Google and NASA Air Quality Partnership – A Collaboration Using GEOS-CF Data and Google Earth Engine

Callum Wayman[1,3], K. Emma Knowland[2,3], Christoph Keller[2,3], Joseph Ardizzone[1,3], Steven Pawson[3]

[1]Science Systems and Applications, Inc.

[2]Morgan State University

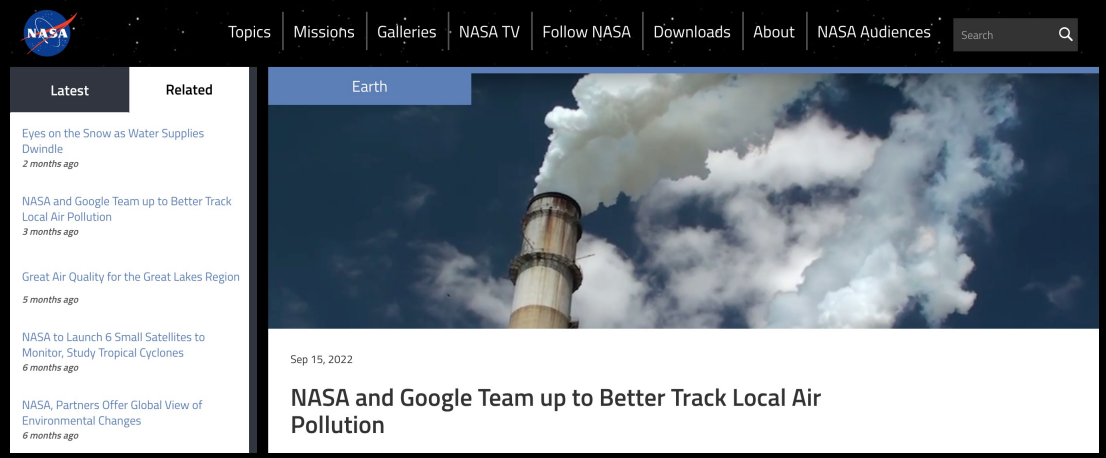
[3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center

Google NASA Air Quality Partnership

NASA and Google have expanded their partnership to create data and tools that help with pollution mitigation and decision making on a local government scale

The goal is to use the technologies available at NASA and Google to create city-scale data estimates and forecasts of air pollutants such as NO_2 derived from the GEOS Composition Forecast (GEOS-CF) model

Efforts are also being led by Pawan Gupta to create a downscaled MERRA-2 PM2.5 product



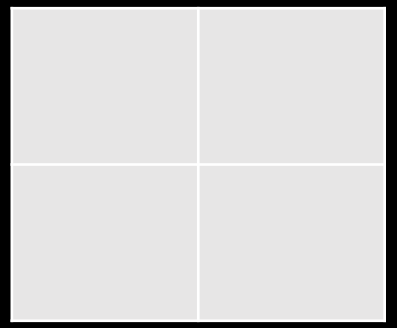
Featured on the NASA homepage!

Bias Correction and Downscaling – A Machine Learning Approach

PI: Pawan Gupta (USRA/MSFC, pawan.gupta@nasa.gov)

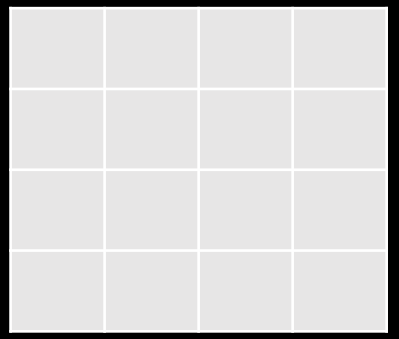
Inputs

MERRA2 GRID



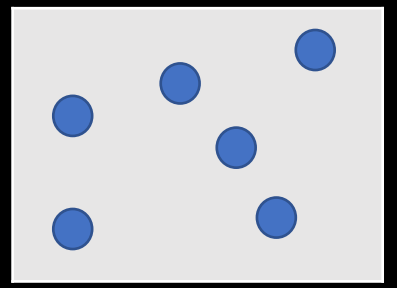
0.5x0.625
PM2.5,
AOD, Met
NetCDF

MODIS GRID



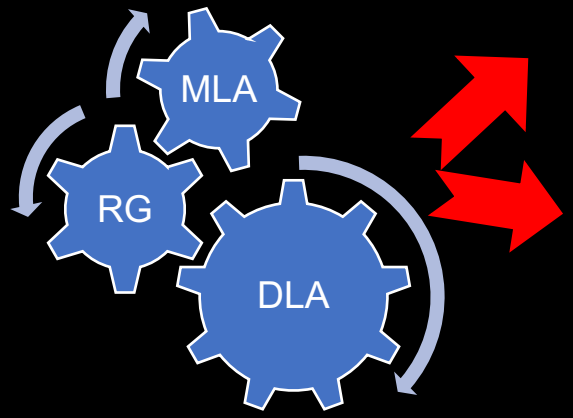
0.01x0.01
AOD
HDF

Ground Data



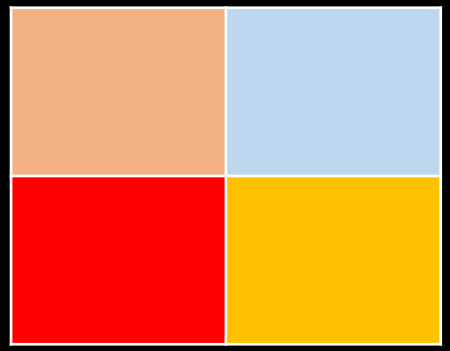
Points
PM2.5
csv

Statistical & AI (DL/ML) Models

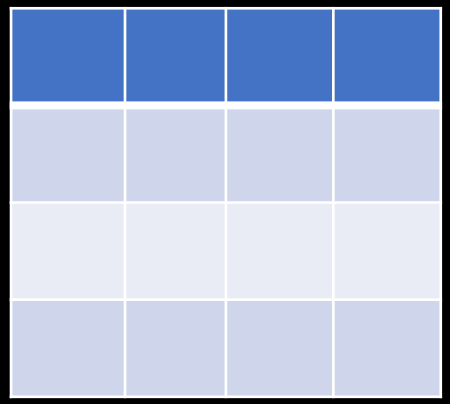


Target Data (SEDAC -1 km, AirNow Point Data)

Outputs



Bias Corrected PM2.5 at MERRA2 Grids

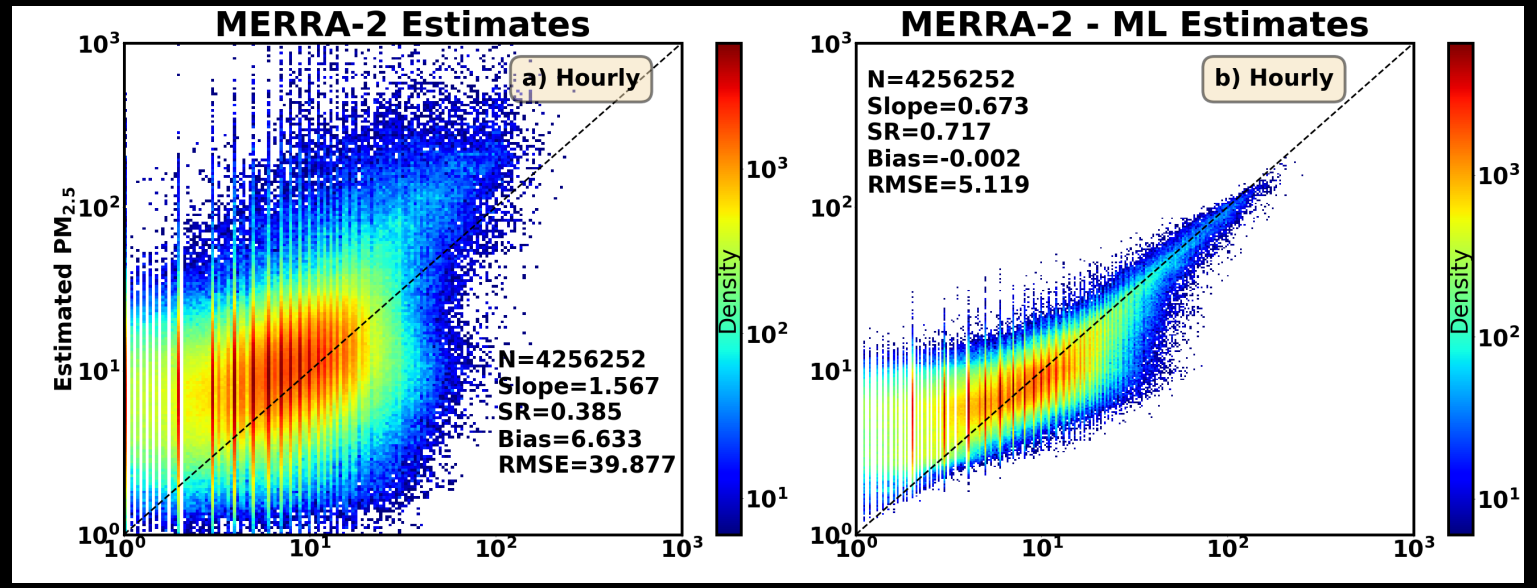


0.01x0.01
PM2.5
NetCDF,
Cloud
Optimized
GeoTiff

Eventually ML model and code should work on GEE while reading inputs from GEE (i.e. MERRA2, MCD19A2)

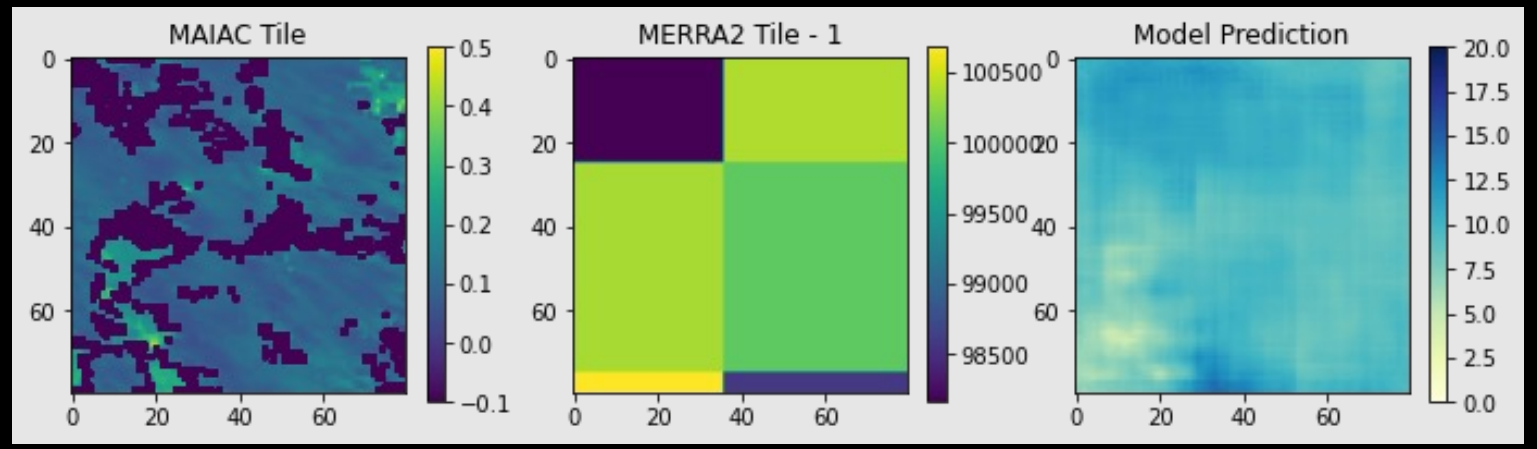
Preliminary Results

- Machine Learning Approach reduces the bias in PM2.5 w.r.t. EPA's measurements (Sayeed et al., 2022, under revision) - https://www.essoar.org/doi/abs/10.1002/essoar.10511086.1



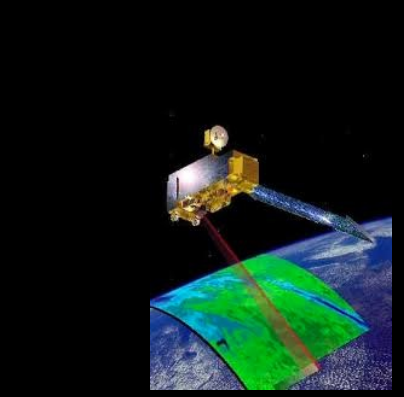
Lead: Alqmah Sayeed (USRA/MSFC)

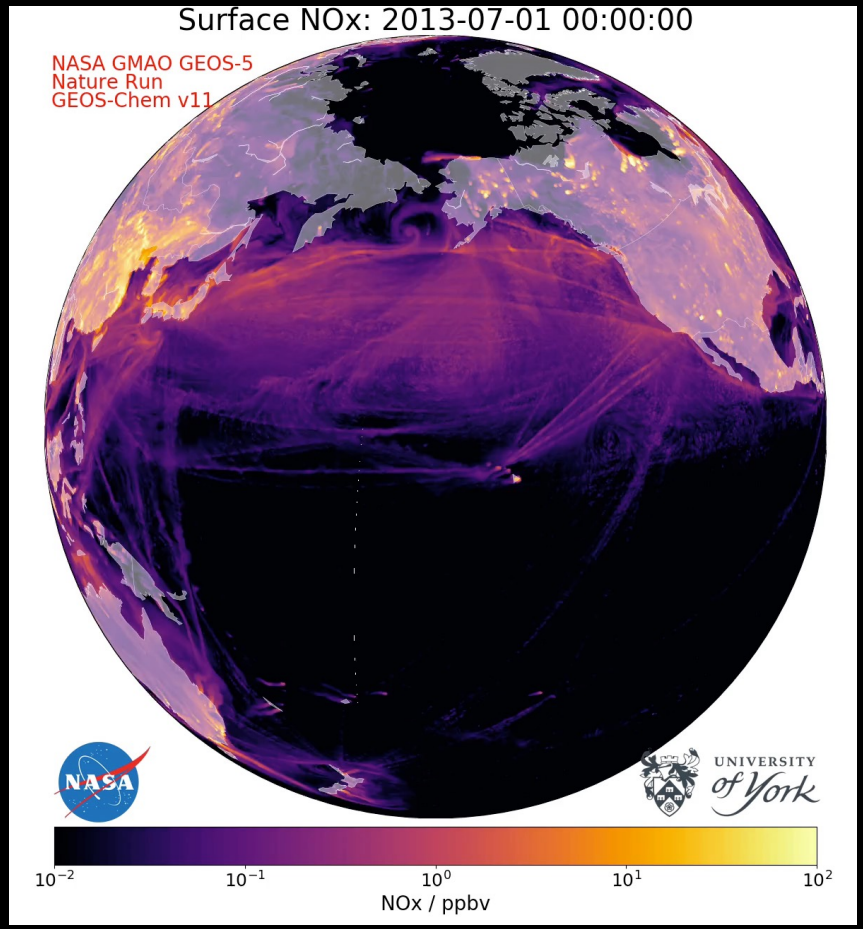
- The coarse resolution MERRA2 PM2.5 (middle panel) is downscaled to a higher resolution (last panel) using satellite aerosol data (first panel) – preliminary results



Lead: Michael Von Phole (USRA/Ames)

We can now simulate atmospheric composition with unprecedented level of detail

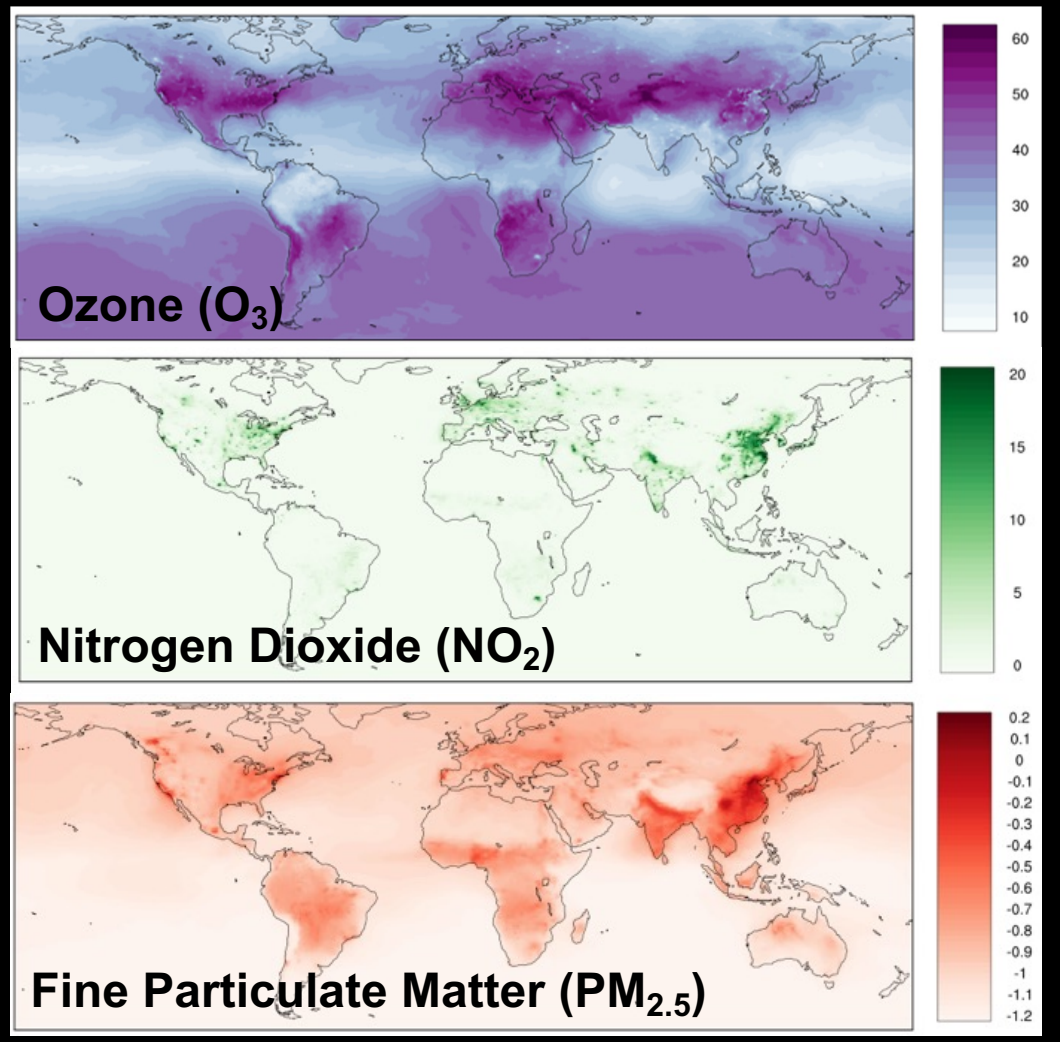




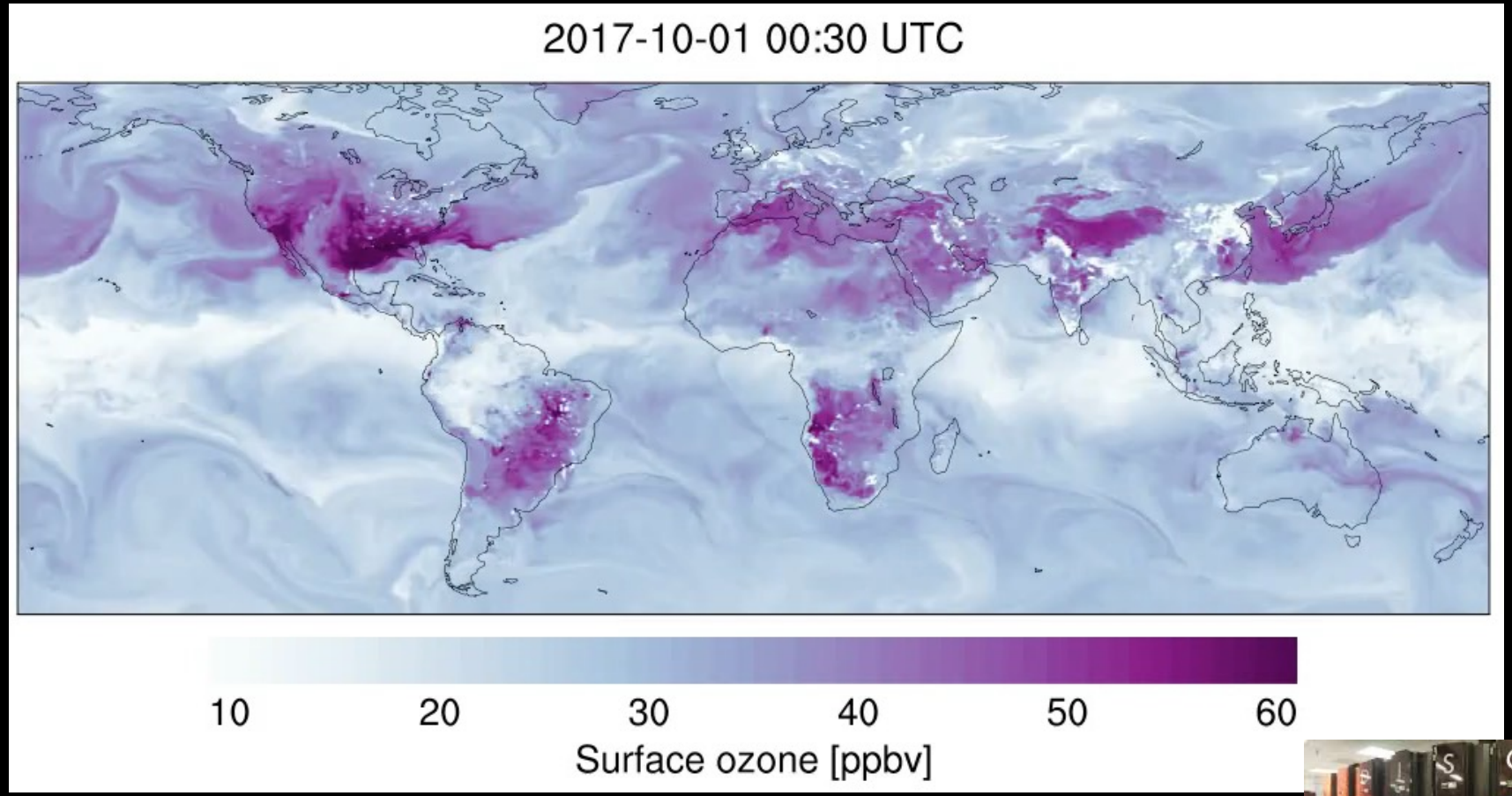
NASA GEOS Composition Forecast

GEOS Weather

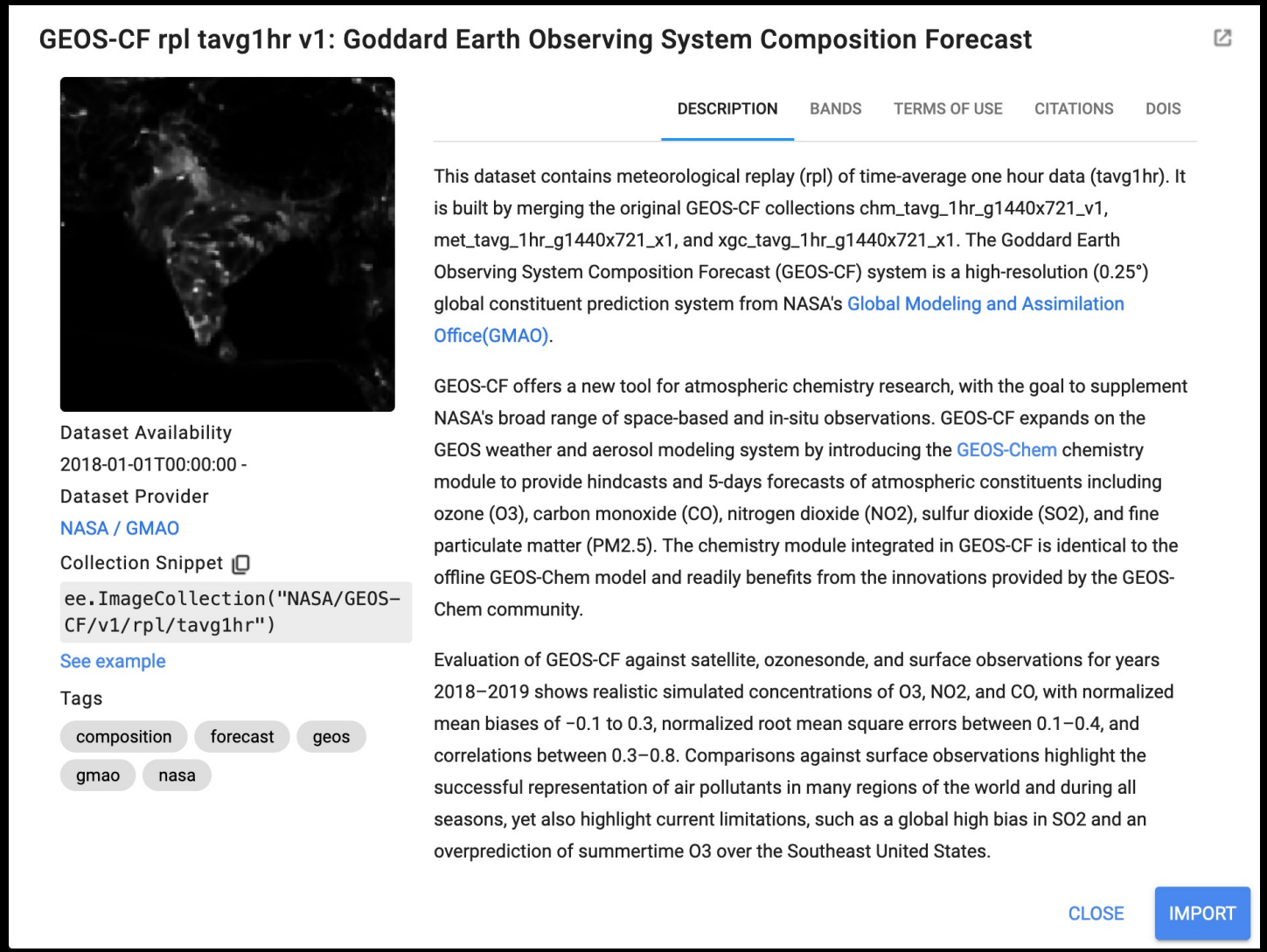
GEOS Chemistry



Producing the GEOS-CF forecasts is computationally expensive!



- 56 million grid cells (25x25 km^2, 72 levels), 250 chemical species
- ~3500 CPUs, 1 simulation days / hour, 1TB of output per day



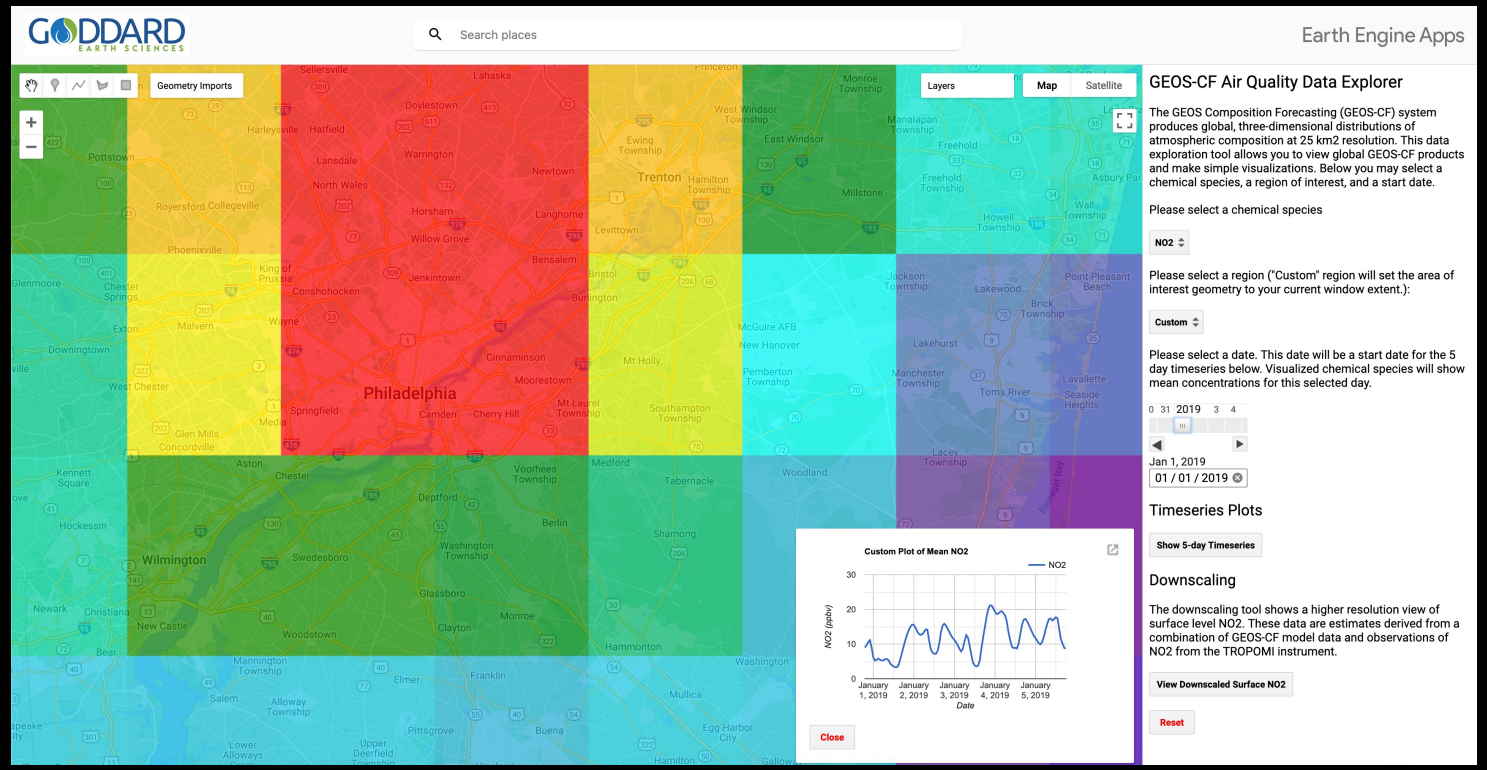
Accessing GEOS-CF through Google Earth Engine

Google Earth Engine (GEE) provides a rich open-access data repository to users

The data catalog contains metadata and documentation that allows anyone to manipulate and visualize data sets

The NASA and Google AQ partnership is providing a framework for GMAO data to be made available on GEE

Accessing GEOS-CF through Google Earth Engine



https://callumwayman.users.earthengine.app/view/geoscfexplorer

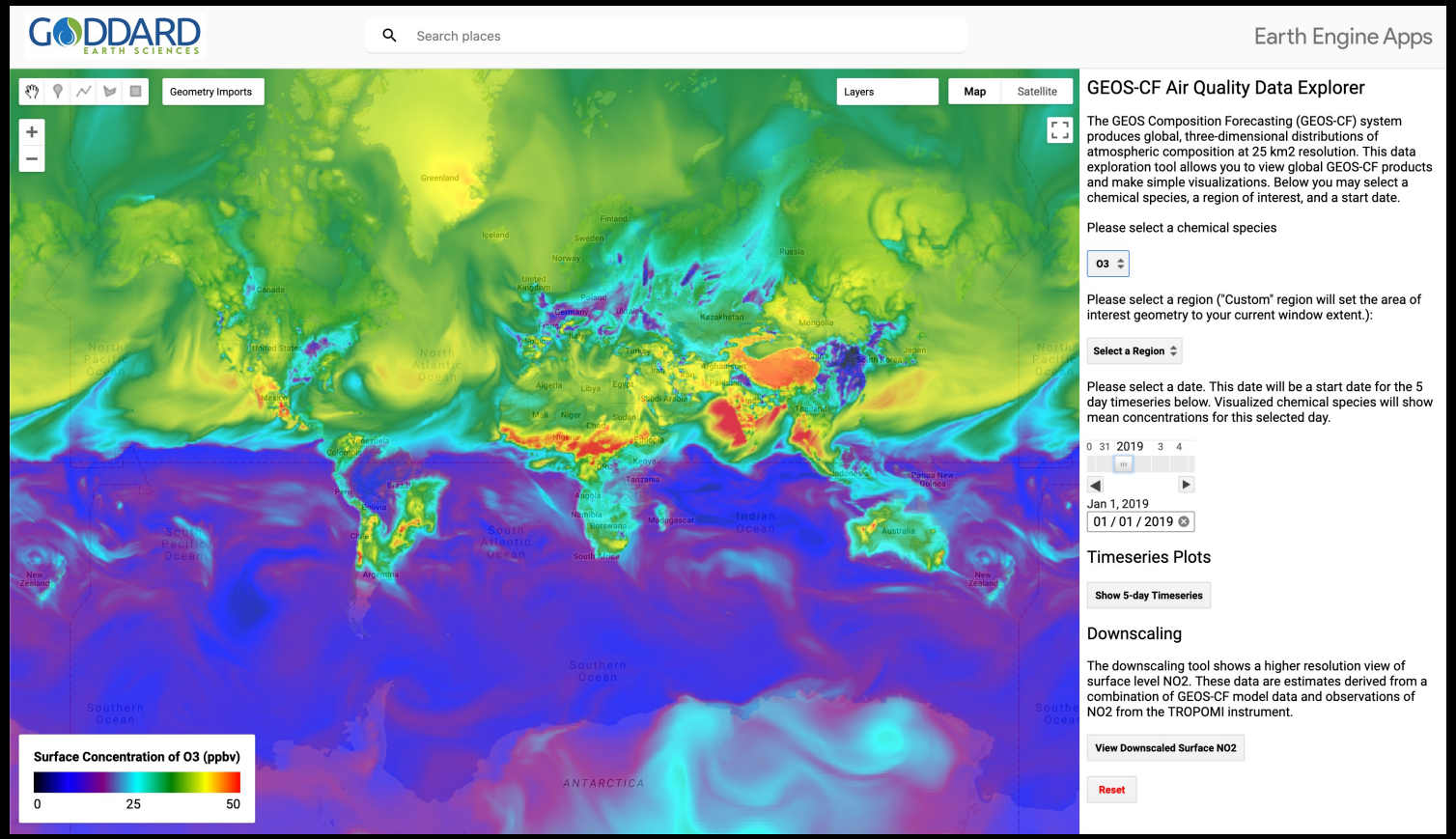
Currently the GEOS-CF 15-minute high-temporal frequency (“htf”) replay collection and the hourly time-averaged (tavg1hr) replay collection have been ingested in the GEE data repository

These collections consist of surface and column chemistry and meteorology fields

The htf and tavg collections are both publicly available through the Earth Engine data repository

Our partners at Google are currently ingesting forecast data sets which will be used for various machine learning applications

Dynamic Applications with GEE



https://callumwayman.users.earthengine.app/view/geoscfexplorer

GEE applications can be produced and easily exported

Dynamic interactive web-map tools will allow for greater flexibility in the ways in which the GMAO can provide data

Open-source data provides opportunities for anyone to produce tools using GMAO data

We are incorporating GMAO data and TROPOMI observations into these applications to provide localized surface level estimates of NO_2 with downscaling techniques

TROPOMI Observed Air Quality

https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-5P/Introducing_Sentinel-5P

The TROPOMI instrument is the single payload on the Sentinel-5P spacecraft

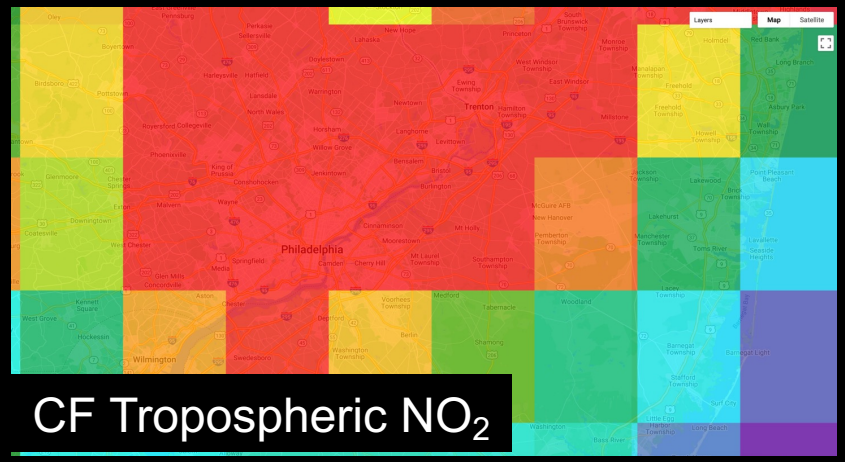
This instrument is dedicated to a mission of atmospheric monitoring

TROPOMI observations of tropospheric NO2 have an approximate 7x3.5 km^2 pixel resolution

https://sentinel.esa.int/web/sentinel/user-guides/sentinel-5p-tropomi

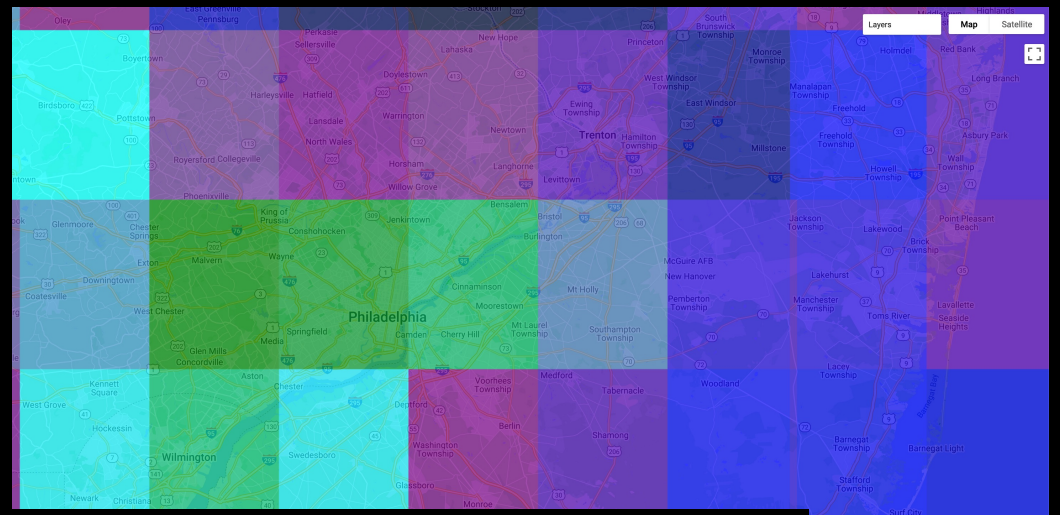
Using Observations with Model Data to Downscale GEOS-CF

CF Surface NO_2



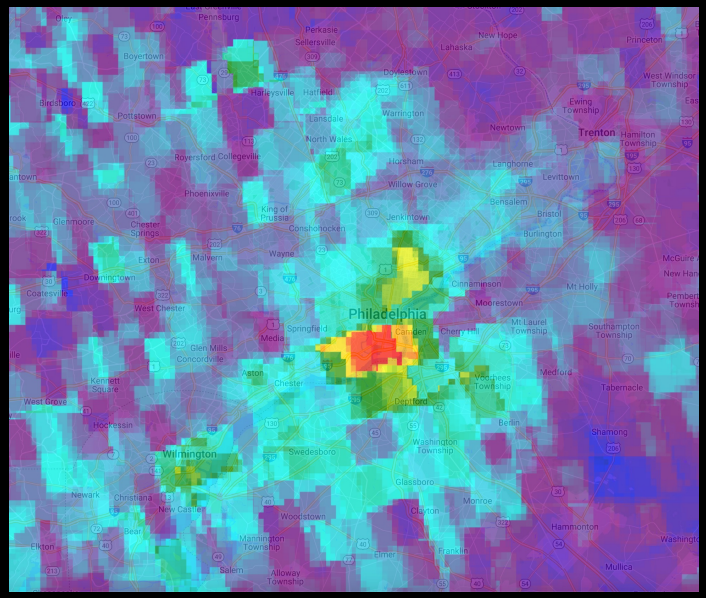
CF Tropospheric NO_2

=



CF Surface/Troposphere NO_2 Ratio

Using Observations with Model Data to Downscale GEOS-CF

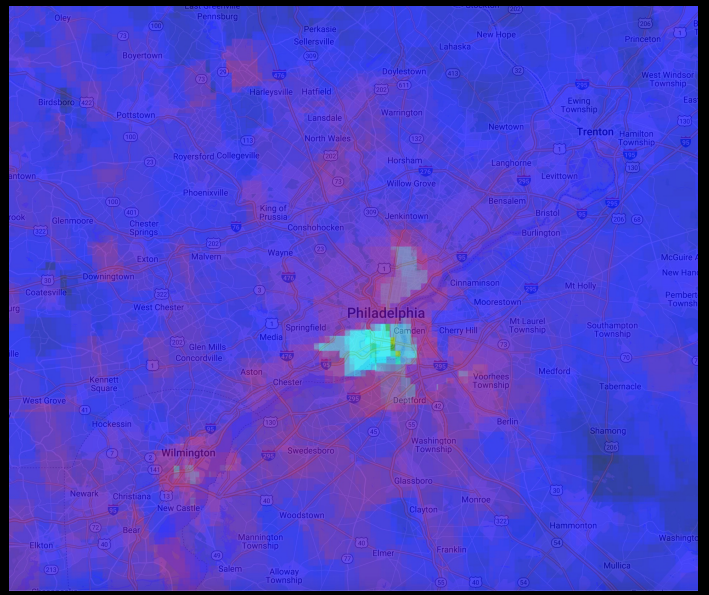


TROPOMI Tropospheric NO_2

*

CF Surface/Troposphere NO_2 Ratio

=

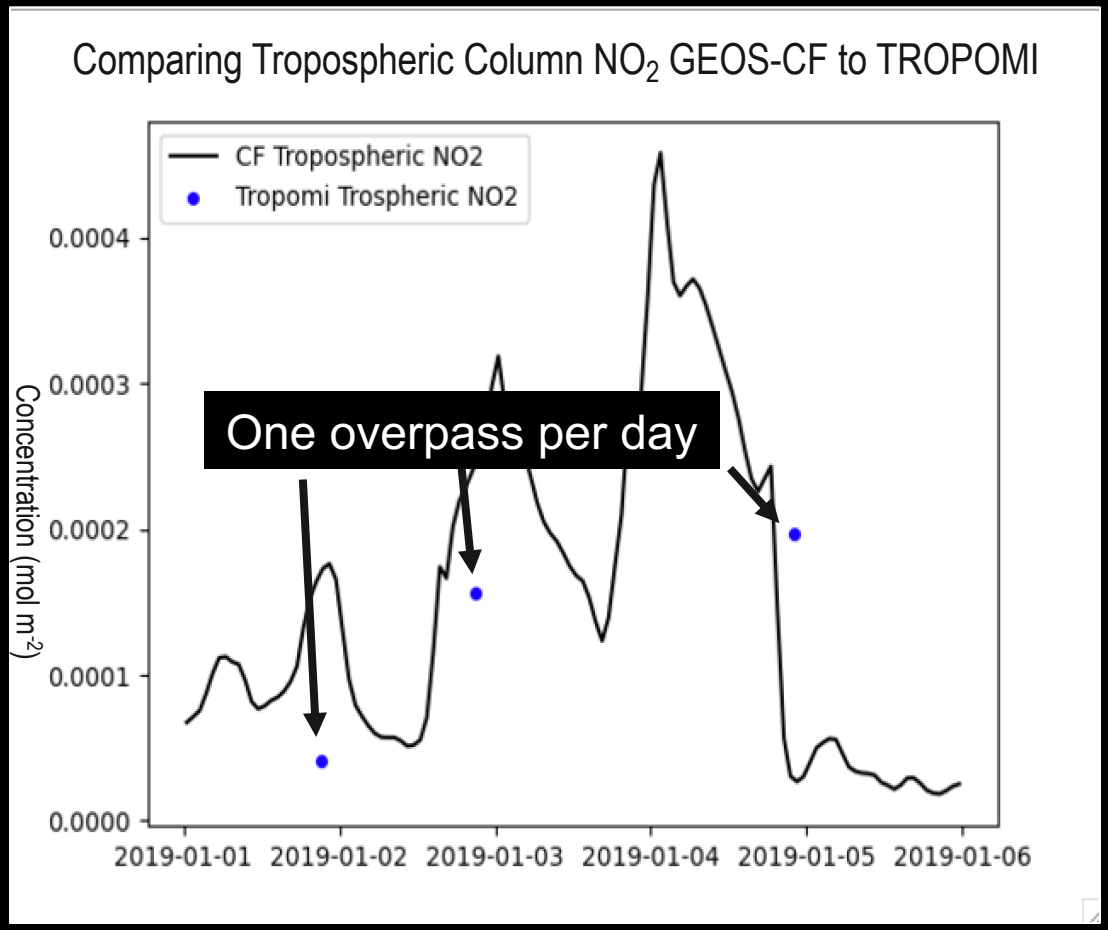


Surface Scaled TROPOMI NO_2

Using Earth Engine Built-In Machine Learning

Google Earth Engine has built-in machine learning algorithms such as the random forest classifier

Our next step is to incorporate other data sets into machine learning models to downscale GEOS-CF imagery on time steps without TROPOMI overpasses



Incorporating Proprietary Data

After incorporating GEOS-CF meteorology and TROPOMI data, Google proprietary data sets will be incorporated into the machine learning model

Google has developed some rich data sets which provide point source air quality concentrations using Google Street View vehicles in several cities around the world

In addition, the Open Buildings data set and the National Landcover Dataset may be used to better understand how buildings, pieces of infrastructure, and changing landuse are contributing to atmospheric NO2 concentrations

Summary

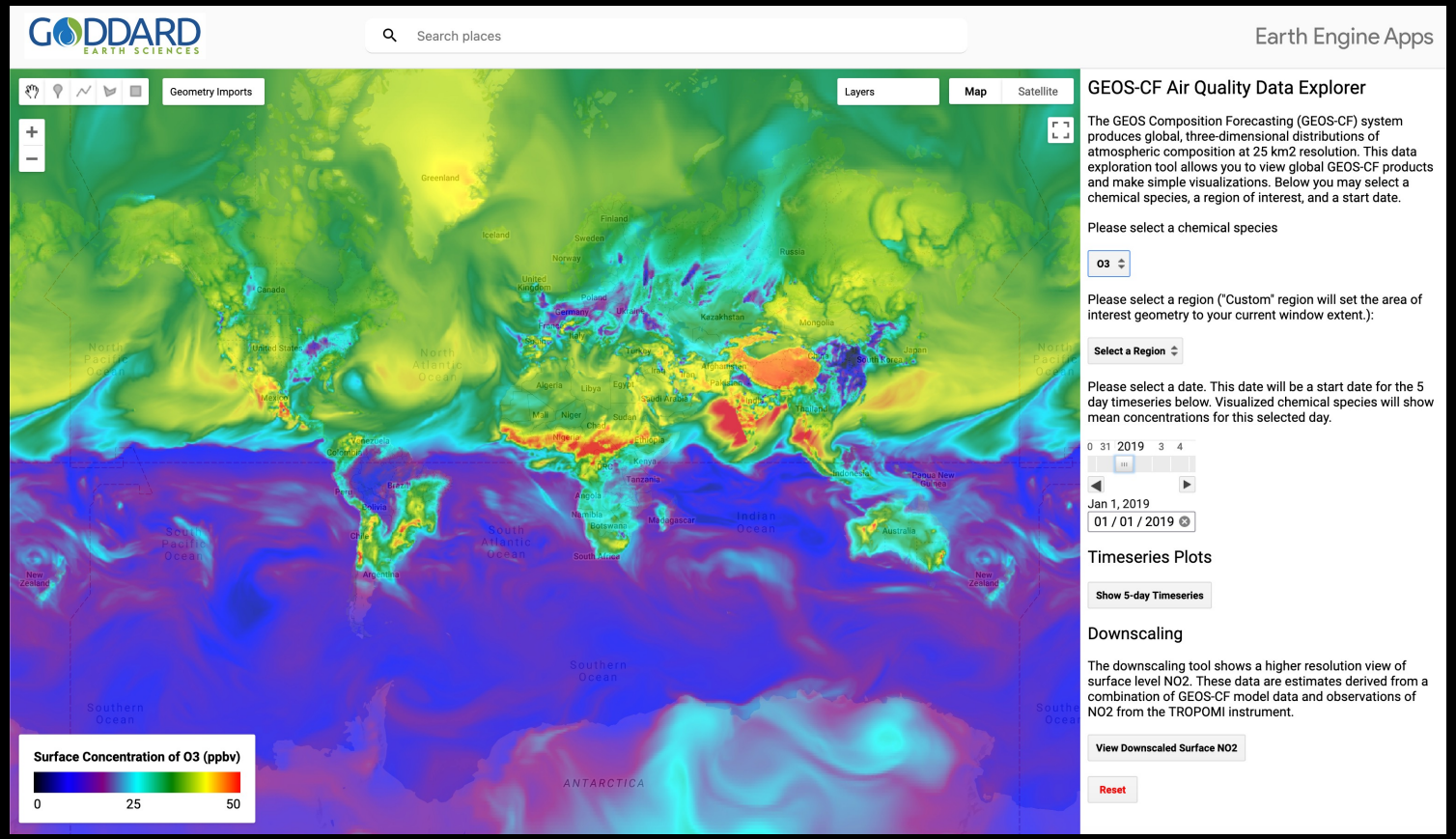
The combined data infrastructures of Google and NASA provide an excellent foundation for new tools and applications

Simple Ratio techniques have been used to produce higher spatial resolution surface estimates of NO_2

Next steps:

- Use machine learning within GEE to spatially downscale the GEOS-CF data to finer grids
- Access proprietary Google data at the street-level
  - This could improve ability to locate high concentrations of NO2 within cities.

Questions?



callum.r.wayman@nasa.gov