</td>
<td>From MERRA</td>
<td>± factor of 7.39</td>
<td></td>
</tr>
<tr>
<td>Water, ice cloud OD₇₅₅</td>
<td>2</td>
<td>0.0125</td>
<td>± factor of 6.05</td>
<td></td>
</tr>
<tr>
<td>Aerosol type 1,2 σₐ</td>
<td>2</td>
<td>0.9</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Water cloud x₁₀</td>
<td>1</td>
<td>0.75</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Ice cloud x₁₀</td>
<td>1</td>
<td>Just below tropopause</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Aerosol type 1,2 σₐ</td>
<td>2</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Water cloud σₐ</td>
<td>1</td>
<td>0.1</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Ice cloud σₐ</td>
<td>1</td>
<td>0.04</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>UTLS aerosol OD₇₅₅</td>
<td>1</td>
<td>0.006</td>
<td>± factor of 6.05</td>
<td>Introduced in B8</td>
</tr>
<tr>
<td>Albedo mean land</td>
<td>1 per band</td>
<td>Prior calc.</td>
<td>1.0</td>
<td>Land</td>
</tr>
<tr>
<td>Albedo slope land</td>
<td>1 per band</td>
<td>0.0</td>
<td>0.0005</td>
<td>Land; units of 1/cm⁻¹</td>
</tr>
<tr>
<td>Albedo mean ocean</td>
<td>1 per band</td>
<td>0.02</td>
<td>[0.2, 0.2, 1e⁻³]</td>
<td>Ocean</td>
</tr>
<tr>
<td>Albedo slope ocean</td>
<td>1 per band</td>
<td>0.0</td>
<td>1.0</td>
<td>Ocean; units of 1/cm⁻¹</td>
</tr>
<tr>
<td>SIF mean</td>
<td>1</td>
<td>Prior calc.</td>
<td>0.008</td>
<td>Land</td>
</tr>
<tr>
<td>SIF slope</td>
<td>1</td>
<td>0.0018</td>
<td>0.0007</td>
<td>Land; units of 1/cm⁻¹</td>
</tr>
<tr>
<td>Wind speed</td>
<td>1</td>
<td>From prior meteorology</td>
<td>5 m s⁻¹</td>
<td>Ocean</td>
</tr>
<tr>
<td>Dispersion shift</td>
<td>1 per band</td>
<td>0.0</td>
<td>0.4 of channel FWHM</td>
<td></td>
</tr>
<tr>
<td>Dispersion stretch</td>
<td>1 per band</td>
<td>0.0</td>
<td>1 pm/channel</td>
<td>OCO-2 only</td>
</tr>
<tr>
<td>EOF amplitudes</td>
<td>3 per band</td>
<td>0.0</td>
<td>10.0</td>
<td>1 per band for B3.5 &amp; earlier</td>
</tr>
</tbody>
</table>

Table 3. Significant ACOS retrieval algorithm changes.

<table>
<thead>
<tr>
<th>GOSAT B2.10</th>
<th>GOSAT B3.3</th>
<th>GOSAT B3.4</th>
<th>GOSAT B3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian aerosol profiles</td>
<td>Residual fitting</td>
<td>Updated ocean surface</td>
<td>MERRA aerosol types</td>
</tr>
<tr>
<td>Sigma pressure levels</td>
<td>1 hPa Pₘₚ prior uncertainty</td>
<td>Band 2 spectral range</td>
<td>Spectroscopy update</td>
</tr>
<tr>
<td>Prior CO₂ profile change</td>
<td>Prior OD₇₅₅ = 0.05</td>
<td>Spectroscopy update</td>
<td></td>
</tr>
<tr>
<td>Spectroscopy update</td>
<td>Spectroscopy update</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected X₇₅₅AK</td>
<td>Fluorescence fit land gain H</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOSAT B7.3</th>
<th>OCO-2 B7</th>
<th>OCO-2 B8</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 EOFs per band</td>
<td>Restricted band ranges</td>
<td>Spectroscopy update</td>
</tr>
<tr>
<td>2hPa Pₘₚ prior uncertainty</td>
<td>4hPa Pₘₚ prior uncertainty</td>
<td>UTLS aerosol</td>
</tr>
<tr>
<td>Updated cloud ice properties</td>
<td></td>
<td>L1B improvements</td>
</tr>
</tbody>
</table>

instrument (such as wavelengths of the various channels, noise model, and instrument line shape functions), slightly different spectral ranges are fit for each instrument. Generally, this is because the trusted calibrated range of OCO-2 spectra is slightly smaller than that of GOSAT, due to the differences in design of the OCO-2 grating spectrometer versus the GOSAT Fourier transform spectrometers. Additionally, while all channels in each band in the given spectral ranges are used for GOSAT, some band channels are masked out for OCO-2. This is due to either underlying bad pixels in the detector arrays or to transient cosmic rays that induce temporary spurious readings in random channels. Both of these processes are described in detail in Crisp et al. (2017).

3.1 Aerosol-related changes

Starting with version B2.10, the 20-layer optical depth retrieval used for clouds and aerosols was replaced with a Gaussian-shaped vertical profile for each of the retrieved scattering particle types. As of version 8, two cloud types, two lower-atmosphere aerosol types, and one stratospheric aerosol are used. The new cloud and aerosol profile treatment is similar to that of Butz et al. (2009) but specifies the aerosol concentration ρₐₑᵣ as a function of x, the pressure relative to
Table 4. ACOS retrieval differences between GOSAT and OCO-2.

<table>
<thead>
<tr>
<th>Category</th>
<th>GOSAT</th>
<th>OCO-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiance used</td>
<td>Estimated total intensity</td>
<td>OCO-2 single polarization</td>
</tr>
<tr>
<td>EOFs, band ranges</td>
<td>Wavenumber space</td>
<td>Channel space</td>
</tr>
<tr>
<td>Fit O$_2$A band offset ?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SIF prior</td>
<td>0</td>
<td>From IMAP retrieval</td>
</tr>
<tr>
<td>Per-band dispersion parameters</td>
<td>Offset only</td>
<td>Offset, slope</td>
</tr>
<tr>
<td>Band 1 fitted range</td>
<td>758.1–772.2 nm</td>
<td>759.2–771.5 nm</td>
</tr>
<tr>
<td>Band 2 fitted range</td>
<td>1597.4–1618.1 nm</td>
<td>1598.1–1617.9 nm</td>
</tr>
<tr>
<td>Band 3 fitted range</td>
<td>2042.1–2079.0 nm</td>
<td>2047.8–2079.9 nm</td>
</tr>
<tr>
<td>Channel mask</td>
<td>None</td>
<td>Bad samples, spikes</td>
</tr>
</tbody>
</table>

The change to a sigma-level pressure system was incorporated at about the same time as the shift to Gaussian aerosol profiles. Instead of fixed pressure levels, the pressure levels scale with the surface pressure:

$$p_i = a_i \, p_{\text{surf}},$$  \hfill (2)$$

where the $a_i$ are chosen such that the total number of pressure boundaries is 20, and the layers have roughly equal pressure widths. The top-most model level is set to 0.01 hPa.

The optical properties of the four scattering types remained unchanged from versions B2.9 to B3.4 and are described in O’Dell et al. (2012). However, the use of two fixed aerosol types, types “2b” and “3b” from the Kahn et al. (2001) climatology, did not accurately represent the true global variability of aerosol on the length scales and timescales probed by GOSAT and OCO-2. Beginning with build 3.5, the aerosol types were changed to be location- and time-dependent, with the prior type information coming from the aerosol climatology of the Modern-Era Retrospective analysis for Research and Applications (MERRA, Rienecker et al., 2011). The MERRA aerosol field is driven by the Georgia Tech/Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model (Chin et al., 2002), and modified by assimilating aerosol optical depth from the MODIS instruments onboard the Terra and Aqua satellites (Colarco et al., 2010). MERRA contains five broad aerosol types: dust (DU), sea salt (SS), sulfate (SO), and black and

![Figure 2. Prior Gaussian profiles of the lower tropospheric aerosol types (red), water cloud (blue), ice cloud (purple), and stratospheric aerosol (green). The local aerosol optical depth (AOD) per unit pressure at 755 nm is plotted as a function of the relative pressure. The lower tropospheric aerosol prior optical depth is not fixed as for the other types, but rather is taken from a climatology described in the text.](image-url)
organic carbon (BC and OC, respectively). Dust and sea salt are each tracked in five separate size bins. Organic carbon and black carbon are tracked in both hydrophobic and hydrophilic categories. In addition to the carbonaceous types, sulfate aerosol and sea salt are also hydrophilic and hence have optical properties that depend on the local relative humidity (RH).

For the aerosol prior in the ACOS retrieval, we primarily sought to specify the typical dominant aerosol types present (in terms of their contribution to the optical depth in the OCO-2 bands) in a given location at a given time of year. Monthly aerosol fields were derived from the MERRA model for the year 2010, and are used for all years in the ACOS retrieval. We aggregated the 15 MERRA types, eight of which have RH-dependent optical properties, into the five aggregated types listed above. We used typical density weightings and relative humidity values to create the optical properties for these aggregated types, as described in Crisp et al. (2017). Their optical properties, including extinction efficiency, single scattering albedo, and asymmetry parameter, are shown in Fig. 3. The organic carbon and sulfate aerosol are generally similar in their optical properties, though their single scattering albedos diverge somewhat in the CO$_2$ bands. The sea salt, water cloud, and dust optical properties are relatively similar across the OCO-2 spectral range.

At each sounding location, the two aggregated aerosol types with the highest mean monthly values of the OD$_{755}$ are selected to be retrieved by the L2FP algorithm. In previous algorithm versions, the total prior OD$_{755}$ was set to 0.15, apportioned equally among four scattering types (water cloud, ice cloud, and two tropospheric aerosol types). However, it was found this was generally too high to allow a fit near OD$_{755} = 0$ for scenes that were almost entirely free of aerosol. This “clear-sky bias” was seen in early simulation tests (O’Dell et al., 2012). The prior OD$_{755}$ is now set to 0.0125 for each cloud type, and set from the MERRA aerosol climatology for each tropospheric aerosol type as the average OD$_{755}$ of that type (at a particular location and month). There is some evidence that the tropospheric aerosol priors are occasionally still too high; methods for specifying the aerosol prior are a continuing topic of investigation.

The cloud ice optical properties were updated in version 7. Before that, they were based on the band-averaged model developed by Baum et al. (2005) primarily for the MODIS instrument and known as the MODIS Collection 5 model. This
cloud ice model considered an ensemble of size-dependent non-spherical ice crystal habits in random orientation. As ice crystal surface roughness was later shown to significantly affect scattering by ice crystals, and simulations with roughened model particles were more consistent with satellite observations of ice cloud polarized reflectances (Yang et al., 2013), we updated the cloud ice optical properties to correspond to the MODIS Collection 6 model, which describes scattering by severely roughened hexagonal column ice crystal aggregates (Baum et al., 2014). This update also fixed several minor issues in the previous cloud ice model, such as those resulting from linear interpolation of the optical properties from MODIS wavelength bands to OCO-2, and those relating to truncation of the phase function.

3.1.1 The need for a stratospheric aerosol

When validating version 7 XCO₂ retrievals, it was discovered via comparisons to both TCCON and models that most ocean soundings in the most southerly 10 degrees of latitude exhibited a high bias of 1–3 ppm during the austral winter (Wunch et al., 2017). Figure 4 shows the bias appears in the Southern Hemisphere winter over the Wollongong TCCON station. The bias is seen in soundings over ocean but not land. The bias is also apparent relative to the Lauder TCCON station (Figs. 11 and A1 from Wunch et al., 2017). Comparisons of OCO-2 soundings to models (Fig. 19) show the bias as a quasi-zonal band over the Southern Hemisphere oceans, again with the larger bias occurring in the Southern Hemisphere winter. There is also evidence of a similar but weaker band of high bias in the Northern Hemisphere. For 2015, it was hypothesized that small aerosol particles may have been injected into the UTLS by the explosive eruptions of the Calbuco (22–30 April 2015) and Wolf (late May 2015) volcanos in southern–central Chile and the Galapagos Islands, respectively. The presence of an aerosol layer with visible optical depths around 0.01 was later confirmed with observations from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observatory (CALIPSO) and the Ozone Mapping Profile Suite (OMPS) satellites (Bègue et al., 2017). These optical depths are small, but have a large impact on the radiances, especially in the O₂A band, due to their high altitude.

It was recognized that our version 7 retrieval algorithm had no way to accommodate the spectral signature of small stratospheric aerosol particles, which have a significantly larger effect on the O₂A band than either of the CO₂ bands due to the small size parameter, i.e., the ratio of the size of the scattering particle to the spectral wavelength. The spectral signature would essentially appear as a radiance offset in the O₂A band. As a first test, we ran hundreds of retrievals on a single sounding that had a large positive bias in the operational retrieval, using slightly different first-guess values for each retrieval. Essentially, a continuum of solutions was found (Fig. 5). On one end of retrieval space, an approximately correct value of surface pressure was found by inserting a thicker ice cloud, which contains larger particles starting in the stratosphere, and therefore has a strong effect on all three bands (see Fig. 3). This type of solution produced a poor χ² in the strong CO₂ band, typically > 2. On the other end of the continuum were solutions where the sulfate layer, which was placed near the surface in the prior, was moved high up into the atmosphere. This solution regime had a much lower reduced χ² (around 1.5) in the strong CO₂ band and an XCO₂ that was typically 3–4 ppm lower, and much more in line with TCCON and model estimates. In these cases, the amount of upper atmosphere cloud ice retrieved was also reduced, as its role was taken over by the sulfate.

These tests indicated that a more realistic solution would often be found if the retrieval could push the prior sulfate into the upper atmosphere, though this seldom occurred. The amount of sulfate needed in the upper atmosphere in these cases is small, approximately 0.01 optical depth at 755 nm. That value is consistent with other observations (Bègue et al., 2017). In addition to actual small particles in the UTLS, the OCO-2 instrument has a documented problem which produces a similar impact on the O₂A and spectrum. As described in Crisp et al. (2017, Sect. 6.5), a very thin layer

![Figure 5](image-url)
of ice appears to build up on the OCO-2 Focal Plane Arrays (FPAs) over time. As this ice layer grows to a thickness similar to the anti-reflective coating thickness (tens of nanometers), it enhances the surface reflectance on the O$_2$A band FPA, producing a scattered light contribution of 0.1 to 0.2%. Much smaller effects are seen on the CO$_2$ detectors. The ice layer is sublimed off every 3–6 months when the instrument goes through a decontamination cycle. While attempts have been made to remove this scattered light contribution in the version 8 calibrated radiance (L1B) product, it is likely that some residual signal remains. Because this is primarily a radiance offset in the O$_2$A band alone, it produces a signal similar to a small UTLS aerosol, and hence would also be mitigated by including a stratospheric aerosol in the retrieval. During algorithm testing of the stratospheric aerosol using version 7 L1B radiances (which contained the scattered light signature), we found that the amount of UTLS aerosol retrieved indeed correlated with the decontamination cycles, lending credence to this hypothesis.

Thus, in version 8 an additional sulfate aerosol was included in the retrieval state vector. For simplicity, a sulfate type identical to the lower-atmosphere type in terms of optical properties was used. Only the total optical depth of the stratospheric sulfate is retrieved, while its Gaussian height and width are kept fixed. This solution treats both actual small particles in the UTLS as well as the radiometric offsets that accompany the real O$_2$A band scattered light signal. Our testing of the version 8 algorithm showed that including this state vector element not only reduced the Southern Ocean bias, but also reduced the negative tropical ocean bias and positive bias over higher northern-latitude lands that were also apparent in Fig. 19. A more complete comparison of version 7 and version 8 validation statistics is given in Sect. 5.

### 3.2 Spectroscopy-related changes

There have been substantial changes between the molecular cross sections used in the earliest ACOS versions and those used in B8. We continue to use in-house lookup tables of absorption coefficients (ABSCO) parameterized as a function of temperature, pressure, wavelength, and water vapor mixing ratio for each of the main absorbing gases in the OCO-2 bands: O$_2$, CO$_2$, and water vapor (H$_2$O). Successive versions of these tables have been refined by incorporating new laboratory results and theoretical models for increasingly accurate absorption coefficients. The ABSCO version used in the B8 algorithm is ABSCO v5.0 (Drouin et al., 2017; Oyafuso et al., 2017); B7 used the previous ABSCO version, v4.2.

The ABSCO v5.0 O$_2$A band tables represent a major step forward from previous ABSCO versions. Earlier ABSCO versions integrated the highest-quality spectroscopic input from a range of studies that had focused on fitting different parameters independently. (See, for example, Thompson et al., 2012, and references therein.) The ABSCO v5.0 tables are based on self-consistent multispectral fits to laboratory spectra that include line mixing, speed-dependent Voigt line shape parameters, and collision-induced absorption (CIA). This self-consistency, and the use of laboratory spectra covering a range of pressures, temperatures, and measurement techniques, are key features of the approach. The O$_2$ spectral line parameters, line mixing, and CIA used in ABSCO v5.0 are described in Drouin et al. (2017). Parameters for broadening of O$_2$ by H$_2$O are from the study by Drouin et al. (2014).

The impact of the latest multi-spectrum fitting update in the O$_2$A band is shown in terms of the accuracy of the retrieved surface pressure in Fig. 6. Panel (a) shows the retrieved surface pressure minus the prior for ABSCO version 4.2, which was used in version 7 of the algorithm, while panel (b) shows the same for ABSCO v5.0, used in version 8. The main improvements seen are that the retrieved surface pressures in version 8 are essentially unbiased with respect to the meteorological prior over land, and that the land and ocean differences are reduced and centered closer to zero. We also note that for OCO-2, no additional line strength scaling was required in the O$_2$A band, as has been necessary for all previous ACOS/GOSAT versions (see, e.g., Crisp et al.,

---

Figure 6. Retrieved minus prior surface pressure for a large selection of OCO-2 soundings, using the oxygen-A band spectroscopy model from both (a) ABSCO v4.2 and (b) ABSCO v5.0, as described in the text. ABSCO v5.0 spectroscopy leads to a more consistent retrieval of surface pressure over both land and ocean surfaces.
For ACOS-GOSAT B3.5 retrievals, an O$_2$ scaling factor of 1.0125 was found to be beneficial, perhaps because of slight instrumental differences between OCO-2 and GOSAT.

The ABSCO v5.0 tables for the 1.61 and 2.06 µm CO$_2$ bands use line parameters and line mixing models derived from self-consistent, multispectral fits by Devi et al. (2016) and Benner et al. (2016), respectively. The parameters are derived from fits to laboratory spectra at multiple pressures and temperatures and the computation incorporates a speed-dependent Voigt line profile with nearest-neighbor line mixing. Earlier versions of the ABSCO tables (Benner et al., 1995; Devi et al., 2007) were based entirely on room temperature multi-spectrum fitting, with theoretical temperature dependences of the line shape and line mixing parameters. The updated spectroscopy includes analyses of spectra recorded at temperatures from 170 to 296 K, representing a significant advance. Parameters for broadening of CO$_2$ by H$_2$O are from Sung et al. (2009) for the 4.3 µm CO$_2$ band and extrapolated to OCO-2’s CO$_2$ bands. Validation of the ABSCO v5.0 tables using up-looking TCCON spectra is described in Drouin et al. (2017) and Oyafuso et al. (2017). We note one important difference between the reference databases and our CO$_2$ absorption coefficients. We found it necessary to incorporate additional absorption in the center of the 2.06 µm band. This additional absorption was parameterized in order to reduce errors in retrievals with TCCON up-looking spectra. Further details can be found in Thompson et al. (2012) and Oyafuso et al. (2017).

Because the laboratory spectra underlying ABSCO currently are only good to roughly 1% absolute accuracy of line intensities, the algorithm allows for overall scaling factors for each of the two CO$_2$ bands. For the 1.61 µm band, the ABSCO v5.0 tables include a uniform scaling to bring the intensities from the Devi et al. (2016) multispectrum fit into line with reference intensity measurements (estimated accuracy ∼ 0.2%) from Polyansky et al. (2015). Oyafuso et al. (2017) show that this pre-scaling of the ABSCO using reference laboratory measurements results in good consistency between single-band up-looking X$_{CO_2}$ retrievals from ground-based FTS spectra and the X$_{CO_2}$ values reported by TCCON (which are themselves calibrated to agree with reference airborne profiles). Reference intensity measurements are not available for the 2.06 µm band at the current time. In tests within the OCO-2 Level 2 algorithm, using OCO-2 radiances, a scaling of 1.004 for the 2.06 µm ABSCO table was found to yield the best agreement between single-band retrievals performed using this band compared with single-band retrievals performed using the 1.61 µm ABSCO table as described above.

Finally, ABSCO v5.0 tables incorporate H$_2$O line parameters from the HITRAN 2012 compilation (Rothman et al., 2013). We use an unofficial, modified version of the MT_CKD continuum, supplied by Eli Mlawer (Mlawer et al., 2012). This continuum version offers a compromise between previous versions of MT_CKD and measurements by Ptashnik et al. (2011), and falls relatively close to measurements by Mondelain et al. (2013). The subsequently released MT_CKD 3.2 has been tested and shown to be a modest improvement over the unofficial version incorporated into ABSCO v5.0, with negligible changes to X$_{CO_2}$ but noteworthy improvements to the water column determination. Methane is not currently included in the B8 (or previous versions) of the forward model, as the impact of methane absorption was found to be negligible for X$_{CO_2}$ retrievals performed using the OCO-2 spectral ranges.

### 3.3 Residual fitting using empirical orthogonal functions (EOFs)

ACOS B3.3 introduced a new way to deal with large spectral residuals caused by imperfect spectroscopy, and solar model and instrument characterization, which were previously treated using a simple “empirical noise” parameterization (Crisp et al., 2012). In contrast, the new approach fits scaling factors to fixed spectral residual patterns for each band.

These patterns are the EOFs that result from a singular value decomposition of spectral residuals from training retrievals. Training scenes were selected to be largely devoid of cloud and aerosol effects, such that residual patterns due to unfitted clouds and aerosols are not a large contributor to the resulting EOF patterns. The EOFs are constructed such that the residuals $r_{s,b}$ of each sounding $s$ and band $b$ can be approximately represented as a linear combination of the EOF patterns:

$$r_{s,b} = \sum_{j=1}^{N_{\text{EOF}}} c_{j,s,b} e_{j,b},$$

where the vectors $e_{j,b}$ are the EOFs for each band. For a diverse set of training retrievals, a matrix $\mathbf{M}$ is created for each spectral band and populated by the residuals of the spectral fits within that band. Training sets typically included more than 10,000 soundings.

Each matrix $\mathbf{M}$ is then decomposed into its eigenvectors using traditional singular value decomposition:

$$\mathbf{M} = \mathbf{U} \mathbf{W} \mathbf{V}^T,$$

with the columns of $\mathbf{U}$ spanning an orthonormal basis of the most persistent spectral residual vectors observed in the training data set. 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 $\mathbf{W}$).

Application of this EOF technique substantially reduces the spectral residuals, yielding values of the relative rms of the residuals of ∼ 0.1% for each band, and reduced $\chi^2$ values near unity. For GOSAT, only the first EOF was found to be necessary. For OCO-2 B7 (and B8) target-mode observations, better agreement with TCCON X$_{CO_2}$ was found.
when the largest three EOFs were employed. The first two EOFs for GOSAT, as well as the first three EOFs for OCO-2, are shown in Fig. 7 for each spectral band. The EOFs for each of the eight spatial footprints sampled by OCO-2 are extremely similar, though they are solved for independently due to the slightly different spectroscopic response of each. The first EOFs for OCO-2 and GOSAT are very similar for each band, indicating that common forward model errors such as spectroscopy and top-of-atmosphere solar flux, rather than instrument-specific effects, are driving the EOF patterns. The first EOF is also very similar to the mean residual pattern, and typically accounts for 50%–60% of the variance in the residuals. The second and third EOFs typically account for only 1%–3% of the variance in the residuals, with higher-order EOFs accounting for even less. For the O$_2$A band, the second EOF appears as a Doppler shift of the first. In the weak CO$_2$ band, the second EOF appears to be due to water vapor lines, while the third EOF appears to be a Doppler shift of the first. Higher-order EOFs often exhibit additional instrument artifacts (such as unidentified bad spectral samples) and forward model errors related to water vapor, as well as other effects that are difficult to interpret.

The EOF formulation was modified in B8 in several ways. First, the EOFs were defined in terms of radiance per unit noise rather than pure radiance, in order to be consistent with the cost function metric that is minimized during the retrieval itself. However, the structure of the EOFs in B8 was much the same as in B7. Second, improved filtering of spectral samples that are contaminated by noisy or dead detector pixels greatly reduced their impact on the EOF patterns. Finally, a manual re-ordering of the EOFs was performed for each of OCO-2’s eight spatial footprints, because the standard ranking by variance would occasionally flip the EOF patterns in different footprints. This mattered because sometimes the third and fourth EOFs would change places (and the current algorithm only fits the first three EOFs). This ensured that, to the extent that the EOFs of each footprint roughly matched, the same three EOF patterns for each footprint are fit within the L2FP retrieval.

### 3.4 Surface model

The forward model for ACOS L2FP retrievals uses one of two surface models, depending on the location of the footprint. Water surfaces are simulated as a linear combination of a Cox–Munk ocean surface (Cox and Munk, 1954) and Lambertian reflector. This surface has seven parameters: wind speed, and a Lambertian albedo at a reference wavenumber with a linear spectral slope term in each of the three spectral bands. The prior wind speed is taken from the resampled meteorology (either ECMWF or GEOS5 FP-IT, as discussed in the next section), as for the other meteorological parameters. The strong CO$_2$ band Lambertian albedo is fixed to 0.02; the other six terms are fit in an essentially unconstrained fashion. This approach leads to the fitted Lambertian albedos gener-

---

**Figure 7.** Spectral patterns of the EOFs for the O$_2$A band (a), weak CO$_2$ band (b), and strong CO$_2$ band (c), for both GOSAT B3.5 (green) and OCO-2 B7 footprint 4 (black). For reference, the light gray trace in each panel shows the modeled spectrum (not to scale). Some stronger water vapor absorption lines (blue vertical lines) in the weak and strong CO$_2$ bands correlate with features in the second and third EOFs.
ally staying small and positive, the latter of which is currently required by our radiative transfer module.

Through ACOS B7, land surfaces were assumed to be purely Lambertian, with an albedo and albedo spectral slope retrieved for each band. The Lambertian surface assumes that the bidirectional reflectance distribution function (BRDF; the ratio of the radiance in the reflected direction to the irradiance from the incident direction; Schaeppman-Strub et al., 2006) is a constant that is often specified as a scalar albedo. Since independent fits are done within each of the three OCO-2 spectral bands, this yields six state variables for land footprints.

Analysis of B7 OCO-2 target-mode observations showed that the retrieved Lambertian albedo and aerosol optical depth sometimes exhibited dependence on the sensor zenith angle for observations of the same surface location. This indicated that the true surface BRDF has dependence on the observation angles. A more physically justified approach would use a non-Lambertian model for the surface BRDF. For trace gas retrievals, a Lambertian surface assumption introduces no errors in the absence of multiple scattering between the surface and atmosphere; in this case, the retrieved albedo is interpreted as the surface reflectance at the primary scattering geometry (Sun–surface–satellite). However, over brighter surfaces with some atmospheric scattering, the assumed BRDF could in principle affect the retrieval via the interaction of the retrieved aerosol, surface pressure, and gas concentrations. Therefore, in B8 it was decided to change the surface model for land footprints to a non-Lambertian surface model. This model assumes a fixed BRDF shape and assumes the surface is azimuthally symmetric, but allows for spectral dependence of the amplitude between and within each of our three bands; full details of the BRDF model are given in Appendix B. While this model does often show reduced correlation between view zenith and the retrieved BRDF amplitude, the retrieved X$_{CO_2}$ (as well as most other state vector parameters) shows very little change versus a version of the B8 retrieval run with a Lambertian surface. Therefore, while B8 does use a non-Lambertian BRDF parameterization, a Lambertian surface appears to work equally well. This fact may be a consequence of the strong filtering used in B8, which tends to remove soundings with multiple scattering. Future applications of the ACOS L2FP algorithm to cases with higher AOD may be more strongly impacted by the non-Lambertian BRDF.

3.5 Additional retrieval algorithm changes

In addition to these changes, a number of additional (mostly minor) changes have also been made to the ACOS L2FP algorithm since B2.9. In B2.10, the prior CO$_2$ profile was changed to match that used by TCCON, which was more realistic than our previous prior formulation; as of B8, this corresponds to the GGG2014 version (Toon and Wunch, 2014). Generally speaking the TCCON CO$_2$ prior profile is relatively simple: it is a function of latitude, altitude, and date only. It includes a simple formulation of the seasonal cycle and currently assumes a fixed secular increase of 0.52 % yr$^{-1}$ (or 2.08 ppm yr$^{-1}$ at 400 ppm). There is no land–ocean or other meridional dependence. It requires specifying the tropopause height, and has simple formulations for the profile in the boundary layer, free troposphere, and stratosphere. A small mistake in the X$_{CO_2}$ averaging kernel was also fixed in B2.10; this was caused by inconsistent assumptions regarding the pressure-dependent gas absorption cross sections throughout our retrieval code, which led to an obvious “kink” in the averaging kernel that had long been visually evident (see, e.g., Fig. 2 of Connor et al., 2008). We use the 2016 version of the Toon solar transmittance spectrum (available at https://mark4sun.jpl.nasa.gov/toon/solar/solar_spectrum.html, last access: 4 December 2018) (Toon, 2014). Changes in the prior covariance matrix for CO$_2$ (O’Dell et al., 2012) were also considered, but rejected, as tests using alternate covariance matrices showed insignificant performance improvements.

In B3.3, solar-induced chlorophyll fluorescence (SIF) fitting over land surfaces was introduced. This change was introduced to combat a bias in X$_{CO_2}$ that results from not fitting for fluorescence when it is present, due to its impact on the O$_3$A band. This problem and our fluorescence fitting scheme are described in detail in Frankenberg et al. (2012). Briefly, we fit for the mean and slope of the fluorescence at the top-of-canopy as a function of wavelength; the mean is expressed as a fraction of the continuum radiance level at a wavelength of 755 nm. The spectral dependence of the fluorescence is taken to be linear. Additional minor algorithm changes to B3.3 included reducing the prior surface pressure uncertainty from 4 to 1 hPa for GOSAT, which is likely to be a more accurate representation of the true prior surface pressure uncertainty for the majority of scenes (see, e.g., Salstein et al., 2008). This has the added benefit of reducing the interference error between SIF, aerosols, zero-level offset, and surface pressure. In B3.4, fitting for SIF was turned off for GOSAT medium gain observations over land, as these regions are nearly all desert with very little biological activity. For OCO-2, the SIF prior is taken from the official SIF retrieval, as described in Sun et al. (2018), because SIF retrievals from individual soundings are meaningful for OCO-2 due to its relatively high SNR.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Land</th>
<th>Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface pressure (hPa)</td>
<td>0.07 ± 0.73</td>
<td>0.05 ± 0.43</td>
</tr>
<tr>
<td>T2m (K)</td>
<td>0.4 ± 2.7</td>
<td>0.3 ± 0.6</td>
</tr>
<tr>
<td>T@700 hPa (K)</td>
<td>0.0 ± 0.8</td>
<td>-0.2 ± 0.8</td>
</tr>
<tr>
<td>TCWV (kg m$^{-2}$)</td>
<td>0.1 ± 1.8</td>
<td>0.4 ± 2.4</td>
</tr>
<tr>
<td>Surface wind speed (m s$^{-1}$)</td>
<td>0.4 ± 1.3</td>
<td>-0.4 ± 0.9</td>
</tr>
</tbody>
</table>
In B8, the prior height of the cirrus cloud layer relative to the surface pressure was moved slightly, from the fixed value of \( x = 0.3 \) to just below the tropopause height (which is a relatively strong function of latitude). The calculation of the tropopause height itself was also refined in B8, which also improved the calculation of the prior CO\(_2\) profile. Finally, the prior meteorology was changed in B8 from ECMWF to GEOS5-FP-IT (Suarez et al., 2008; Lucchesi, 2013). Some statistics regarding differences in surface pressure, temperature, water vapor, and surface wind speed between the two models for retrieved GOSAT soundings are given in Table 5; the corresponding difference statistics for OCO-2 soundings are nearly identical. Only soundings passing the O\(_2\)A band prescreener are included. In general, the two models are very similar, with for instance 95% of all soundings having a surface pressure difference of less than 1.5 hPa. Surface pressure probably affects our retrieved X\(_{\text{CO}_2}\) the most, as it is used not only in the retrieval, but also in the bias correction, where differences in the prior surface pressure will lead to a first-order change in the bias-corrected X\(_{\text{CO}_2}\). Currently, the “noise” from the surface pressure difference between these two models would amount to roughly 0.6 hPa, or about 0.25 ppm in X\(_{\text{CO}_2}\), which is quite a bit less than our noise-driven error (\( \sim 1.4 \) hPa on average) and regional biases (\( \sim 2.4 \) hPa on average). Retrieved surface pressure errors are discussed in more detail in Sect. 4.3.4.

### 4 Retrieval filtering and bias correction

All soundings passing the prescreening criteria (Table 1) are processed with the L2FP retrieval algorithm. Of these, some 10%–20% fail to converge to a solution, typically because of unscreened clouds or other factors that cannot properly be modeled in the retrieval. Some fail simply because of the nonlinear nature of the problem – in general, there is no perfect way to minimize the cost function. Of the 80%–90% of soundings that do converge to a minimum in the cost function, typically three to six iterations are required. Despite our best efforts to prefilter problematic soundings, there are inevitably some retrievals with X\(_{\text{CO}_2}\) errors that exceed those predicted by theory. Ideally, the X\(_{\text{CO}_2}\) errors would be normally distributed, with errors consistent with the 1\(\sigma\) a posteriori uncertainty on X\(_{\text{CO}_2}\) from the retrieval (see, e.g., Rodgers, 2000), but often there are retrievals with systematically biased X\(_{\text{CO}_2}\) and/or larger-than-expected scatter. This problem is partially mitigated by applying a bias correction, which can reduce both scatter on smaller spatial scales and biases on larger spatial scales. However, problematic soundings still remain. A quality-filtering procedure then attempts to remove these soundings with larger-than-expected differences from our truth metrics. For GOSAT, this process was described in O’Dell et al. (2012) and Crisp et al. (2012). The problem of biases is dealt with via a linear bias correction (Wunch et al., 2011a). In this section, we describe both filtering and bias-correction procedures for X\(_{\text{CO}_2}\) for B8 retrievals only, unless otherwise noted. A similar procedure was used for GOSAT data as well as OCO-2 B7, but the procedures were more mature and robust for B8.

#### 4.1 Truth proxy training data sets

Both filtering and bias correction require a training data set, which consists of soundings for which we have both the OCO-2 retrieved X\(_{\text{CO}_2}\) as well as a reliable, independent estimate of X\(_{\text{CO}_2}\). The latter we call a “truth proxy”. We used four such data sets: TCCON, models, models in the Southern Hemisphere only, and a new validation method for OCO-2 called the “small-area approximation”. Table 6 lists the truth proxies used in version 8, while Fig. 8 shows the spatial distribution of the truth proxy data sets matched to actual OCO-2 soundings.
Table 6. XCO₂ truth proxies for retrieval evaluation.

<table>
<thead>
<tr>
<th>Name</th>
<th>Nₕ land</th>
<th>Nₕ ocean glint</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCCON</td>
<td>228k</td>
<td>88k</td>
<td>Geometric colocation requirement, GGG2014</td>
</tr>
<tr>
<td>Multi-model median</td>
<td>1132k</td>
<td>861k</td>
<td>Median of six models (see text)</td>
</tr>
<tr>
<td>Models_SHA</td>
<td>260k</td>
<td>251k</td>
<td>Same as above for lat &lt; 20° S</td>
</tr>
<tr>
<td>Small-area approximation (SAA)</td>
<td>795k</td>
<td>505k</td>
<td>Areas &lt; 100 km along-track</td>
</tr>
</tbody>
</table>

4.1.1 TCCON-based truth proxy

The most direct truth proxy is the comparison to TCCON, which currently has 25 operational stations globally, but with heavy representation in North America, Europe, Asia, and Oceania. For the OCO-2 B8 evaluation, the latest version of TCCON retrievals was employed (GGG2014, Wunch et al., 2015). Many schemes have been used to match air masses observed by satellites to those viewed from TCCON stations. Examples include a geographic-centric scheme (Cogan et al., 2012; Inoue et al., 2013; Oschepkov et al., 2013; Kulawik et al., 2016), a scheme based on the potential temperature at 700 hPa (Keppel-Aleks et al., 2011; Wunch et al., 2011b), model-based selection (Guerlet et al., 2013), and geostatistical selection (Nguyen et al., 2014). These more sophisticated techniques were primarily used because GOSAT had fairly sparse data and required relatively loose matching criteria to yield sufficient numbers of matched observations. This is less of a problem with OCO-2 at lower and mid latitudes, with its higher spatial sampling density. High-latitude validation with TCCON remains challenging, where OCO-2 data are still sparse.

Table 7 lists the TCCON sites used as truth proxies in this work. Training data ranges correspond to the quality-filtering and bias-correction procedures described in Sect. 4.2 and 4.3, respectively. Validation data ranges correspond to the basic and bias-correction procedures described in Sect. 4.2 and 4.3, still sparse. High-latitude validation of a problem with OCO-2 at lower and mid latitudes, with sufficient numbers of matched observations. This is less of a problem with OCO-2 at lower and mid latitudes, with its higher spatial sampling density.

4.1.2 Small-area approximation truth proxy

To supplement TCCON, we used a method new for OCO-2 called the “small area approximation”, or SAA². The SAA relies on the high spatial resolution of OCO-2 footprints (1.3 × 2.3 km²), and the relatively long decorrelation length of CO₂ concentration in the atmosphere (500–1000 km; see, e.g., Chevallier et al., 2017, Fig. 1). Specifically, this approximation assumes that for a given overpass of an area not larger than 100 km in spatial extent, XCO₂ can be considered uniform over the area. True XCO₂ variability was evaluated by Worden et al. (2017) by examining output from the GEOS-5 7 × 7 km² “nature run”. It was found to be typically less than 0.1 ppm per 100 km areas away from strong known sources, thus justifying our small area assumption. In fact, this error is considerably lower than can be obtained by any of the other truth metrics. The major drawback of this method is that it is insensitive to biases due to variables that vary slowly on these small scales, such as those related to viewing geometry and some surface and aerosol parameters.

4.1.3 Model-based truth proxies

The third validation data set is based on results from global carbon flux inverse models, and is referred to as the “multi-model median”. In order to evaluate OCO-2 retrievals against a posteriori results from an array of models, and to avoid the biases in one particular model, a suite of six models sampled at the OCO-2 sounding locations and times was used. Table 8 provides a summary of the models that were used. The models generally differed in their prior flux assumptions, prior flux uncertainty, transport model, initial conditions, spatial resolution, and inverse method, but had one commonality in that all assimilated in situ CO₂ concentration data. Because of these differences, the models often yielded a post-

---

²Not to be confused with the South Atlantic Anomaly.
Table 7. TCCON stations used in this work.

<table>
<thead>
<tr>
<th>TCCON station</th>
<th>Training date range</th>
<th>Validation date range</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ny Ålesund, Spitzbergen, Norway</td>
<td>Not used</td>
<td>May 2015–May 2017</td>
<td>Notholt et al. (2017)</td>
</tr>
<tr>
<td>Tsukuba, Japan</td>
<td>Sep 2014–Feb 2017</td>
<td>Sep 2014–May 2017</td>
<td>Morino et al. (2016a)</td>
</tr>
<tr>
<td>Wollongong, Australia</td>
<td>Sep 2014–Nov 2016</td>
<td>Sep 2014–May 2017</td>
<td>Griffith et al. (2014b)</td>
</tr>
</tbody>
</table>

Table 8. Models used in this work.

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Land biosphere</th>
<th>Inverse method</th>
<th>Transport</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAMS</td>
<td>15r2</td>
<td>ORCHIDEE</td>
<td>4D-Var</td>
<td>LMDZ</td>
<td>Chevallier et al. (2010)</td>
</tr>
<tr>
<td>Univ. Edinburgh</td>
<td>v2.1</td>
<td>CASA</td>
<td>EnKF</td>
<td>GEOS-Chem</td>
<td>Feng et al. (2009)</td>
</tr>
<tr>
<td>Jena CarboScope</td>
<td>s04_v3.8</td>
<td>Special</td>
<td>4D-Var</td>
<td>TM3</td>
<td>Rödenbeck (2005)</td>
</tr>
<tr>
<td>CarbonTracker</td>
<td>CT2015, CT-NRT_v2016-1</td>
<td>CASA</td>
<td>EnKF</td>
<td>TM5</td>
<td>Peters et al. (2007), with updates documented at <a href="https://carbontracker.noaa.gov">https://carbontracker.noaa.gov</a> (last access: 4 December 2018)</td>
</tr>
<tr>
<td>TM5-4DVar</td>
<td>2016</td>
<td>CASA</td>
<td>4D-Var</td>
<td>TM5</td>
<td>Basu et al. (2013)</td>
</tr>
<tr>
<td>OU</td>
<td>2016</td>
<td>CASA</td>
<td>4D-Var</td>
<td>TM5</td>
<td>Crowell et al. (2018b)</td>
</tr>
</tbody>
</table>

For each matched OCO-2 sounding, the model median was computed from all available models for that sounding. If any model XCO2 value differed by more than 1.5 ppm from the model median, that sounding was excluded from our training data set. This requirement helped ensure that at the very least, all the models were generally consistent with each other for a given sounding in our training set. Generally the root-mean-squared difference of the model XCO2 values was less than 0.7 ppm for any given sounding satisfying this requirement. The median of the model-predicted XCO2 for soundings satisfying this criterion was then taken as the truth estimate.

rriori XCO2 fields that disagreed to some extent, with differences ranging from a few tenths of a ppm to several ppm as discussed below. We used model output that covered a minimum period from September 2014 through December 2015, though a few models (CarbonTracker, TM5-4DVar) extended into March 2016. To compare against the models, for simplicity we computed only true XCO2 values from the a posteriori CO2 concentrations, rather than averaging-kernel-corrected values. Previous authors have shown that this effect is typically small, on the order of a few tenths of a ppm (Wunch et al., 2011a; Inoue et al., 2013; Lindqvist et al., 2015).
it was assumed that the Southern Hemisphere (25–55°S) could be taken to be meridionally uniform in $X_{\text{CO}_2}$ at any given time, with a latitudinal gradient of $-1$ ppm from 25 to 55°S, and the change in mean $X_{\text{CO}_2}$ over time could be prescribed with a linear secular trend (taken to be 1.9 ppm yr$^{-1}$). This served reasonably well for the GOSAT retrievals at that time, which exhibited rather large errors. However, the SHA has the primary shortcoming that meridional anomalies can sometimes exceed 0.5–1.0 ppm and are typically larger over land versus ocean. We find that substituting the model median instead of the zonally corrected mean used in Wunch et al. (2011a) results in error variances of the approximation 3–4 times lower, when comparing against any particular model as truth. Therefore, in order to maintain a connection to the truth metric of Wunch et al. (2011a), in this work we adopt the modified SHA called “Model_SHA”. This is simply the model median, discussed above but only used in the Southern Hemisphere below a latitude of 20°S.

4.2 Quality filtering

The construction of the operational OCO-2 filtering and bias correction for version 8 only, and make notes where version 7 differed significantly. The filtering procedure yields two quantities. The first is a binary flag denoted the $X_{\text{CO}_2}$ quality flag, which requires that a series of parameter-based tests are all passed. The second is a graded set of “warn levels”, which assigns each retrieval an integer value from 0 (most likely to yield accurate $X_{\text{CO}_2}$) to 5 (least likely to yield accurate $X_{\text{CO}_2}$). A genetic algorithm (Mandrake et al., 2013) finds combinations of variables that are best at predicting variance reduction in $X_{\text{CO}_2}$ over both small areas ($\lesssim 10 \times 80 \text{ km}^2$) and in the Southern Hemisphere (south of 25°S). In this document, we focus only on the quality flag filtering.

Filtering is accomplished by first identifying variables that cause the largest $\delta X_{\text{CO}_2}$, where $\delta X_{\text{CO}_2}$ is defined as the retrieved–true $X_{\text{CO}_2}$, the latter of which is evaluated for a given truth proxy. This was done sequentially, by identifying the single variable responsible for the largest fraction of the variance in $\delta X_{\text{CO}_2}$. We then created a simple threshold-based filter for this variable. After application of the filter, this process was repeated multiple times until it appeared that the majority of problematic data were removed. Because bias correction affected this procedure, a preliminary filter set was first created, after which a preliminary bias correction was developed. The preliminary bias correction was then applied, $\delta X_{\text{CO}_2}$ was updated accordingly, and the filters were re-derived using this bias-corrected $\delta X_{\text{CO}_2}$. Generally this had only a minor effect on the filters, and often served to increase the fraction of data passed through filtering.

Selection of thresholds for particular filters was somewhat subjective: generally bias was regarded as more problematic than scatter, but both were considered. Variables were typically selected as filters if they were correlated with bias greater than about 0.5 ppm, or significant scatter (greater than about 2 ppm). The filtering variables and thresholds were derived separately for land (combined nadir and glint) and ocean soundings. The final values of the filtering thresholds for the $X_{\text{CO}_2}$ quality flag are given in Appendix A. Filter-

Figure 9 shows maximum difference from the model median for both the Northern Hemisphere winter (December, January, February; DJF) and summer (June, July, August; JJA). Most soundings passed our “model-agreement” requirement over ocean at all times and over land in DJF, where the bulk of the land biosphere is quiet and hence $X_{\text{CO}_2}$ is more robustly modeled. In JJA, however, a substantial fraction of land soundings fail this test, in particular over Northern Hemisphere regions such as Asia. Tests showed that our results were not strongly sensitive to the agreement threshold chosen.

Finally, Wunch et al. (2011a) used a truth proxy called the “Southern Hemisphere Approximation” (SHA) in which it was assumed that the Southern Hemisphere (25–55°S) is more robustly modeled. In JJA, however, a substantial fraction of land soundings fail this test, in particular over Northern Hemisphere regions such as Asia. Tests showed that our results were not strongly sensitive to the agreement threshold chosen.

Figure 9 shows maximum difference from the model median for both the Northern Hemisphere winter (December, January, February; DJF) and summer (June, July, August; JJA). Most soundings passed our “model-agreement” requirement over ocean at all times and over land in DJF, where the bulk of the land biosphere is quiet and hence $X_{\text{CO}_2}$ is more robustly modeled. In JJA, however, a substantial fraction of land soundings fail this test, in particular over Northern Hemisphere regions such as Asia. Tests showed that our results were not strongly sensitive to the agreement threshold chosen.

Finally, Wunch et al. (2011a) used a truth proxy called the “Southern Hemisphere Approximation” (SHA) in which it was assumed that the Southern Hemisphere (25–55°S) was developed. The preliminary bias correction was then applied, $\delta X_{\text{CO}_2}$ was updated accordingly, and the filters were re-derived using this bias-corrected $\delta X_{\text{CO}_2}$. Generally this had only a minor effect on the filters, and often served to increase the fraction of data passed through filtering.
Figure 10. $\delta X_{\text{CO}_2}$ versus select filtering variables for land (nadir + glint) data, using TCCON as a truth proxy. Shown are the mean bias in each parameter bin for both raw (black circles) and bias-corrected (light blue circles) $X_{\text{CO}_2}$, as well as the standard deviation of the bias-corrected $\delta X_{\text{CO}_2}$ (dark blue diamonds). The histogram of each parameter is shown in gray. The vertical black dashed lines denote filtering thresholds for the $X_{\text{CO}_2}$ quality flag, while the thin red solid lines show filtering thresholds for the warn levels. The quality flag filters are applied cumulatively from left to right and top to bottom. The fraction passing at each step, as well as the rms error of the bias-corrected $X_{\text{CO}_2}$, are shown in the upper right corner of each panel. Please see Table A1 for a complete definition of all of the filter variables.
ing variables selected and their thresholds were the same or similar, regardless of the particular truth proxy used.

An example of this sequential filtering approach is shown in Fig. 10, which shows the X\textsubscript{CO}\textsubscript{2} error versus filtering parameters for nadir and glint land soundings, using TCCON as the truth proxy. Overall, the results were found to be robust for all our truth proxies. Just a few variables do the bulk of the filtering. For both land and ocean, the CO\textsubscript{2} and H\textsubscript{2}O ratios computed by the IMAP-DOAS preprocessor account for a significant fraction of the total filtering. These variables represent the ratio of the total column CO\textsubscript{2} (H\textsubscript{2}O) as derived from the weak CO\textsubscript{2} band to that from the strong CO\textsubscript{2} band. As discussed at length in Taylor et al. (2016), values of these gas ratios that deviate significantly from unity indicate the presence of significant atmospheric scattering. As shown in Fig. 10, ratios significantly away from the median values can result in both large scatter and large biases. Another robust finding is that biases are associated with large absolute values of the retrieved–prior surface pressure (d\textsubscript{P}) for both the Level-2 and ABP preprocessor retrievals. All of these variables (CO\textsubscript{2} and H\textsubscript{2}O ratios and surface pressure) are most likely diagnosing scattering-induced errors due to improperly modeled clouds and aerosols.

Two variables associated with small-scale variability are also associated with increased scatter: the standard deviation of the surface altitude within OCO-2’s field-of-view, and another parameter called “Max_Declocking”, which is determined independently for each of the three OCO-2 bands. The latter is related to a slope in the observed radiance within an individual sounding’s field-of-view, and is determined from OCO-2’s color slices as discussed in Crisp et al. (2017). The scatter associated with surface elevation appears to be related to an instrument-to-spacecraft offset specification error, which results in small (several hundred meters) pointing errors, which is improved in the next data version (version 9) and allows for relaxation of this filter (Kiel et al., 2018).

Another interesting variable that can result in both bias and scatter is the tropospheric lapse rate of the retrieved CO\textsubscript{2} profile, called co2_grad_del. It is determined by the difference in retrieved CO\textsubscript{2} between the surface and the retrieval pressure level at 0.7 times the surface pressure, minus this same

Figure 11. Same as Fig. 10 but for ocean glint measurements, where the truth proxy is the multi-model median.
quantity for the prior:

\[
\text{co}_2_{\text{grad.del}} = [c(1) - c_{0.7})] - [c_{\text{ap}}(1) - c_{\text{ap}}(0.7)],
\]

where \(c(x)\) and \(c_{\text{ap}}(x)\) respectively denote the retrieved and a priori CO$_2$ dry air mole fraction at relative pressure \(x\).

The reason why this variable is strongly associated with bias and scatter is still being investigated; it may be due to CO$_2$ spectroscopy errors, or some other factor. There is also a filter associated with dark surfaces; scenes with a strong CO$_2$ band albedo less than 0.05 consistently exhibit a bias in retrieved X$_{\text{CO}_2}$, and are thus excluded. Note this will tend to flag most snow- and ice-covered surfaces (such as over Greenland and Antarctica), which are highly absorbing at wavelengths longer than about 2 \(\mu\)m. It also tends to exclude dark forests such as in the Amazon. There are also filters associated with the retrieved slope of the strong CO$_2$ band albedo, the fit quality in the CO$_2$ bands, and a number of retrieved aerosol variables. Of particular note is the total retrieved optical depth associated with our larger aerosol types: dust, water cloud, and sea salt (DWS). High values of DWS are associated with negative biases in X$_{\text{CO}_2}$ over land, and it is used as both a filter and bias-correction variable. Although ice is also a large type, it is confined to the upper atmosphere in our retrieval and has its own dedicated filter.

Similar variables are used for filtering over water surfaces (Fig. 11), though note that almost no aerosol-related variables are used. This may be because water surfaces have relatively uniform optical properties, such that the retrieved variables indirectly associated with cloud and aerosol scattering, such as the CO$_2$ and H$_2$O ratios and the slope of the strong CO$_2$ band albedo, are more effective than over land, obviating the need for additional aerosol-related filtering. It may also be because most downward-propagating, forward scattered light is absorbed by the ocean surface, so the pathways for aerosol contamination are significantly less than over land, as noted by Butz et al. (2013).

As seen in the upper left panel of Fig. 11, the dominant filtering variable for water-glint soundings is the slope of the strong CO$_2$ band albedo. This is the slope of the retrieved Lambertian albedo in that band, which is generally small and is added onto the reflectivity coming from the primary Cox and Munk surface, which is a function of wind speed only. Negative slopes are strongly associated with X$_{\text{CO}_2}$ bias, which appears indicative of either cloud ice or sea salt aerosol scattering, both of which yield a negative slope in these units\(^3\). Large positive values of this slope are likely associated with contamination by sulfate aerosol or other small particle types. The sensitivity of this variable to cloud and aerosol scattering has been confirmed with simulations. About 10\% of water-glint soundings are flagged by this filter.

\(^3\)The units of the albedo slope are in per unit wavenumber, increasing with wavenumber.

After filtration, about 31\% of land soundings and 55\% of water soundings pass the X$_{\text{CO}_2}$ quality flag\(^4\). As depicted in Fig. 12, the pass rates are not uniformly distributed around the globe. Over land, very bright and dark surfaces are preferentially filtered out, as well as locations with many low clouds such as the Amazon, which are sometimes missed by our prefilters (Taylor et al., 2016). Nearly all soundings over ice surfaces are filtered out, because the albedo of ice is very low at 2 \(\mu\)m, hence yielding low signal-to-noise. The higher quality of water soundings is likely due to higher uniformity of water surfaces in glint mode, higher and more uniform SNR in all three bands, and fewer surface-atmosphere scattering mechanisms. Over both land and water, soundings at higher solar zenith angles are also removed at a higher rate by our quality flag. This is most likely due to the relatively large effects of scattering on our retrievals for these geometries, specifically when the fraction of the light received at the detector from atmospheric scattering is a larger fraction of the total. Over water, approximately 70\% of soundings pass at lower viewing angles, while nearly all soundings fail at high viewing angles.

### 4.3 Bias correction

After filtering, systematic biases remain in retrieved X$_{\text{CO}_2}$, which must be corrected in order to minimize errors. The OCO-2 bias correction contains three main pieces: parametric, footprint-level, and global biases. Parametric biases are functionally related to a given parameter associated with a given sounding. Examples of this could be surface pressure, albedo quantities, or retrieved aerosol quantities. Footprint-level biases are corrected to ensure that each of OCO-2’s eight sensors, or “footprints”, yield the same X$_{\text{CO}_2}$ value when observing similar scenes. This is not always the case due to small calibration errors in the eight individual footprints. The final step of the bias correction removes any global mean bias that may remain. The overall bias-correction equation is then written as

\[
X_{\text{CO}_2,\text{bc}} = \frac{X_{\text{CO}_2,\text{raw}} - C_{\text{P}}(\text{mode}) - C_{\text{F}}(j)}{C_{\text{0}}(\text{mode})},
\]

where \(C_{\text{P}}\) is the mode-dependent parametric bias, \(C_{\text{F}}\) is the footprint-dependent bias for footprints \(j = 1...8\), and \(C_{\text{0}}\) represents a mode-dependent global scaling factor. The following subsections discuss each of these corrections in detail.

#### 4.3.1 Bias correction: parametric biases

The most complex but most important of the three aspects of the bias correction is inferring biases dependent upon different...
Table 9. Land bias-correction parameters, their coefficients, and percentage of the variance explained, for different truth proxies and observing modes.

<table>
<thead>
<tr>
<th>Truth proxy</th>
<th>Mode</th>
<th>$N$</th>
<th>$dP$</th>
<th>$\text{co}_2\text{ grad}_\text{del}$</th>
<th>DWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCCON</td>
<td>Nadir</td>
<td>92k</td>
<td>-0.38</td>
<td>-0.028</td>
<td>-8.8</td>
</tr>
<tr>
<td></td>
<td>Glint</td>
<td>68k</td>
<td>-0.38</td>
<td>-0.026</td>
<td>-6.4</td>
</tr>
<tr>
<td></td>
<td>Target</td>
<td>245k</td>
<td>-0.29</td>
<td>-0.023</td>
<td>-7.8</td>
</tr>
<tr>
<td>SAA</td>
<td>Nadir</td>
<td>242k</td>
<td>-0.37</td>
<td>-0.031</td>
<td>-9.5</td>
</tr>
<tr>
<td></td>
<td>Glint</td>
<td>251k</td>
<td>-0.36</td>
<td>-0.030</td>
<td>-9.6</td>
</tr>
<tr>
<td>Models</td>
<td>Nadir</td>
<td>281k</td>
<td>-0.34</td>
<td>-0.029</td>
<td>-9.3</td>
</tr>
<tr>
<td></td>
<td>Glint</td>
<td>300k</td>
<td>-0.34</td>
<td>-0.027</td>
<td>-9.0</td>
</tr>
<tr>
<td>Models_SHA</td>
<td>Nadir</td>
<td>87k</td>
<td>-0.35</td>
<td>-0.032</td>
<td>-8.8</td>
</tr>
<tr>
<td></td>
<td>Glint</td>
<td>87k</td>
<td>-0.36</td>
<td>-0.029</td>
<td>-10.3</td>
</tr>
<tr>
<td>B8 adopted</td>
<td>All</td>
<td></td>
<td>-0.36 ± 0.028</td>
<td>-0.029 ± 0.0027</td>
<td>-8.5 ± 1.1</td>
</tr>
<tr>
<td>Reference value</td>
<td></td>
<td>0.0</td>
<td>15.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>B7 adopted</td>
<td>All</td>
<td></td>
<td>-0.30</td>
<td>-0.028</td>
<td>-7 to -11*</td>
</tr>
</tbody>
</table>

* B7 used \text{ln}(DWS) rather than DWS in its bias correction.

ent retrieval parameters. Most near-infrared X$_\text{CO}_2$ retrievals have required this, for both GOSAT (Wunch et al., 2011a; Cogan et al., 2012; Guerlet et al., 2013) and OCO-2 (Reuter et al., 2017; Wu et al., 2018) measurements. A nontrivial fraction of the bias comes from the retrieval algorithm itself, as shown in the simulation-based study of O’Dell et al. (2012), in which the instrument model and spectroscopy were perfect yet biases still emerged in the retrievals. Previous versions of the ACOS algorithm applied to GOSAT have shown dependencies on the surface albedo in the CO$_2$ bands, $dP$ (retrieved minus prior surface pressure), $\text{co}_2\text{ grad}\_\text{del}$, the retrieved ice cloud height, and other variables. The parametric bias correction has the form of a multiple linear regression, following Wunch et al. (2011a):

$$C_P = \sum_i c_i (p_i - p_{i,\text{ref}}),$$

Figure 12. Fraction of L2FP processed soundings passing a quality filter (a, b) and total number of good-quality soundings per month per $6^\circ \times 6^\circ$ boxes (c, d), for the Northern Hemisphere winter (DJF) and summer (JJA). The number plots have a logarithmic color scale, and grid boxes with no data are shown in gray.
where \( c_i \) are the regression coefficients, \( p_i \) are the selected parameters, and \( p_{i,\text{ref}} \) are convenient reference values. We note that the reference values are nontrivial in that they interact with the last term in the bias correction, the global scaling factor. Ideally, the reference value will be the value of the parameter at which that parameter does not bias the retrieved \( X_{\text{CO}_2} \), but this is impossible to disentangle from the global scaling factor. Wunch et al. (2011a) took the parameter reference values to be the estimated global mean value of each parameter. Here we do not require this, though for some variables, the estimated global mean is used.

In order to identify the variables of interest, we used all four truth proxies and identified combinations of one, two, three, and four variables that removed the most variance, for each observing mode and over both land and water. Variables that remove less than 5% of the variance are not included, as overfitting is a potential danger here. Typically the different truth proxies agree on the most important variables but disagree on the variables that explain just a few percent of the variance or less. As shown in Table 9, it was found that three fit parameters were required over land, and that their values did not strongly depend on observing mode. These variables were \( dP \) (retrieved minus prior surface pressure), \( \text{co2\_grad\_del} \), and finally DWS, which as stated previously is the combined retrieved optical depth of dust, water cloud, and sea salt aerosol. DWS represents the retrieved optical depth of large particles in the lower-to-middle troposphere in the retrieval. While ice cloud particles are large, they are placed in the upper troposphere in the retrieval, and all other aerosol types in the retrieval are much smaller.

In Table 9, the coefficients of each parameter inferred from each truth proxy and observing mode typically agree to within 10%–20%. The final result represents a combination of the average of these individual values, but was also driven by consensus amongst the scientists involved. Table 9 also gives the approximate uncertainty on each parameter, which is estimated as the standard deviation of the estimates from the different truth proxies and viewing modes. Also shown is the B7 bias correction, which was very similar, though it used \( \ln(\text{DWS}) \) instead of DWS. Figure 13 shows the result of the multiple regression for these three variables against the model-based truth proxy, for all nadir and glint soundings.

### Table 10. Ocean glint bias-correction parameters for different truth proxies.

<table>
<thead>
<tr>
<th>Truth proxy</th>
<th>( N )</th>
<th>( dP ) max (( \text{co2_grad_del}, -6 ))</th>
<th>Coefficient (% variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCCON</td>
<td>72k</td>
<td>−0.24 (30%)</td>
<td>0.078 (5%)</td>
</tr>
<tr>
<td>SAA</td>
<td>385k</td>
<td>−0.22 (60%)</td>
<td>0.092 (10%)</td>
</tr>
<tr>
<td>Models</td>
<td>610k</td>
<td>−0.25 (35%)</td>
<td>0.105 (8%)</td>
</tr>
<tr>
<td>Models_SHA</td>
<td>157k</td>
<td>−0.12 (16%)</td>
<td>0.134 (21%)</td>
</tr>
<tr>
<td>B8 adopted</td>
<td></td>
<td>−0.23 ± 0.06</td>
<td>0.090 ± 0.024</td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td>0.0</td>
<td>−6.0</td>
</tr>
<tr>
<td>B7 Adopted</td>
<td></td>
<td>−0.08</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Figure 14. Same as Fig. 13 but for ocean glint measurements again using the model mean truth proxy.
over land. In general, $dP$ explains about 30% of the variance over land, $\text{co}_2\text{\_grad\_del}$ about 20%, and DWS roughly 5%–10%.

A similar procedure was followed for glint soundings over water. For this observing mode, only $dP$ and $\text{co}_2\text{\_grad\_del}$ were needed; all other variables explained only a very small fraction of the variance, and were not consistent among truth proxies. The fit to these two variables for the model-based truth proxy is shown in Fig. 14. A true linear regression will not work for $\text{co}_2\text{\_grad\_del}$. Instead of fitting a nonlinear form, we instead fit against the variable $\max(\text{co}_2\text{\_grad\_del}, -6)$. This gives essentially the fit as shown in the figure, where the best-fit line to the bias increases with increasing $\text{co}_2\text{\_grad\_del}$ until a value of $-6$, above which the fitted bias is held constant. As shown in Table 10, the different truth proxies again yielded similar results to within about 20%, with the exception of “Models_SHA”, which was an outlier. The reason for this is unknown, though we speculate that the actual parameter variability in the Southern Hemisphere is too small to obtain sensible slopes. Therefore, this truth proxy was excluded in the calculation of the final coefficients for glint soundings over water. Also, it is worth noting that the $dP$ coefficient in B7 was roughly 3 times smaller than the value of $-0.23$ adopted for B8. This was driven by inconsistencies in the B7 truth proxy data sets and a very small training data set, which yielded an unrealistically small value. Later analyses showed that B7 probably should have used a higher value, more in line with the B8 result.

4.3.2 Bias correction: footprint biases

After fitting for the parametric biases, the dependence of the $X_{CO_2}$ bias on footprint was evaluated. As with the parametric biases, the footprint biases were evaluated using the suite of truth proxies and for each observing mode separately (land nadir, land glint, and ocean glint). For all frames that contained all eight footprints, the difference of each footprint from the mean of its frame was calculated, with the result being the estimated set of footprint biases for each truth proxy. Note that this was done after application of the parametric bias correction. The resulting biases were quite consistent across truth proxies; thus, the results across truth proxies were averaged. As shown in Fig. 15, there was virtually no dependence on viewing mode, and no obvious land–water differences. It appears that the footprint-level biases are truly instrument-related, and thus do not seem to strongly depend on other factors. Therefore, a single set of footprint-level biases was used. The adopted footprint biases for footprints 1–8 were $(-0.36, -0.15, -0.16, -0.14, 0.02, 0.33, 0.13, 0.34)$ ppm, with an uncertainty of roughly ±0.03 ppm (1σ) on each. The reason for the general increase in this bias with increasing footprint is not known. Finally, while the fraction of the variance explained by footprint-level biases (∼2%) is small compared to that explained by the parametric biases, they are straightforward to evaluate and could have an effect on local-scale analyses, and are therefore removed.

4.3.3 Bias correction: global scaling

Despite the corrections described above, there is still an overall $X_{CO_2}$ bias on the order of 1–2 ppm relative to the true atmosphere. As shown in Eq. (6), the denominator term $C_0$ represents a global bias correction, and is parametrized as a function of viewing mode. As for the quality filtering and parametric bias correction, we found that the land scaling factor is roughly the same for nadir, glint, and target modes, such that a single scaling can be used for all land soundings. Ocean glint required a slightly different global scaling factor.
Figure 17. Annual mean biases and changes from the prior for all quality-flag “good” soundings from September 2014 through September 2017. (a–c) Parametric biases for $dP$, co2_grad_del, and DWS. (d) Sum of the parametric bias terms — note the color scale change. (e) Departure of the raw retrieved $X_{\text{CO}_2}$, including the global scaling correction, from the prior, $X_{\text{CO}_2,\text{raw}}/C_0 - X_{\text{CO}_2,\text{ap}}$. (f) Departure of the bias-corrected $X_{\text{CO}_2}$ from the prior, $X_{\text{CO}_2,\text{bc}} - X_{\text{CO}_2,\text{ap}}$.

Table 11. Global scaling factors (divisors) for OCO-2 version 8.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Method</th>
<th>$N$</th>
<th>Divisor $C_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land target</td>
<td>TCCON</td>
<td>138</td>
<td>0.9958</td>
</tr>
<tr>
<td>Land nadir</td>
<td>TCCON</td>
<td>313</td>
<td>0.9962</td>
</tr>
<tr>
<td>Land glint</td>
<td>TCCON</td>
<td>277</td>
<td>0.9960</td>
</tr>
<tr>
<td>Ocean glint</td>
<td>TCCON</td>
<td>236</td>
<td>0.9956</td>
</tr>
<tr>
<td>Ocean glint</td>
<td>Coastlines*</td>
<td>536</td>
<td>0.9955</td>
</tr>
<tr>
<td>Ocean glint</td>
<td>Model bootstrap*</td>
<td>500,000</td>
<td>0.9954</td>
</tr>
<tr>
<td>Land (all)</td>
<td>Adopted</td>
<td></td>
<td>0.9958</td>
</tr>
<tr>
<td>Ocean glint</td>
<td>Adopted</td>
<td></td>
<td>0.9955</td>
</tr>
</tbody>
</table>

* Using a value of 0.9958 for land soundings.

The B8 global scaling was determined primarily from several hundred direct OCO-2 overpasses of TCCON stations. We followed the geometric colocation method of Wunch et al. (2017), with the exception that sites in the Southern Hemisphere required the same latitude and longitude colocation thresholds as sites in the Northern Hemisphere. The TCCON value for a given overpass was determined as the mean of the observations within $\pm 2$ h of the OCO-2 overpass. At least 3 valid TCCON and 20 quality-flag “good” OCO-2 soundings were required per overpass. For each viewing mode, the slope $m$ of the best-fit line passing through the origin was calculated with the method of York et al. (2004):

$$X_{\text{CO}_2,\text{tccon}} = \frac{X_{\text{CO}_2,\text{OCO2}}}{C_0},$$
which provides the best-fit value of \( C_0 \). The results for each viewing mode are shown in Table 11 under Method “TC-
CON”. It can be seen that the three sets of land overpasses yielded a scaling factor \( C_0 \) consistent with each other to
within their respective errors. The errors are representative
differences of the fitted slope due to both retrieval errors
as well as linear fit differences (for instance, using a least-
squares fit versus a least-absolute-deviation fit), and were
typically \( \pm 0.0003 \) (\( \sim 0.12 \) ppm). Because target-mode ob-
servations were better colocated, the target observation value
of 0.9958 was adopted for all land observations.

For ocean glint, the global scaling factor was estimated
with three methods: direct overpasses of TCCON stations
and two independent “bootstrap” methods, using coastline
crossings and models. Direct overpasses of TCCON stations
yielded a value of 0.9956, slightly lower than the adopted
land value but still consistent to within errors. For the coast-
line bootstrap method, a set of several hundred small ar-

areas centered on coastlines were identified in which it was
possible to ratio the mean value of \( X_{\text{CO}_2} \) over water to the
mean value of \( X_{\text{CO}_2} \) over land. This yielded a mean wa-
ter/land ratio of 0.9997 \( \pm 0.0001 \), meaning that ocean values
were slightly lower than land values. Multiplying this wa-
ter/land ratio by the land scaling value yielded an ocean scal-
ing value of 0.9955. Finally, the multi-model median truth
proxy was used, wherein the slope of OCO-2 versus models
was calculated, sampled at good-quality OCO-2 sounding lo-
cations. It was found that land required a scaling of 0.9950
and ocean a scaling of 0.9946, suggesting a water/land ra-
tio for OCO-2 of roughly 0.9996 \( \pm 0.0001 \), similar to that of
the coastline crossing value, and thus an ocean scaling value
of 0.9954. Note that the absolute comparison of mod-
els relative to OCO-2 was not used here, as spin-up issues
and averaging kernel corrections (ignored in this analysis)
could yield a spurious global offset between the models and
OCO-2. Therefore, only relative land–ocean differences were
used to infer \( C_0 \) in this “Model Bootstrap” method. As seen
in Table 11, the “bootstrap” methods were remarkably con-
sistent with each other and with the direct TCCON overpass
method, all suggesting an ocean scaling of roughly 0.9955.

All of the methods described above for determining the
global scaling of \( X_{\text{CO}_2} \) show remarkable consistency with
each other, giving confidence that the overall scale of OCO-2
data is known to within a few tenths of a ppm, and the land–
ocean difference is known to better than 0.2 ppm. One caveat
is that the direct TCCON overpass comparisons did not ac-
count for the averaging kernel correction in this analysis.
After the release of OCO-2 \( X_{\text{CO}_2} \) B8, subsequent analysis
showed that this correction typically lowered the value of
TCCON by \( \sim 0.1 \) ppm on average relative to OCO-2. There-
fore, there is additional possible uncertainty in the overall
magnitude of OCO-2 data by this amount.

4.3.4 Bias-correction evaluation

With all three sources of bias now characterized, the overall
role of the bias correction in the retrieval can be evaluated.
Figure 17 is an attempt to show this at the mean annual scale.
Panels (a)–(c) show the biases due to the three different bias
terms: \( dP \), \( \text{co2\_grad\_del} \), and DWS. Of these three, \( dP \) is
the strongest. This is apparent in panel (d), which shows the
sum of the parametric biases (note the scale change between
columns). Panel (e) shows the mean change from the prior
in the raw retrieval, after correcting for the global bias term
only (which is likely due to spectroscopic and instrumental
bias). The change from the prior to the bias-corrected \( X_{\text{CO}_2}
\) is shown in panel (f), which combines these two terms.

The parametric bias, while smaller than the difference be-
tween the raw retrieved and prior \( X_{\text{CO}_2} \), is of a comparable
order of magnitude. Ideally, the parametric bias would be
much smaller. As noted above, the parametric bias is domi-
nated by the \( dP \) term. This term reflects the fact that the error
in \( X_{\text{CO}_2} \) is reduced using the prior rather than retrieved sur-
face pressure. That is, the ACOS-retrieved slant column of air
is dominated by systematic errors that are not reflected in the
estimated \( \text{CO}_2 \) slant column. We have identified two easily
rectifiable sources of such error, both related to how the B8
surface pressure prior fields were derived. One of these was
caused by a slight error in the knowledge of the pointing of
the OCO-2 instrument that induces spurious small-scale er-
ror in the estimated prior surface pressure in regions of high
topographic variability. The second was caused by sampling
the GEOS5 FP-IT surface pressure at the wrong time of day,
by up to several hours. Both of these effects have been cor-
corrected for in the version 9 OCO-2 \( X_{\text{CO}_2} \) data, described in
detail in Kiel et al. (2018).

Even after correcting the prior, the OCO-2 surface pres-
ure retrievals still contain significant regional biases. These
biases have a largely zonal structure, with a negative bias of
up to \( \sim -5 \) hPa at the highest latitudes and a positive bias of
several \( h\)Pa in the tropics, with an overall mean bias of
roughly \( +2 \) hPa. The standard deviation of the surface pres-
ure bias (relative to the GEOS5 FP-IT prior) for quality-
filtered soundings is roughly 2.8 hPa. By binning nearby
soundings, it is found that more than 2.4 hPa of this variation
comes from systematic error scales greater than 1\( ^\circ \). One pos-
sible hypothesis is that this systematic error is due to \( \text{O}_2 \)
sorption cross-sectional errors and how they manifest them-
selves in the retrieval. For instance, an incorrect parameter-
ization of the temperature dependence of absorption could
yield errors similar to those observed. Future updates to our
spectroscopy may reduce this bias, and are an active area of
research within the ACOS team.

Currently, however, these multi-hPa systematic errors in
the retrieved surface pressure are likely larger than errors in
the GEOS5 FP-IT prior, which are believed to be on the or-
der of 1–2 hPa. For instance, three reanalysis surface pres-
ure sets were recently compared, and it was found that the
Figure 18. TCCON validation for OCO-2 X\textsubscript{CO\textsubscript{2}}, versions B7 (a, c, e) and B8 (b, d, f), for land nadir (a, b), land glint (c, d), and ocean glint (e, f) observations. Each symbol represents the overpass–mean comparison for one site overpass, with the total number of overpasses per site given in parentheses. Thus each symbol represents tens to hundreds of OCO-2 observations co-averaged. Quality-filtered and bias-corrected X\textsubscript{CO\textsubscript{2}} is shown for OCO-2, along with the averaging-kernel correction. The solid line denotes the one-to-one line, and the dashed line is the line of best fit. The collocation strategy is described in the text. Shown in the upper left of each panel are the total number of overpasses (N), the mean, standard deviation (σ), and rms of the OCO-2 minus TCCON differences, the slope of best fit, and the $R^2$ of the two data sets.
differences were as small as 0.5 hPa in the tropical oceans and became larger at higher latitudes, such as in the Southern Ocean, where rms differences were on the order of 2 hPa (Jucks et al., 2015, Sect. 4.2.1). This is the basic reason behind the artificially high (4 hPa) 1σ uncertainty on the prior surface pressure currently used in the OCO-2 retrieval. If future versions of the retrieval (including spectroscopy updates) yield a more unbiased retrieval of surface pressure, it may reduce the role of dP in the bias correction and lead to a more accurate retrieval of XCO₂.

The figure reveals several additional interesting features. The co2_grad_del bias is of a smaller magnitude than the dP bias, and has the strongest effect in tropical forests and a more diffuse effect elsewhere, including the tropical and southern oceans. The DWS bias has the largest effects over northern Africa, temperate Eurasia, and Australia, where large dust particles are prevalent. The total parametric bias pattern can sometimes cancel the departure from the prior, such as over Australia, indicating that the original departure was likely spurious. However, the overall large positive departure in the northern middle-to-high latitudes is not strongly affected by the bias correction, and thus appears to be a feature well captured by the native retrieval itself. This latter feature seems to imply that the interhemispheric gradient is too weak in the TCCON prior. The GGG2014 TCCON CO₂ prior relies exclusively on the age of the air to create the interhemispheric gradient of CO₂. The next release of the TCCON software includes an age-independent term in the CO₂ priors representing the source–sink imbalance, which will roughly double the interhemispheric gradient (Geoff Toon, personal communication, 2018).

5 Brief evaluation of OCO-2 XCO₂

Thus far, we have completely described the mechanics of the current ACOS retrieval and methodology, but have yet to evaluate the actual algorithm performance. As B7 XCO₂ was validated in detail in Wunch et al. (2017), in this section we focus primarily on the differences (mainly improvements) between B7 and B8. Figure 18 shows the relationship between colocated TCCON and OCO-2 XCO₂ observations, for both B7 and B8 OCO-2 retrievals. Operational quality filtering and bias correction have been applied for each version. We used the same colocation strategy as described in Sect. 4.1.1, but with the additional requirement that fossil fuel emissions from the 1 km 2013 ODIAC database (Oda and Maksyutov, 2011), smoothed with a 5 km Gaussian smoother, be less than 300 g m⁻² month⁻¹ at the location of the OCO-2 soundings; this eliminated OCO-2 soundings in the vicinity of strong fossil fuel sources. All good-quality OCO-2 soundings within each overpass were averaged (thus, one symbol on each plot denotes one overpass). Only overpasses with at least 10 such OCO-2 soundings were included. The TCCON XCO₂ values are averages of all good-quality TCCON soundings at that site within ±2 h of the OCO-2 overpass.

Based on these figures, B8 appears generally superior to B7 in terms of agreement with TCCON. First, there are more good-quality data in B7, at least for glint observations. This is primarily due to improvements in the prescreeners. In terms of accuracy, the scatter and outliers of XCO₂ over both land and ocean are reduced in B7, especially over ocean, for which the outliers were driven by Southern Hemisphere observations (in particular the high southern-latitude sites, Lauder and Wollongong). The bias over land in B7 was significant, with a 0.3 ppm difference between land nadir and land glint. The overall apparent bias over land of 0.2–0.3 ppm is partly due to neglecting the averaging kernel effect when solving for the global divisors in the bias correction, and partly due to methodological differences in how we calculate the global bias in the first place. While the R² values are significantly higher than those reported in Wunch et al. (2017), this is due to the extended length of the data record used here. Coupled with the secular increase in CO₂, this leads to a larger dynamic range in XCO₂ and hence more signal and higher correlations than when using a shorter data record. Finally, we note that in B8 (and to some extent B7) there is generally a negative slope between TCCON and OCO-2. This appears to be due to a trend in OCO-2 XCO₂, relative to TCCON (not shown), in which OCO-2 appears to be losing XCO₂ at a rate of 0.1–0.2 ppm yr⁻¹, though this trend is not statistically significant. The apparent trend may be due to an OCO-2 calibration effect, and is currently under study.

The general improvement in B8 over B7 can also be seen in comparisons to models. Figure 19 shows the comparison of both versions to the six-model suite as discussed in Sect. 4.1.3. The large-scale differences between OCO-2 and model XCO₂ are generally reduced in the latest version. The positive bias with respect to models over the Southern Hemisphere mid-latitude oceans is nearly removed in B8, as are some negative biases over Australia and the Sahara. Positive biases in the Northern Hemisphere mid-to-high latitudes over land are also reduced. We also see that the latitudinal extent of the ocean glint data is greater in version 8, especially in DJF; again, this is primarily due to updates in the prescreeners. A large-scale negative difference between OCO-2 relative to the model median persists in the tropical oceans, particularly over the Pacific and Indian basins, and is currently under study.

There are clear deficiencies remaining in the B8 OCO-2 XCO₂ product. First, note again that Fig. 18 showed overpass average statistics. As each overpass had at least 10 OCO-2 soundings, and each sounding typically had < 1 ppm posterior uncertainty, if the errors were all independent, the overpass average errors would be less than 0.3 ppm. Even taking into account TCCON and colocation errors, it is likely that the overpass-level (small-area) errors are still much larger than this due to correlated errors in the OCO-2 retrievals. A more detailed validation of OCO-2 B8, along with how
its errors are correlated and how they integrate down, is the subject of a forthcoming paper (Kulawik et al., 2018). Other error sources also are still plainly visible in the OCO-2 data. For instance, the topography-related biases noted by Wunch et al. (2017) still exist in B8, but have recently been tracked to a misspecification of the satellite-to-ground pointing vector, and will be corrected in a forthcoming version 9. Also, there are still cloud-related errors in the OCO-2 data, for instance as noted by Massie et al. (2016). It is believed that these are often related to three-dimensional cloud effects, for instance as discussed in Merrelli et al. (2015).

6 Summary and outlook

As described in this paper, the OCO-2 retrieval algorithm for $X_{CO_2}$ has been evolving more or less continuously over the last decade, and is beginning to achieve accuracies that enable ground-breaking carbon cycle science. The latest version (B8) has the lowest biases and highest throughput of any version yet, though regionally coherent biases still remain at a significant level ($\sim 1$ ppm). A number of choices and assumptions go into the algorithm, and due to its complexity in terms of the number of variables it must accommodate, further research is needed to improve it. These assumptions relate to a number of factors, such as clouds and aerosols, surface pressure, spectroscopy, and potential instrument problems (e.g., scattered light, drifting calibration), to name a few. Advances in these areas will form the basis of future algorithm improvements. These issues affect not only OCO-2, but potentially all current and future sensors relying on this technology to measure $X_{CO_2}$, such as the TanSat (Yang et al., 2018), OCO-3 (Eldering et al., 2018), GOSAT-2 (Nakajima et al., 2012), MicroCarb (Pascal et al., 2017), and GeoCarb (Moore et al., 2018) missions.

To fully exploit space-based short-wave infrared measurements of reflected sunlight at high spectral resolution for studies of the carbon cycle, improvements must be made in several areas. First, satellite-based $X_{CO_2}$ retrieval biases must be further reduced. Next, in order to make and validate retrieval improvements, we must have validation data that are more accurate than the satellite retrievals in the first place. Satellite retrievals are currently pushing that limit; for instance, the stated station-level calibration accuracy for TCCON is currently 0.4 ppm (Wunch et al., 2010). Calculations of OCO-2 station-level differences with TCCON show that the mean absolute bias at all stations with at least five valid overpasses is 0.4 ppm, right at this level. Finally, we must have inversion and data assimilation systems that can make maximum use of these data. This requires, for example, minimizing transport model error, currently an active area of research (see, e.g., Basu et al., 2018; Schuh et al., 2018).

Significant progress has been made in the past decade in the retrieval of $X_{CO_2}$ from SCIAMACHY, GOSAT, and OCO-2 radiances. This paper shows that this progress conti...
continues. Given the numerous sensors planned for development and launch in the near term, the future of passive remote sensing of CO$_2$ remains bright.

*Code and data availability.* The OCO-2 L2 Full Physics Code is open source and available on Github https://github.com/nasa/RtRetrievalFramework (last access: 4 December 2018), and the User’s Guide, for it is available at http://nasa.github.io/RtRetrievalFrameworkDoc/ (last access: 4 December 2018). All of the OCO-2 data products are publicly available through the NASA Goddard Earth Science Data and Information Services Center (GES DISC) for distribution and archiving (http://disc.sci.gsfc.nasa.gov/OCO-2; last access: 4 December 2018). TCCON data were obtained from the TCCON data archive hosted by CaltechDATA, and are available from https://tccondata.org/.
Appendix A: XCO₂ quality flag definitions

The ACOS XCO₂ quality flags use a number of variables. Each has an upper and lower threshold. These variables and thresholds are given in Table A1.

Table A1. XCO₂ quality flag definition for version 8.

<table>
<thead>
<tr>
<th>variable</th>
<th>meaning</th>
<th>land filter</th>
<th>ocean filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>co2_ratio</td>
<td>Ratio of Band 2 to Band 3 CO₂ column from the IDP* algorithm</td>
<td>[1, 1.025]</td>
<td>[0.997, 1.018]</td>
</tr>
<tr>
<td>h2o_ratio</td>
<td>Ratio of Band 2 to Band 3 H₂O column from the IDP* algorithm</td>
<td>[0.88, 1.01]</td>
<td>[0.88, 1.01]</td>
</tr>
<tr>
<td>dP</td>
<td>Retrieved minus prior surface pressure (hPa)</td>
<td>[-6, 14]</td>
<td>[-4, 10]</td>
</tr>
<tr>
<td>dP_ABP</td>
<td>Retrieved minus prior surface pressure from ABP algorithm (hPa)</td>
<td>[-10, 13]</td>
<td>[-50, 10]</td>
</tr>
<tr>
<td>co2_grad_del</td>
<td>Retrieved vertical gradient in CO₂ (ppm) (see text for details)</td>
<td>[-80, 100]</td>
<td>[-20, 30]</td>
</tr>
<tr>
<td>albedo_slope_sco2</td>
<td>Retrieved slope of the Lambertian component of the surface albedo (cm⁻¹)</td>
<td>[-1.8 × 10⁻⁴, 10⁻³]</td>
<td>[5 × 10⁻⁶, 7 × 10⁻⁵]</td>
</tr>
<tr>
<td>rms_rel_wco2</td>
<td>Relative rms of Band 2 fit residuals (%)</td>
<td>&lt; 0.22</td>
<td>&lt; 0.30</td>
</tr>
<tr>
<td>Max_Declocking_wco2</td>
<td>See text for details</td>
<td>&lt; 0.75</td>
<td>&lt; 0.2</td>
</tr>
<tr>
<td>Max_Declocking_sco2</td>
<td>See text for details</td>
<td>&lt; 0.3</td>
<td></td>
</tr>
<tr>
<td>eof33rel</td>
<td>Retrieved relative amplitude of 3rd EOF of Band 3</td>
<td></td>
<td>[-0.3, 0.25]</td>
</tr>
<tr>
<td>wind speed</td>
<td>Retrieved surface wind speed (m s⁻¹)</td>
<td></td>
<td>[1.5, 25]</td>
</tr>
<tr>
<td>Altitude SD</td>
<td>Standard deviation of the surface elevation in the FOV (m)</td>
<td>&lt; 60</td>
<td></td>
</tr>
<tr>
<td>Band 3 albedo</td>
<td>Retrieved albedo SCO2 band</td>
<td>[0.05, 0.6]</td>
<td></td>
</tr>
<tr>
<td>S31</td>
<td>Continuum signal band 3 rel. to band 1</td>
<td>[0.03, 0.4]</td>
<td></td>
</tr>
<tr>
<td>τIC</td>
<td>Retrieved optical depth of ice cloud</td>
<td>&lt; 0.04</td>
<td>&lt; 0.035</td>
</tr>
<tr>
<td>τ</td>
<td>Retrieved total aerosol + cloud optical depth</td>
<td>&lt; 0.5</td>
<td></td>
</tr>
<tr>
<td>τDU + τWA + τSS</td>
<td>Retrieved optical depth of three large types</td>
<td>&lt; 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(dust, water cloud, and sea salt)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>τWA</td>
<td>Retrieved optical depth of water cloud</td>
<td>[0.0005, 0.1]</td>
<td></td>
</tr>
<tr>
<td>τSS</td>
<td>Retrieved optical depth of sea salt</td>
<td>&lt; 0.125</td>
<td></td>
</tr>
<tr>
<td>HIC</td>
<td>Retrieved relative pressure height of ice cloud</td>
<td>[-0.5, 0.45]</td>
<td></td>
</tr>
<tr>
<td>τSU + τOC</td>
<td>Retrieved sulfate + organic carbon optical depth</td>
<td>&lt; 0.3</td>
<td></td>
</tr>
<tr>
<td>τOC</td>
<td>Retrieved organic carbon optical depth</td>
<td>&lt; 0.08</td>
<td></td>
</tr>
<tr>
<td>τST</td>
<td>Retrieved stratospheric aerosol optical depth</td>
<td>&lt; 0.02</td>
<td></td>
</tr>
</tbody>
</table>

* IMAP-DOAS preprocessor.
Appendix B: Land surface BRDF parameterization

Since OCO-2 observations outside of target mode contain only single observation geometries, it is unrealistic to attempt to retrieve the BRDF shape on a per-observation basis. This fact, combined with the similar improvement seen in the different trial BRDFs, suggested that a single fixed BRDF shape could be used for all land footprints. The selected BRDF shape is a particular parameter set for the Rahman–Pinty–Verstraete (RPV) kernel (Rahman et al., 1993) that has been used as an initial guess for spectral multi-angle polarimetric aerosol remote sensing (Dubovik et al., 2011). This fixed BRDF shape is used within the physical forward model, and a similar set of two state variables is applied independently to each band to allow for the BRDF amplitude to have a linear spectral variation across the band. Since the RPV kernel assumes the surface is azimuthally symmetric, the absolute values of the azimuth angles are unimportant and the kernel function can be expressed in terms of the azimuth angle difference. Thus, the BRDF model used in the algorithm is a function of three angular variables, and can be expressed as

\[ \rho(\theta_i, \theta_t, \Delta \phi) = [w + s(v - v_0)] F(\theta_i, \theta_t, \Delta \phi; C), \]  

(B1)

where \( \theta_i \) and \( \theta_t \) are the zenith angles in the incident and reflected directions, \( \Delta \phi \) is the relative zenith angle, \( w \) and \( s \) are the BRDF “weight” and “weight slope”, \( F \) is the fixed BRDF shape (the RPV kernel), and \( C \) are the fixed BRDF shape parameter values. The weight \( w \) and weight slope \( s \) are the retrieved variables, one each per band, with the linear variation computed in wavenumber space relative to per-band reference values \( (v_0) \). The RPV kernel function (Rahman et al., 1993) has three parameters: \( \rho_0 \), the hotspot parameter; \( \Theta \), the asymmetry parameter; and \( k \), the anisotropy parameter. The functional form, as implemented in the VLIDORT routines used by the OCO-2 forward model, uses the exact form as given in Rahman et al. (1993).

The fixed values used for the parameters are \( \rho_0 = 0.05 \), \( \Theta = -0.1 \), and \( k = 0.75 \). Figure B1 shows the RPV kernel shape for these parameters.

While the retrieval works with the weight and slope variables, \( w \) and \( s \), due to the normalization of the RPV kernel these are not the most convenient or intuitive quantities. Therefore, we evaluate the actual BRDF kernel function for the primary observation geometry (incident direction from the Sun, reflected direction toward the sensor), and then scale by the retrieved weight and weight slope, to obtain the BRDF reflectance and reflectance slope. In the absence of atmospheric scattering, these values will be equal to the retrieved albedo and albedo slope using a Lambertian assumption. Therefore, in much of the high-quality retrieval output from the Version 8 algorithm, the reported BRDF reflectance and reflectance slope values will be similar to the albedo and albedo slope values reported in the Version 7 results. Observations with relatively higher amounts of aerosols or other complicating effects would be expected to have larger differences between the BRDF reflectance and Lambertian albedo.

![Figure B1. The shape of the RPV kernel for all reflection angles, for a 45° solar incidence angle. The incident direction is at 180°, meaning the solar irradiance is directed inwards from the left side.](image-url)
Author contributions. CO and AE were involved in nearly all aspects of this work. PW provided critical guidance on nearly all aspects of the work, throughout all stages. DC provided guidance on instrument effects on the algorithm and feedback on the algorithm. MG provided project leadership and algorithm guidance. BF designed and implemented many tests and performed substantial data analysis. CF was involved in earlier phases of algorithm development, in particular with the IMAP-DOAS retrieval, SIF retrieval, and EOF development. MK contributed to the bias correction and analyzed TCCON target-mode data. HL provided the updated cloud ice optical properties. LM created the testing data sets and warning levels, and contributed to filtering and bias correction. AM contributed substantial analysis and wrote the BRDF section. VN guided implementation of the BRDF surface model. RN performed data analysis, in particular with regards to aerosol effects. GO led the target-mode operational effort and organized the validation activities. VP led the spectroscopy effort and wrote the spectroscopy section. TT edited early drafts of this paper and led the pre-filtering activities and analysis. DW provided critical analysis of TCCON/OCO-2 comparisons and advice regarding TCCON data use. BD and FO made important contributions to the ABSSCO spectroscopy. AC operationalized the code and assisted in daily operations. JM and MS coded the L2 algorithm and processed testing data sets. DB, SB, FC, SC, LF, and PP provided model data used in the filtering, bias-correction, and validation efforts. MD, OG, DG, FH, LI, RK, IM, JN, HO, CP, CR, MKS, KS, RS, YT, OU, and VV provided TCCON data used in the filtering, bias-correction, and validation efforts.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Part of this work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA) for the Orbiting Carbon Observatory 2 project. Work at Colorado State University and the Geology and Planetary Sciences Department at the California Institute of Technology was supported by subcontracts from the OCO-2 project. CarbonTracker CT2015 and CT-NRT.v2016-1 results provided by NOAA ESRL, Boulder, Colorado, USA, from the website at https://carbontracker.noaa.gov (last access: 4 December 2018). Jena CarboScope inversion results provided by Christian Rödenbeck and the Max Planck Institute for Biogeochemistry, from http://www.bgc-jena.mpg.de/CarboScope/ (last access: 4 December 2018). Special thanks to Geoff Toon for illuminating discussions regarding the TCCON CO2 prior. We also thank Julia Marshall and one anonymous referee for their reviews of the manuscript, which greatly improved its quality.

Edited by: Helen Worden
Reviewed by: Julia Marshall and one anonymous referee

References


Gunson, M. and Eldering, A.: OCO-2 Level 2 bias-corrected XCO₂ and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing V8r Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/XR7ZWYSKP7D0, 2017.


Hase, F., Blumenstock, T., Dohe, S., Gross, J., and Kiel, M.: TCCON data from Karlsruhe (DE), Release GGG2014.R1, TCCON Data Archive, hosted by Caltech-


Morino, I., Yokozeki, N., Matsuzaki, T., and Horikawa, M.: TCCON


