

# Making Satellite Datasets Accessible for Everyone

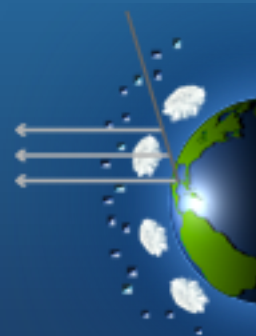
A look into my NASA Internship - summer 2015

Aaron Scott

University of North Dakota

# NASA Internship

- NASA/Univ. of Virginia HPC workshop
- Langley Research Center - Hampton, VA
- Atmospheric Science Data Center (ASDC)



**Atmospheric  
Science  
Data Center**



## Atmospheric Science Data Center

# Overview

- “Responsible for processing, archival, and distribution of NASA Earth science data”
  - Radiation budget, clouds, aerosols, and tropospheric chemistry
- > 45 projects
- > 1700 archived data sets
- > 3 Petabytes (PB) of data



## Atmospheric Science Data Center

# Support

- NASA operates several different satellite missions
  - CALIPSO
  - CERES
  - MISR
  - MOPITT
- ASDC works with science teams
  - Provide data support and Processing
  - Archiving and subsetting



## Atmospheric Science Data Center

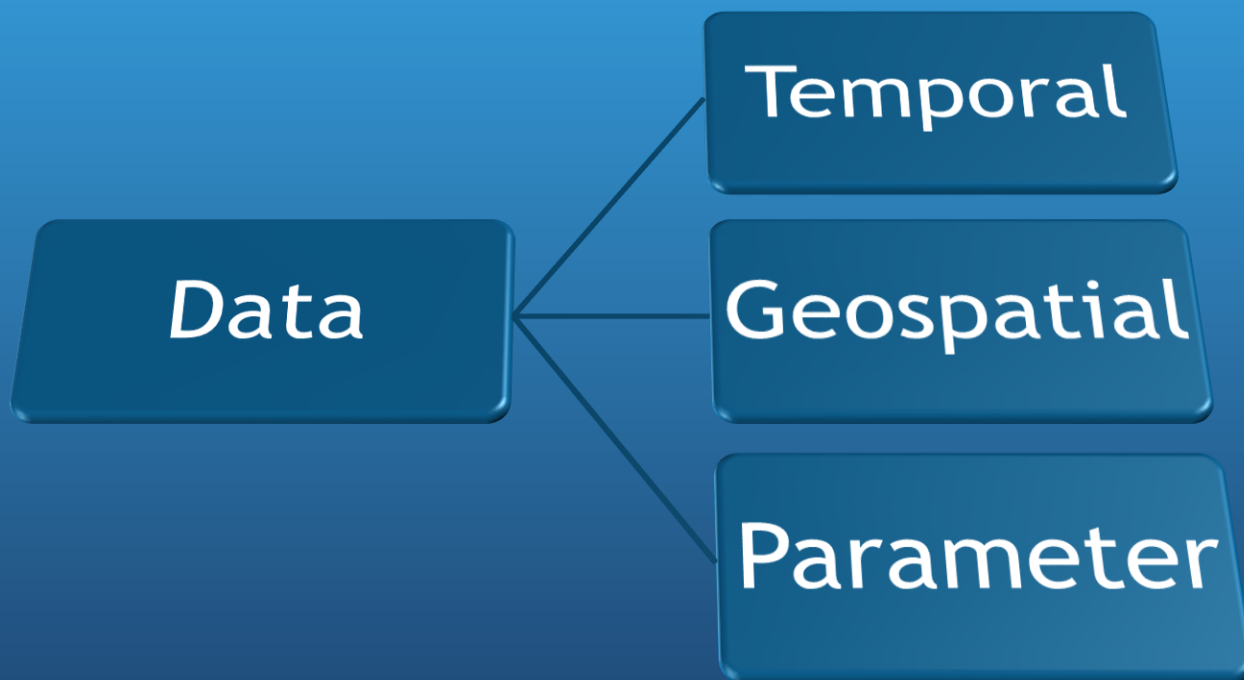
# Datasets

- Satellite datasets are inherently large
  - Data archives are growing daily
- Researchers need accessibility
- Let the scientists do science!
  - Analyze data not process
- Researchers only want necessary data



**Atmospheric  
Science  
Data Center**

# Subsetting





**Atmospheric  
Science  
Data Center**

# Subsetting

- Subset - a set that is part of a larger set





**Atmospheric  
Science  
Data Center**

# Subsetting

- Subset - a set that is part of a larger set







**Atmospheric  
Science  
Data Center**

# Subsetting

- Subset - a set that is part of a larger set





**Atmospheric  
Science  
Data Center**

# Making Data Useful

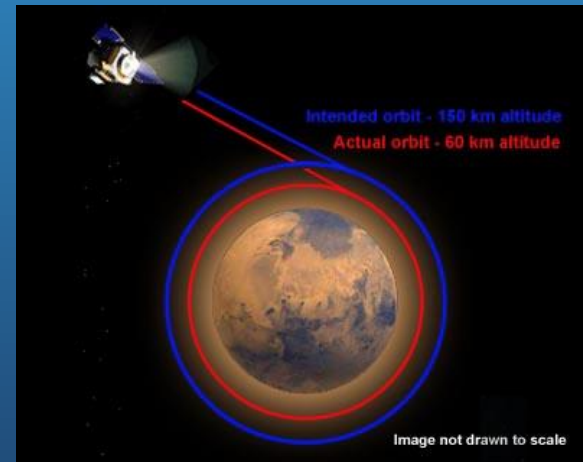
- Data collection without communication is ... pointless?
- Data needs to be in some format that is useful
- Standard names
- Standard dimensions
- Units



## Atmospheric Science Data Center

# Making Data Useful

- And units are important!
- Mars Climate Orbiter
- Navigation error
  - English and metric units
  - 60 miles off trajectory



Source:  
[http://personal.victoria.ac.nz/stephen\\_marshall/SE/Failures/SE\\_MCO.html](http://personal.victoria.ac.nz/stephen_marshall/SE/Failures/SE_MCO.html)



## Atmospheric Science Data Center

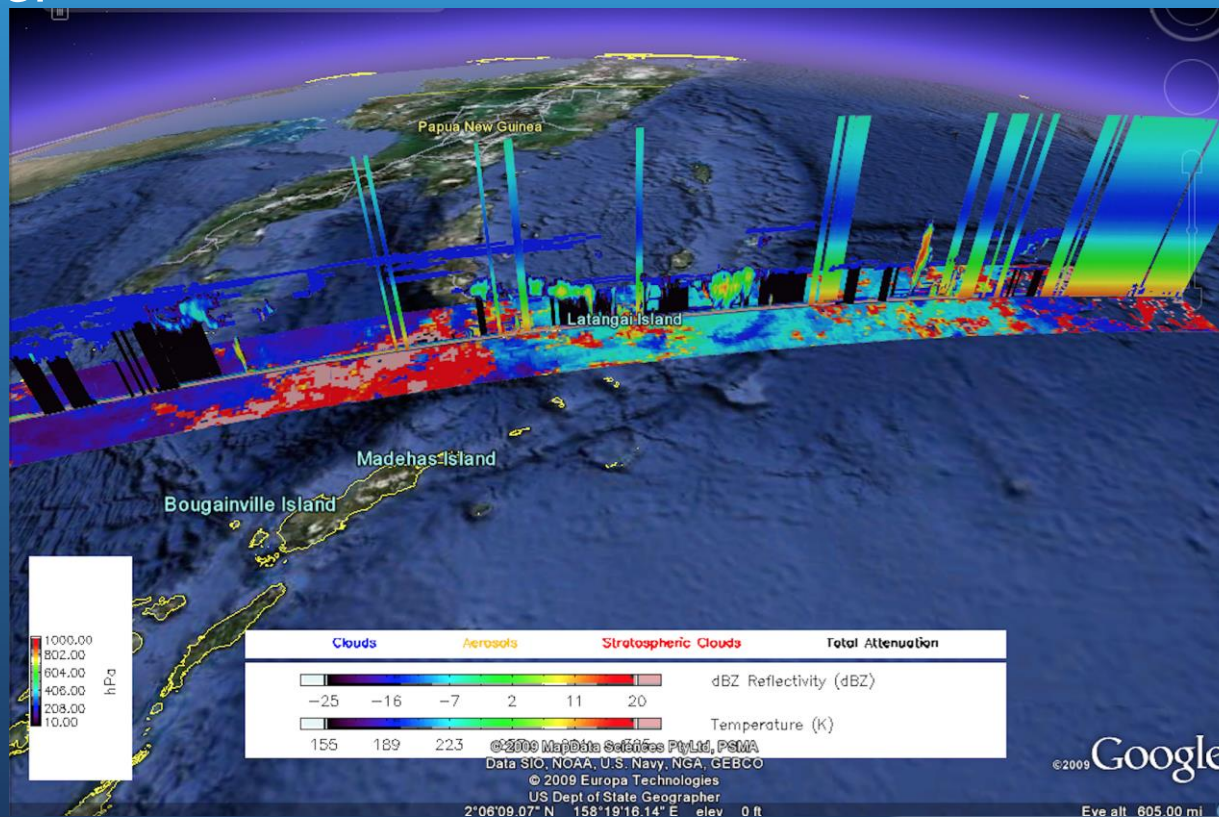
# Making Data Useful

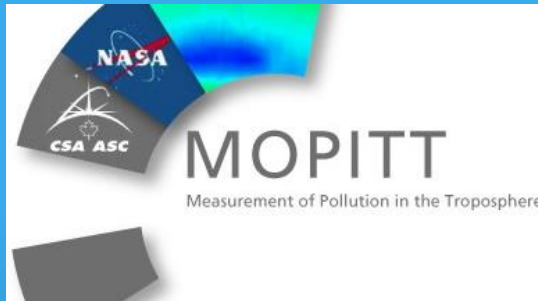
- CF (Climate and Forecast) Compliance
- Metada standards
  - Data that tells us about data (units, etc.)
- Standard names
- Standard Dimensions

# Atmospheric Science Data Center

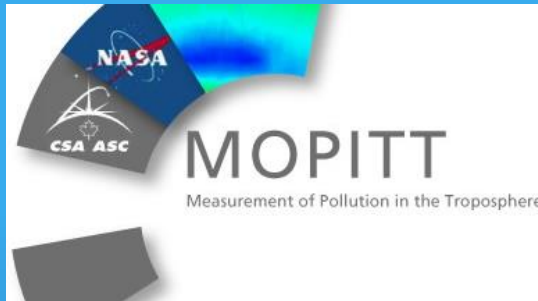
## Making Data Useful

- Data compliance makes using data and visualization easier

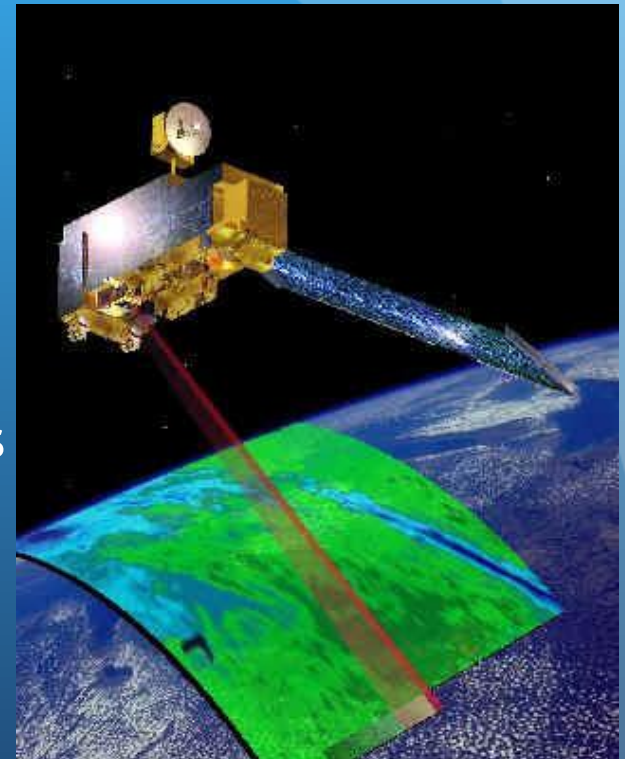




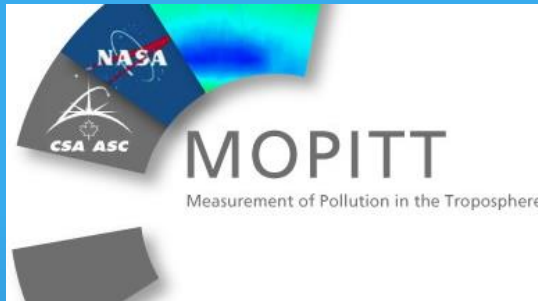
- Measurement of Pollution in the Troposphere
- Launched on TERRA in 1999
- Canadian Space Agency (CSA)
- National Center for Atmos. Research (NCAR)/NASA
- Measures carbon monoxide and methane



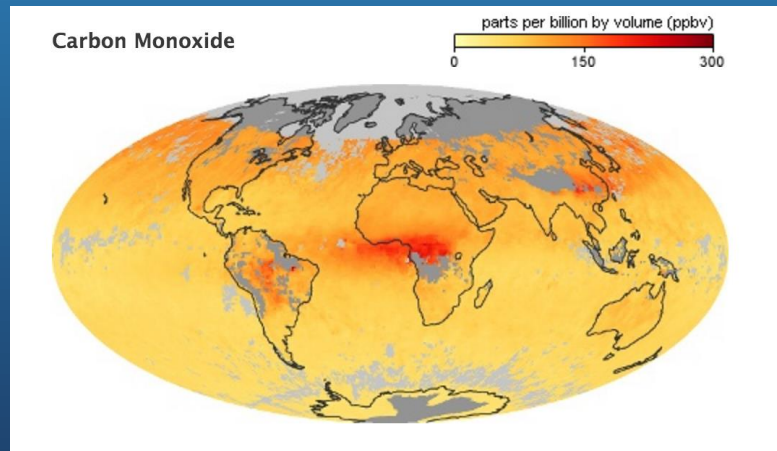
- Carbon monoxide profiles
  - Horizontal Resolution = 22 km
  - Vertical Resolution = 3 km
- High enough resolution to track the gas to certain cities or other sources
- Swath = 640 km wide



Source:  
<http://www.atmosp.physics.utoronto.ca/MOPITT/home.html>

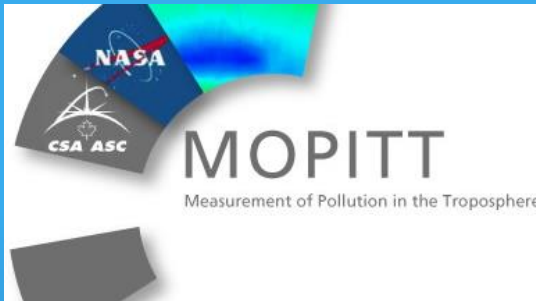


- Carbon monoxide is a colorless, odorless, and poisonous gas
- Burning of fossil fuels
- Leads to atmospheric conditions such as smog



MOPITT Monthly Average for December 2015





### MOPITT Spatial Subset with Day-only Solar Zenith Angle Filter

Data granule:

MOP02T-20130811-L2V10.1.1.prov.hdf

Bounding Box (lon-lat pairs):

POLYGON((43 -12, 43 -26, 51 -26, 51 -12, 43 -12))

The original data granule (24 hours of data) contains 193192 observations.  
The subsetted result file contains 1265 observations.





**Atmospheric  
Science  
Data Center**

# HDF5/NetCDF4

- Hierarchical Data Format (HDF)
- Network Common Data Form (NetCDF)
- Common data structures used in the atmospheric sciences



## Atmospheric Science Data Center

### MOPITT Subsetter Algorithm

- MOPITT data subsetter at ASDC was upgraded to incorporate HDF5/NetCDF4
- Algorithm follows CF compliance
- Python code used as development language



# Atmospheric Science Data Center

## MOPITT Subsetter Algorithm

Data

- Not CF
- Large file



Algorithm

- Python code reads in xml order file



Data Subset

- CF
- Smaller file

# MOPITT Subsetter Algorithm

## Step 1: Select a MOPITT data product and parameter(s)

The data subset options include being able to select data product types and choose a down-selection of parameters to extract. Data parameters have been grouped together based on type and are listed in the left most window.

If you choose not to select additional parameters, then a default package is put into the resulting output file. This default package includes parameters of interest to most MOPITT users including geolocation information, retrieved CO products, a priori CO profiles and retrieval averaging kernels. Any selected additional parameter grouping(s) are shown in the center window. The far right window displays the complete listing of the data parameters that will be included in your output file.

Users wanting all data variables in the file must select all parameter groupings.

This subsetting tool currently operates on MOPITT Version 5 Level 2 products.

In many situations, characteristics of MOPITT daytime and nighttime products are quite different. Over land, retrieval quality is typically greater for daytime overpasses than nighttime overpasses. Users may subset MOPITT daytime and nighttime products using the Solar Zenith Angle Filter.

Choose Solar Zenith Angle Filter:

Day Night **Both**

Choose Data Product:

MOPITT V5 Derived CO (Near and Thermal Infrared Radiances)

Choose Parameter Group(s):

Retrieved Byproducts  
Scene Information  
Cloud Diagnostics  
Auxiliary Diagnostics  
Measured Radiances

⋮ Add +  
⋮ Add all +  
+ Remove ⋮  
+ Remove all ⋮

Selected Parameter Group(s):

Data variables included in the output file:

Pressure Grid  
Surface Pressure  
Retrieved CO Mixing Ratio Profile  
Retrieved CO Surface Mixing Ratio  
Retrieved CO Total Column  
A Priori CO Mixing Ratio Profile  
A Priori CO Surface Mixing Ratio  
Solar Zenith Angle  
Retrieval Averaging Kernel Matrix  
Degrees of Freedom for Signal

- Web based ordering platform
- It's like "take out" for scientists!

# MOPITT Subsetter Algorithm

## Step 2: Select a temporal range (optional)

Use the temporal options to narrow your search to a specific temporal domain. If you do not make a temporal selection, the default is to search the complete range of time in which the satellite has acquired data. If you limit your search to a specific time domain, the search will return all available data that intersect with your selected time range.

### Calendar dates

#### Calendar dates

The MOPITT instrument began operations on March 3, 2000. The dates reflected in the calendar will represent the range in which the satellite has been in data acquisition mode.

To change the date range, you may either click on the text area and select a predefined range from the drop-down menu or enter your own date with the format "yyyy-mm-dd".

From  to

# MOPITT Subsetter Algorithm

## Step 3: Select a geospatial range (optional)

Use the geospatial options to narrow your search to a specific geospatial area. If you do not make a geospatial selection, the default is to search the whole globe. If you limit your search to a specific area, the search will return all available data that intersect with your selected area.

### User-defined bounding box

#### User-defined bounding box

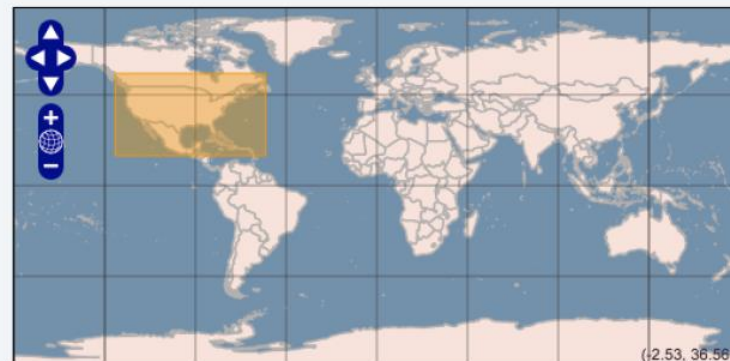
Modify the geospatial fields to specify your area of interest, or use your mouse directly on the map to draw a bounding box by clicking and dragging. The map uses latitude/longitude bounds (north, south, east, and west) to define the area of a box. If you use the mouse to draw the area on the map, the fields are filled in automatically, based on the box drawn.

Latitudes and longitudes are in Decimal Degrees (DD) format.

Use: '+' for north latitudes or east longitudes; use '-' for south latitudes or west longitudes. Example: +40.68, -74.04

To cross the anti-meridian, left must be greater than right. Example: (left) +148.64, (right) -115.73

Top  
 55.54687  
 Left -129.792      Right -55.2612  
 Bottom  
 14.76562



# MOPITT Subsetter Algorithm

File Window Tools Help

Recent Files /Users/akscott1/asdc\_data/mopitt/mop02J/MOP02J-20130901-L2V Clear Text

MOP02J-20130901-L2V16.2.3.he5

- HDFEOS
  - ADDITIONAL
  - SWATHS
    - MOP02
      - Data Fields
        - RetrievedCOTTotalColumn

	0	1
0	1.365608...	1.242986...
1	1.364728...	7.778927...
2	1.422757...	1.248851...
3	1.392825...	8.201487...
4	1.412119...	9.776761...
5	1.799319...	1.843219...
6	1.457688...	1.036670...
7	1.763774...	1.554489...
8	1.386726...	7.414713...
9	1.236879...	1.305489...
10	1.416555...	8.412195...
11	1.368372...	7.413908...
12	1.354735...	6.218542...
13	1.389725...	6.227862...
14	1.419861...	6.653859...
15	1.409381...	6.266876...
16	1.311287...	5.573917...
17	1.613217...	1.323089...
18	1.465825...	1.132263...
19	1.290484...	5.048060...
20	1.325890...	5.724920...
21	1.378089...	9.261203...
22	1.300528...	9.679804...
23	1.315486...	7.219176...
24	1.245209...	1.201353...
25	1.230626...	8.598101...
26	1.277704...	6.204622...
27	1.113296...	4.950213...

RetrievedCOTTotalColumn (47096, 12)  
32-bit floating-point, 191756 x 2  
Number of attributes = 2  
Units = mol/cm<sup>2</sup>  
\_FillValue = -9999.0

Log Info Metadata

File Window Tools Help

Recent Files tt1/mopitt\_python\_output/MOP02J-20130901-L2V16.2.3.he5\_Subset Clear Text

MOP02J-20130901-L2V16.2.3.he5\_Subset.hdf5

- HDFEOS
  - ADDITIONAL
  - SWATHS
    - MOP02
      - Data Fields
        - RetrievedCOTTotalColumn
      - Geolocation Fields

	0	1
0	1.365608...	1.242986...
1	1.364728...	7.778927...
2	1.422757...	1.248851...
3	1.392825...	8.201487...
4	1.412119...	9.776761...
5	1.799319...	1.843219...
6	1.457688...	1.036670...
7	1.763774...	1.554489...
8	1.386726...	7.414713...
9	1.236879...	1.305489...
10	1.416555...	8.412195...
11	1.368372...	7.413908...
12	1.354735...	6.218542...
13	1.389725...	6.227862...
14	1.419861...	6.653859...
15	1.409381...	6.266876...
16	1.311287...	5.573917...
17	1.613217...	1.323089...
18	1.465825...	1.132263...
19	1.290484...	5.048060...
20	1.325890...	5.724920...
21	1.378089...	9.261203...
22	1.300528...	9.679804...
23	1.315486...	7.219176...
24	1.245209...	1.201353...
25	1.230626...	8.598101...
26	1.277704...	6.204622...
27	1.113296...	4.950213...

RetrievedCOTTotalColumn (21928576, 14)  
32-bit floating-point, 18062 x 2  
Number of attributes = 2  
Units = mol/cm<sup>2</sup>  
\_FillValue = -9999.0

Log Info Metadata

- View of data before and after passed through algorithm
- HDFView software





**Atmospheric  
Science  
Data Center**

# Conclusions

- Making data more accessible requires:
  - Communication
  - Collaboration
  - Governance
- The efforts of the ASDC and similar organizations will continue to work to provide accessible data that can be used by everyone
- Furthering science starts with proper data management

# Acknowledgements

- Walt Baskin - NASA Mentor
- ASDC
- ND Space Grant Consortium - THANK YOU!