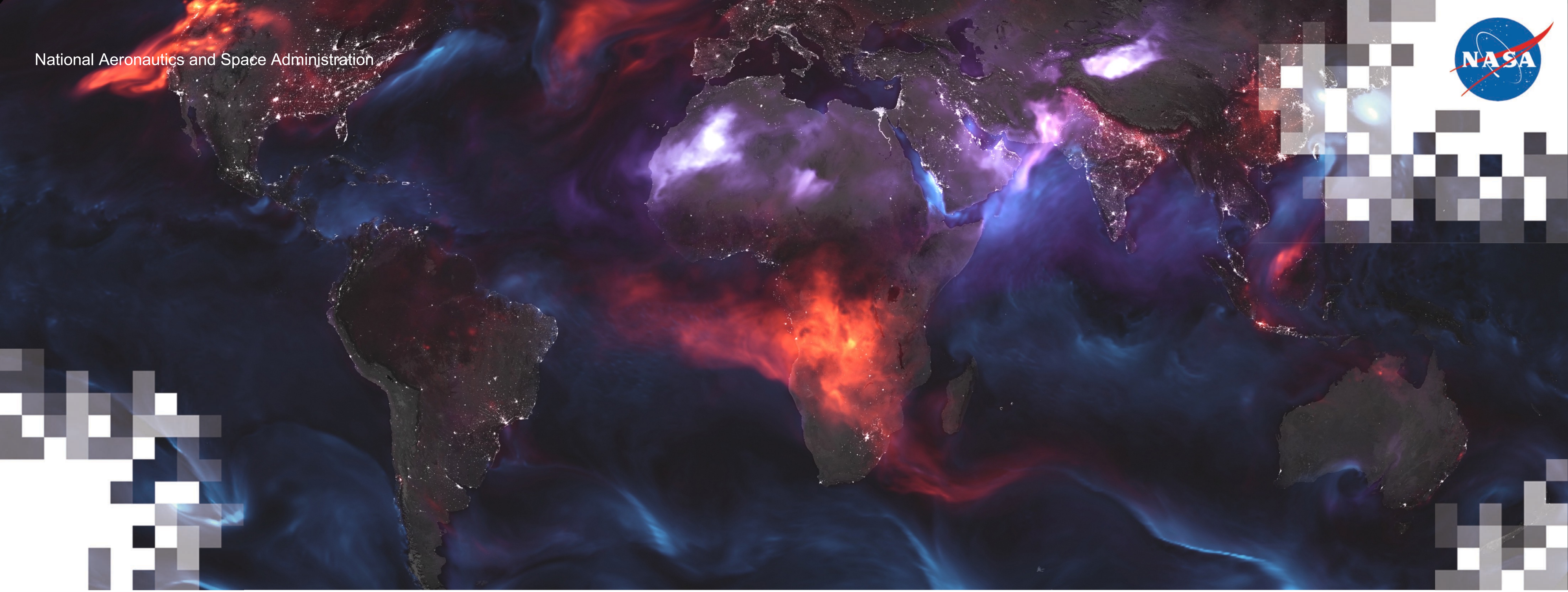


Introduction to Python Tools for Visualization and Analysis

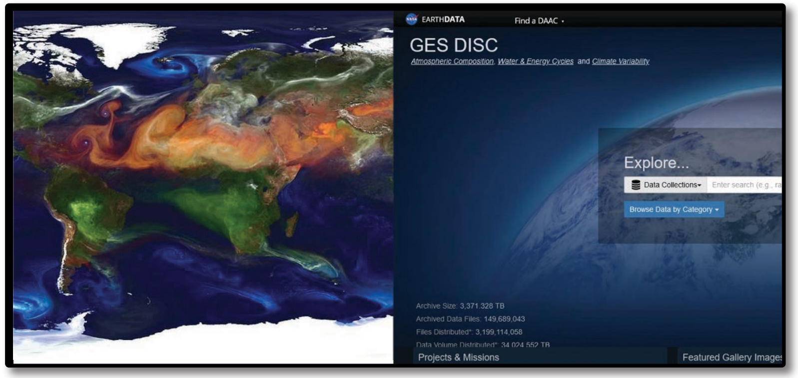
Tools for Analyzing NASA Air Quality Model Output

Pawan Gupta, Melanie Follette-Cook, Sarah Strode

February 24, 2020

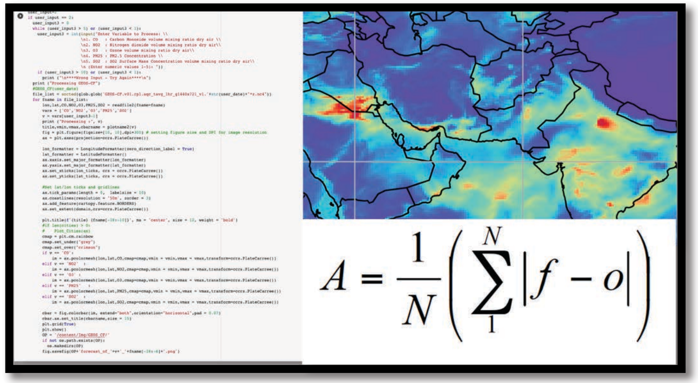
Training Outline

Part 1: February 22, 2022



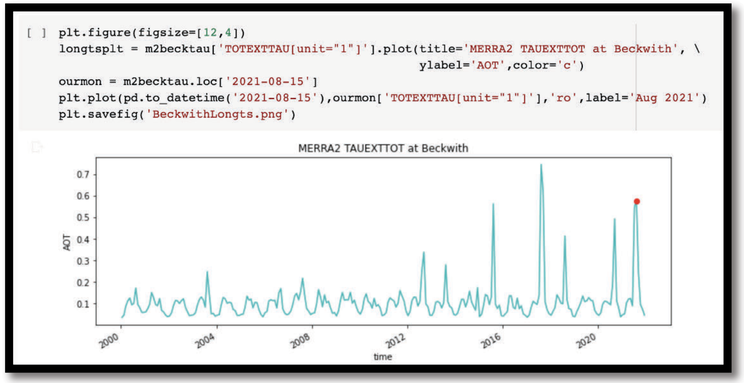
Review of NASA Air Quality Forecasts and Reanalysis

Part 2: February 24, 2022



Introduction to Python Tools for Visualization and Analysis

Part 3: March 1, 2022



Interpreting Model Output for Air Quality Assessment

Sarah Strode

Melanie Follette-Cook

Pawan Gupta

Learning Objectives

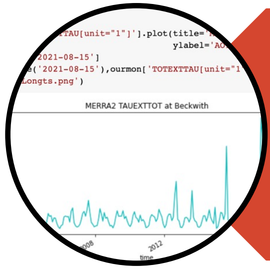
By the end of this session participants will be able to:

- Download, read, subset, and map GEOS-FP, GEOS-CF, and MERRA2 output using python scripts
- Extract model output at a given ground location
- Save the model output in a .csv file
- Learn spatial and temporal collocation methods and examples of comparing model output with satellite observations
- Learn methods of validation with surface observations

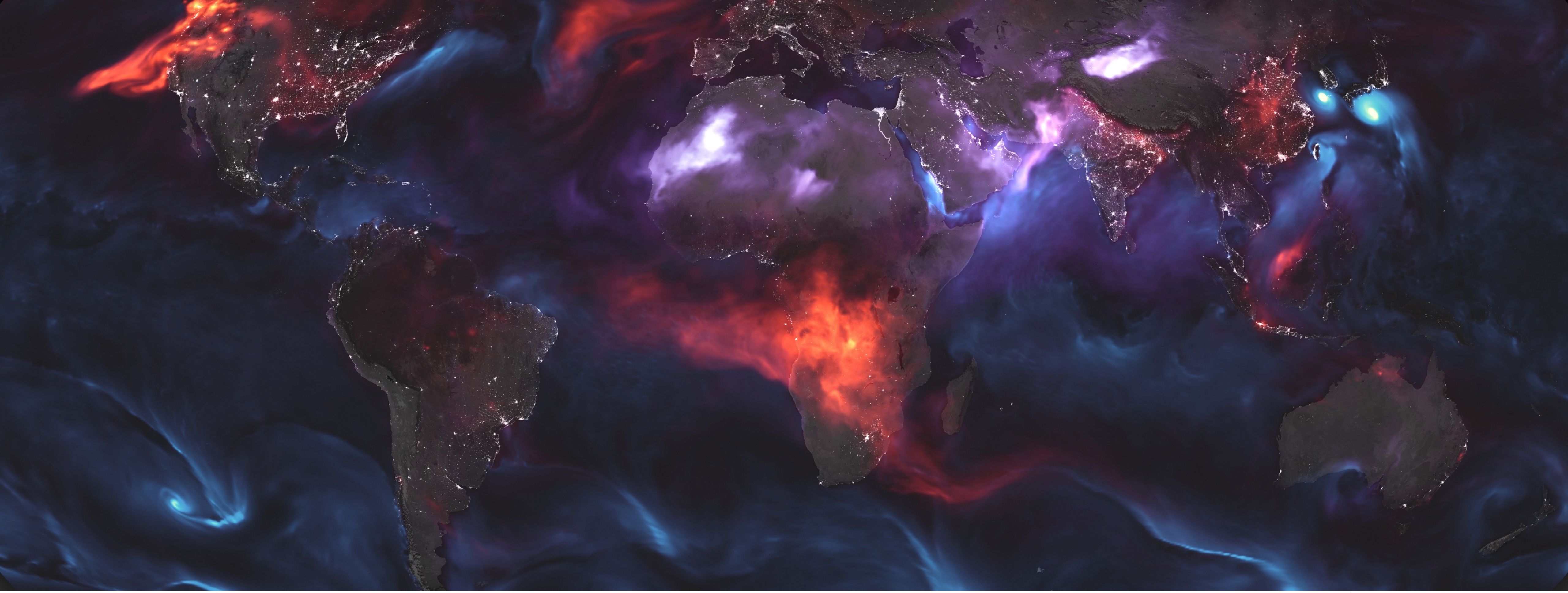
Before We Start - Prerequisites

Google Account

Earthdata Login



Download Data & Codes



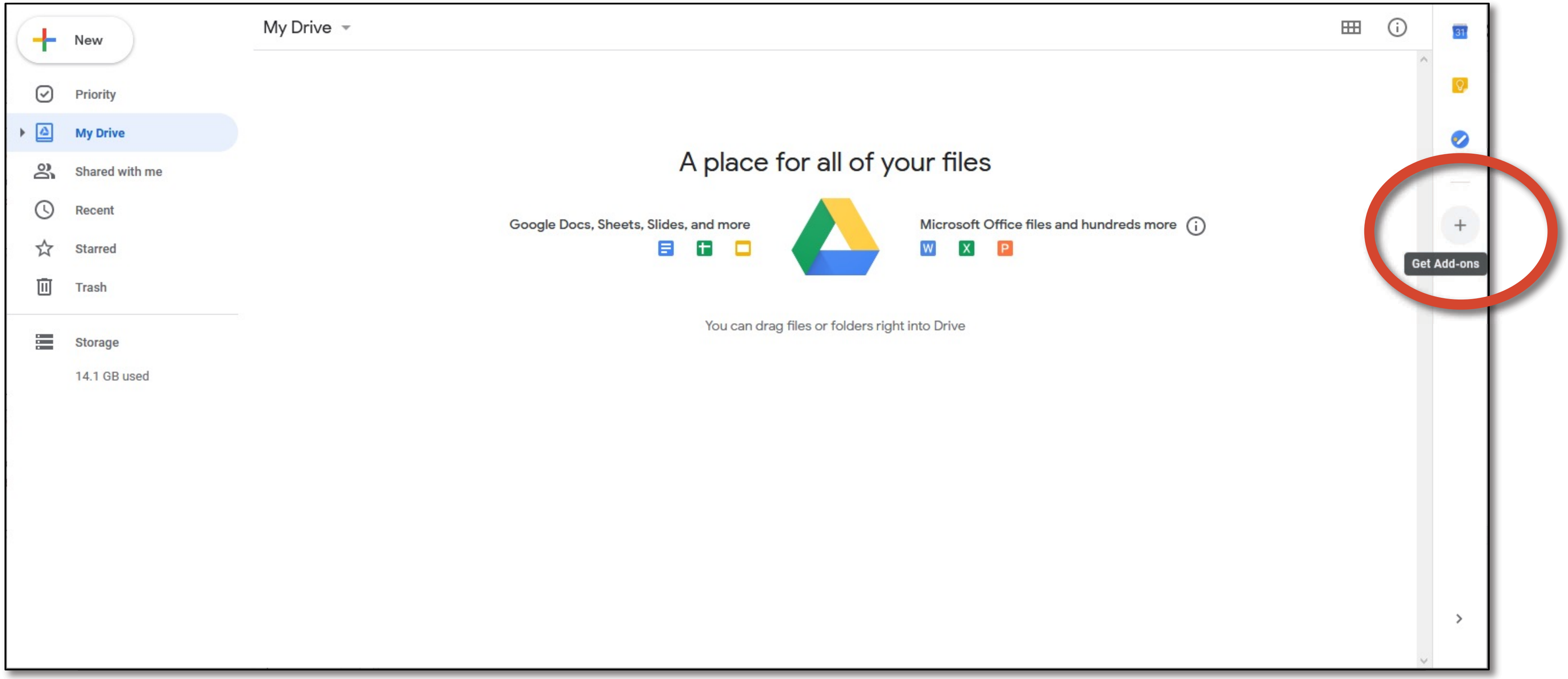
Download Data and Codes & Install Google Colaboratory Add-on and Add Notebooks to ‘Colab Notebooks’ Folder on Google Drive

Get the Data and Codes for Part 2

- Download the codes and data to your local Google drive and install Google Colab
- Data and Code Link – Click here
- Make sure you have a Google account – if you do not have one, create one, it's free – click here
- Follow the steps to copy the data and codes

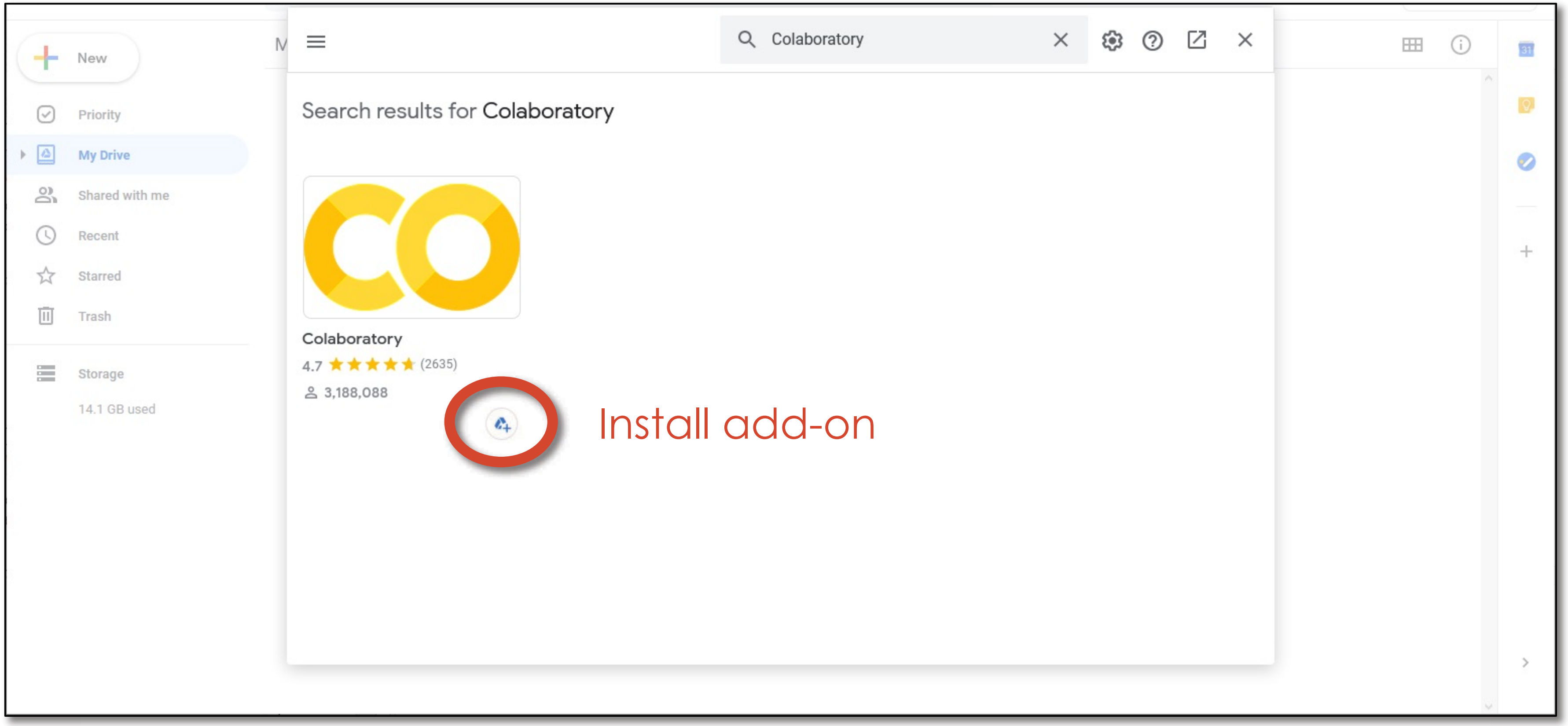
Add Your Notebooks to Google Drive

Step 1: Go to drive.google.com and click the + on the right to add add-ons.



Add Your Notebooks to Google Drive

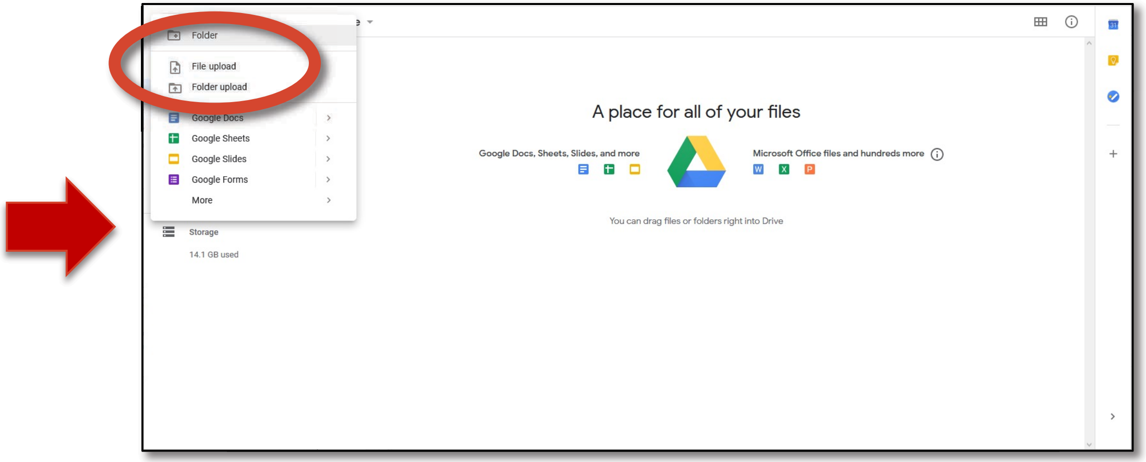
Step 2: Search for "Colaboratory" and install



Add Your Notebooks to Google Drive

Step 3: Add Notebook to Google Drive by dragging over files, or clicking New → File Upload

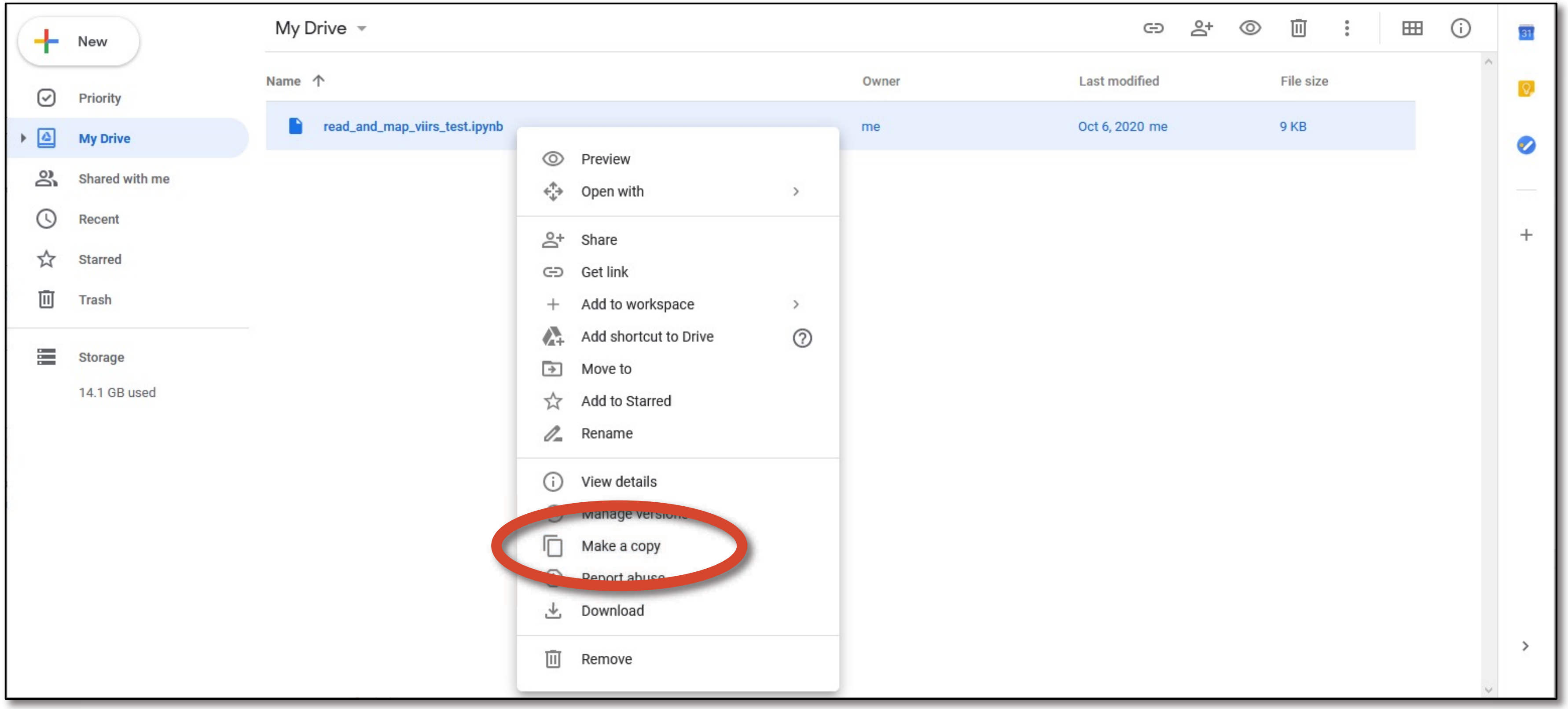
* If you already had Colaboratory installed, add the file to your Colab Notebooks folder. *



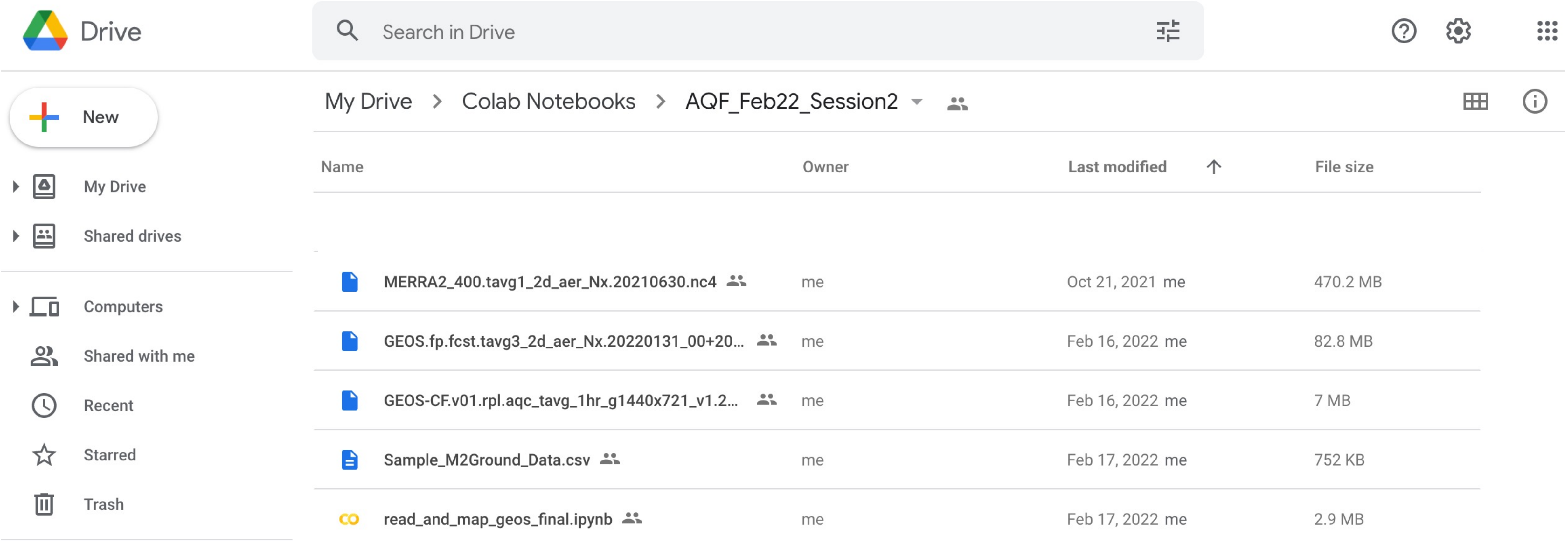
Add Your Notebooks to Google Drive

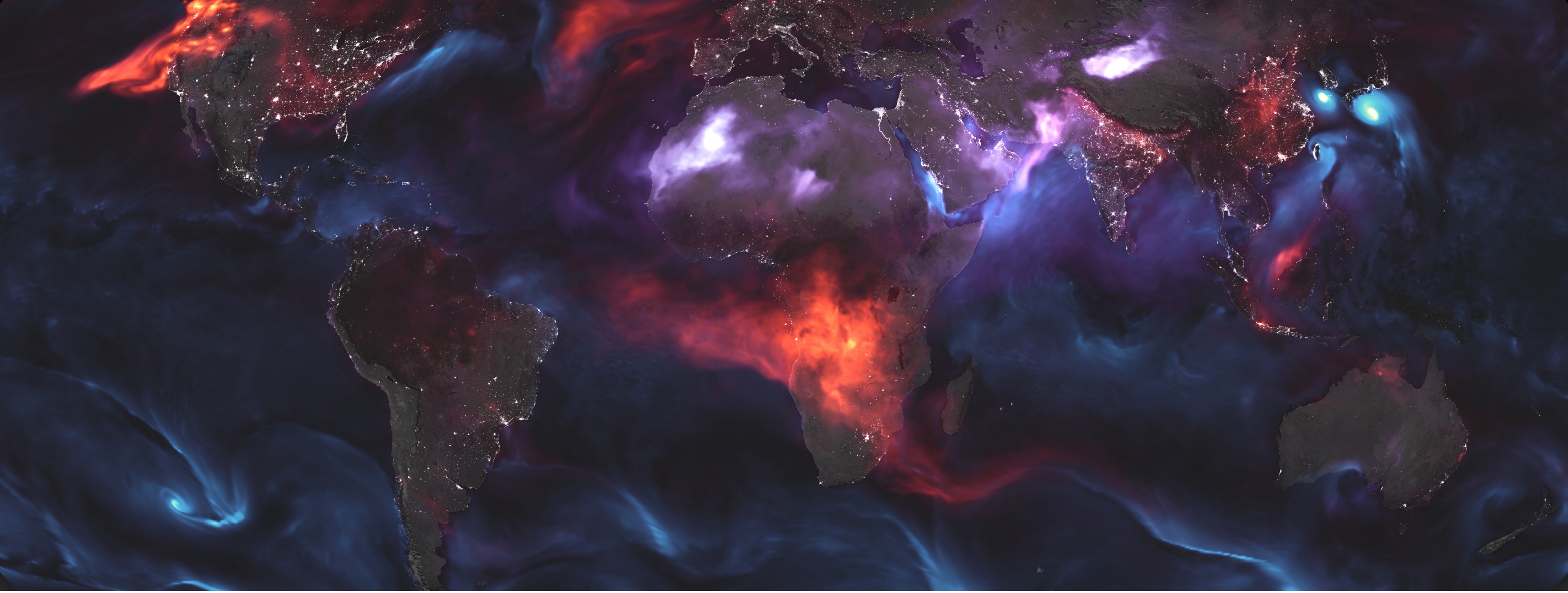
Step 3a: Right-click on your file and click "Make a copy". *This will create the Colab Notebooks folder in your Google Drive. The file copy will be inside this folder.*

* This step is only necessary if you had to install Colaboratory. *



Google Colab Ready Look

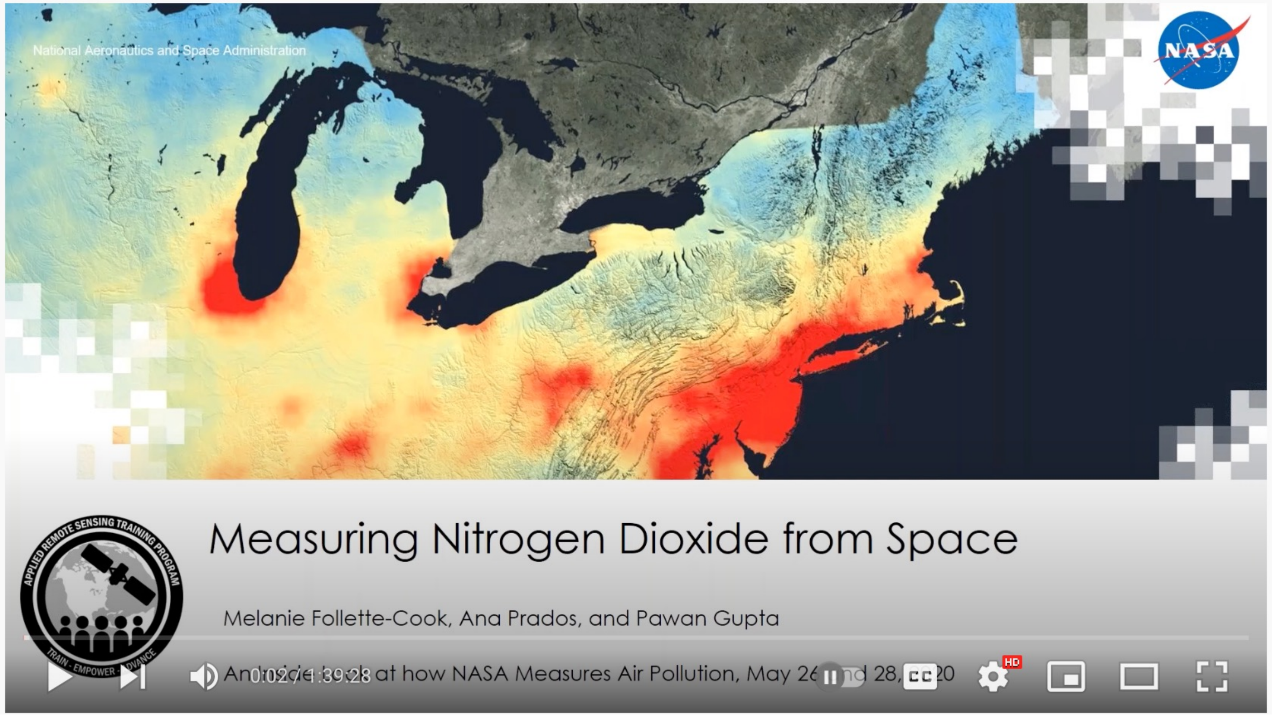




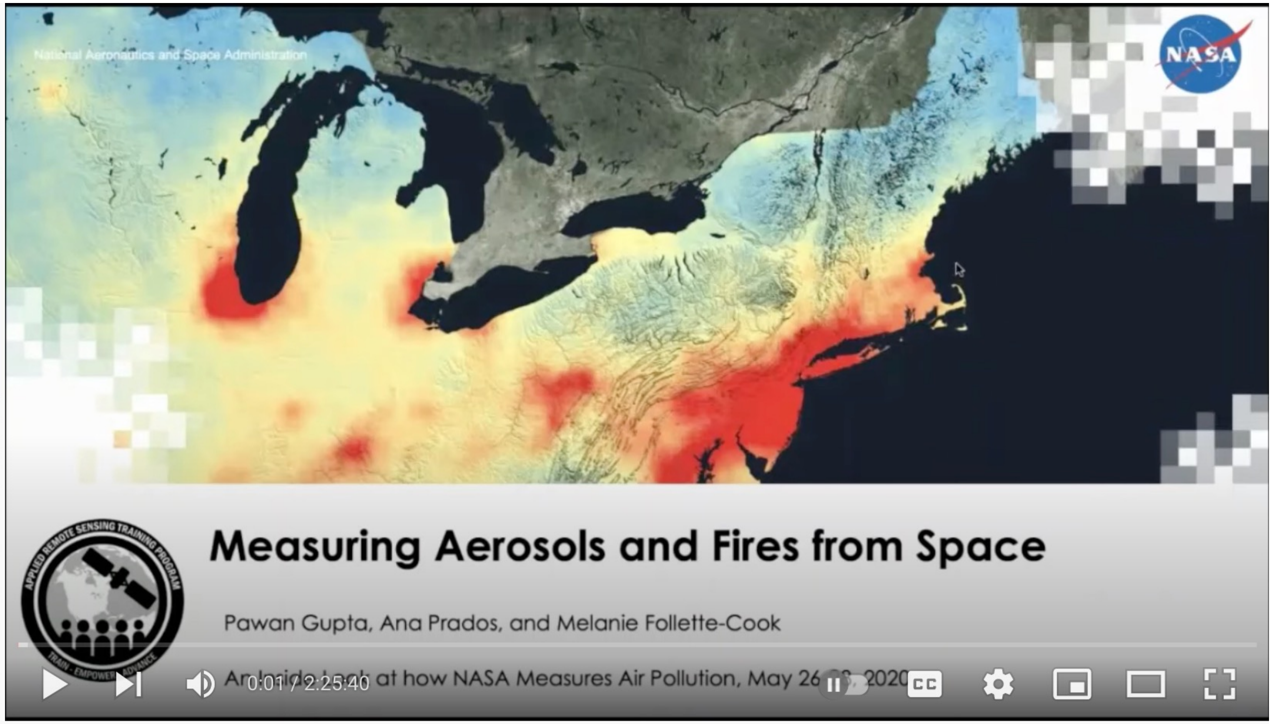
Satellite-Model Intercomparison

ARSET Training on Satellite Aerosols and NO_2 Datasets

https://appliedsciences.nasa.gov/join-mission/training/english/arset-inside-look-how-nasa-measures-air-pollution



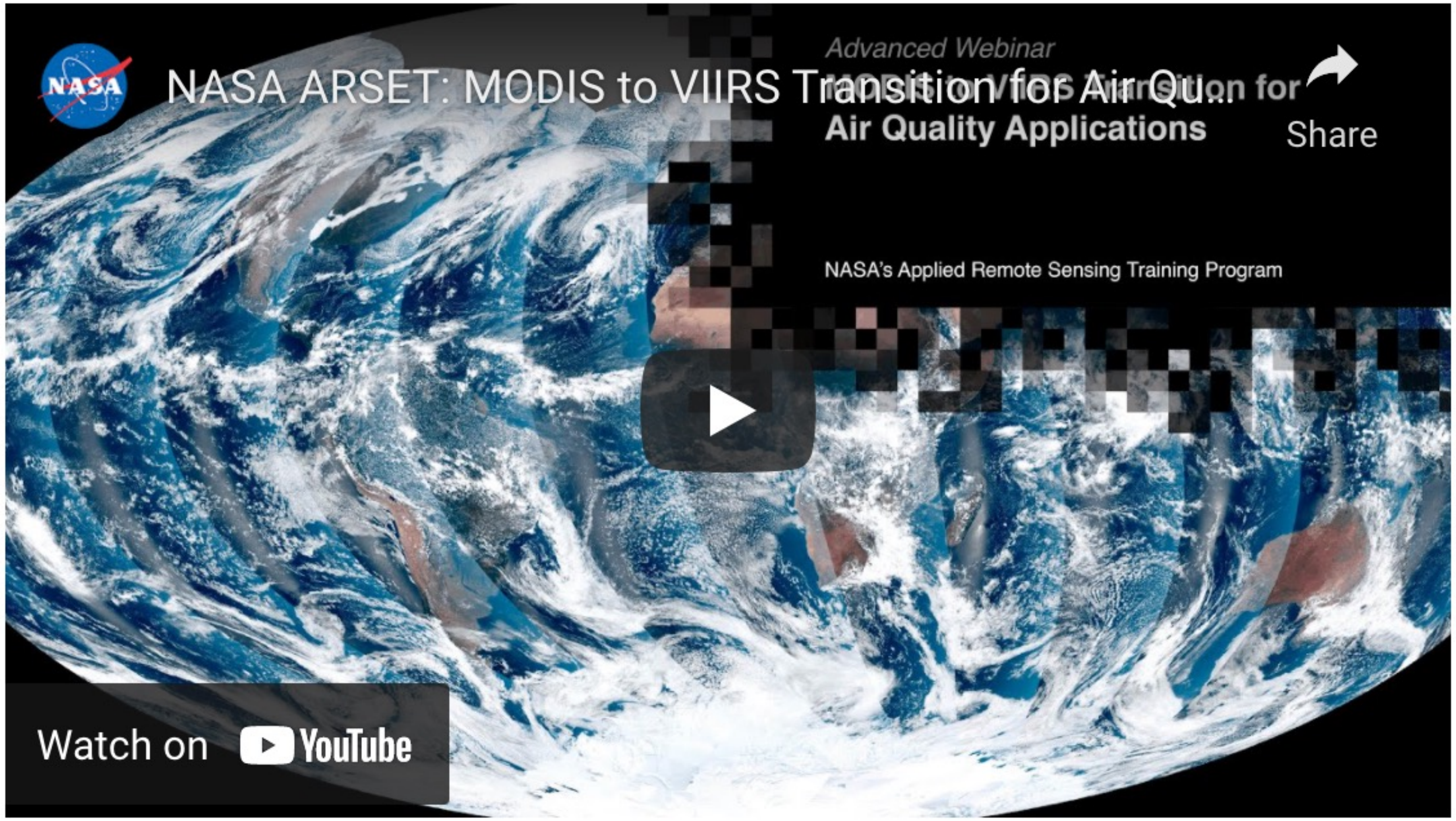
https://www.youtube.com/watch?v=truE3iCaGt8



https://www.youtube.com/watch?v=VpUzhXs8TX8

Python Tools for Satellite Data Analysis

https://appliedsciences.nasa.gov/join-mission/training/english/arset-modis-viirs-transition-air-quality-applications



Popular repositories

VIIRS_NASA Public

Jupyter Notebook ☆ 7 9

https://github.com/NASAARSET

Reminder – Spatial Resolution

Latitude X Longitude

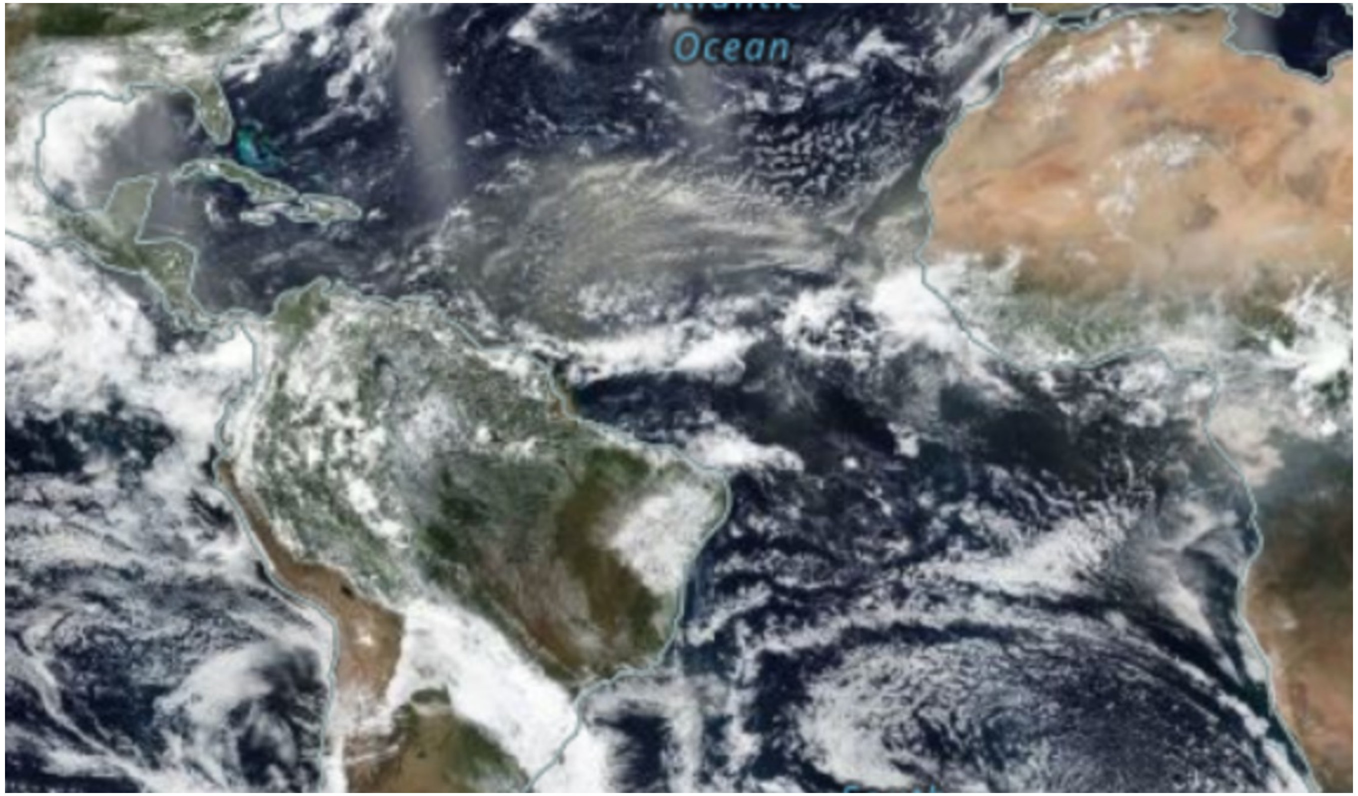
GEOS-FP
- 0.25° x 0.325° (~25 km)
- 721 x 1152 grid cells

GEOS-CF
- 0.25° x 0.25° (~25 km)
- 721 x 1440 grid cells

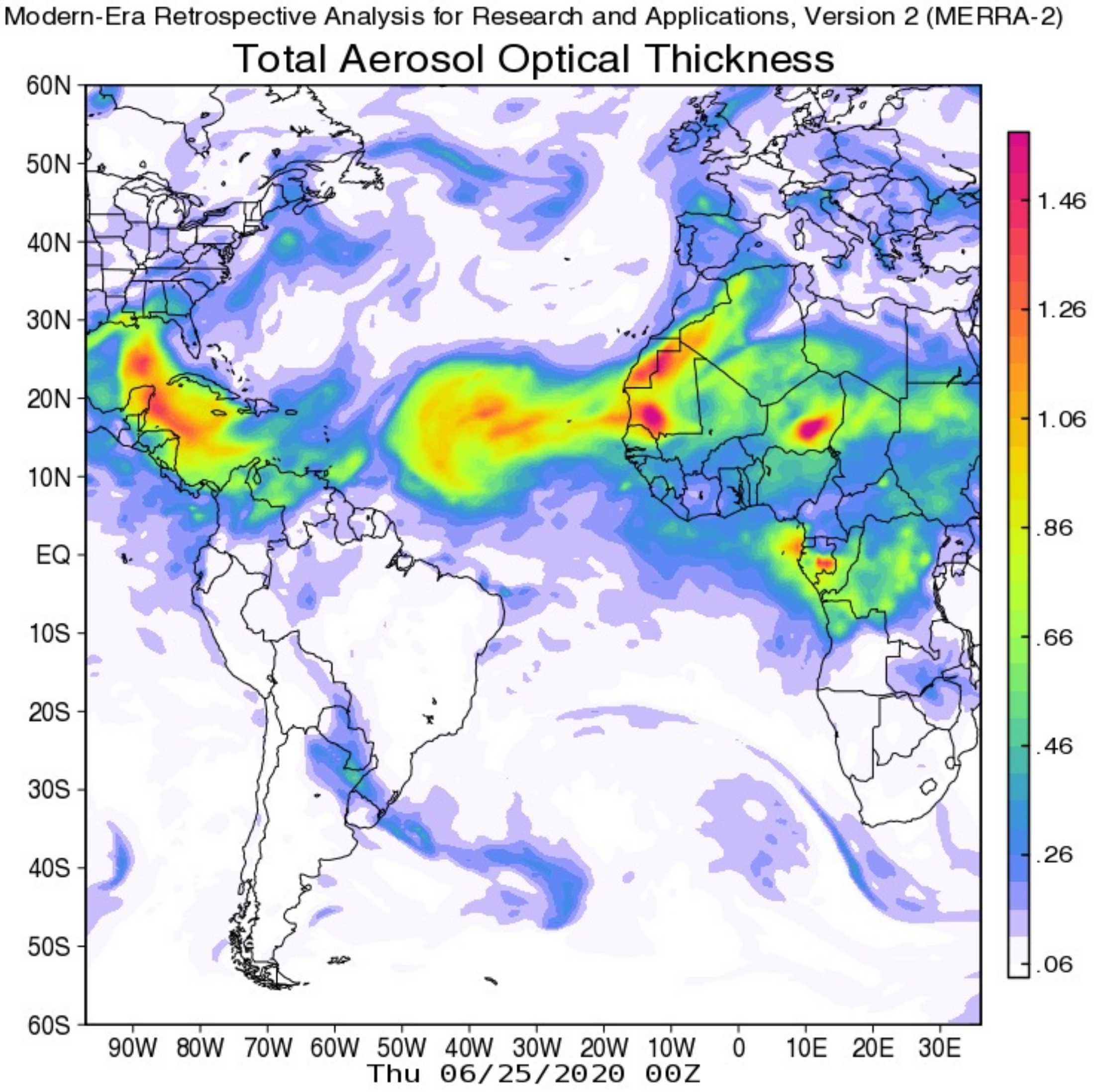
MERRA2
- 0.5° x 0.625° (~50 km)
- 361 x 576 grid cells

Qualitative Comparisons – True Color Image vs. Model Output

June 25, 2020



https://worldview.earthdata.nasa.gov/



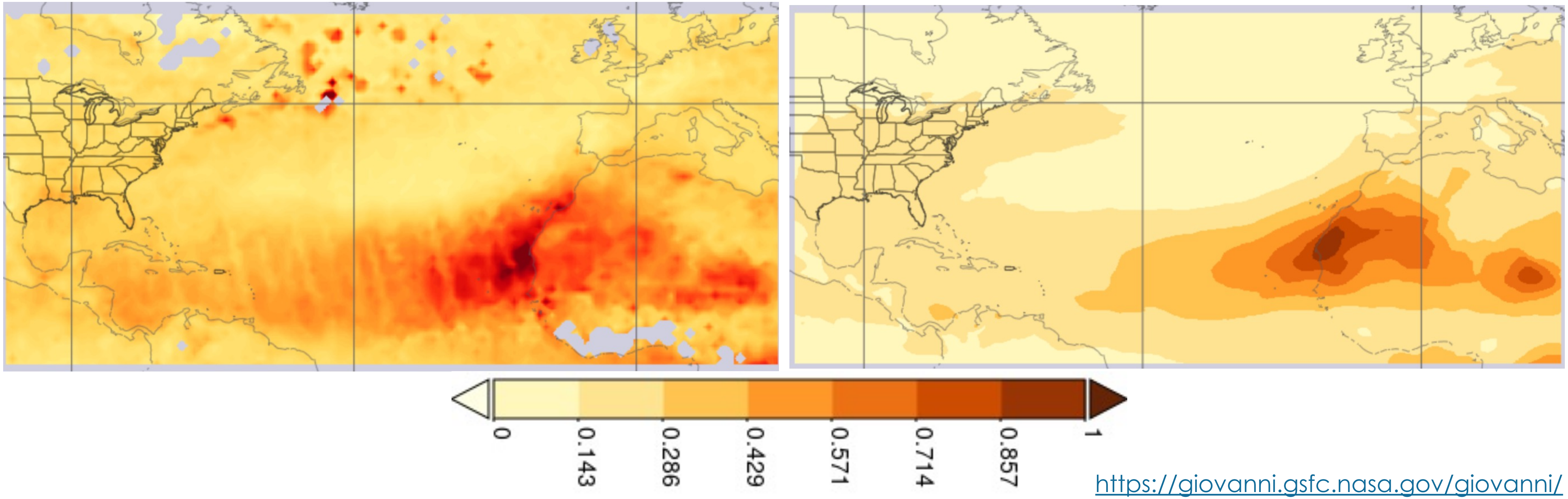
https://fluid.nccs.nasa.gov/reanalysis/

Comparisons with satellite imagery can show if the model is capturing broad spatial patterns.

Spatial Patterns Comparison

July 2020 – Dust Outflow over Atlantic

MODIS-Aqua Aerosol Optical Depth (550 nm) | **MERRA2 Aerosol Optical Depth (550 nm)**



https://giovanni.gsfc.nasa.gov/giovanni/

Comparisons with satellite derived geophysical parameter (i.e., AOD) can provide confidence on model's capability to capture larger spatial patterns and magnitude.

Satellite vs. Model – Quantitative Comparison

Know your data.

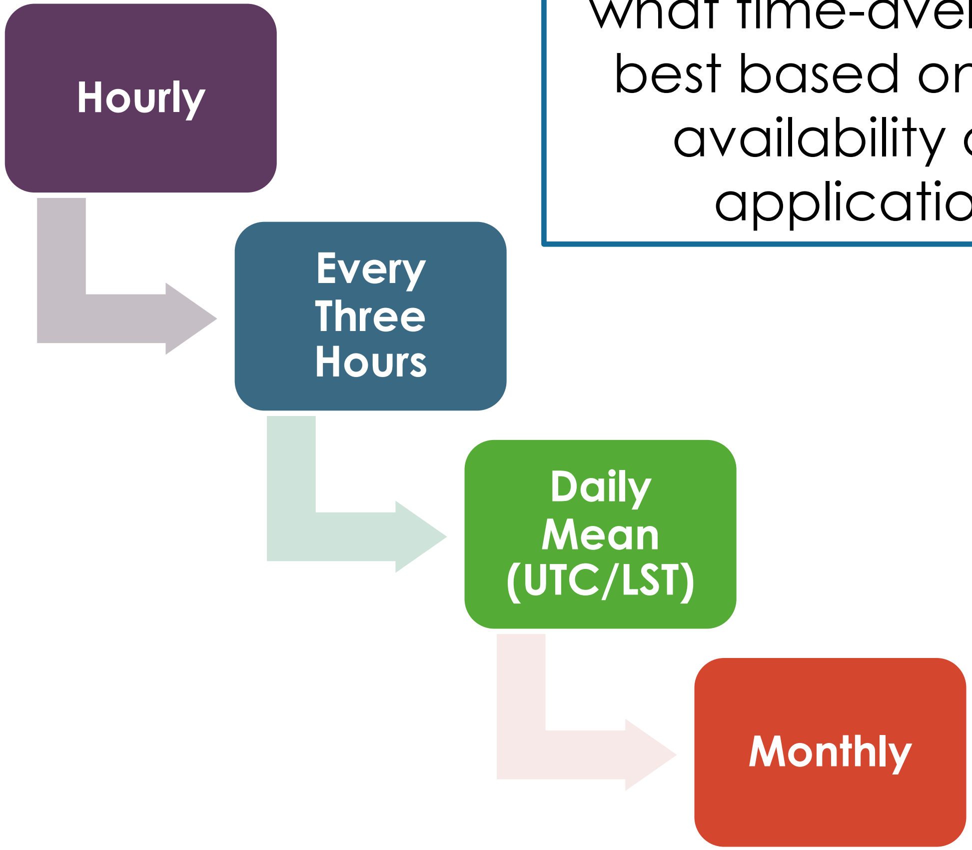
- Satellite
  - Instantaneous and typically once per day from LEO, but GEO can provide more frequent measurements
  - Only available in cloud free conditions
  - Can be averaged over time (hourly, daily, monthly, etc.)
  - Varying pixel size for level 2 data
  - Level 3 data are gridded and averaged over time

- Model
  - Instantaneous and averaged over time
  - Forecasts are typically hourly, but analysis and reanalysis can be averaged over time
  - Global model outputs are available everywhere irrespective of cloud cover
  - GEOS outputs are in fixed, angular grids

Satellite vs. Model – Quantitative Comparison

Temporal Matching

- Ensure that both model and satellite data correspond to the same date and time
- Time matching should be done as close as possible
- Most satellite data and model outputs are reported in UTC but
- **Know your data time zone. It is critical to ensure that both data and model output are reported in the same time zone (Note - Python has datetime function to convert between time zones).**

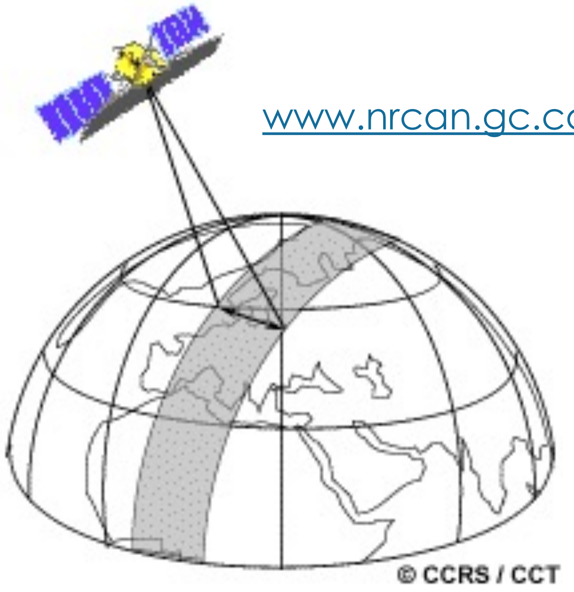


Users must choose what time-average is best based on data availability and application.

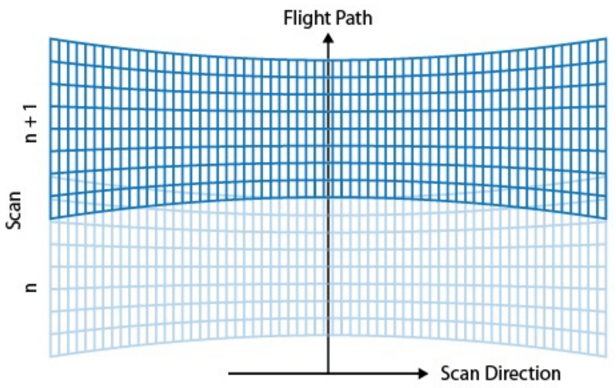
Satellite vs. Model – Quantitative Comparison

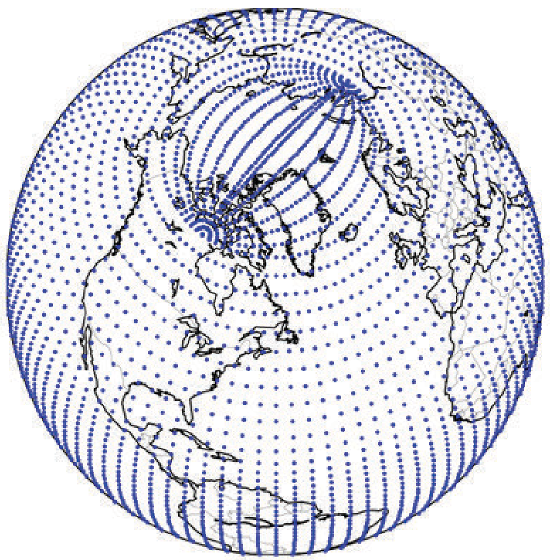
Spatial Matching (Resolutions)

- Typically, satellite data and model output have different spatial resolutions.
- **Know your data geolocation system (whether the lat/lon corresponds to the center vs. the corner of the pixel/grid)**
- For example:
  - MODIS AOD → 1km, 3km, 10km, 1° resolution
  - GEOS outputs → 0.25°x0.325°, 0.25°x0.25°, or 0.5° X0.625° resolution
- Satellite data can be swath data (not fixed grid size) or gridded (fixed angular grid – level 3 data)
- Therefore, **it is important to resample both data sets at the same spatial grid resolution**



www.nrcan.gc.ca

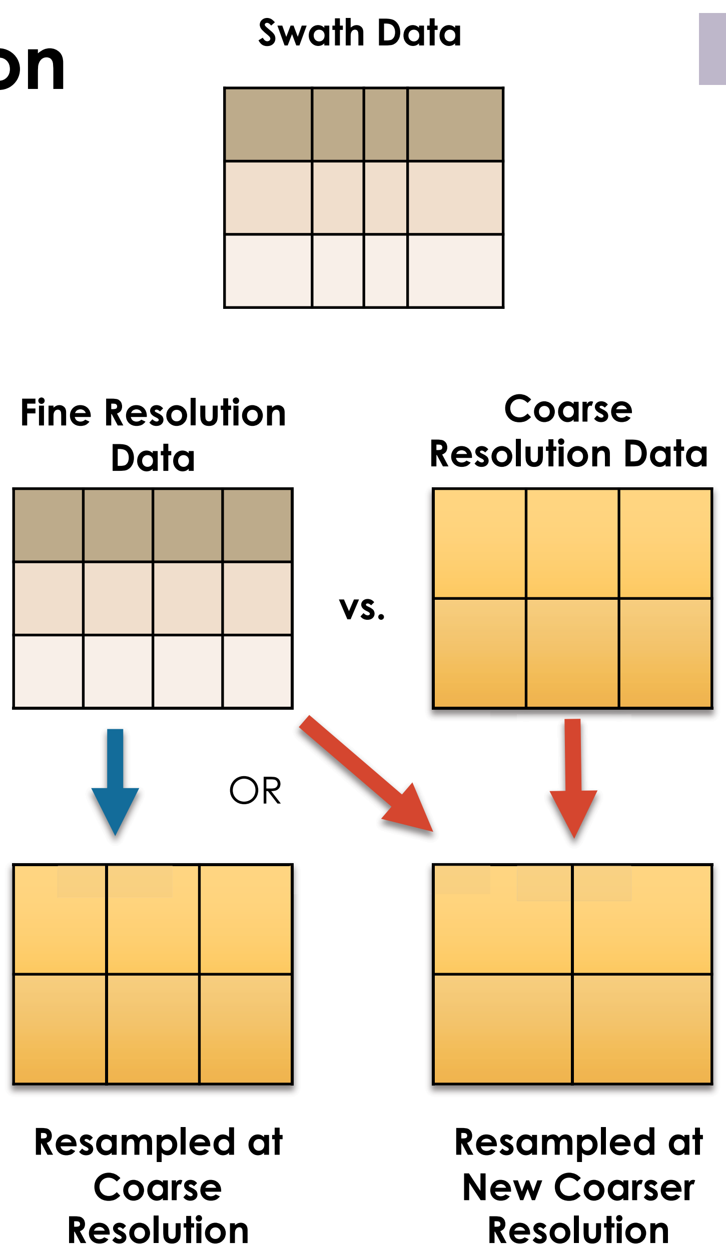




Satellite vs. Model – Quantitative Comparison

Spatial Scales (Resolutions)

- Both data and model output can be averaged over a larger areas for comparison
  - For example, over country, state, or a bounding box or a polygon
- Can be resampled to a new resolution, typically done at coarser equal angle grids
- Finer resolution (smaller grid size) can be resampled over coarser resolution grid (larger grid size)
- Other collocation algorithms such as linear interpolation or nearest neighbor can be used to bring two datasets to the same resolution – **choice of method depends on your end goal**



Tools for Comparisons

Using Level 3 Data - https://giovanni.gsfc.nasa.gov/giovanni/

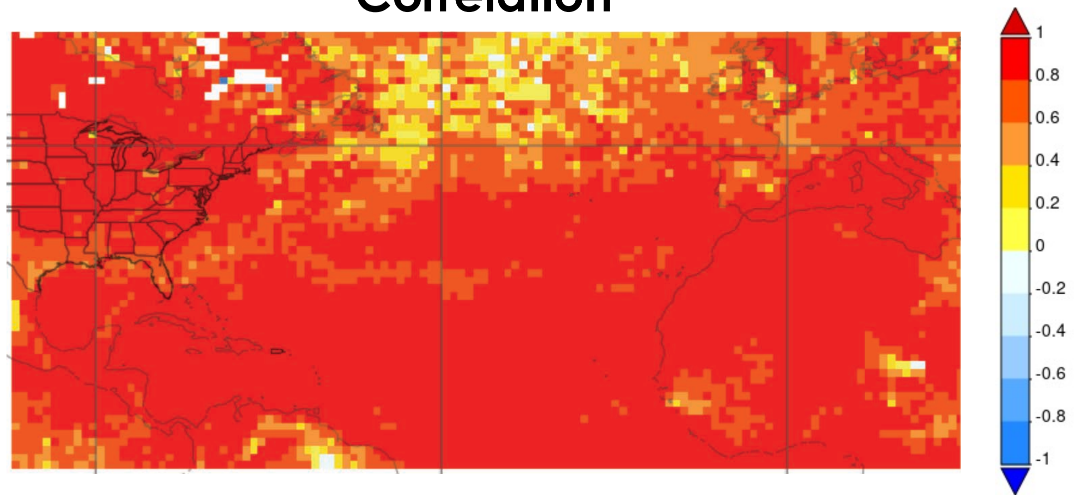
- NASA GIOVANNI allows users to compare level 3 gridded sets from different sources.
- Currently it has most of the satellite gridded (L3) and MERRA-2 variables.
- Using GIOVANNI, you can:

Maps
- Time Averaged Map
- Map, Recurring Averages
- Time Averaged Overlay Map
- Map, Accumulated
- Animation
  Limited to: **365 time steps**
- Map, Difference of Time Averaged

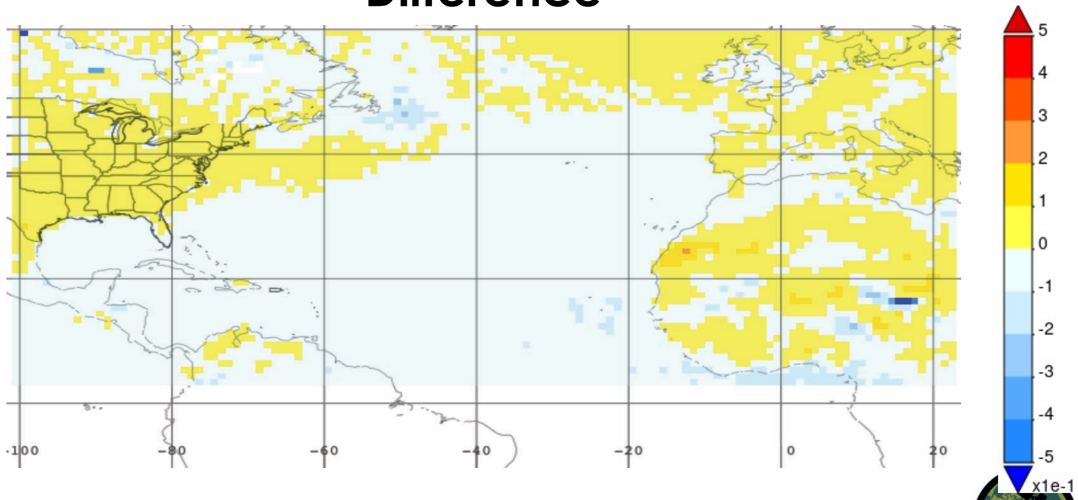
Comparisons
- Map, Correlation
- Scatter, Area Averaged (Static)
- Scatter (Interactive)
  Limited to: **30000 points**
- Scatter (Static)
- Scatter, Time-Averaged (Interactive)
  Limited to: **30000 points**

Time Series
- Time Series, Area-Averaged Differences
- Time Series, Area-Averaged
- Hovmoller, Longitude-Averaged
- Hovmoller, Latitude-Averaged

Correlation



Difference



MERRA2 AOD vs MODIS-AQUA AODs for 2020-2021

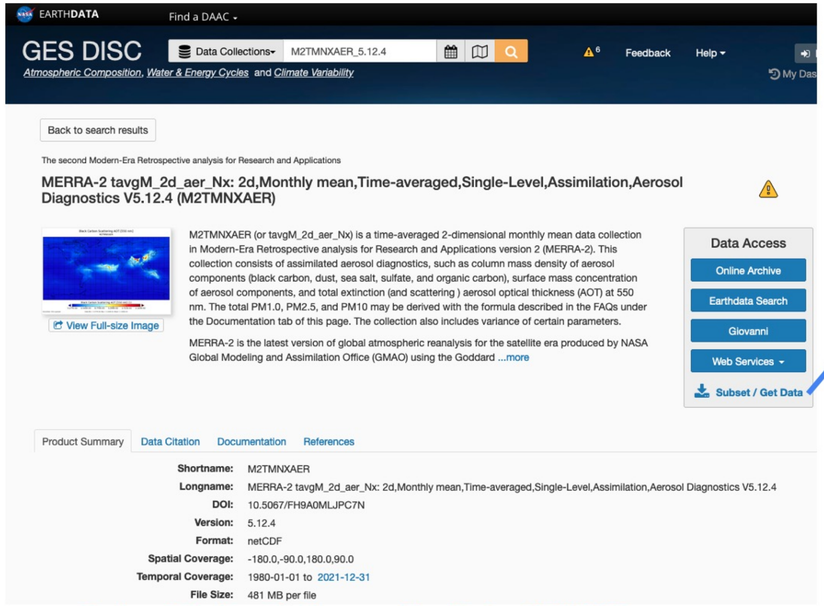
Tools for Comparisons

Regridding/Resampling

- GES DISC Subsetter and Regridder – Refer to Part 1 of this webinar
  - MERRA2 data
  - OMI data

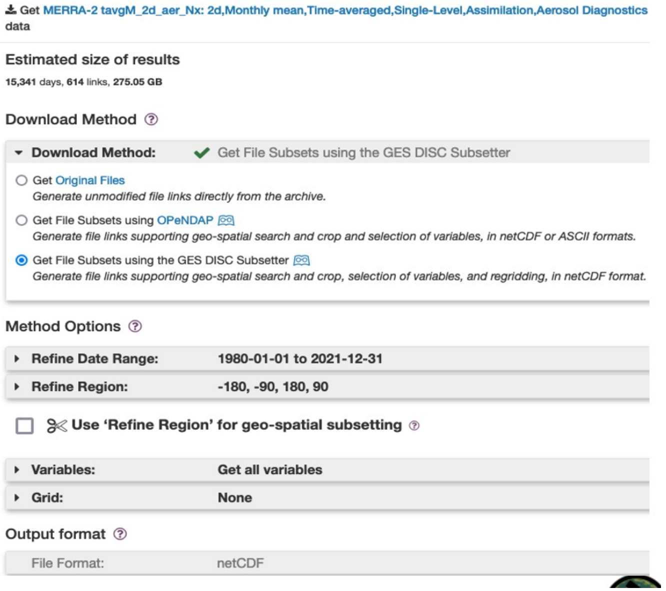
Python sample code to regrid the satellite and model data

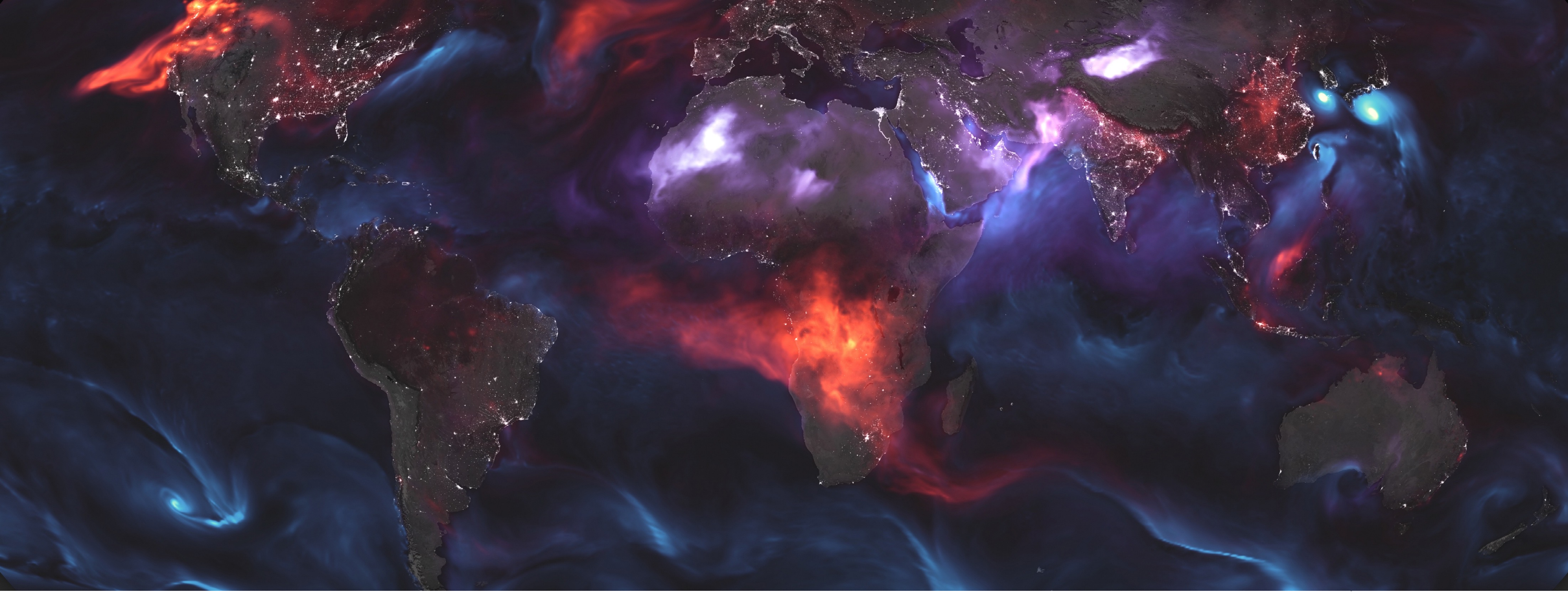
b. Demo How to use the Level 3 and 4 Subsetter and Regridder



The above is the dataset landing page of the collection M2TMNXAER:
https://disc.gsfc.nasa.gov/datasets/M2TMNXAER_5.12.4/summary?keywords=M2TMNXAER_5.12.4

- **Subset / Get Data** → subset, regrid, and download data, and compute daily statistics (mean, minimum, maximum) on-the-fly



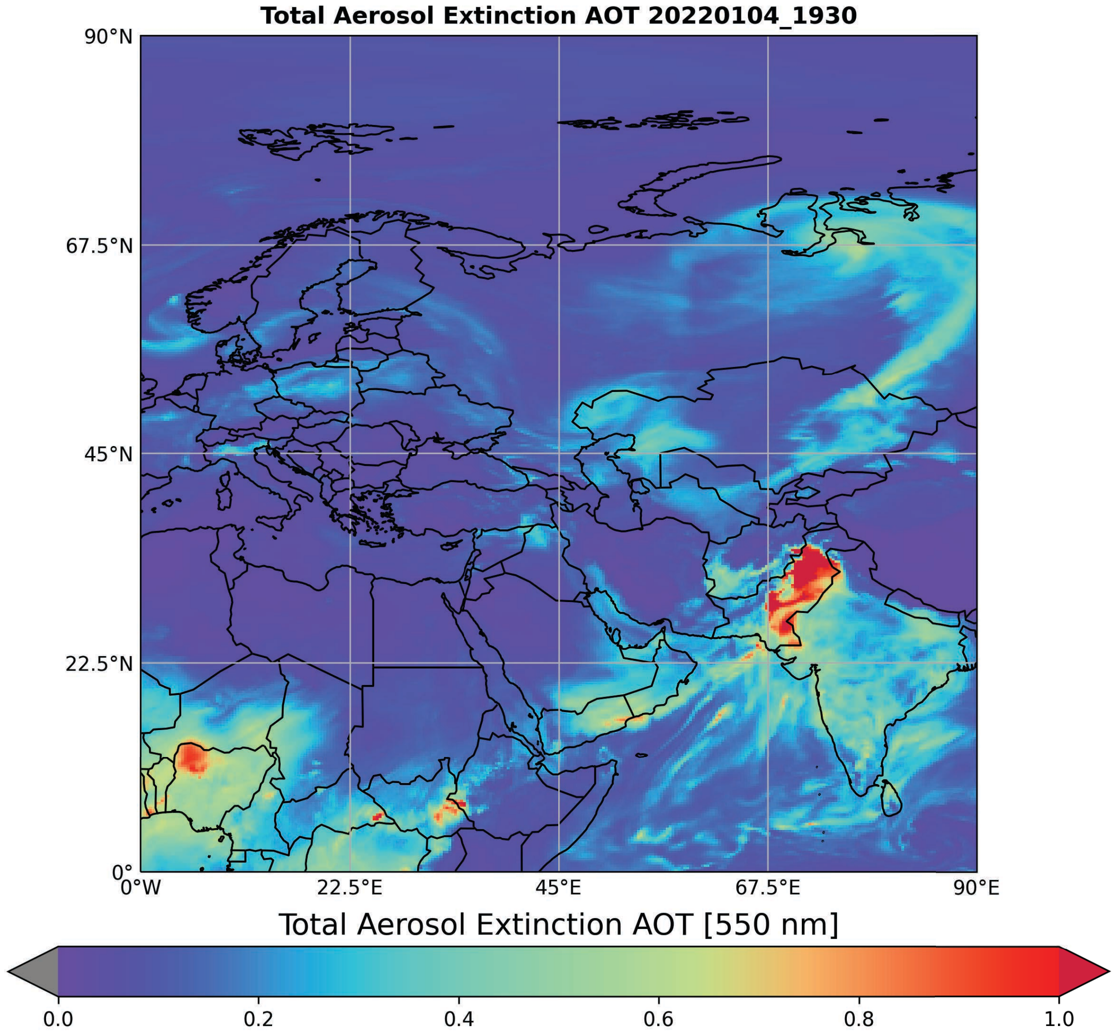


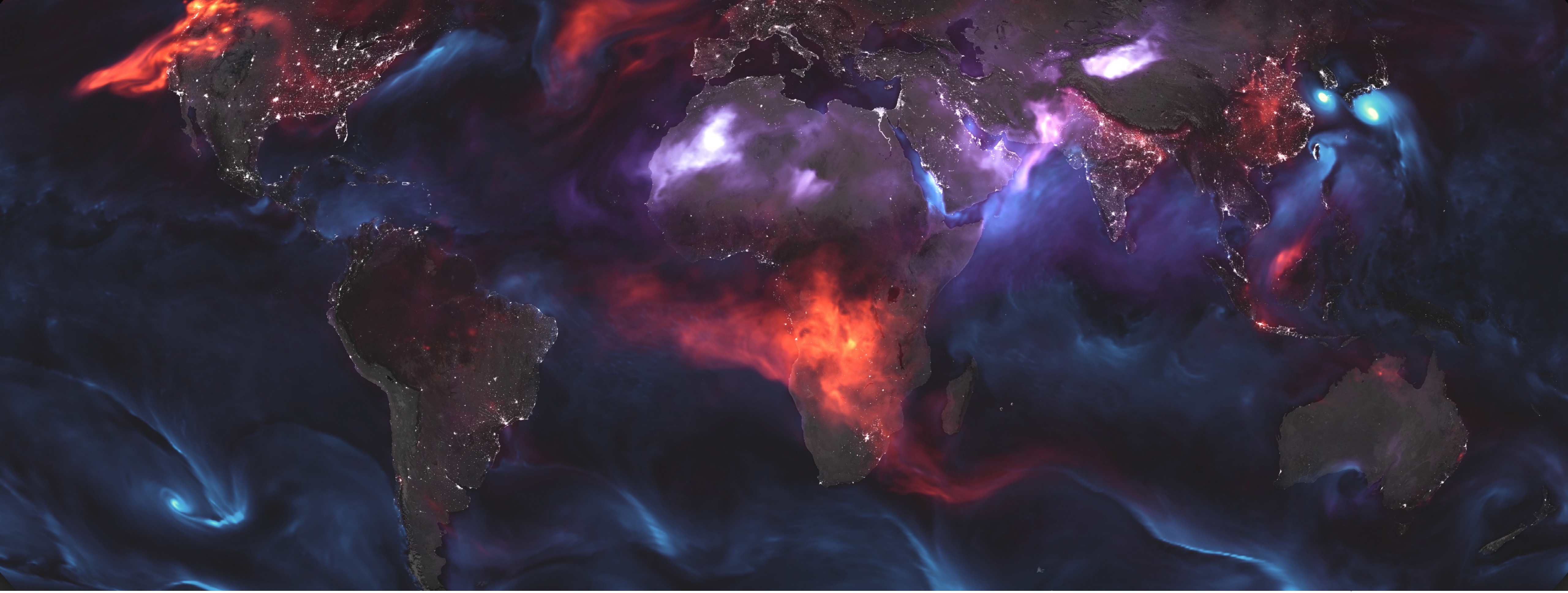
Python Demo – Reading & Mapping

Python Jupyter Notebook – read_and_map_geos_final.py

This notebook can:

- Read GEOS-FP, GEOS-CF, MERRA-2 files in NetCDF format
- Download data for a given date
- Map the data for a selected variable and parameter
- Export select variables to .csv file
- Extract data over a given location

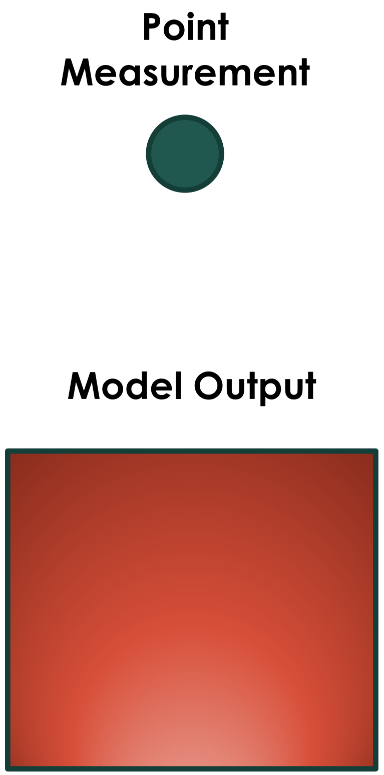




Validation and Inter-Comparison

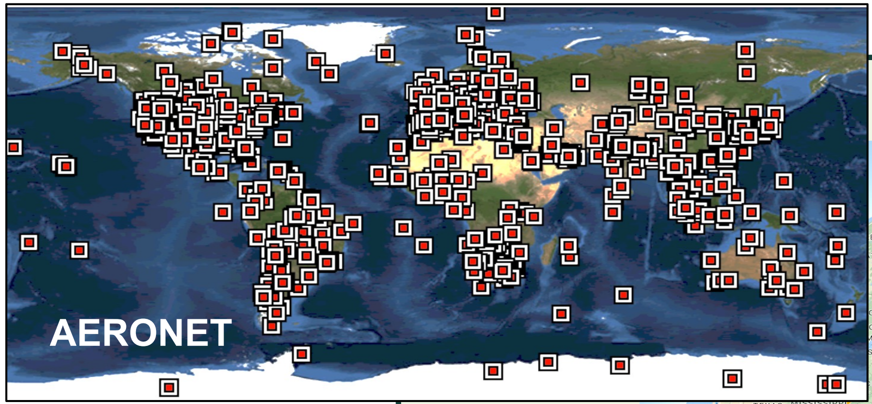
Validation & Challenges

- Independent validation is required to assess the accuracies and uncertainties of model outputs (reanalysis, analysis, or forecasts).
- The Earth-Atmosphere system is dynamic in nature – continuously changing.
- The frequency of ground measurements can vary in space (i.e., geographically) and time (i.e., length of data availability).
- Sometimes we rely on measurements collected during field campaigns.
- Often ground measurements are from stationary locations and thus represent smaller areas (point measurement) vs. model output, which is averaged over a larger spatial area (spatial resolution/grid size).

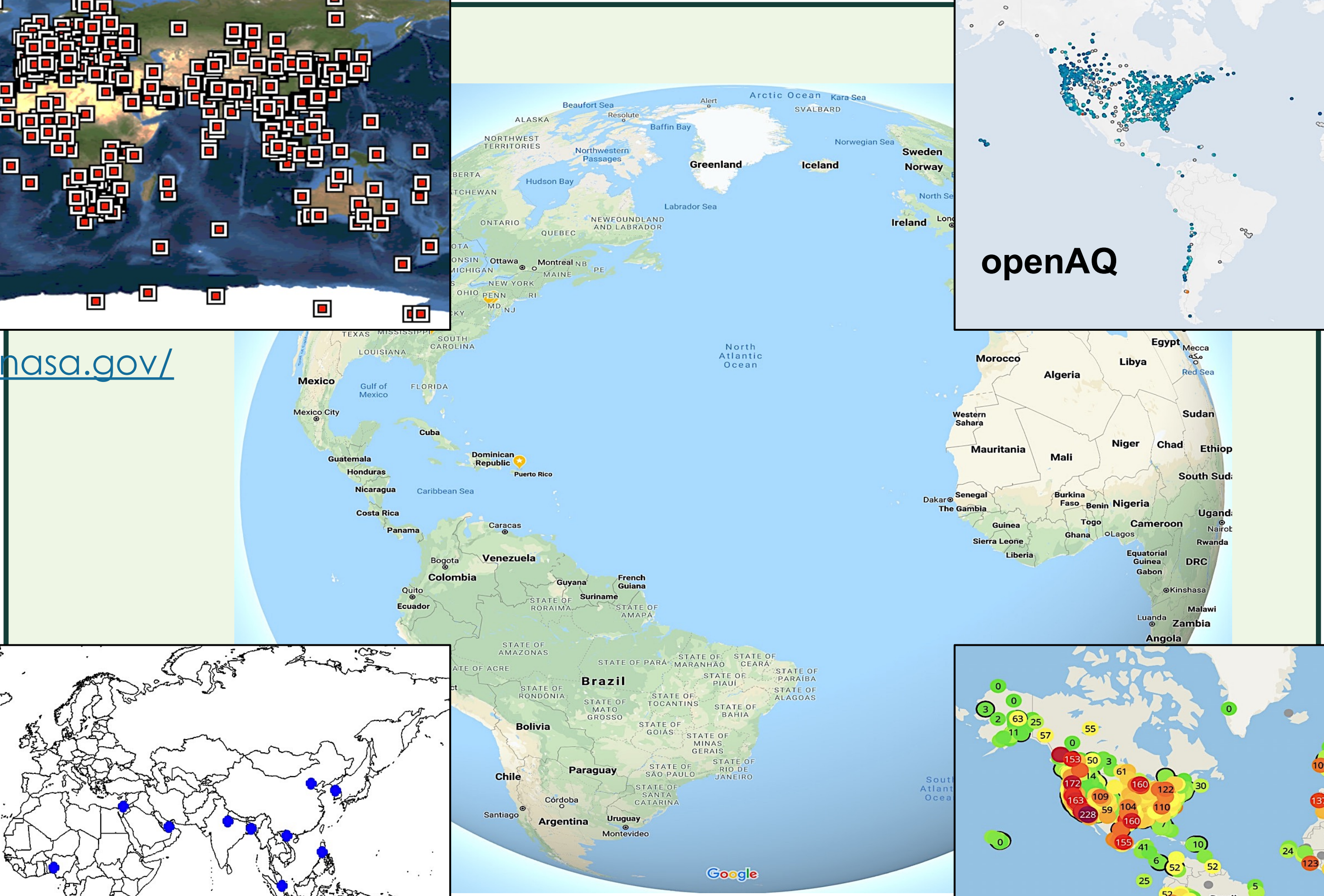


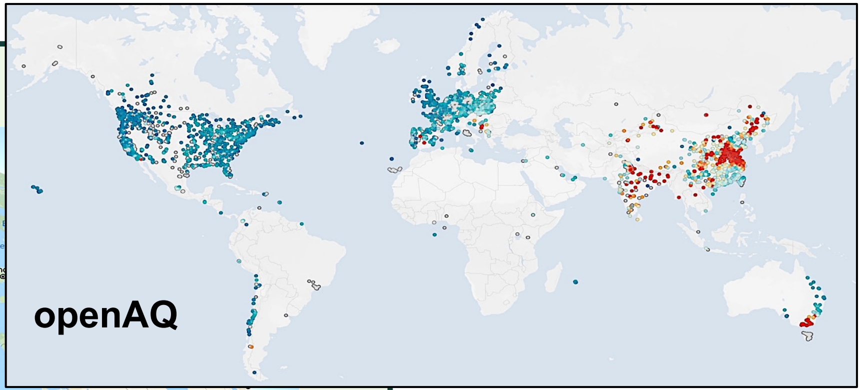
Reading suggestion on validation - https://doi.org/10.1002/2017RG000562

Global Open Ground Networks/Data Sources for Air Quality

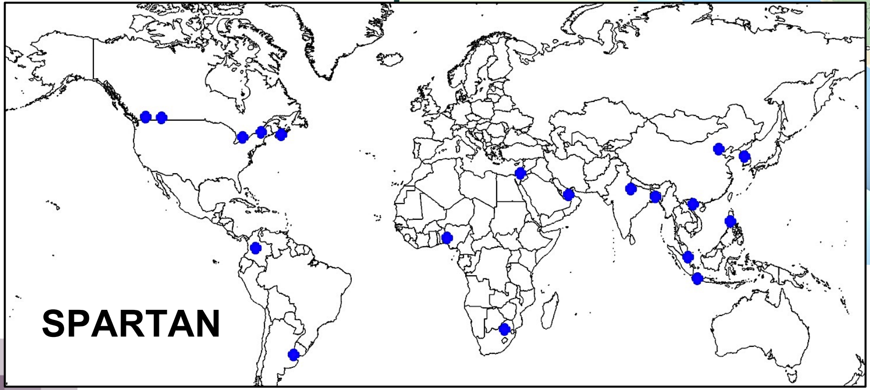


https://aeronet.gsfc.nasa.gov/

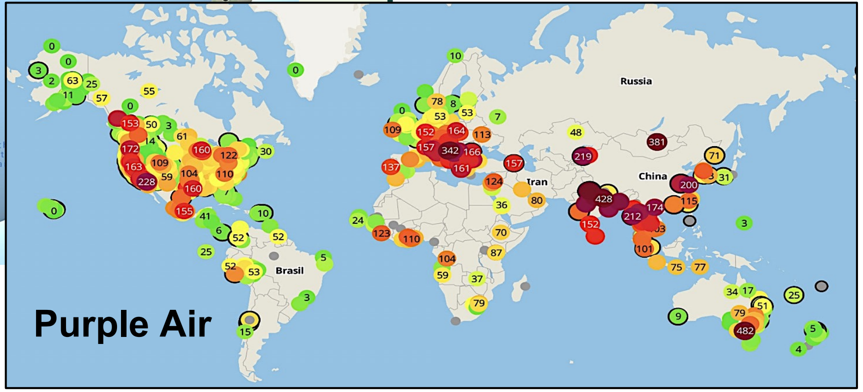




https://openaq.org/



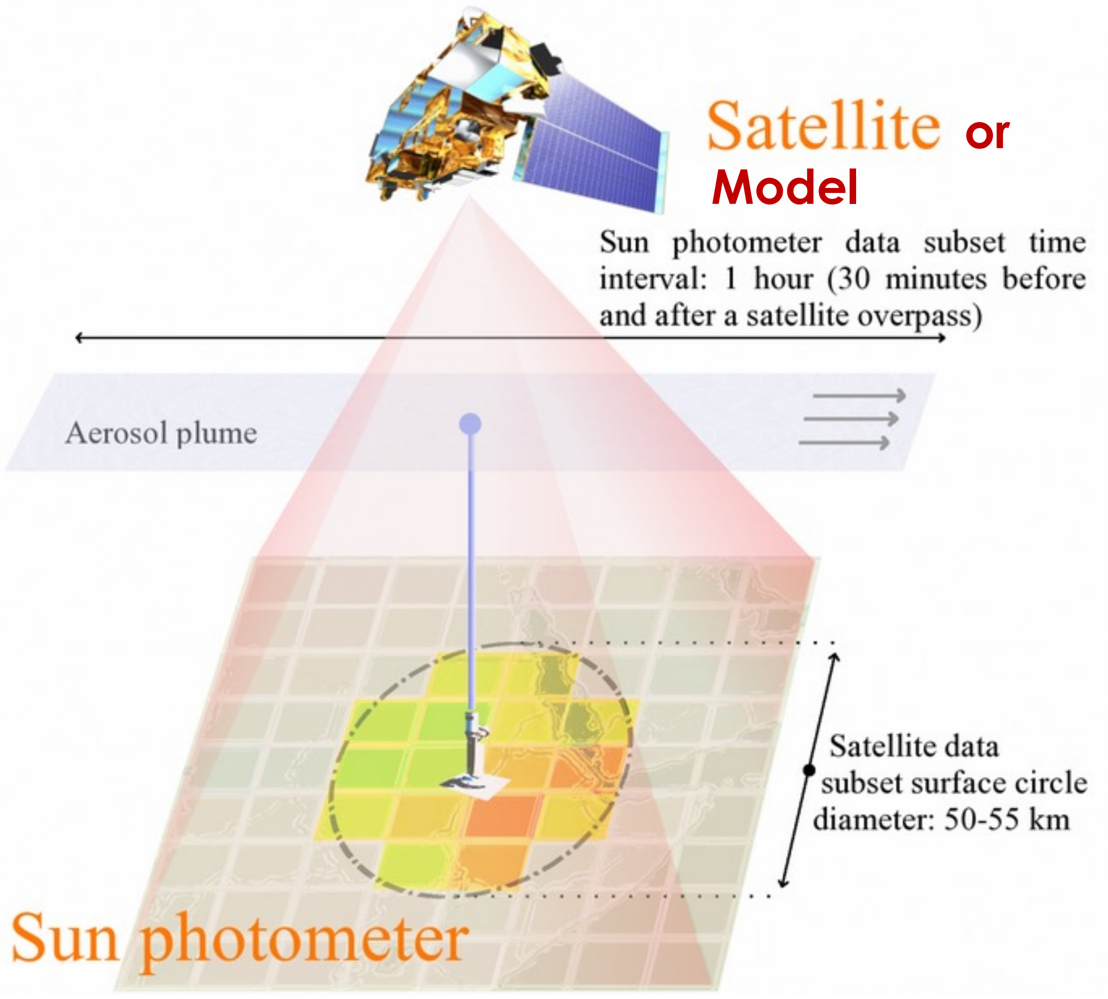
https://www.spartan-network.org/



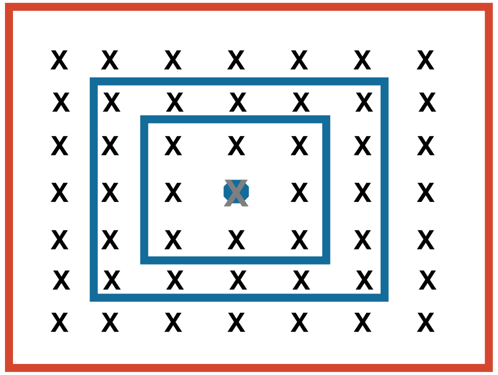
https://map.purpleair.com/?mylocation

Step 1: Spatial Collocation

- **Need**
  - Model/Satellite output files (e.g., GOES-FP, GEOS-CF, MERRA2, MODIS)
  - Need latitude and longitude of ground station
- **Spatial Matching**
  - Find Nearest Neighbor – Model grid cell or pixel closest to ground location
  - Average model/satellite data around ground station for:
    - 3x3 pixels/grids
    - 5x5 pixels/grids
    - All pixels within certain search radius
    - Save statistics (mean, median, std, count)
    - Include date/time of model outputs
- Repeat for each file/date and generate a time series



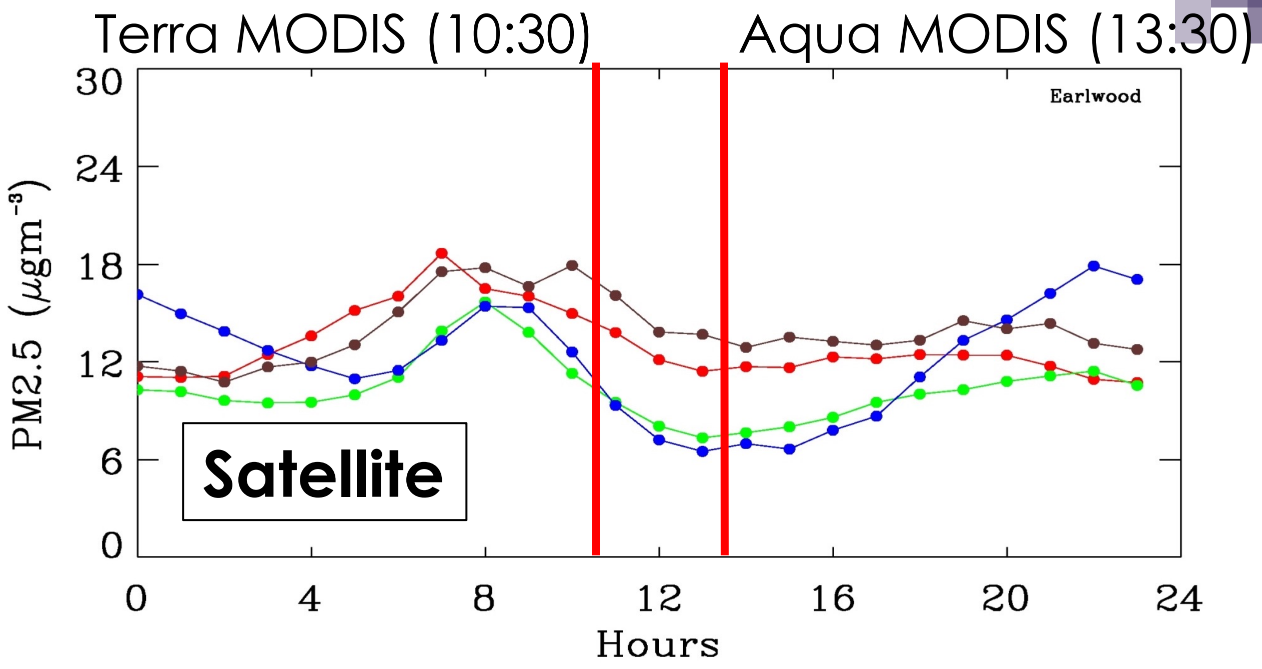
Or
Air Quality Monitor

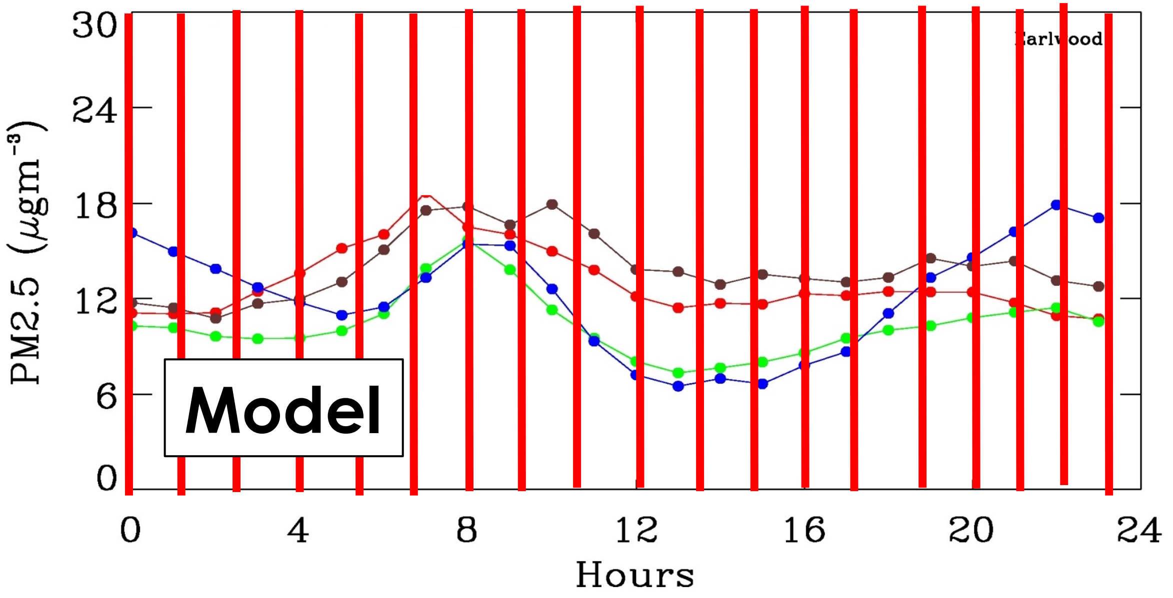


See – python function getpointdata

Step 2: Temporal Collocation

- **Need**
  - Ground data file with date/time information
  - Extracted model/satellite data from step 1
- **Temporal Matching**
  - Ensure both ground and model outputs' date/time information is in same time zone (local time or UTC)
  - Read the two data files into the code
  - Pick the ground measurement (e.g., $PM_{2.5}$, NO_2, or AOD) closest to satellite overpass or model output time
  - Depending on temporal resolution of data (both ground/model), you can average over e.g., one hour, three-hours, or a day to match the two datasets
  - In case of GEOS outputs, we will match the model output to the nearest hourly ground measurement





Evaluating the Forecasts – Numerical

- **Accuracy**
  - Mean closeness between forecast and observed value

Forecast

Observation

$$A = \frac{1}{N}\left(\sum_{1}^{N}\left|f - o\right|\right)$$

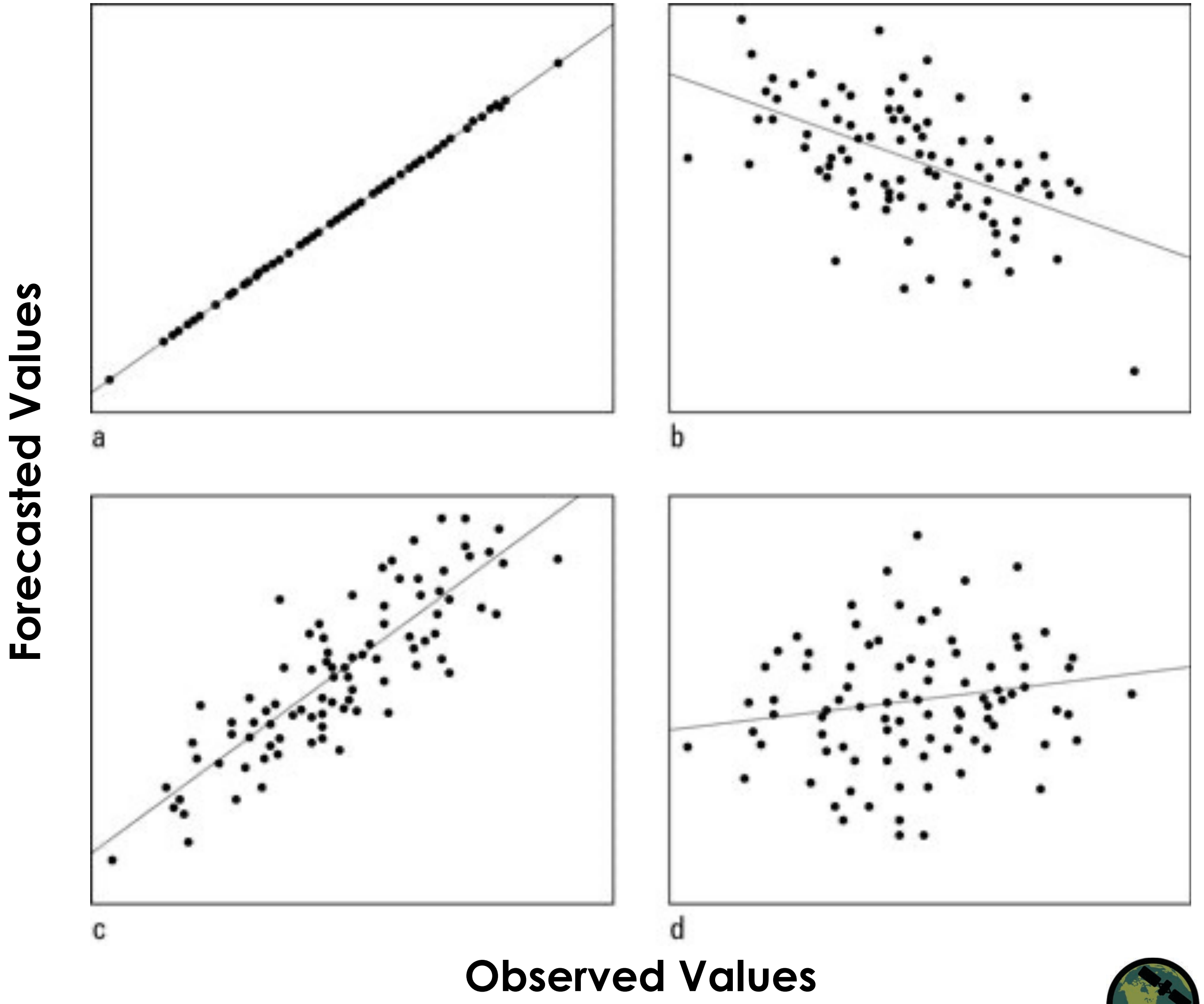
Number of Data Points

- **Bias**
  - On average, an indicator of under or over estimation by forecast

$$B = \frac{1}{N}\left(\sum_{1}^{N}(f - o)\right)$$

Evaluating the Forecasts - Numerical

- **Correlation**
  - Degree of relationship between forecast and observed value

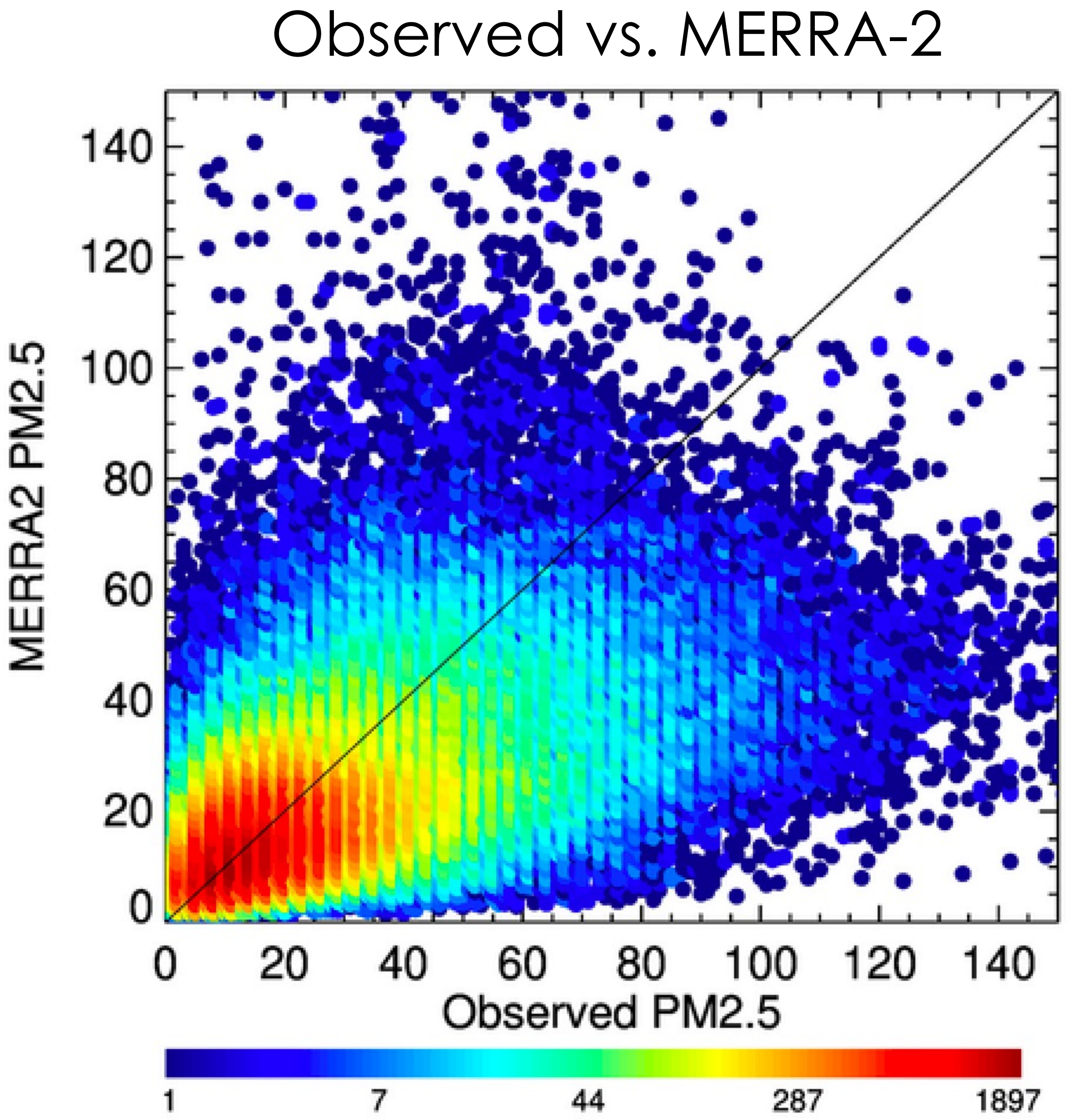


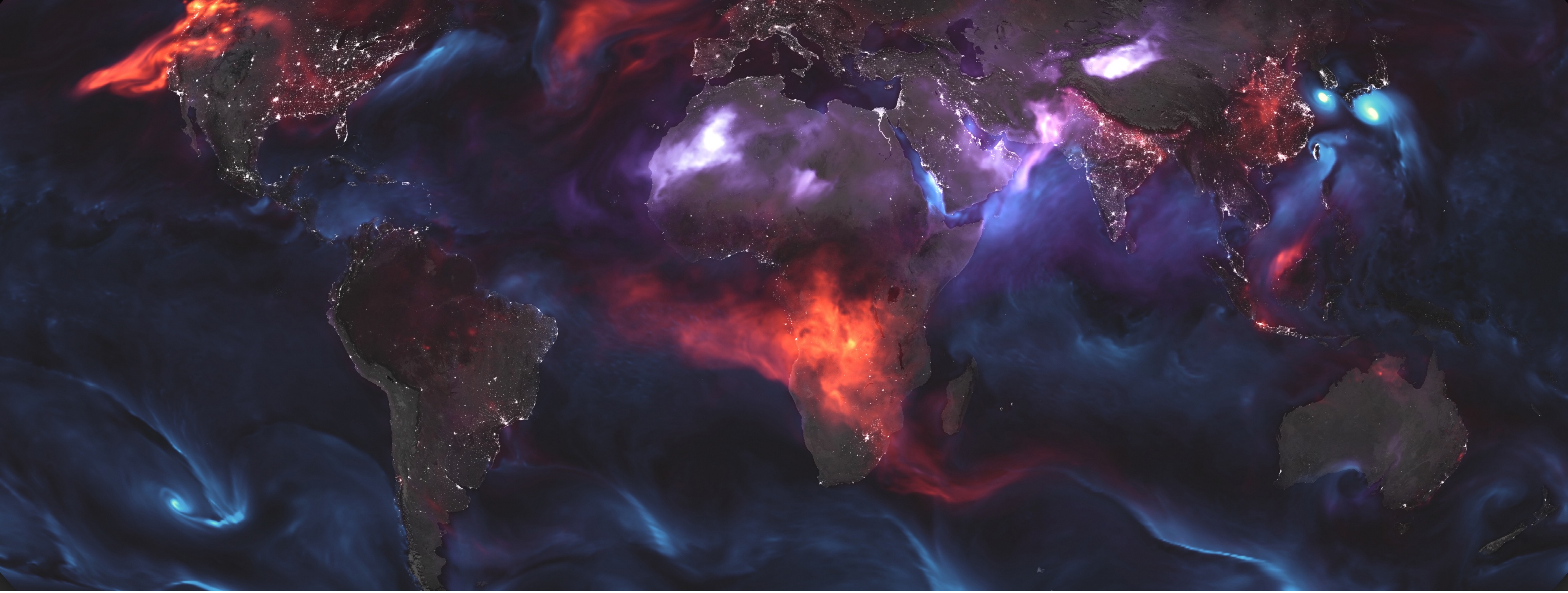
- **Categorical**

Observed \ Forecasted	Good	Moderate	Unhealthy for Sensitive Groups	Unhealthy
Good	k	l	m	n
Moderate	o	p	q	r
Unhealthy for Sensitive Groups	s	t	u	v
Unhealthy	w	x	y	z

Analysis - Density Scatter Plot

- A two-dimensional histogram
- It shows the number of points in each region of the plot (i.e. density)
- The number of points within area of the 2-D space is counted and represented with a color.
- There are number of ways in which density of points can be estimated and presented - https://www.python-graph-gallery.com/2d-density-plot/





Python Demo – Density Scatter Plot

Python Jupyter Notebook Task Flow

Set Up Environment

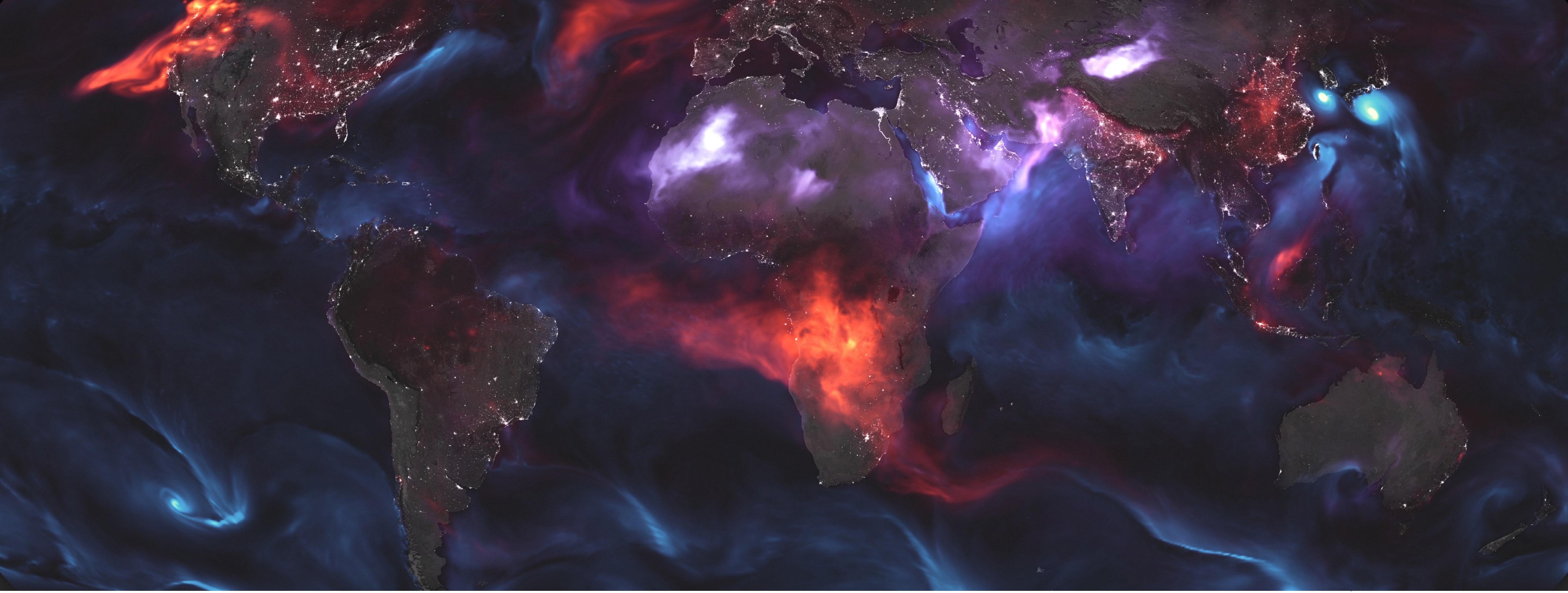
Mount Google Drive → Install Python Packages → Import Required Packages

Analytical Functions

- Map
- Output CSV File
- Output at a Location
- Regrid the Data and Map
- Density Scatter Plot
- Read GEOS-FP
- Read GEOS-CF
- Read MERRA2

Additional Available Functions

- MERRA2 Download
- GEOS-FP Download
- GEOS-CF Download

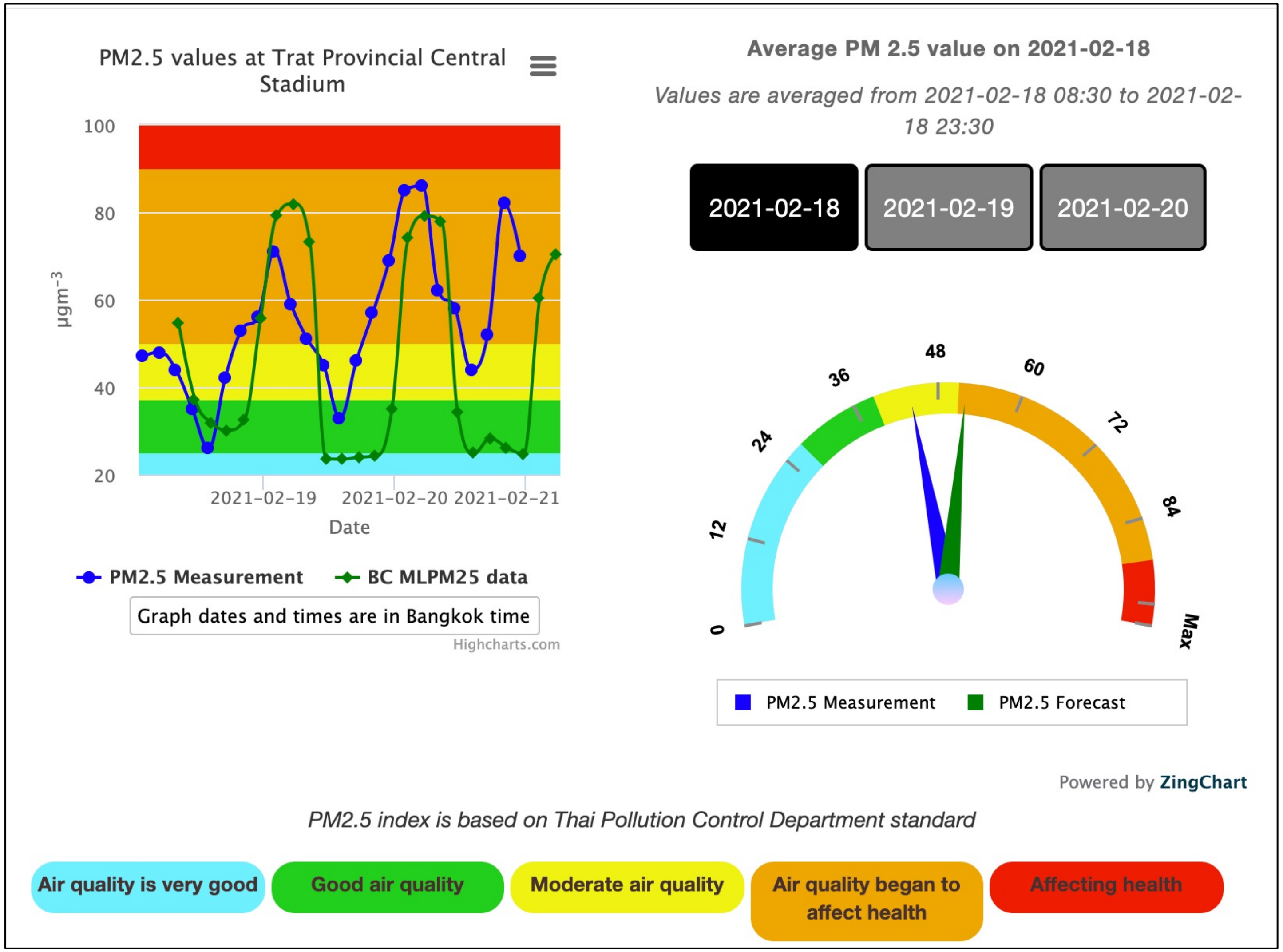


Possible Additional Analyses

Analysis – Diurnal and AQI

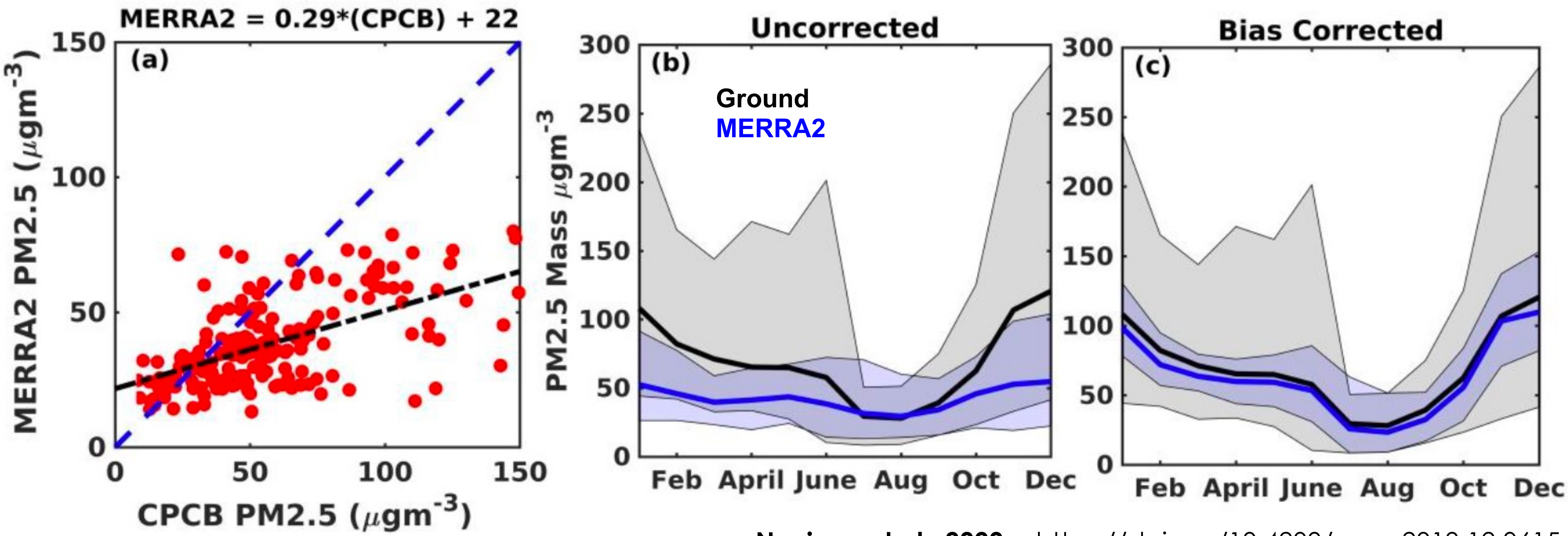
- Model outputs (forecasts or analysis) can also be evaluated for:
  - Air Quality Index (category)
  - 24-Hour Mean
  - Diurnal changes

Stay tuned to part 3 of this webinar for some examples



https://aqatmekong-servir.adpc.net/en/map/

Analysis – MERRA-2 Assessment Example

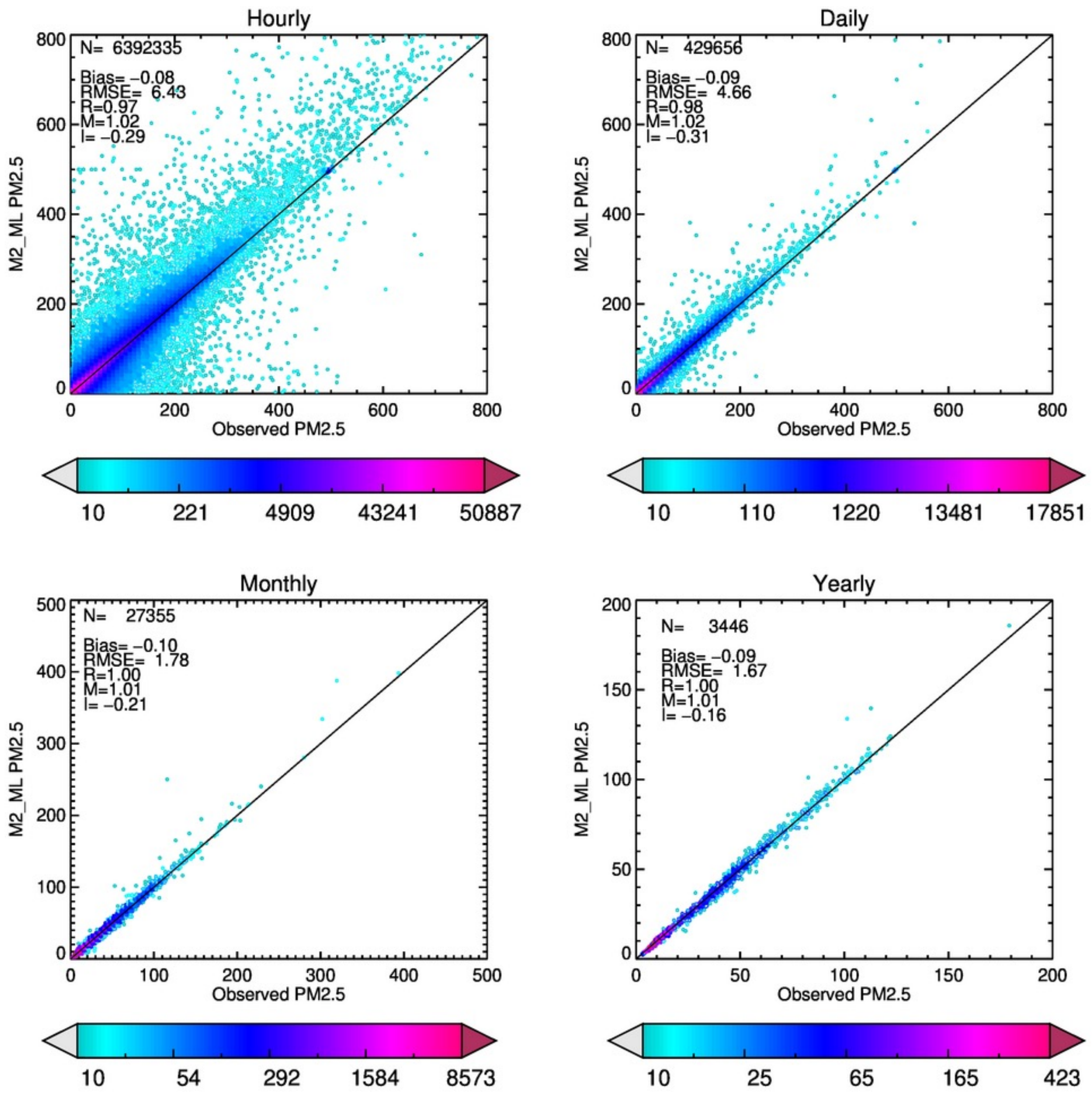


Navinya et al., 2020 - https://doi.org/10.4209/aaqr.2019.12.0615

Ground observations can be used to bias correct reanalysis and forecasts.

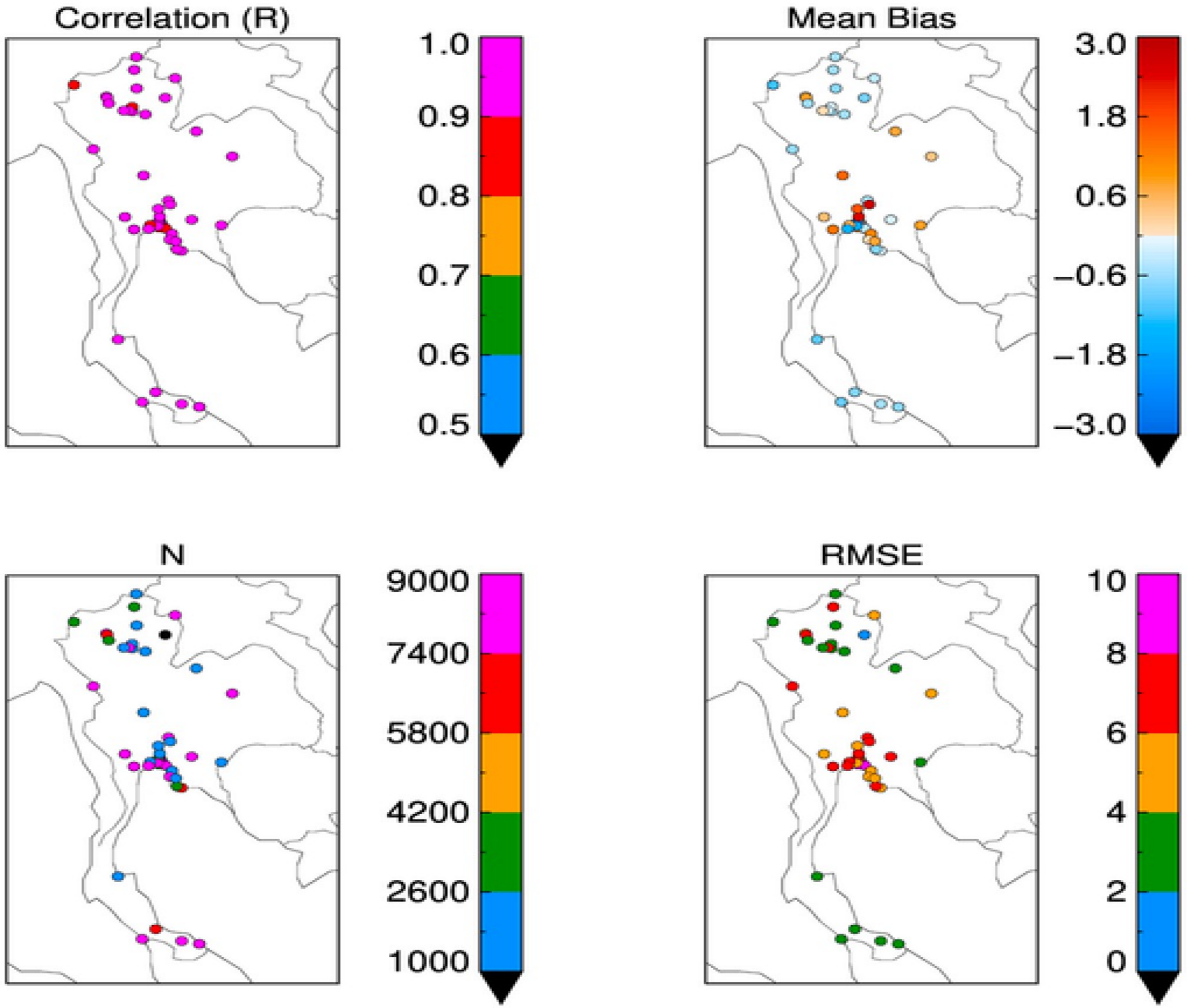
Analysis – Temporal Averaging

- Observations with high temporal resolution (e.g., hourly or daily) can be highly variable and noisy. But also critical for real-time air quality information.
- Averaging over longer time periods reduces variability in the data.
- Longer time averaging windows are best when examining long term trends or seasonal changes.



Analysis – Map the Statistics

- The model outputs can be evaluated over individual ground stations
- The comparison statistics can be mapped using color coded points
- Maps help understand the spatial distribution of model performance
- Maps can help identify specific patterns in model performance



Gupta et al., 2021, https://aaqr.org/articles/aaqr-21-05-oa-0105

Summary – Take Home

- NASA GOES Provides:
  - MERRA-2 – A reanalysis with satellite data assimilation
  - GEOS FP – 3-hourly aerosol forecast with satellite data assimilation system
  - GEOS-CF – Hourly forecast of PM2.5 and trace gases using GEOS-Chem chemistry
- Multiple satellites provide valuable parameters to evaluate model outputs.
- There are several ground networks available for validating model outputs.
- Python tools provided here are just examples, and users must check their accuracy and modify them to fit their analysis needs.
- The analysis types are some more popular examples, and users can explore other potential analyses while evaluating model performance.
- **And finally, know your data, know the model, and most importantly, know your applications or science questions before beginning any analysis.**

Thank You!