OCO-2 GEOS L3 XCO2 Product User's Guide

Document Version: 1.0
Date: 14 March 2022

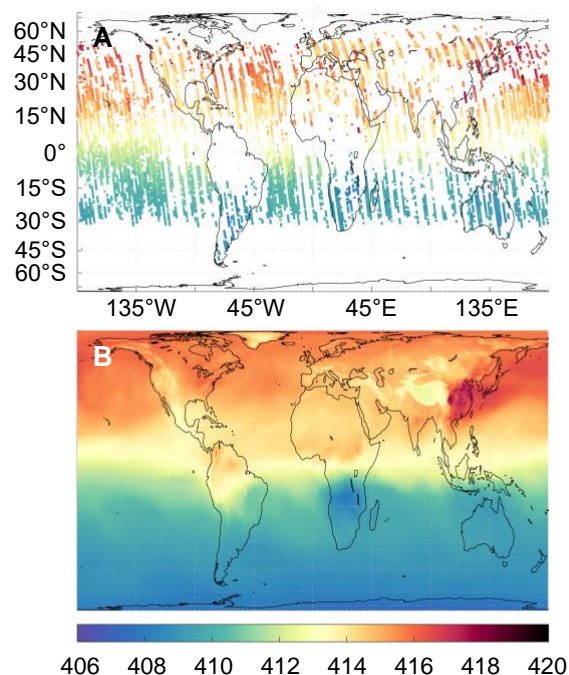# Table of Contents

1. Introduction

This document provides a brief description of NASA's Orbiting Carbon Observatory 2 (OCO-2; *Crisp et al.*, 2004; *Eldering et al.*, 2017a) Goddard Earth Observing System (GEOS) Level 3 (L3) gridded, gap-filled, column-averaged, dry-air mole fraction of carbon dioxide (XCO_2) fields. These fields are produced by assimilating OCO-2 retrievals into GEOS with the Constituent Data Assimilation System (CoDAS). Data are provided globally at 0.5° x 0.625° resolution at both daily and monthly frequencies from 1 January 2015 to present, typically with a two to three-month latency.

2. Description of the OCO-2 GEOS L3 Product

OCO-2 provides spatially resolved estimates of XCO_2 based on the column integrated number densities of carbon dioxide and molecular oxygen, which are retrieved from near infrared spectra in sunlit, clear-sky conditions. More details on the OCO-2 mission can be found in *Eldering et al.* (2017b) with additional information on the retrieval process available in *O'Dell et al.* (2018). The OCO-2 team produces several widely used data products classified as Level 1B (L1B; calibrated spectral radiances) and Level 2 (L2; orbital track retrievals), which are fully described in the OCO Version 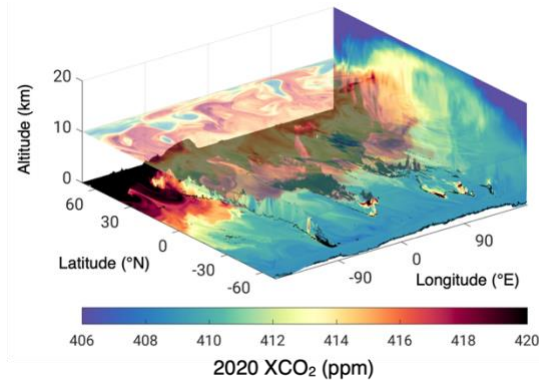10 (v10) User's Guides available at the data archive pages (https://doi.org/10.5067/6O3GEUK7U2JG and https://doi.org/10.5067/E4E140XDMPO2).

Though the OCO-2 mission provides the highest quality space-based XCO_2 retrievals to date, the L2 data are characterized by large gaps in coverage due to OCO-2's narrow 10-km ground track and an inability to see through clouds and thick aerosols. Several different methods have been explored to produce spatially complete L3 (gridded) XCO_2 fields including averaging, kriging, and data assimilation. Here, we describe fields produced using a data assimilation technique commonly referred to as state estimation within the geophysical literature. Data assimilation synthesizes simulations and observations, adjusting the state of atmospheric constituents like CO_2 to reflect observed values, thus gap-filling observations when and where they are unavailable based on previous observations and short

Figure 1. Snapshots of OCO-2 L2 soundings and assimilated OCO-2 GEOS L3 fields: (**A**) 16 days of OCO-2 L2 XCO_2 soundings on 1–16 April 2020 and (**B**) the 16-day average of assimilated OCO-2 GEOS L3 XCO_2 fields over the same period. Data assimilation combines satellite observations (A) with a weather-resolving atmospheric model to form gridded, time-varying, three-dimensional fields, from which averages (B) and uncertainties (below) follow. Units are parts per million (ppm).

transport simulations by GEOS. Compared to other methods, data assimilation has the advantage that it makes estimates based on our collective scientific understanding, notably of the Earth's carbon cycle and atmospheric transport (Figure 1; more details in *Weir et al.*, 2021b).

Figure 2. Transects of instantaneous OCO-2 GEOS CO_2 at the surface (bottom), 10 km above sea level (top, transparent), and along the International Dateline (right) on 9 April 2020, 00:00 UTC. By reproducing the global, high-resolution, vertical and temporal variability of CO_2, the assimilation system can synthesize heterogenous data types across drastically different scales, e.g., satellite retrievals and in situ measurements from surface stations and aircraft.

OCO-2 GEOS L3 data are produced by ingesting OCO-2 L2 retrievals every 6 hours with GEOS CoDAS, a modeling and data assimilation system maintained by NASA's Global Modeling and Assimilation Office (GMAO). GEOS CoDAS uses a high-performance computing implementation of the Gridpoint Statistical Interpolation (GSI; *Wu et al.*, 2002) approach for solving the state estimation problem. GSI finds the analyzed state that minimizes the three-dimensional variational (3D-Var) cost function formulation of the state estimation problem. In particular, it ingests column retrievals of trace gas abundances taking into account both their vertical sensitivity, i.e., averaging kernel, and a priori assumptions. When and where data are unavailable, e.g., in cloudy scenes, gaps are filled using 6-hour GEOS simulations and all previous observations in a statistically optimal way (*Jazwinski*, 1970). The atmospheric circulation in GEOS is constrained by the millions of remote sensing and in situ observations every hour included in the Modern Era Retrospective analysis for Research and Application, version 2 (MERRA-2; *Gelaro et al.*, 2017). This accurate representation of transport patterns is critical for interpreting measured variations that reflect a combination of nearby and distant surface fluxes due to the long lifetime of CO_2.

The assimilated OCO-2 GEOS L3 product uses the same bias correction as the OCO-2 L2 product, applies additional quality control (QC) flags, and inflates the reported error to reflect cross-track variability of XCO_2. The additional QC flags screen for 1) soundings over snow and ice, 2) glint angles greater than 80°, 3) tracks with less than 4 footprints, 4) soundings with reported uncertainties less than 10^{-3} parts per million (ppm). Given a cross-track standard deviation of σ_t and a reported retrieval uncertainty of σ_r, we use an inflated uncertainty of

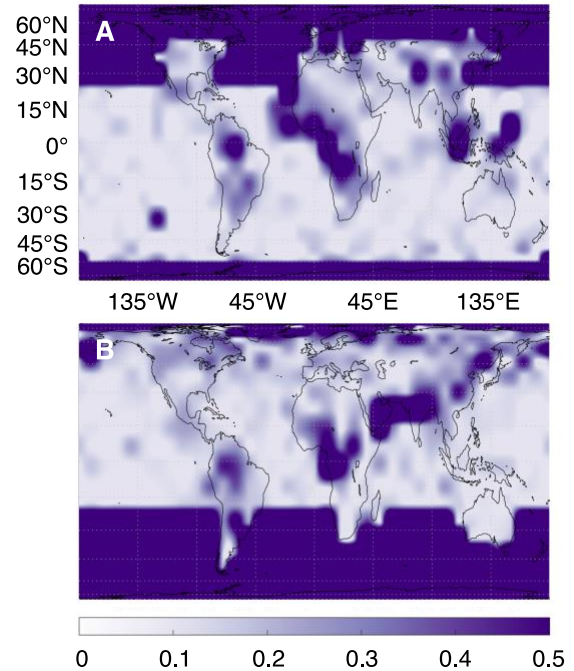$$\sigma' = \sqrt{\sigma_r^2 + \sigma_t^2}$$

for all soundings in the track.

Prior to any assimilation of XCO_2 data, GEOS CoDAS products are informed by a high-quality, observationally-derived estimate of surface-atmosphere CO_2 flux called the Low-order Flux Inversion (LoFI; *Weir et al.*, 2021a). Fossil fuel emissions are prescribed based on the

Open-source Data Inventory for Anthropogenic CO_2 (ODIAC), which combines national total emissions estimates with satellite observations of nighttime lights to produce 0.1° global emissions maps (*Oda and Maksyutov*, 2011). Emissions from fires come from the Quick Fire Emissions Dataset (QFED; *Darmenov and da Silva*, 2015), which estimates emissions in near real time using Moderate Resolution Imaging Spectrometer (MODIS) fire radiative power data. Land-atmosphere exchange is derived from the Carnegie-Ames-Stanford-Approach–Global Fire Emissions Database version 3 (CASA-GFED3; *Randerson et al.*, 1997) model that uses Advanced Very High Resolution Radiometer (AVHRR) Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI) 3g (*Pinzon and Tucker*, 2014) and MERRA-2 meteorological data to estimate monthly net primary production and heterotrophic respiration carbon fluxes (https://doi.org/10.5067/VQPRALE26L20). Ocean fluxes use a simple technique to restore inter-annual variability to the *Takahashi et al.* (2014) climatology. Finally, an additional empirically derived sink is applied to ensure realistic atmospheric growth rates. A unique feature of the flux collection used by GEOS CoDAS is the ability to run retrospectively, using year-specific satellite-derived estimates, or in near real time using a projected atmospheric growth rate and data from previous years. Simulations using LoFI tend to perform comparably to modern flux inversions (*Weir et al.*, 2021a; *Peiro et al.*, 2022) and serve as the baseline in the assessment of assimilation skill in Section 3.

For production of OCO-2 GEOS L3 data, GEOS CoDAS runs on a horizontal grid with a nominal horizontal resolution of 50 km and 72 terrain-following vertical levels from the surface up to 0.01 hPa (see Figure 2). Its relatively fine spatial resolution enables it to reproduce surface and aircraft observations with high fidelity (see Section 3 and *Weir et al.*, 2021a, 2021b; *Bell et al.*, 2020, *Campbell et al.*, 2020; *Zhang et al.*, 2022, *Peiro et al.*, 2022). Three-dimensional CO_2 fields are vertically integrated to produce two-dimensional XCO_2 fields which are provided at both daily and monthly temporal resolution. Data are provided at the same horizontal resolution and in the same format as MERRA-2 products to facilitate interoperability.

Figure 3. Snapshots of OCO-2 GEOS L3 analysis precisions: (**A**) 1 January 2015 (**B**) 1 July 2015. Units are ppm and the range was selected to highlight variability from 0 to 0.5 ppm.

Uncertainty quantification

Daily and monthly random errors, i.e., precisions, are calculated using the a posteriori *Desroziers et al.* (2005) diagnostics. Those diagnostics converge in the mean to observation-space projections of the background, analysis, and observation error covariances. These diagnostics are only representative of column, i.e., XCO_2 errors because they are expressed in observation space, not state space. Furthermore, they assume Gaussianity, which

3

is an assumption needed in practice, but almost surely false. Since these diagnostics reference themselves, they are not as useful for estimating systematic errors—systematic errors are thus not reported here. Analyses against independent data (e.g., Section 3) serve as a better indicator of systematic error.
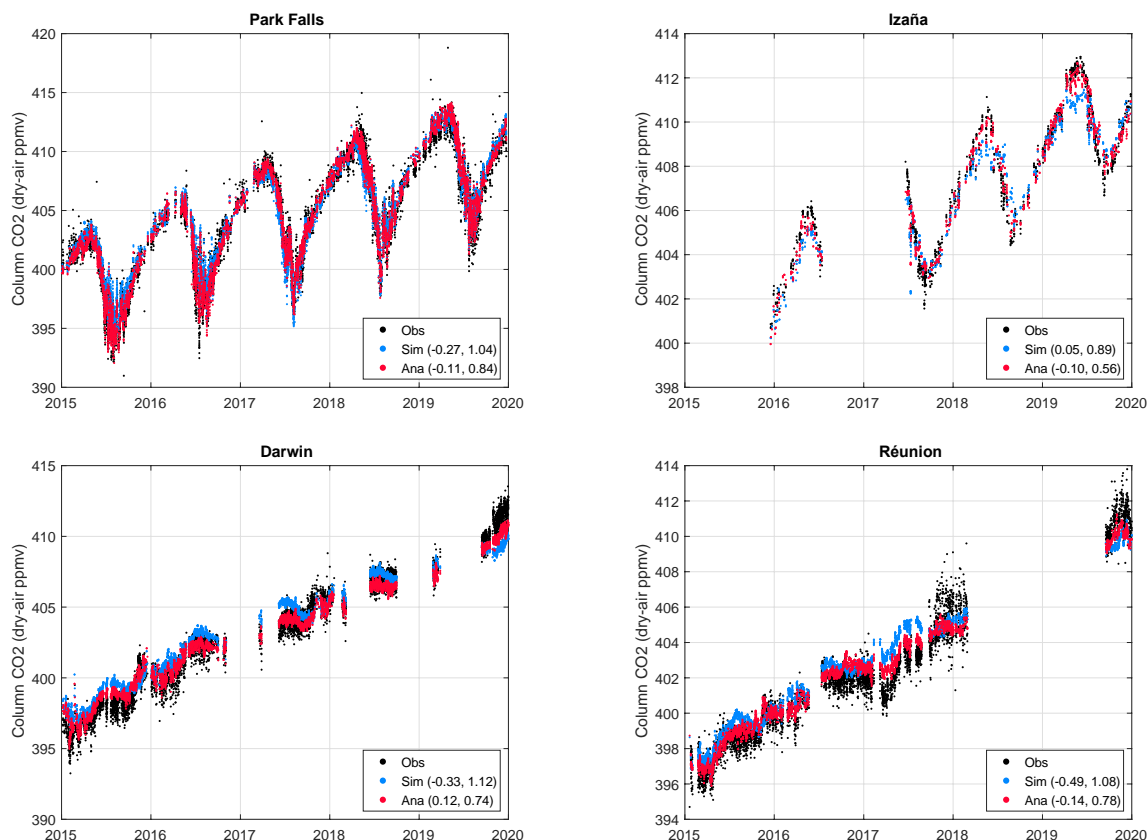
Since the OCO-2 swath is just 10 km, we must bin soundings over a very wide spatial and temporal extent to have enough samples for the statistics of the Desroziers diagnostics to converge. We combine all soundings for a given month into 8° x 10° bins. Averages over these bins are used to estimate the daily error statistics. The bin size was selected based on experimentation to balance the need to resolve important geophysical features (higher resolution) with that of statistical convergence (lower resolution). For bins in which there are no soundings, we set the daily analysis error standard deviation to 1 ppm which is roughly consistent with the results in *Weir et al.* (2021a), *Peiro et al.* (2022), *Zhang et al.* (2022), and the evaluation in Section 3. We then interpolate the daily random error statistics onto the 0.5° x 0.625° analysis grid. For two examples, see Figure 3. The reported monthly random errors are simply the average of the daily random errors. This average produces a number that is quite small with the important note that for monthly statistics we expect the systematic error (i.e., bias) to play a greater role than the random error.

Users should understand that during Arctic and Antarctic night, there is no observational coverage from OCO-2. Nevertheless, since GEOS can transport increments, it is possible for the assimilation to inform regions without sunlight. In the evaluation below, we see that assimilation improves GEOS comparisons to independent data near the South Pole even during its winter. For this reason, we do not currently provide a data indicator. Instead, if the user needs an indication of OCO-2 L2 sounding coverage, we suggest either consulting the L2 files or using the reported random errors of the OCO-2 GEOS L3 analysis to determine the precision of the "integrated" data constraint of all ingested products, not just OCO-2 L2 soundings, on XCO_2 at a given place and time.
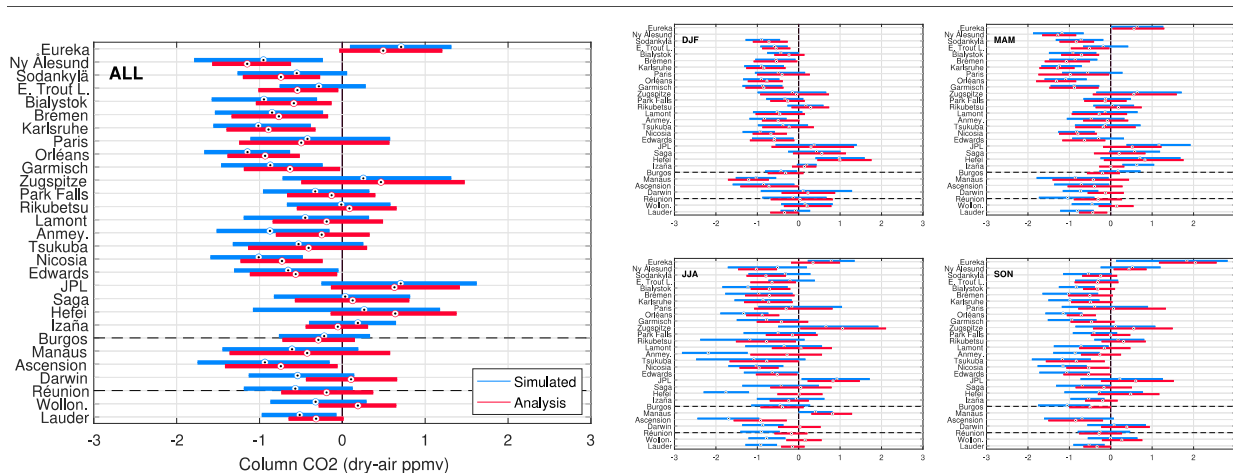
3. Independent Data Evaluation

Here we evaluate OCO-2 GEOS L3 XCO_2 against Total Carbon Column Observing Network (TCCON; *Wunch et al.*, 2011) and the analysis three-dimensional CO_2 (Figure 2), from which XCO_2 is computed, against NOAA ObsPack GlobalView+ v6.1 collection of in situ data (*Schuldt et al.*, 2021). The ObsPack comparison focuses on surface data from the marine boundary layer and aircraft campaign data from NASA's Atmospheric Tomography (ATom; *Wofsy et al.*, 2018), Atmospheric Carbon and Transport–America (ACT-America; *Davis et al.*, 2021) and Arctic Carbon Atmospheric Profiles (Arctic-CAP; *Sweeney et al.*, 2022) flights from the Arctic-Boreal Vulnerability Experiment (ABoVE; *Miller et al.*, 2019), the National Science Foundation's O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study (ORCAS; *Stephens et al.*, 2018), and Japan's Comprehensive Observation Network for Trace Gases by Airliner (CONTRAIL; *Machida et al.*, 2008). Figures showing the current status of this evaluation are given below. Updates and more comprehensive comparisons of GEOS products to these and other datasets are available on the GMAO Carbon fluid page https://fluid.nccs.nasa.gov/carbon/.
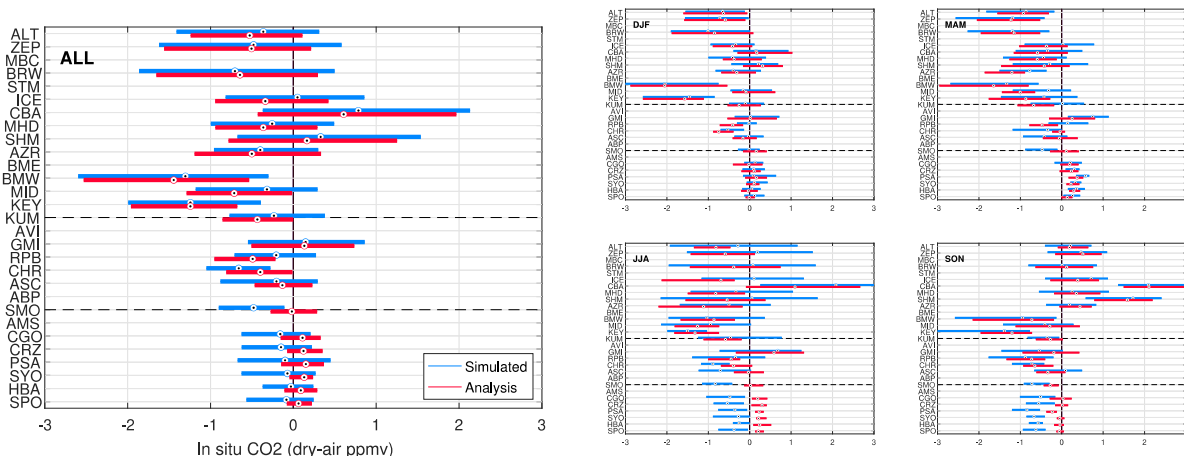
Figure 4. Evaluation against TCCON XCO$_2$ observations (black dots) of the baseline GEOS LoFI simulation (blue dots) and the OCO-2 GEOS L3 analysis (red dots). Numbers in parentheses are the mean and root-mean-squared-difference (RMSD) of the observation - gridded product differences. The analysis shows clear improvement across sites, notably in RMSD. These comparisons are summarized for all TCCON sites and by season in Figure 5.
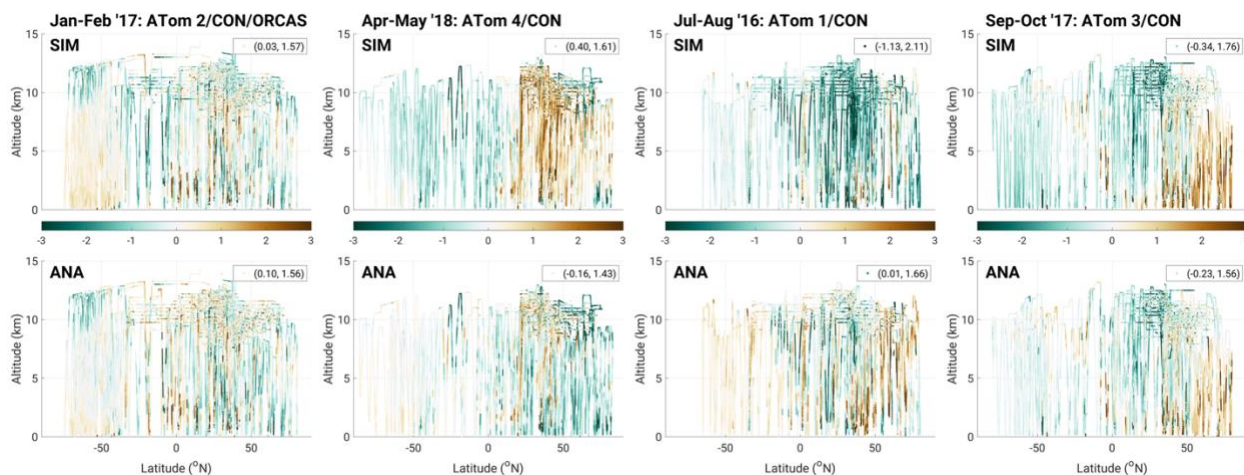

Figure 5. Box and whisker plots of TCCON observation - simulation (blue) and observation - analysis (red) differences for all months (ALL), December, January, February (DJF), March, April, May (MAM), June, July, and August (JJA), and September, October, and November

5

(SON). Boxes span the 25th to 75th percentiles with the median indicated by a circled dot. On the vertical axis, sites are ordered by increasing latitude with the Tropics of Cancer and Capricorn indicated with dashed lines. See Figure 4 for more details about the simulation and analysis. The analysis shows small, yet consistent, improvements upon the simulation over all time periods with the most notable, and surprising, improvements over the Southern Hemisphere in austral winter (MAM and JJA).
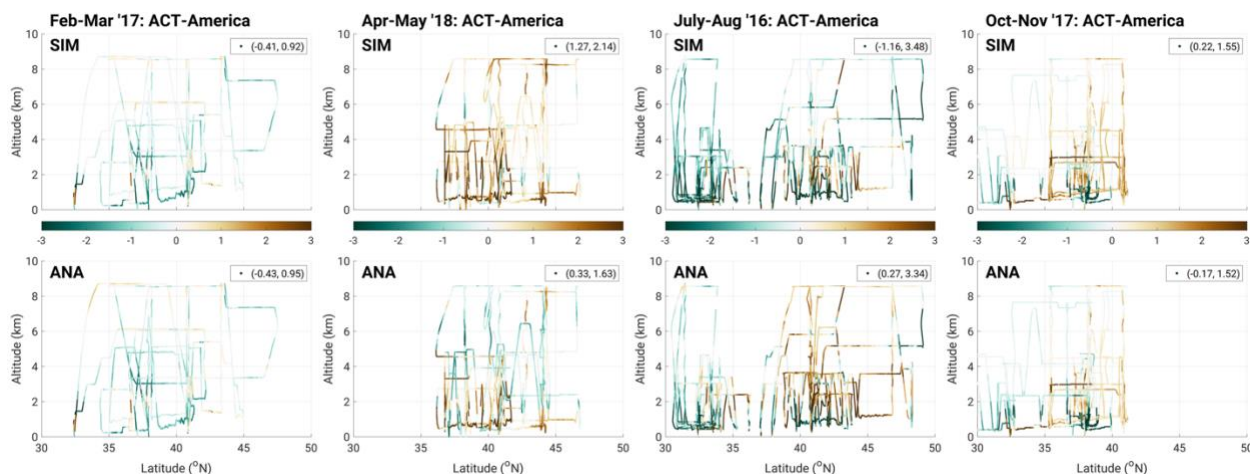

Figure 6. Same as for Figure 5, but for NOAA marine boundary layer measurements. As in the column, in the surface comparison shows the analysis shows small, yet consistent, improvements upon the simulation over all time periods with the most notable, and surprising, improvements over the Southern Hemisphere in austral winter (MAM and JJA).
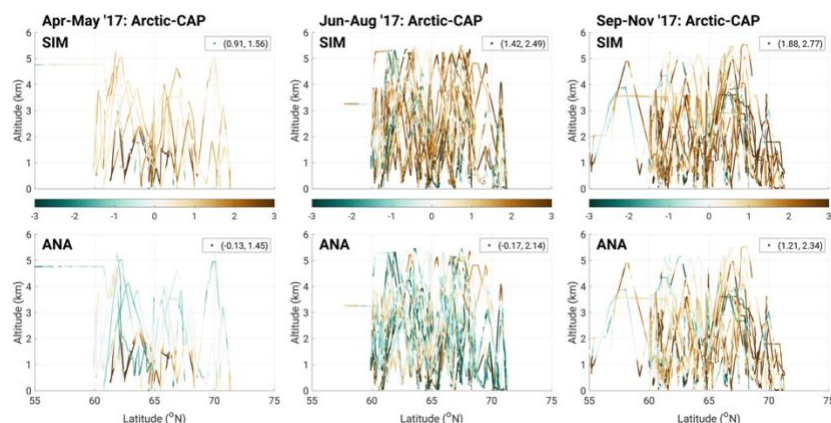

Figure 7. Latitude by altitude cross-sections of ATom aircraft campaign comparisons. Campaigns are ordered seasonally from left to right: DJF (ATom 2, Winter 2017), MAM (ATom 4, Spring 2018), JJA (ATom 1, Summer 2016), SON (ATom 3, Fall 2017). The top row shows observations - simulation values and the bottom shows observation - analysis values. As in Figure 4, numbers in parentheses indicate the mean and root-mean-squared-difference (RMSD). When available, CONTRAIL and ORCAS data for the same seasons are shown as well. The

analysis shows consistent improvements in three seasons (MAM, JJA, and SON) and comparable skill (DJF) in one season.

Figure 8. Same as for Figure 7 for ACT-America and zoomed-in over North America. Again, the analysis shows consistent improvements in three seasons (MAM, JJA, and SON) and comparable skill (DJF) in one, when there is limited sunlight and far less OCO-2 data.

Figure 9. Same as for Figure 7 for Arctic-CAP and zoomed-in over the Arctic. As in the previous figures, the analysis shows consistent improvements in MAM, JJA, and SON.

4. File Naming Conventions

The standard full name for OCO GEOS L3 products will follow the format described below:
[satellite]_[data product]_[time resolution]_[date]_[build ID]Ar.nc4

For all files, **satellite** = 'oco2' and **data product** = 'GEOS_L3CO2'. **Time resolution** describes the frequency at which variables are written within the file and can be 'month' or 'day'. **Date** defines the day or month contained in the file and has the form yyyymm for monthly files, and the form yyyymmdd for daily files. More details on filenames are given in the description of

available datasets below. Note that the OCO L2 products use a 2-digit year while we use a 4-digit year identifier.

5. Available Data

Data are provided in netCDF format in the following collections:

oco2_GEOS_L3CO2_day: OCO-2 Level 3 Daily XCO2

> **Frequency:** *Daily, containing 1 daily value*
> **Dimensions:** *longitude=576, latitude=361, time=1*
> **Granule Size:** *~3.2 MB*
> **Short name**: *oco2_GEOS_L3CO2_day*
> **Filename:** *oco2_GEOS_L3CO2_day_**yyyymmdd**_B10206Ar.nc4 where **yyyymmdd** reflects*

the 4-digit year, month, and day of the date whose contents are reported in the file.
> **doi**: 10.5067/Y9M4NM9MPCGH

Science Variables

Name	Dim	Description	Units
XCO2	tyx	CO2 Dry-Air Column Average (analysis)	mol mol-1
XCO2PREC	tyx	CO2 Dry-Air Column Average Precision	mol mol-1

oco2_GEOS_L3CO2_month: OCO-2 Level 3 Monthly XCO2

> **Frequency:** *Monthly, containing 1 monthly value*
> **Dimensions:** *longitude=576, latitude=361, time=1*
> **Granule Size:** *~3.2 MB*
> **Short name**: *oco2_GEOS_L3CO2_month*
> **Filename:** *oco2_GEOS_L3CO2_month_**yyyymm**_B10206Ar.nc4 where **yyyymm** reflects the*

4 digit year and month whose contents are reported in the file.
> **doi**: 10.5067/BGFIODET3HZ8

Science Variables

Name	Dim	Description	Units
XCO2	tyx	CO2 Dry-Air Column Average (analysis)	mol mol-1
XCO2PREC	tyx	CO2 Dry-Air Column Average Precision	mol mol-1

6. Contacts

Lesley Ott (lesley.ott@nasa.gov)
Brad Weir (brad.weir@nasa.gov)

7. Acknowledgement of Funding

8. References

Bell, E., O'Dell, C. W., Davis, K. J., Campbell, J., Browell, E., Scott Denning, A., ... & Weir, B. (2020). Evaluation of OCO-2 XCO_2 variability at local and synoptic scales using lidar and in situ observations from the ACT-America campaigns. *J. Geophys. Res.: Atmos.*, *125*, e2019JD031400. https://doi.org/10.1029/2019JD031400

Campbell, J. F., Lin, B., Dobler, J., Pal, S., Davis, K., Obland, M. D., ... & Kochanov, R. (2020). Field evaluation of column CO_2 retrievals from intensity-modulated continuous-wave differential absorption lidar measurements during the ACT-America campaign. *Earth and Space Sci.*, *7*, e2019EA000847. https://doi.org/10.1029/2019EA000847

Crisp, D., Atlas, R. M., Breon, F. M., Brown, L. R., Burrows, J. P., Ciais, P., ... & Schroll, S. (2004). The orbiting carbon observatory (OCO) mission. *Adv. Space Res.*, *34*, 700–709. https://doi.org/10.1016/j.asr.2003.08.062

Darmenov, A. S., & da Silva, A. M. (2015). The Quick Fire Emissions Dataset (QFED): Documentation of versions 2.1, 2.2 and 2.4. *NASA Technical Report Series on Global Modeling and Data Assimilation*, *38*, NASA/TM-2015-104606. https://gmao.gsfc.nasa.gov/pubs/docs/Darmenov796.pdf

Davis, K. J., Browell, E. V., Feng, S., Lauvaux, T., Obland, M. D., Pal, S., ... & Williams, C. A. (2021). The atmospheric carbon and transport (ACT)-America mission. *Bull. Am. Meteorol. Soc.*, *102*, E1714–E1734. https://doi.org/10.1175/BAMS-D-20-0300.1

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *QJRMS*, *131*, 3385–3396. https://doi.org/10.1256/qj.05.108

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., ... & Weir, B. (2017a). The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science*, *358*, eaam5745. https://doi.org/10.1126/science.aam5745

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., ... & Yoshimizu, J. (2017b). The Orbiting Carbon Observatory-2: First 18 months of science data products. *Atmos. Meas. Tech.*, *10*, 549–563. https://doi.org/10.5194/amt-10-549-2017

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... & Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.*, *30*, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press, Inc., New York, New York, USA.

Miller, C. E., Griffith, P. C., Goetz, S. J., Hoy, E. E., Pinto, N., McCubbin, I. B., ... & Margolis, H. A. (2019). An overview of ABoVE airborne campaign data acquisitions and science opportunities. *Environ. Res. Lett.*, *14*, 080201. https://doi.org/10.1088/1748-9326/ab0d44

Oda, T. and Maksyutov, S. (2011). A very high-resolution (1km x 1km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmos. Chem. Phys.*, *11*, 543–556. https://doi.org/10.5194/acp-11-543-2011

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., ... & Velazco, V. A. (2018). Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmos. Meas. Tech.*, *11*, 6539–6576. https://doi.org/10.5194/amt-11-6539-2018

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., ... & Baker, I. (2022). Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7. *Atmos. Chem. Phys.*, *22*, 1097–1130. https://doi.org/10.5194/acp-22-1097-2022

Pinzon, J. E., & Tucker, C. J. (2014). A non-stationary 1981–2012 AVHRR $NDVI_{3g}$ time series. *Remote Sens.*, *6*, 6929–6960. https://doi.org/10.3390/rs6086929

Schuldt, K., Mund, J., Luijkx, I. T., Aalto, T., Abshire, J. B., Aikin, K., … & van den Bulk, P. (2021). Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957–2019, obspack_co2_1_GLOBALVIEWplus_v6.1_2021-03-01, NOAA Global Monitoring Laboratory. http://doi.org/10.25925/20201204

Stephens, B. B., Long, M. C., Keeling, R. F., Kort, E. A., Sweeney, C., Apel, E. C., ... & Watt, A. S. (2018). The O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study. *Bull. Am. Meteorol. Soc.*, *99*, 381–402. https://doi.org/10.1175/BAMS-D-16-0206.1

Sweeney, C., Chatterjee, A., Wolter, S., McKain, K., Bogue, R., Newberger, T., ... & Miller, C. E. (2020). Atmospheric carbon cycle dynamics over the ABoVE domain: an integrated analysis using aircraft observations (Arctic-CAP) and model simulations (GEOS). *Atmos. Chem. Phys. (accepted)*. https://doi.org/10.5194/acp-2020-609

Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., Ho, C., Newberger, T., ... & Munro, D. R. (2014). Climatological distributions of pH, pCO_2, total CO_2, alkalinity, and $CaCO_3$ saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chem.*, *164*, 95–125.

Weir, B., Ott, L. E., Collatz, G. J., Kawa, S. R., Poulter, B., Chatterjee, A., ... & Pawson, S. (2021). Bias-correcting carbon fluxes derived from land-surface satellite data for retrospective and near-real-time assimilation systems. *Atmos. Chem. Phys.*, *21*, 9609–9628. https://doi.org/10.5194/acp-21-9609-2021

Weir, B., Crisp, D., O'Dell, C. W., Basu, S., Chatterjee, A., Kolassa, J., ... & Ott, L. E. (2021). Regional impacts of COVID-19 on carbon dioxide detected worldwide from space. *Science Adv.*, *7*, eabf9415. https://doi.org/10.1126/sciadv.abf9415

Wofsy, S. C., Afshar, S., Allen, H. M., Apel, E., Asher, E. C., Barletta, B., ... & Wennberg, P. (2018). ATom: Merged atmospheric chemistry, trace gases, and aerosols. ORNL DAAC, Oak Ridge, Tenn., USA. https://daac.ornl.gov/ATOM/guides/ATom_merge.html

Wu, W. S., Purser, R. J., & Parrish, D. F. (2002). Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weath. Rev.*, *130*, 2905–2916. https://doi.org/10.1175/1520-0493(2002)130%3C2905:TDVAWS%3E2.0.CO;2

Wunch, D., Toon, G. C., Blavier, J. F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., ... & Wennberg, P. O. (2011). The total carbon column observing network. *Philos. Trans. Royal Soc. A*, *369*, 2087–2112. https://doi.org/10.1098/rsta.2010.0240

Zhang, L., Davis, K. J., Schuh, A. E., Jacobson, A. R., Pal, S., Cui, Y. Y., ... & Basu, S. (2022). Multi-Season Evaluation of CO_2 Weather in OCO-2 MIP Models. *J. Geophys. Res.: Atmos.*, *127*(2), e2021JD035457. https://doi.org/10.1029/2021JD035457