Dataset Documentation

EAGER: MethaneAIR

Award #1856426, Amount: $277,144, Period of support: February 1, 2019–January 31,2022

Authors

Christopher Chan Miller[1,2,3]	cmiller@g.harvard.edu
Jonathan E. Franklin[1]	jfranklin@g.harvard.edu
Sébastien Roche[1,2]	sroche@g.harvard.edu
Jonas S. Wilzewski[1,2]	jwilzewski@g.harvard.edu
Kang Sun[5,6]	kangsun@buffalo.edu
Xiong Liu[2]	xliu@cfa.harvard.edu
Kelly Chance[2]	kchance@cfa.harvard.edu
Amir H. Souri[2,4]	ahsouri@gmail.com
Eamon Conway[2]	econway@northeastern.edu
Jenna Samra[2]	jsamra@cfa.harvard.edu
Jacob Hawthorne[2]	jacob.hohl@cfa.harvard.edu
Carly Staebell[5,6]	carlysta@buffalo.edu
Apisada (Ju) Chulakadabba[1]	achulakadabba@g.harvard.edu
Maryann Sargent[1]	mracine@fas.harvard.edu
Joshua S. Benmergui[1,7]	benmergui@g.harvard.edu
Bruce C. Daube[1]	bdaube@fas.harvard.edu
Yang Li[8]	Yang_Li3@baylor.geu
Josh Laughner[9]	josh.laughner@jpl.nasa.gov
Steven C. Wofsy[1]	wofsy@g.harvard.edu

[1]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA.
[2]Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA
[3]Climate Change Research Centre, University of New South Wales, Kensington, NSW, Australia
[4]Kostas Research Institute for Homeland Security, Northeastern University, Burlington, MA, USA
[5]Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY, USA
[6]Research and Education in Energy, Environment and Water Institute, University at Buffalo, Buffalo, NY, USA
[7]Environmental Defense Fund, New York, NY
[8]Department of Environmental Science, Baylor University, Waco, TX
[9]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Correspondence:
Principal Investigator
Steven C. Wofsy, Harvard University wofsy@g.harvard.edu
Professor of Atmospheric and Environmental Chemistry
Harvard University
24 Oxford St
Cambridge, MA 02138

617-495-4566
https://seas.harvard.edu/person/steven-wofsy

Table of Contents

1.0 Data Set Description

Introduction

MethaneAIR (“MAIR”) is a dual imaging spectrometer observing reflected sunlight at 1.65 μm (CH_4 and CO_2) and 1.27 μm ($O_2\ {}^1\Delta_g$, to help determine the optical path). It measures absorption spectra with high spectral resolution, fine spatial resolution, wide swath, and high signal-to-noise ratio, enabling high-contrast, fine-grained images of CH_4 column mean dry mole fraction. MAIR duplicates the spectroscopy of the upcoming MethaneSAT satellite (“MSAT”, launch Q1 2023). Development of the MAIR sensor and associated retrieval algorithms was supported by MethaneSAT LLC (a subsidiary of the Environmental Defense Fund).

In 2019, NSF supported science demonstration flights of MAIR with an EAGER grant, carried out in August 2021 on the NCAR Gulfstream V. Two engineering test flights in Nov. 2019 (TF01, TF02) were successful, but the flight series was terminated due to avionics failures on the GV. The series was rescheduled for March 2020, but the US shut down that month due to COVID-19. Flights rescheduled for June 2021 were pushed to late July/August of 2021 to accommodate another mission that had been displaced by unavailability of the C-130. After a short test flight (TF03), research flights RF04-RF09 targeted oil and gas infrastructure in Texas, New Mexico, and North Dakota. There were also 21 flight segments on RF04 and RF05 to compare emission rates of methane derived from MethaneAIR data to controlled releases carried out near Odesa, TX, by the Stanford group of Adam Brandt.

Engineering test flight TF10 was carried out at the end of the flight series. The spectrometer was inverted to view the upward-looking port of the G-V, recording the spatial pattern of airglow in the 1.27 μm band.

This MethaneAIR data release includes the parameters given in Table 1. All research flights (RF04-RF09) are included in the current data release.

Data version number and date

Version 01, 28 February 2023

Data Status

Final

Time period covered by the data

Nov. 2019 – August 2021

Physical location of the measurement or platform

NCAR Gulfstream V

Data Frequency

0.1 Hz

Data sources and Restrictions

Sources: See Table 1. Restrictions : None

Table 1. Variables in the Level 2 NetCDF File for MethaneAIR

xmx = cross track number of pixels = 256 (172 filled, others are missing).
tmx = along track number of pixels (unlimited)
zmx = number of vertical layers in the retrieval = 19

Variable	Units	Source	Description
lon(xmx, tmx)	Degrees E	Avionics[1]	Longitude (pixel)
lat(xmx, tmx)	Degrees N	Avionics[1]	Latitude
zsurf(xmx, tmx)	km	DEM	Surface Elevation
sza(xmx, tmx)	Degrees	Avionics[1]	Solar Zenith Angle
aza(xmx, tmx)	Degrees	Avionics[1]	Relative Azimuth Angle
vza(xmx, tmx)	Degrees	Avionics[1]	Viewing Zenith Angle
z_obs(tmx)	km	Avionics[1]	Altitude of observation
tau(tmx)	Hours (0 = 1/1/1985)	Avionics[1]	Time (GEOS Tau Format)
clon(xmx, tmx, cmx)	Degrees E		Pixel intersection corners
clat(xmx, tmx, cmx)	Degrees W		Pixel intersection corners
psurf0(xmx, tmx)	hPa	GeosFP	Surface Pressure *a priori*
alb0(xmx, tmx)		Computed from 1622.5nm radiance	Albedo = pi*radiance/cos(SZA)
u(xmx, tmx)	m s^{-1}	Geos FP	Zonal wind
v(xmx, tmx)	m s^{-1}	Geos FP	Meridional wind
rms(xmx, tmx)		Retrieval	RMS of spectral fit—note 5
n_iter(xmx, tmx)		Retrieval	Number of iterations
cost_func(xmx, tmx)		Retrieval	Cost function at convergence
ch4_dofs(xmx, tmx)		Retrieval	CH_4 Profile DoFS
co2_dofs(xmx, tmx)		Retrieval	CO_2 Profile DoFS
h2o_dofs(xmx, tmx)		Retrieval	H_2O Column DoFS
Tshft_dofs(xmx, tmx)		Retrieval	Temp. Prof Offset DoFS
psrf_dofs(xmx, tmx)		Retrieval	Surface Pressure DoFS
isrfsqz_w1_dofs(xmx, tmx)		Retrieval	CO_2 Window ISRF DoFS
isrfsqz_w2_dofs(xmx, tmx)		Retrieval	CH_4 Window ISRF DoFS
xch4_0(xmx, tmx)	mole/mole	GGG 2020	XCH_4 prior
xco2_0(xmx, tmx)	mole/mole	GGG 2020	XCO_2 prior
xch4(xmx, tmx) – *note 2*	mole/mole	Retrieval	Retrieved XCH4, not bias corrected
pix_area_m2(xmx, tmx)–note 3	m^2	Avionics[1]	Geometric pixel area
ch4_vcd(xmx, tmx)	molecules cm^{-2}	Retrieval	CH_4 Vert. Col. Density
ch4_vcd0(xmx, tmx)	molecules cm^{-2}	GGG 2020	Prior CH_4 Vert. Col. Density
co2_vcd(xmx, tmx)	molecules cm^{-2}	Retrieval	CO_2 Vert. Col. Density
co2_vcd0(xmx, tmx)	molecules cm^{-2}	GGG 2020	Prior CO_2 Vert. Col. Density
h2o_vcd(xmx, tmx)	molecules cm^{-2}	Retrieval	H_2O Vert. Col. Density
h2o_vcd0(xmx, tmx)	molecules cm^{-2}	GGG 2020	Prior H_2O Vert. Col. Density
air_vcd(xmx, tmx)	molecules cm^{-2}	Retrieval	Air Vert. Col. Density
air_vcd0(xmx, tmx)	molecules cm^{-2}	GGG 2020	Prior Air Vert. Col. Density
ch4_pvcd0(xmx, tmx, zmx)	molecules cm^{-2}	GGG 2020	Prior CH_4 Layer-by-layer VCD
A_ch4(xmx, tmx, zmx)		Retrieval	CH_4 Col. Averaging Kernel
co2_pvcd0(xmx, tmx, zmx)	molecules cm^{-2}	GGG 2020	Prior CO_2 Layer-by-layer VCD
A_co2(xmx, tmx, zmx)		Retrieval	CO_2 Col. Averaging Kernel
h2o_pvcd0(xmx, tmx, zmx)	molecules cm^{-2}	GGG 2020	Prior H_2O Layer-by-layer VCD
air_pvcd0(xmx, tmx, zmx)	molecules cm^{-2}	Retrieval	Prior Air Layer-by-layer VCD

Tshft(xmx, tmx)	K	Retrieval	Retrieved Temp. Prof. Shift
psurf(xmx, tmx)	hPa	Retrieval	Surface pressure
isrfsqz_w1(xmx, tmx)		Retrieval	CO_2 Window ISRF Squeeze Factor
isrfsqz_w2(xmx, tmx)		Retrieval	CH_4 Window ISRF Squeeze Factor
tropp(xmx, tmx)	hPa	Geos FP	tropopause pressure
granule_id(tmx)			Granule (30s) index counter
xch4_bias_corr_v2(xmx, tmx) – note 4		Retrieval	XCH_4 corrected for cross track bias.
xch4_bias_corr_v2:bias_model = "pls"			X. Track Bias Correction method
xch4_bias_corr_v2:pls_ncomp = 19LL			Number of Components used by PLS X. Track Bias Correction method

Notes:

xmx = cross track direction

1. Final position and pixel information uses GV avionics information plus satellite images to correct to ~ 100m pixel location accuracy.
2. xch4, xch4_bias_corr_v2 represent the column mean mole fraction of CH_4 relative to dry air.
3. Geometric single pixel is approximately 25 m along track x 5 m across track from 40,000 feet altitude. The point spread function is approximately 12m across, 35m along track. **NB**: **The data in the file aggregates 5 cross track pixels for a geometric pixel of approximately 25m x 25m from 40,000 feet.**
4. The derived xch4 has a stable cross-track pattern of bias associated with nanoscale inhomogeneities in the slit. Slow drift in this bias pattern is associated with temperature changes in the instrument. These are removed using PLS regression against the retrieved ISRF squeeze factors over the time course of each flight.
5. Standard deviation of the residual fit normed to the mean radiance at each pixel.

2.0 Instrument Description

MethaneAIR is a push broom airborne imaging instrument designed and built with support from MethaneSAT, LLC, a subsidiary of the Environmental Defense Fund. MethaneAIR detects CH_4 absorption around 1.65 µm and CO_2 absorption near 1.61 µm with one spectrometer. A second spectrometer is dedicated to detect the O_2 absorption in the $^1\Delta$ band around 1.27µm, to provide an additional light-path constraint for cloud screening or in a full-physics retrieval. Specifications are summarized in Table 2.

The data in this release is obtained using the CO_2-Proxy method. Here the CH_4 column mean dry mole fraction (XCH_4) is obtained by retrieving the vertical column densities (VCDs) of CO_2 and CH_4 using the 1.65µm band spectrometer. The spectral fit algorithm optimizes these on a 19-layer vertical grid assuming a non-scattering atmosphere. Scattering is accounted for indirectly by normalizing the CH4 VCD against CO2, whose a priori concentration is better known:

$$XCH_4 = XCO_2^{prior} \times VCD\_CH_4 / VCD\_CO_2.$$

The spectral fit algorithm uses molecular cross sections and apriori CH_4/CO_2 profiles from the GGG2020 [https://doi.org/10.5194/amt-2022-267]. The remaining a priori data are mostly from the GEOS-FP reanalysis. (https://gmao.gsfc.nasa.gov/GMAO_products/NRT_products.php). CO_2 and CH_4 VCDs are jointly retrieved in spectral windows **1595-1618** and **1629-1654nm**.

A geometric single pixel is approximately 25 m along track x 5 m across track from 40,000 feet altitude. The point spread function is approximately 12m across, 35m along track. **The data in the files aggregates 5 cross track pixels for a geometric pixel of approximately 25m x 25m from 40,000 feet.** The swath width is about 4.5 km

The standard deviation of this retrieval for a flat field of XCH_4 is approximately 40 ppb (for the 5x aggregated pixel). When gridded to 20m x 20m the standard deviation of a flat field is about 20 ppb, reflecting the oversampling in the spatial domain.

Table 2. Instrument Description

Optical and detector		Spectral and spatial	
Focal length (mm)	25	O_2 passband (nm)	1236–1319
F-number	3.5	O_2 dispersion (nm per pixel)	0.08
Entrance slit width (µm)	34	O_2 spectral FWHM (nm)	0.22
Optical transmittance (%)	≥ 37	CH_4 passband (nm)	1592–1680
Polarization sensitivity (%)	≤ 25	CH_4 dispersion (nm per pixel)	0.1
FPA dimensions (spectral × spatial pixels)	1024 × 1280	CH_4 spectral FWHM (nm)	0.3
Pixel pitch (µm)	12	Plate scale (° per pixel)	0.0275
Quantum efficiency below 1650 nm	> 0.7	Field of view (°)	23.7
Frame rate (Hz)	10	Swath width (km) at 12 km altitude	5.05
Readout noise (electrons), typical	35	Cross-track pixel (m) at 12 km altitude	5.76
Dark current (electrons s^{-1} per pixel)	< 8500	Along-track pixel (m)	≈ 25

Source: Staebell, C., Sun, K., Samra, J., Franklin, J., Chan Miller, C., Liu, X., Conway, E., Chance, K., Milligan, S., and Wofsy, S.: Spectral calibration of the MethaneAIR instrument, Atmospheric Measurement Techniques, 14, 3737–3753, https://doi.org/10.5194/amt-14-3737- 2021, publisher: Copernicus GmbH, 2021.

3. Data Collection and Processing

Data from the imaging spectrometer were collected on board in real time at 10 Hz. The data were uploaded to the cloud after landing and sent to the Harvard large cluster. Avionics data (1 Hz and 25 Hz) from the GV navigational instrumentation was uploaded to the NCAR server and then downloaded to the Harvard cluster.

Data processing steps:

Level 0: apply calibration equations to convert counts into electrons.

Level 1: apply radiometric calibrations to convert Level 0 data to spectra (radiance as a function of wavelength for each pixel on the focal plane array) (Level 1a); perform geometric analysis to map the spectra to pixels on the ground, fine-tuned using high resolution satellite data collected as close in time as possible in our wavelength bands. The resulting geolocated spectra comprise product Level 1b.

Level 2: Carry out the CO_2 proxy retrieval described in the Instrument Description and notes to Table 1. The resulting Level 2 product provides georeferenced XCH4 with approximately 20m x 20m spatial resolution on the ground. See Table 1 for description of the variables derived in the analysis. **Level 2 files for all research flights are provided in this data release.**

Level 3: The Level 2 data are gridded to 20m x 20m pixels using the "snowflake" algorithm of Kang Sun (https://doi.org/10.5194/amt-11-6679-2018), which accounts for the spatial oversampling by this imaging spectrometer. Level 3 data are screened for clouds and low radiance (e.g. water bodies). **Level 3 files will be provided in a subsequent data release.**

4.0 Data Format

Structure

The data are provided in netCDF format, one per flight for Level 2, and one for Level 3.

Naming

The file name provides the date of the flight (also given within the file).

Parameters

Level 2: Table 1 (above) gives the parameter names, units, definitions, and related information.

Level 3: The gridded dataset is divided into non-overlapping segments since the GV can pass over one spot on the ground many times during a flight. The structure of the Level 3 data is given in Table 3.

Table 3. Parameters and net CDF structure of Level 3 (gridded, 20m x 20m) data

netcdf stacked_segments_my_target {
dimensions:
x = 5428 ; (example)
y = 5500 ;
segment = 10 ;
variables:
double xmid(x) ;
double ymid(y) ;
float xmesh(segment, x, y) ;
float ymesh(segment, x, y) ;
float sza(segment, x, y) ;
float vza(segment, x, y) ;
float aza(segment, x, y) ;
float alb0(segment, x, y) ;
float xch4(segment, x, y) ;
float xch4_0(segment, x, y) ;
float xco2_0(segment, x, y) ;
float h2o_vcd(segment, x, y) ;
float rms(segment, x, y) ;
float ch4_dofs(segment, x, y) ;
float co2_dofs(segment, x, y) ;
float psurf0(segment, x, y) ;
float tau(segment, x, y) ;
float xch4_bias_corr_v2(segment, x, y) ;
float isrfsqz_w1(segment, x, y) ;
float isrfsqz_w2(segment, x, y) ;
double slope(x, y) ; note 1
double intercept(x, y) ; note 1
double rval(x, y) ; note 1
double first_tau(x, y) ; note 1
short nvalid(x, y) ; note 1
}

All parameters have the same definitions as in Level 2 (See Table 1).
Notes
1. Additional parameters at Level 3:
Results of the least-squares fit of **xch4_bias_corr_v2** vs **time** (tau) for those grid squares imaged more than once on a flight:

slope = slope of the fit (units: XCH4 per hour)
intercept = XCH4 at time first_tau
first_tau = time of the first valid point in the time series (units: hours since 1/1/1985).
nvalid = number of valid time points in the time series