MEASURES 2017 XCO2 Data Fusion (v3)

Data User's Guide

Hai Nguyen[1], Junjie Liu[1], Susan Kulawik[2], David Baker[3], Jonathan Hobbs[1],

Amy Braverman[1], Matthias Katzfuss[4], Vineet Yadav[1]

February 18, 2022

[1]Jet Propulsion Laboratory, California Institute of Technology

[2]Bay Area Environmental Research Institute

[3]Cooperative Institute for Research in the Atmosphere, Colorado State University

[4]Department of Statistics, Texas AM University

DUG version 0.97

Jet Propulsion Laboratory

California Institute of Technology*

*A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (80NM0018F0527).

1 Introduction

This Data User Guide describes the data products (v3) 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). The products available from this MEaSURES project are listed below:

Table 1: MEaSURES products

Description	ShortName	DOI
OCO-2 Gridded bias-corrected XCO2 and other select fields aggregated as daily files	OCO2GriddedXCO2	`https://doi.org/10.5067/582L7HTJ343N`
Multi-Instrument Fused bias-corrected XCO2 and other select fields aggregated as daily files	MultiInstrumentFusedXCO2	`https://doi.org/10.5067/ZS346LH1NTIS`
OCO-2 10-Second Averaged XCO2 and other select fields*	OCO2_10SA_XCO2	`https://doi.org/10.5067/GRVGCTOKA8TL`
Monthly Gridded Level 4 bias-corrected XCO2 and other select fields aggregated from ACOS	ACOSMonthlyGriddedXCO2	`https://doi.org/10.5067/6NHKPP06K0ZF`

*This product is still in preparation.

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 (Crisp et al., 2010). 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 (O'Dell et al., 2012).

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 (Morino et al., 2011). 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 Crisp et al., 2012, 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 four primary products: 1) OCO-2 Level 4 bias-corrected XCO2 and other select fields aggregated as daily files, 2) Multi-Instrument Fused Level 4 bias-corrected XCO2 and other select fields aggregated as daily files, 3) a 10-seconds averaged product, and 4) Monthly Gridded Level 4 bias-corrected XCO2 and other select fields aggregated from ACOS as Level 4 monthly files. Note that Products 1, 2, and 3) are on daily temporal output resolution, while Product 4 is only monthly output resolution (i.e., there is one output file per month). Products 1), 2), and 4) are produced using a variant of kriging (also known as optimal interpolation), and details can be found in Section 3 of the ATBD. The 10-seconds averaged product is produced at NOAA, and the data and ATBD can be found at `https://climatesciences.jpl.nasa.gov/projects/co2measures/`.

2.1 Data version and quality filter

For our v2 fusion products, we use ACOS Version 9 data, which are produced by the Jet Propulsion Lab at NASA. For the OCO-2 Level 2 data, we also use the Version 10 data. Links to both the datasets and their userguides can be found in the table below:

Table 2: Data Source and User Guides

Data	Source	Data User Guide
OCO-2	https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_10r/summary	https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO2_OCO3_B10_DUG.pdf
ACOS	https://oco2.gesdisc.eosdis.nasa.gov/data/GOSAT_TANSO_Level2/ACOS_L2_Lite_FP.9r/	https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/ACOS_v9_DataUsersGuide.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 2 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 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'. For Product 4, the ShortName is 'ACOSMonthlyGriddedXCO2'.

yyyymmdd is the year, month, and date of the data used as input into our fusion algorithm. Note that for the 'ACOSMonthlyGriddedXCO2' product, this field is truncated to 'yyyymm' since it is a monthly product.

VersionNumber is the version number of the fusion algorithm.

DateTime is the date time, in POSIX format, of the time of fusion production

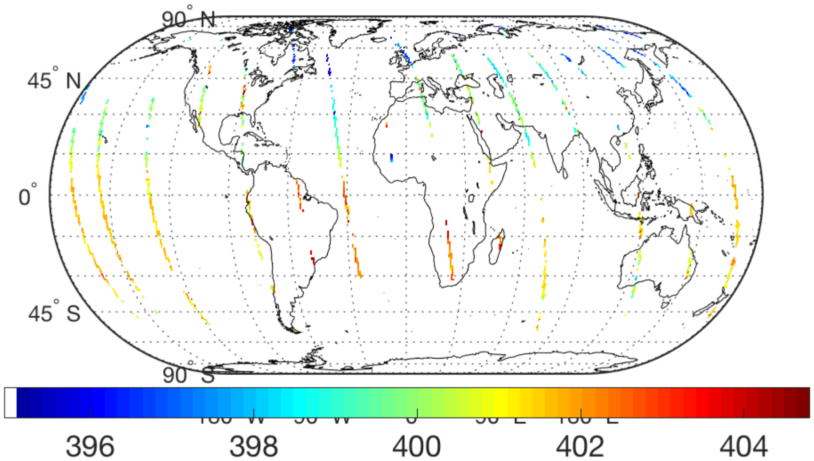
For instance, an example of an output dataset name is 'MultiInstrumentFusedXCO2_20150101_V3_190815145203.nc'.

2.3 Output resolution

The fused products are produced at daily $1^\circ \times 1^\circ$ 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 150 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 has 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



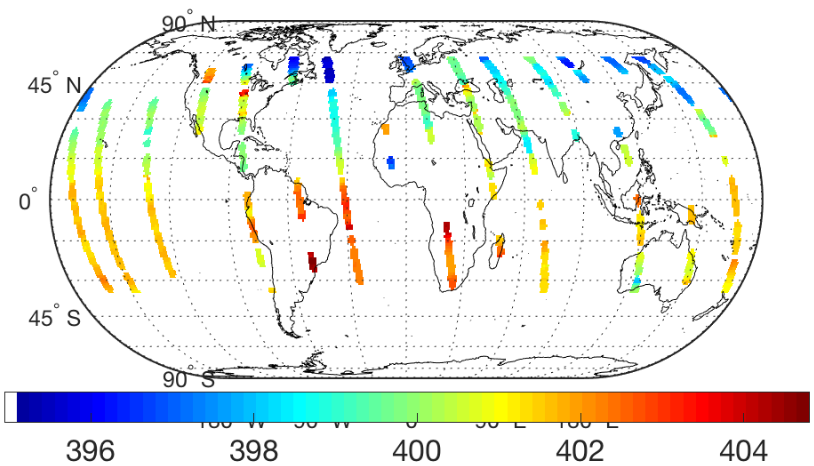


Figure 1: Top: a heat plot of XCO2 from OCO-2 for August 29, 2016. Bottom: a heat plot of the fused output XCO2 for the same date.

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 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 Wunch et al. (2017). One significant factor in determining the residual bias is whether the surface is land or ocean. Therefore, many flux inversion studies opt to assimilate the XCO2 data separately for land and ocean. Consequently, we stratify our fusion products into 4 different modes, as seen in the table below: In the fusion

Table 3: 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

outputs, 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 Table 4.

2.5.1 Sparse matrix format

The precision matrices are stored separately for land and ocean in Dictionary of Keys format, which records the row index, column index, and value of all non-zeros entries along with the matrix shape. For instance, for land observations these information are recorded in the following variables: precision_matrix_row_indices_land, precision_matrix_col_indices_land, precision_matrix_values_land, precision_matrix_col_indices_land, and precision_matrix_size_land. For ocean observations, the names of the variables are the same except with ending flag 'land' being replaced with 'ocean'.

The information above can easily be used to reconstruct the sparse precision matrix in any software of choice. Below is an example in Python:

```
import netCDF4
from scipy.sparse import dok_matrix

# sample OCO-2 output file
f = 'OCO2GriddedXCO2_20190119_v3_1639714315.nc'

# indices are 1-based, but python is 0-based
col = dat[ 'precision_matrix_col_indices_ocean' ][:] - 1
row = dat[ 'precision_matrix_row_indices_ocean' ][:] - 1
N = dat[ 'precision_matrix_size_ocean' ][:]
values = dat ['precision_matrix_values_ocean' ][:]
```

```
# construct a sparse scipy matrix in DOK format
P = dok_matrix((N,N) )

for i in range( len( col )):
    P[row[i], col[i] ] = values[i] # Update element
```

2.6 Inflation factors

Flux inversion studies often make the assumption that the input data (here, either the Level OCO-2 retrievals, 10-seconds averages, or our fusion product) are statistically independent of one another. One natural consequence of this assumption is that the 'information content' of a product depends on the size of the dataset. We currently are studying various choices of 'inflation factors', which would allow flux modelers to normalize the information content to make it more comparable across different products. In this version, we provide a preliminary metric based on the ratio of the variance of the mean estimates.

For this version of the product, we estimated the inflation factor as the ratio of the variance of the mean estimate (with spatial dependence) to that arising from the independence assumption (please see the ATBD for more details). The inflation factor is given by

$$c = \frac{\sigma^2}{\sigma_D^2} = \frac{\sum_i \sum_j \mathbf{R}^{ij}}{\sum_i \sum_j \mathbf{R}_D^{ij}}.$$

In our data products, these inflation factors are calculated separately for land

and ocean fused estimates, then their values are given in 'inflation_factor_land' and 'inflation_factor_ocean'.

Table 4: Name, dimension, and brief description of the variables within the output netCDF files

Name	Dim.	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.
inflation_factor_land	1x1	The sum (elementwise) of the full prediction covariance matrix divided by the sum of the diagonals for land estimates
inflation_factor_ocean	1x1	The sum (elementwise) of the full prediction covariance matrix divided by the sum of the diagonals for ocean estimates

precision_matrix_values_land	N_l x 1	Values of non-zero elements of the precision matrix for land predictions in dictionary of keys (DOK) format. N_l is the number of non-zero values within this precision matrix
precision_matrix_col_indices_land	N_l x 1	Column indices of non-zero elements of the precision matrix for land predictions in dictionary of keys (DOK) format.
precision_matrix_row_indices_land	N_l x 1	Row indices of non-zero elements of the precision matrix for land predictions in dictionary of keys (DOK) format. Indices are 1-based.
precision_matrix_values_ocean	N_o x 1	Values of non-zero elements of the precision matrix for ocean predictions in dictionary of keys (DOK) format. N_o is the number of non-zero values within this precision matrix
precision_matrix_col_indices_ocean	N_o x 1	Column indices of non-zero elements of the precision matrix for ocean predictions in dictionary of keys (DOK) format. Indices are 1-based
precision_matrix_row_indices_ocean	N_o x 1	Row indices of non-zero elements of the precision matrix for ocean predictions in dictionary of keys (DOK) format. Indices are 1-based
precision_matrix_size_land	1x1	The number of fusion grid points over land. This value is necessary to reconstruct the precision matrix from (row,column,value) tuples
precision_matrix_size_ocean	1x1	The number of fusion grid points over ocean. This value is necessary to reconstruct the precision matrix from (row,column,value) tuples
sif_740nm	1x1	Solar induced chlorophyll fluorescence at retrieved wavelength: SIF_740

References

Crisp, D., Boesch, H., Brown, L., Castano, R., Christi, M., Conner, B., Frankenberg, C., McDuffie, J., Miller, C., Natraj, V., O'Dell, C., O'Brien, D., Polonski, I., Oyafuso, F., Thompson, D., Toon, G., and Spurr, R. (2010). OCO (Orbiting Carbon Observatory): Level 2 Full Physics Retrieval Algorithm Theoretical Basis. Version 1.0 Rev. 4, November 10, 2010. JPL, NASA, Pasadena, CA.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L. (2012). The ACOS CO2 retrieval algorithm– Part II: Global XCO2 data characterization. *Atmospheric Measurement Techniques*, 5:687—-707.

Morino, I., Uchino, O., Inoue, M., Yoshida, Y., Yokota, T., Wennberg, P. O., Toon, G. C., Wunch, D., Roehl, C. M., Notholt, J., Warneke, T., Messerschmidt, J., Griffith, D. W. T., Deutscher, N. M., Sherlock, V., Connor, B., Robinson, J., Sussmann, R., and Rettinger, M. (2011). Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra. *Atmospheric Measurement Techniques*, 4(6):1061–1076.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D. (2012). The ACOS CO2 retrieval algorithm – Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, 5(1):99–121.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., et al. (2017). Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON.