# MEASURES 2017 Data Fusion

# Data User's Guide

August 29, 2019

Version 0.8

Jet Propulsion Laboratory

California Institute of Technology

# 1 Introduction

This Data User Guide describes the data products for the Making Earth Science Data Records for Use in Research Environments project (MEaSUREs '17: Records of Fused and Assimilated Satellite Carbon Dioxide Observations and Fluxes from Multiple Instruments). The focus of this user guide is to describe the format, resolution, and contents of the netcdf products. Readers who are interested in the theoretical basis for the product should check the Algorithm Theoretical Basis Document (ATBD), which is available at URL placeholder.

# 2 Mission and product overview

The Orbiting Carbon Observatory-2 (OCO-2) is NASA's first Earth remote sensing instrument dedicated to studying carbon dioxide's global distribution. It was launched on July 2, 2014, and it uses three high-resolution grating spectrometers to acquire observations of the atmosphere in three observation modes: nadir, glint, and target. In nadir mode, the instrument points to the local nadir to collect data directly below the spacecraft. Nadir mode does not provide adequate signal-to-noise ratio over the dark ocean surface, and thus over ocean OCO-2 uses glint mode. In that mode, OCO-2 points its mirrors at bright glint spots where the solar radiation is specularly reflected from the surface. Finally, in target mode the instruments locks its view onto specific surface locations (usually a ground-based TCCON station or observational tower) while flying overhead. OCO-2 has a repeat cycle of sixteen days and a sampling rate of about one million observations per day, making it a high-density

and high-resolution complement to GOSAT. The $CO_2$ concentrations in an atmospheric column are inferred from the observed spectra through optimal estimation (**?**). The outputs are available as 20-dimensional $CO_2$ profiles and column-averaged $CO_2$ concentrations. The latter is derived from the former using a pressure weighting function, which is a 20-dimensional vector of weights derived from local atmospheric conditions. A pressure weighting function is convolved with the 20-dimensional $CO_2$ vector in a linear combination to form the column-averaged estimate (**?**).

GOSAT is a polar-orbiting satellite dedicated to the observation of carbon dioxide and methane, both major greenhouse gases, from space. It flies at approximately 665 kilometers (km) altitude, and it completes an orbit every 100 minutes. The satellite returns to the same observation location every three days (**?**). NASA's Atmospheric $CO_2$ Observations from Space (ACOS) team uses the raw-radiance data from GOSAT to estimate the column-average $CO_2$ mole fraction in ppm, extending from the surface to the satellite over a base area corresponding to the instrument's footprint. In this article, we will be using GOSAT retrievals that are processed by the ACOS team to yield Level 2 column-average $CO_2$ data (see **?**, for more details), which were available to us through NASA's Goddard Earth Sciences Data and Information Services Center. Hereafter, we refer to these as ACOS data. Since the ACOS product is produced at the Jet Propulsion Laboratory by the same team behind the OCO-2 instruments, much of the retrieval characterization (e.g., priors, choice of pressure levels, forward models, etc.) are the same between the two products.

This MEaSUREs project produces three primary products: 1) OCO-2 Level 3 bias-corrected XCO2 and other select fields aggregated as daily files, 2) Multi-Instrument

Fused Level 3 bias-corrected XCO2 and other select fields aggregated as daily files, and 3) a 10-seconds averaged product. Products 1) and 2) are produced using a variant of local kriging (also known as optimal interpolation), and details can be found in Section 3 of the ATBD.

## 2.1 Data version and quality filter

For our fusion products, we use ACOS Version 7 data, which are produced by the Jet Propulsion Lab at NASA. These data are available at `https://disc.gsfc.nasa.gov/datacollection/ACOS_L2_Lite_FP_7.3.html`. For the OCO-2 Level 2 data, we use the Version 9 data, which are available at `https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_9r/summary`. The User Data Guide for ACOS V7.3 can be found at `https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/ACOS_v7.3_DataUsersGuide-RevF.pdf`, and the Data User Guide for OCO-2 Version 9 can be found at `https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO2_DUG.V9.pdf`.

Typically, OCO-2 and ACOS L2 data vary in retrieval quality due to different atmospheric conditions (e.g., contamination of the radiance by clouds or uncertainties in the atmospheric aerosols). Hence, the OCO-2 team recommends that the Level XCO2 data be filtered to eliminate potential 'bad' data. Here, we make use of the 'xco2_quality_flag' quality flag from the Lite products. From the OCO-2 Level 2 Data Quality Guide:

> "xco2_quality_flag [...] is simply a byte array of 0s and 1s. This filter has been

derived by comparing retrieved XCO2 for a subset of the data to various truth proxies, and identifying thresholds for different variables that correlate with poor data quality. It applies a number of quality filters based on retrieved or auxiliary variables that correlate with excessive XCO2 scatter or bias."

For the fusion product, we filter both ACOS and OCO-2 L2 product by selecting only values for which xco2_quality_flag == 0. Both data products employ a bias correction process, which is a post-processing algorithm that applies a small offset to each retrieved XCO2 value to correct for instrument biases. For our fusion, we make use of the bias-corrected XCO2 values from both ACOS and OCO-2 products.

## 2.2 Output naming convention

Our products are based on the OCO-2 Level 2 Lite products, and hence they also have daily resolution. That is, we produce 1 output netcdf file per day. The products have the following naming convention

ShortName_yyyymmdd_VersionNumber_DateTime.nc,

where

**ShortName** is a unique dataset identifier. For Product 1, the ShortName is 'OCO2GriddedXCO2. For Product 2, the ShortName is 'MultiInstrumentFusedXCO2'.

**yyyymmdd** is the year, month, and date of the data used as input into our fusion algorithm

**VersionNumber** is the version number of the fusion algorithm

**DateTime** is the date time, or any part thereof as applicable, of the first data observation in the file

For instance, an example of a output dataset name is 'MultiInstrumentFusedXCO2_20150101_V1_190815145203.nc'.

## 2.3 Output resolution

The fused products are produced at daily $0.5^\circ \times 0.5^\circ$ resolution. Since the fusion are relying upon the daily OCO-2 and GOSAT Level 2 data, we only produce fused estimates at grid cells where there is observed data within 200 km of said location. Practically, this leads to an output grid which mostly replicates the observational swath of the input data (e.g., OCO-2), although with slightly expanded coverage. We demonstrate this by plotting the OCO-2 Level 2 Lite XCO2 product and our fused XCO2 in Figure 1.

## 2.4 Data fusion output modes

The OCO-2 instrument have three primary observation modes: glint, nadir, and target. The nadir mode consists of observations where the surface solar zenith angle is less than 85 degrees, and the glint mode consist of observation at latitudes where the solar zenith angle of the glint spot is less than 75 degrees. Finally, target mode consists of very localized observations are conducted over selected OCO-2 validation sites. The three modes differ in their quality and biases. They also differ in their

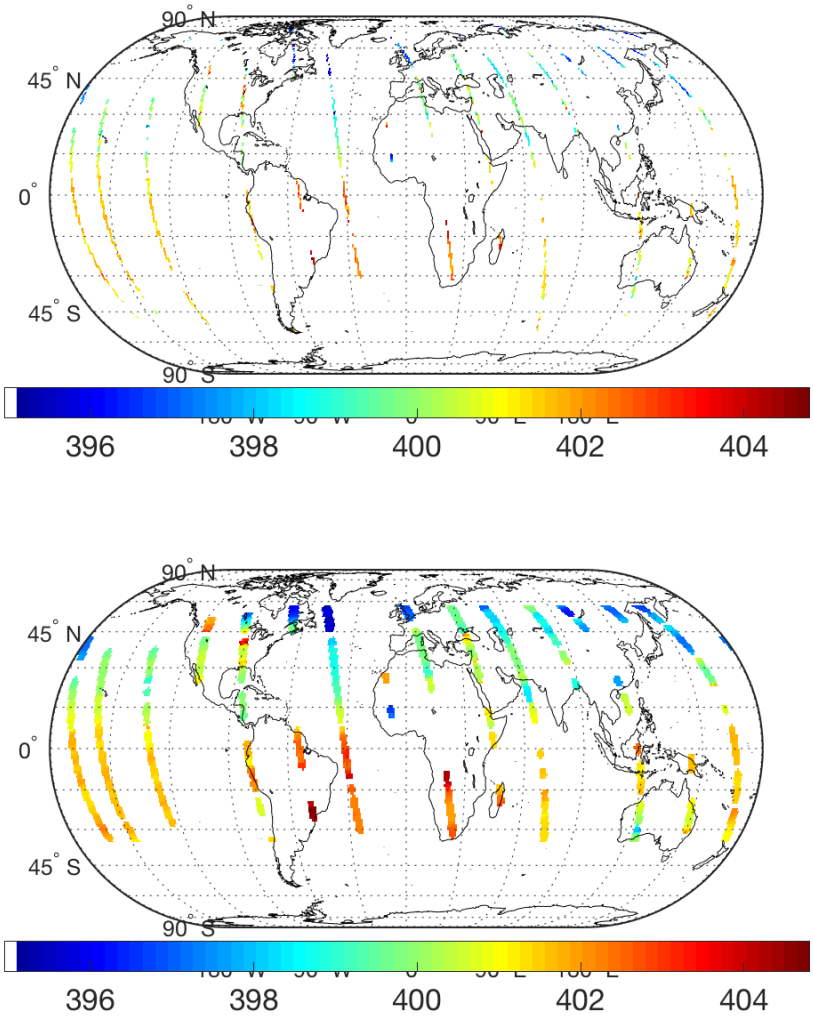


Figure 1: Top: a heat plot of XCO2 from OCO-2 for August 29, 2016. Bottom: a heat plot of the fused output XCO2 for the same date.

spatial coverage. Nadir mode, for instance, is only collected over land, while glint mode can collect observations over both land and ocean.

It has been shown that the bias correction process for ACOS and OCO-2 still

demonstrate residual bias, which depends on surface type, latitude, and scattering by aerosol **?**. One significant factor in determining the residual bias is whether the surface is land or ocean. Therefore, many flux inversion studies opt for assimilate the XCO2 data separately for land and ocean. Consequently, we stratify our fusion products into 4 different products, as seen in the table below: In the fusion outputs,

Table 1: Fusion output modes

| Product | Description |
|---|---|
| Land Only | Uses only Land observations from ACOS and OCO-2 (Land Nadir) |
| Ocean Only | Uses only Ocean observations from ACOS and OCO-2 (Land Glint and Ocean Glint) |
| Land and Ocean | Uses all ACOS observations and OCO-2 Glint and Nadir modes |
| Target | Uses only Target observations from OCO-2 |

these different modes can be identified by the variable 'source_data_mode', which is an integer ranging from 1 to 4, where 'Land Only' = 1, 'Ocean Only' = 2, 'Land and Ocean' = 3, and 'Target' = 4.

## 2.5 Output format

The fusion outputs are in netCDF format, and they include the following variables: longitude, latitude, pressure levels, pressure weighting functions, XCO2, time, prior mean, and column averaging kernel, along with other auxiliary variables. Using the naming convention of the OCO-2 Lite files and the fusion output files, these variables are described in the table below:

Table 2: Variables within fusion output files

| Name | Dimension | Description |
|---|---|---|
| longitude | 1x1 | The longitude at the center of the sounding field-of-view |
| latitude | 1x1 | The latitude at the center of the sounding field-of-view |
| xco2 | 1x1 | The bias-corrected XCO2 (in units of ppm) |
| xco2_uncertainty | 1x1 | The posterior uncertainty in XCO2 calculated by the L2 algorithm, in ppm. |
| time | 1x1 | The time of the sounding in seconds since 1970-01-01 |
| xco2_apriori | 1x1 | The prior XCO2 assumed by the L2 retrieval, in ppm. |
| co2_profile_apriori | 20x1 | The prior mean profile of CO2 in ppm |
| xco2_averaging_kernel | 20x1 | The normalized column averaging kernel for the retrieved XCO2 |
| pressure_levels | 20x1 | The retrieval pressure level grid for each sounding in hPa |
| pressure_weight | 20x1 | The pressure weighting function on levels used in the retrieval |
| date | 7x1 | The full date and time of the sounding in UTC, organized as (year, month, day, hour, minute, second, milliseconds). This information is redundant with that from the time variable. |
| source_data_mode | 1x1 | An integer ranging from 1 to 4, where 'Land Only' = 1, 'Ocean Only' = 2, 'Land and Ocean' = 3, and 'Target' = 4. |