User Guide for Observations for CAMx evaluation with MELODIES MONET

Last Updated: May 2025

This guide includes information on observational data sets that have been tested to work with CAMx model output to date. Example yaml files for comparing CAMx to the different types of observations are available in the NSF NCAR ACOM MELODIES MONET repository (https://github.com/NCAR/MELODIES-MONET-ACOM/tree/SIP_Colorado/sip_specific_tools)

Surface Observations

AirNow

When a user installs the Python package melodies-monet, it comes with a command line tool of the same name: melodies-monet. This tool can be used to download hourly Airnow data from the U.S. EPA by issuing the following command:

```
melodies-monet get-airnow -s 'YYYY-MM-DD' -e 'YYYY-MM-DD'
```

This command will download the data and convert into the format (NETCDF) to be used directly in MELODIES MONET. AirNow data include key air pollutants (e.g. ozone, PM2.5, PM10, NO2, NO, NOx, NOy, CO) and also meteorological variables (e.g. winds, temperature, solar radiation).

BoulderAir

BoulderAir has been collecting air quality measurements at various sites in the Colorado Front Range since 2017 and include some key criteria pollutants (e.g. ozone, NOx), VOCS (including benzene, toluene, ethane, propane, etc.) and also some meteorological measurements. The data are available to CDPHE, for other users a data sharing agreement is needed. The raw data comes in .csv files. Not all parameters are measured at all sites and the content of the data files, the temporal resolution and also the filename convention varies for different sites and also times. Site information is not included in the data files but BoulderAir has provided a separate text file with metadata.

BoulderAir data need to be pre-processed to be consistent with MELODIES-MONET's input format. This can be done by using the BoulderAir preprocessor, as follows:

```
./boulderair.py -c coordinates_file.csv -p path_to_data_XYZ_etc.csv
-v variable1,variable2,...,variablen -r resample_freq -m method -o
output_name.nc
```

The preprocessor will query all selected file names (using wildcard is enabled) and create a NETCDF file with the desired variables that has a format similar to the one used for AirNow thus is ready for use in MELODIES MONET.

Given the limited access of the data, this preprocessor is not available in the public MELODIES MONET repository but only in the password protected ACOM repository (https://github.com/NCAR/MELODIES-MONET-ACOM/tree/SIP_Colorado/sip_specific_tools)

VOC Canister Samples

VOC canister sampling data are accessible from the CDPHE website (https://www.colorado.gov/airquality/air_toxics_repo.aspx#o3). The measurements represent an

average over a 3-hour collecting window (typically 6-9am) with the collection frequency changing per site. Data are available for CAMP and Platteville going back to 2011 and for Greeley and Brighton and Greeley since 2019 and 2021, respectively.

The data are reported in Excel files with individual tabs for different measurement sites and compound groups (non-methane VOCs, methane, carbonyls). Figure 1 provides a snapshot of the Excel data file for 2016 which demonstrates the complexity of the data file. The reader developed for MELODIES MONET reads directly the Excel spreadsheet and extracts the selected variables without the need for a user to preprocess the data. The MELODIES MONET yaml file, however, requires a user to set the number of lines representing the header since this can vary across tabs, and also the latitude and longitude information for the sites. Any desired quality flag filtering can be defined in the YAML file.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Colorado Department of Public Health & Environment - Air Pollution Control Division**												
2	**2016 Ozone Precursor Study - NMOCs**						****** Raw Data from ERG (Jan-Jun), AAC (Sep-Dec) ******						
3	**Jan. 2016 - Dec. 2016 (3-hr. samples, 6a.m. - 9a.m. MST)**							***(NOTE: No data for Jul-Aug)***					
4									U	Under Detection Limit			
5									J	Analyte was detected between the SRL and the SDL.			
6									CE	Not reportable due to a co-eluting compound.			
7	***DECO = Denver-CAMP site, 2105 Broadway***								A-01	Concentration determined by EPA Method TO-15			
8		***(-D1 = Primary, -D2 = Collocated)***							R-F	Replicate exceeds DQO criteria.			
9													
10		**Lab sample**	**Sample**	**Prep**	**Analysis**		**CAS**			**Detection**			
11	**Site**	**ID**	**Date**	**Date**	**Date**	**Analyte**	**Number**	**Result**	**Anote**	**Limit**	**Units**		
12													
13	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1,2,3-Trimethylbenzene	526-73-8	0.345		0.222	ppbC		
14	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1,2,4-Trimethylbenzene	95-63-6	2.06		0.381	ppbC		
15	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1,3,5-Trimethylbenzene	108-67-8	0.509		0.228	ppbC		
16	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1,3-Butadiene	106-99-0	0.331		0.109	ppbC		
17	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Butene	106-98-6	ND	CE, U	0.101	ppbC		
18	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Decene	872-05-9	ND	U	0.381	ppbC		
19	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Dodecene	112-41-4	0.454	U	0.804	ppbC		
20	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Heptene	592-76-7	ND	U	0.094	ppbC		
21	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Hexene	592-41-6	0.308		0.105	ppbC		
22	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Nonene	124-11-8	0.256		0.143	ppbC		
23	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Octene	111-66-0	0.268		0.188	ppbC		
24	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Pentene	109-67-1	0.584		0.084	ppbC		
25	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Tridecene	2437-56-1	ND	U	0.449	ppbC		
26	DECO	6010706-01	1/1/16	1/1/16	1/19/16	1-Undecene	821-95-4	ND	U	0.269	ppbC		
27	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,2,3-Trimethylpentane	564-02-3	0.335		0.131	ppbC		
28	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,2,4-Trimethylpentane	540-84-1	0.963		0.122	ppbC		
29	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,2-Dimethylbutane	75-83-2	0.255		0.128	ppbC		
30	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,3,4-Trimethylpentane	565-75-3	0.497		0.130	ppbC		
31	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,3-Dimethylbutane	79-29-8	0.387		0.098	ppbC		
32	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,3-Dimethylpentane	565-59-3	1.06		0.141	ppbC		
33	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2,4-Dimethylpentane	108-08-7	0.701		0.137	ppbC		
34	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2-Ethyl-1-butene	760-21-4	ND	U	0.094	ppbC		
35	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2-Methyl-1-butene	563-46-2	0.402		0.119	ppbC		
36	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2-Methyl-1-pentene	763-29-1	ND	U	0.126	ppbC		
37	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2-Methyl-2-butene	513-35-9	0.424		0.108	ppbC		
38	DECO	6010706-01	1/1/16	1/1/16	1/19/16	2-Methylheptane	592-27-8	0.428		0.199	ppbC		

Sites_info | NMOC-DECO | NMOC-PVCO | Methane-DECO | Methane-PVCO | Carbonyl-DECO | Carbonyl-PVCO

Figure 1: Example Excel file showing the content of the VOC measurements for 2016 at the Platteville site.

Remote Sensing Observations

Retrievals of atmospheric composition from ground-based or space-borne remote sensing instruments are highly dependent on the specific instrument and assumptions about the physical state of the atmosphere at the time of observation. Compared to in-situ measurements they do not represent a direct measurement of the species of interest but retrieve information on the concentrations from remotely sensed radiances. Thus when using remote sensing products have to make sure to account for the measurement sensitivity and information content and accordingly use information on the dependence of the measurement sensitivity with altitude as represented by either airmass factors or averaging kernels. The accuracy of satellite data is

usually less good than from ground-based remote sensing data because corrections need to be made for contributions from the surface and from other atmospheric constituents contributing to the reflected solar radiation measured by the instrument at the top of the atmosphere (TOA). A paper on the methodologies used on MELODIES MONET for comparing models to different satellite products is in preparation for submission to the journal Geoscientific Model Development (GMD)

PANDORA

The ground-based remote sensing PANDORA instrument (https://pandora.gsfc.nasa.gov) is a UV–VIS spectrometer which retrieves the atmospheric column amounts for ozone, HCHO and NO2 from direct sun measurements. Data can be downloaded from the open-access Pandonia Global Network website (https://data.ovh.pandonia-global-network.org/). The data are organized by site and a user should download the Pandora*.txt files in the ./L2 directory under each site.

For Colorado there are two PANDORA instruments available. Both are located in Boulder, one at CU Boulder with measurements since 2018 and the other at NSF NCAR with measurements available since 2021. File name convention can be found in the Data Product README (chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.pandonia-global-network.org/wp-content/uploads/2025/01/PGN_DataProducts_Readme_v1-8-10.pdf)

The wget command can be used to download data for the two Boulder stations:

```
wget
https://data.ovh.pandonia-global-network.org/BoulderCO-NCAR/Pandora204s1/L2/Pandora204s1_BoulderCO-NCAR_L2_rnvh3p1-8.txt
wget
https://data.ovh.pandonia-global-network.org/BoulderCO/Pandora57s1/L2/Pandora57s1_BoulderCO_L2_rnvh3p1-8.txt
```

TEMPO

The Tropospheric Emissions: Monitoring of Pollution (TEMPO) instrument aboard the Intelsat-40e (IS-40e) commercial satellite is a UV-visible spectrometer and the first space-based instrument to retrieve atmospheric trace gases across North America every hour during daytime (https://tempo.si.edu/). TEMPO captures high-resolution data on ozone, aerosols, nitrogen dioxide and other gases in the atmosphere. Currently available products for tropospheric air quality include the total column amounts of NO2 and HCHO. The geostationary orbit scans most of North America, extending from just south of Mexico city and the southern section of Canada, from the Atlantic to the Pacific, every hour and with a nominal spatial resolution of 2 x 4.75 km^2. Data are available from August 2023 onwards.

MELODIES MONET is set up to compare CAMx to TEMPO Level 2 NO2 and HCHO column data and to the ratio HCHO/NO2.

Tempo data are available in NETCDF format and in addition to the retrieved quantity amounts also include ancillary information needed for comparing to models such as the a priori profiles, cloud fraction, solar zenith angle and airmass factors.

The TEMPO L2 data can be downloaded from NASA EARTHDATA (https://search.earthdata.nasa.gov/search) using the NASA EarthAccess library (https://earthaccess.readthedocs.io). A python download script that utilizes this library is

provided in our ACOM repository (https://github.com/NCAR/MELODIES-MONET-ACOM/blob/SIP_Colorado/sip_specific_tools/scripts/download_tempo.py).
To use the script a user needs to get a login to EarthAccess and have EarthAccess installed. The latter can be done by either using pip (`pip install earthaccess`) or conda (`conda install -c conda-forge earthaccess`). A user will enter their desired time period into the script and the script automatically downloads all available data.

TROPOMI

The TROPOspheric Monitoring Instrument (TROPOMI), onboard the polar orbiting Sentinel-5 Precursor mission, was launched in October 2017. It provides daily retrievals of atmospheric composition with a high signal-to-noise ratio and a relatively high spatial resolution (3.5 × 7 km² at nadir, improved to 3.5 × 5.6 km² since August 6, 2019). TROPOMI provides daily near-global coverage. The current implementation of MELODIES MONET enables comparing models to Level 2 NO2 tropospheric column retrievals. TROPOMI data are accessible via the Sentinel-5P data portal (https://s5phub.copernicus.eu/dhus/#/home) and also the NASA EARTHDATA (https://search.earthdata.nasa.gov/search).
The same script that is used for TEMPO can also be used for downloading TROPOMI data from NASA EARTHDATA. The only change a user has to make is to set the name_key entry to:

```
name_key: str = 'S5P_L2__NO2____HiR'
```

MOPITT

Measurements Of Pollution In The Troposphere (MOPITT) is an instrument aboard the Terra satellite, launched in 1999. MOPITT has an overpass time of ~10:30 LT and achieves global coverage in about 3 days. It provides information on tropospheric CO concentrations at 22 km resolution at nadir. CO is retrieved as total column or profile volume mixing ratios on 10 layers. Averaging kernels and a priori are supplied with the CO retrievals. The most recent retrieval version is version 9. The use of the daytime joint near-infrared/thermal infrared (NIR/TIR) product is recommended as it includes the most retrieval information and increased sensitivity closer to the ground. In MELODIES MONET, the MOPITT daily and monthly Level 3 data can be used in MELODIES MONET.
Data are available from 2000 through 2024 (doi: 10.5067/TERRA/MOPITT/MOP03J.009) and can be searched and downloaded from that link. The same script that is used for TEMPO can also be used for downloading MOPITT data from NASA EARTHDATA. The only change a user has to make is to set the name_key entry to:

```
name_key: str = 'MOP03J'
```

Note: the MOPITT data seem to be currently unavailable with the aforementioned script, which uses the earthaccess tool. Although they are expected to be restored, the timeline for that is unclear. The suggested workaround is to download them manually from https://asdc.larc.nasa.gov/project/MOPITT . This is specific for MOPITT data and the download script should work for any other dataset available from NASA's Earthdata platform.