NASA GEOS Forecasting Capabilities for Air Quality

K. Emma Knowland

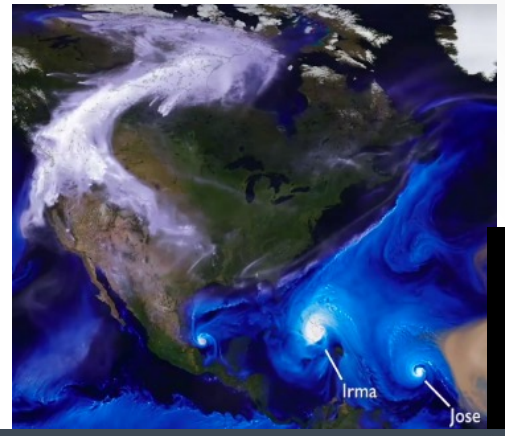
Morgan State University/GESTAR-II

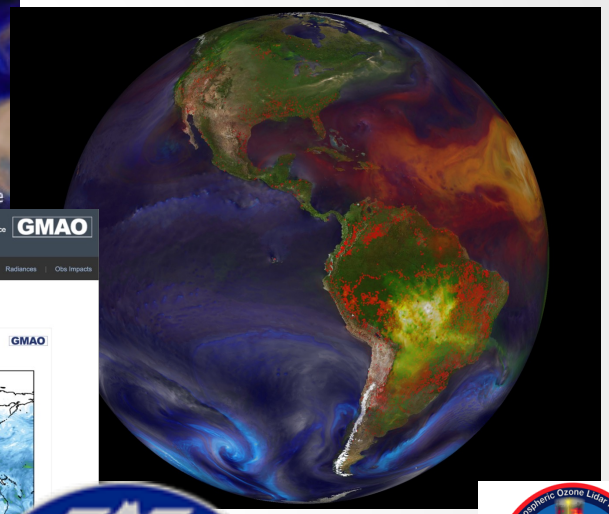
NASA Global Modeling and Assimilation Office (GMAO)

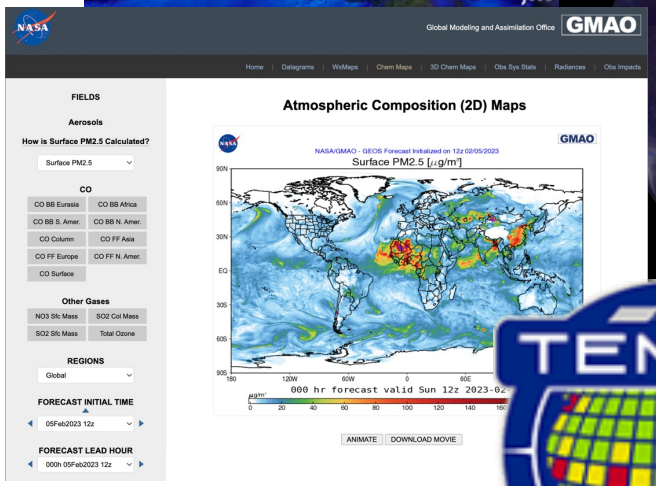
In collaboration with Christoph Keller, Carl Malings and many other scientists from GMAO and other labs at NASA Goddard Space Flight Center

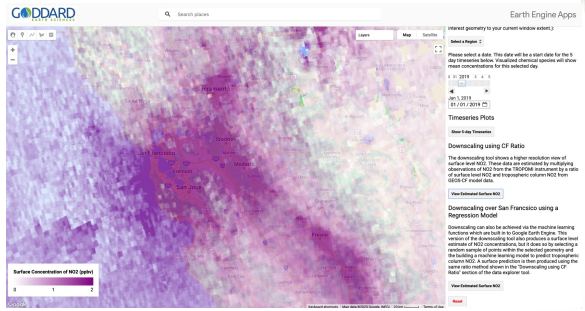
OUTLINE

1. Overview of GEOS current capabilities
   - Numerical Weather Prediction
   - Composition Forecasting
2. How to access GEOS “big data” for research scientists and engaged community members
3. Bias-correcting techniques for Air Quality decision making









K. Emma Knowland
k.e.knowland@nasa.gov

GMAO's core mission is to enhance the value of NASA's observations to understand, analyze and predict changes in the physics, chemistry and biology of the Earth system

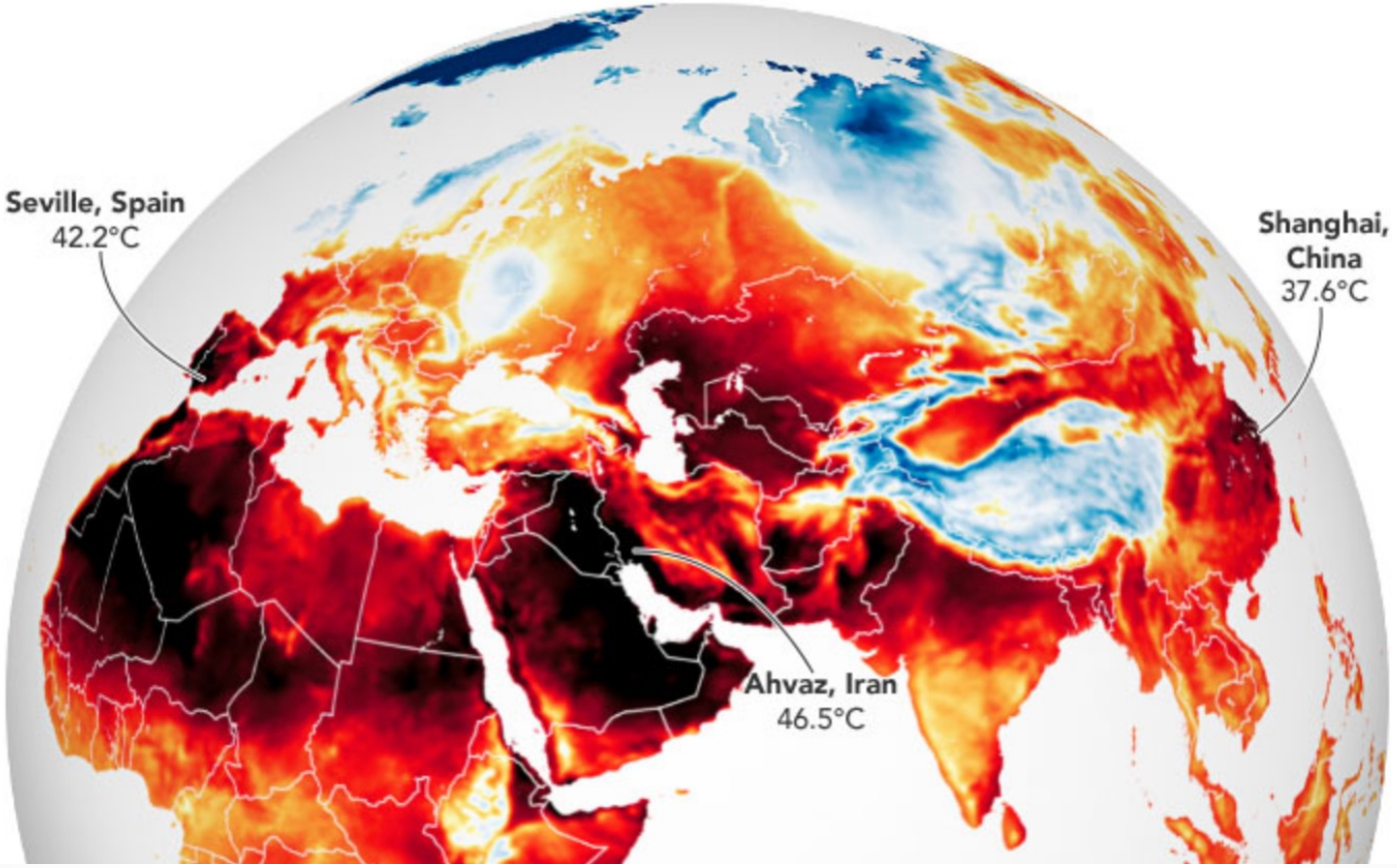
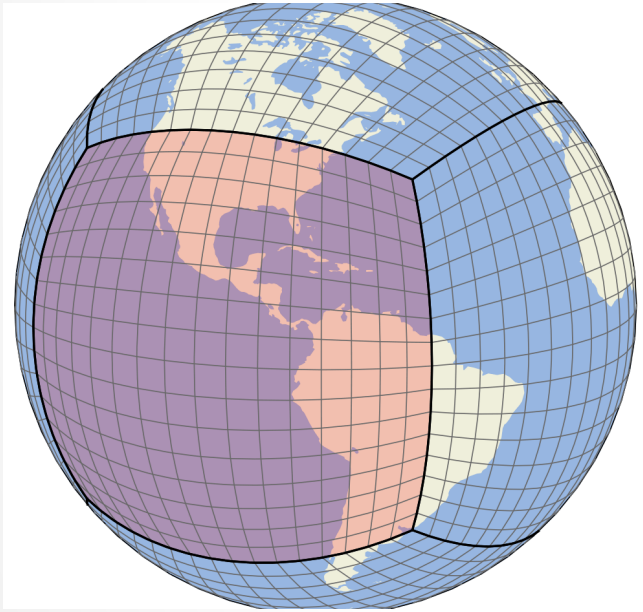


Image by Joshua Stevens,
NASA Earth Observatory
https://earthobservatory.nasa.gov/images/150083/heatwaves-and-fires-scorch-europe-africa-and-asia

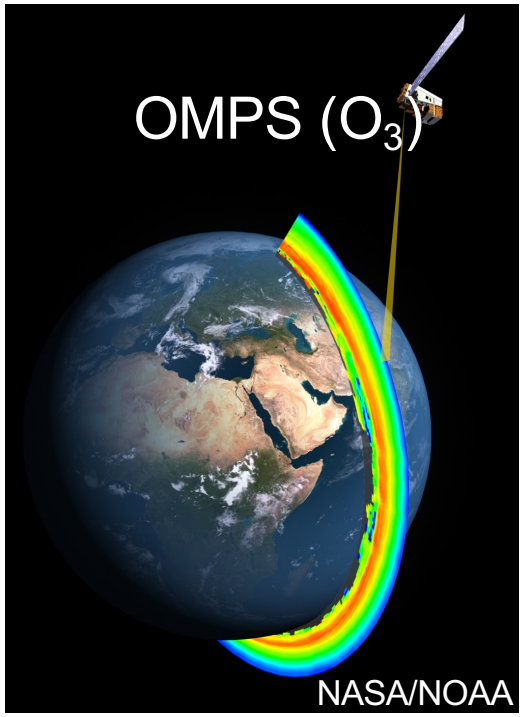
K. Emma Knowland
k.e.knowland@nasa.gov

NASA GMAO has a mature Earth System model

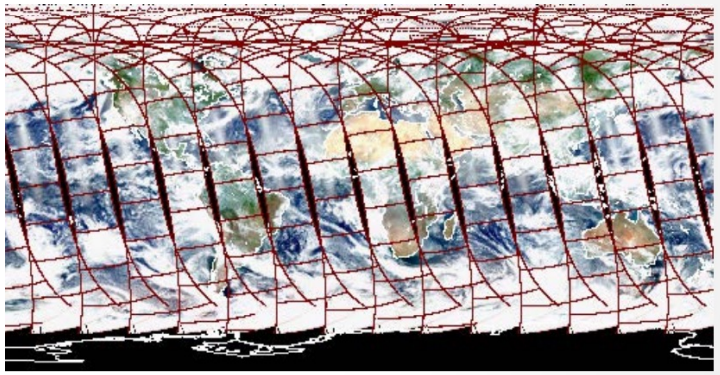
GEOS



Bindle et al., 2021 GMD



www.nasa.gov



https://modis.gsfc.nasa.gov/

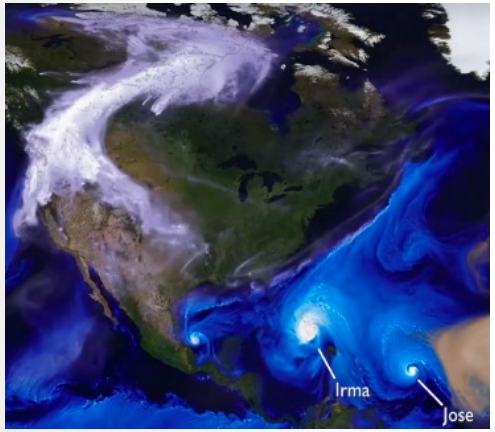
K. Emma Knowland
k.e.knowland@nasa.gov

NASA GMAO global meteorology and chemistry products

GEOS

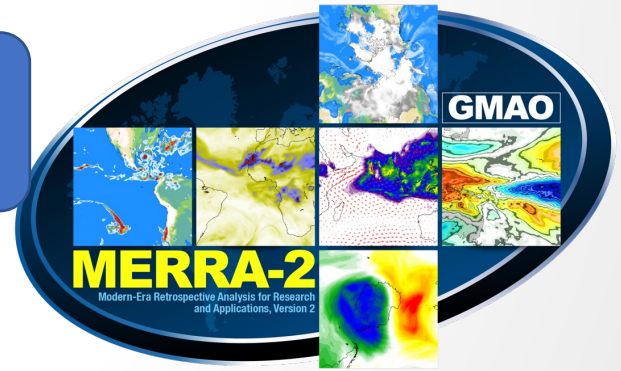
Forecasting

Reanalysis



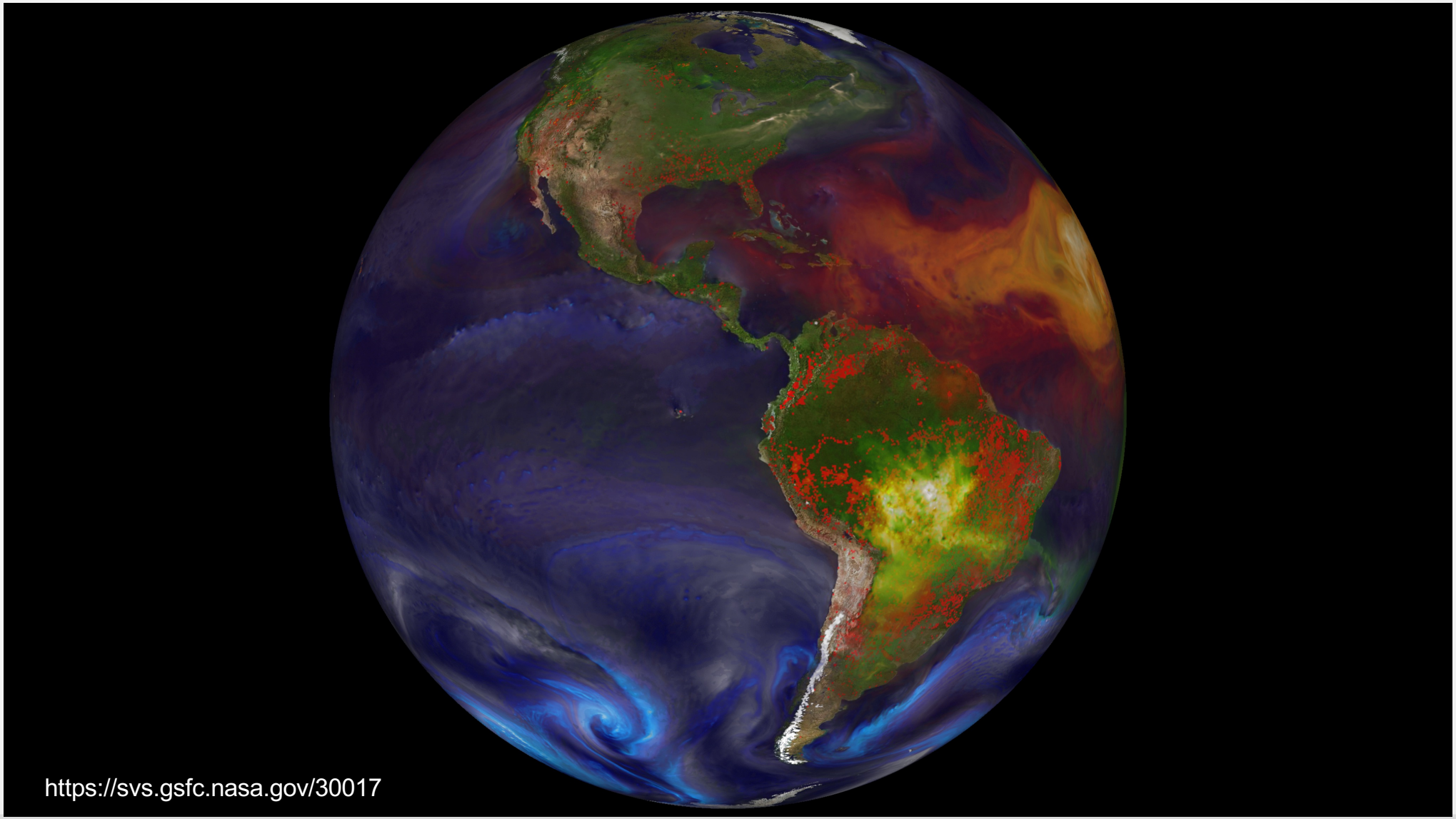
GEOS FP

MERRA-2



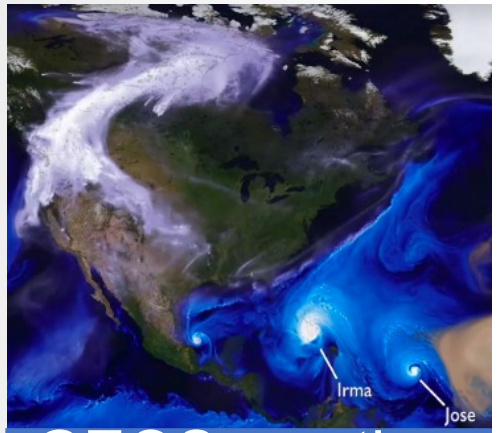
K. Emma Knowland
k.e.knowland@nasa.gov

GEOS NWP capabilities

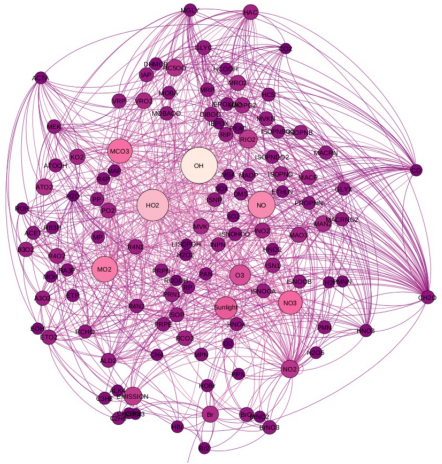


K. Emma Knowland
k.e.knowland@nasa.gov

GEOS Composition Forecast



GEOS weather forecasting



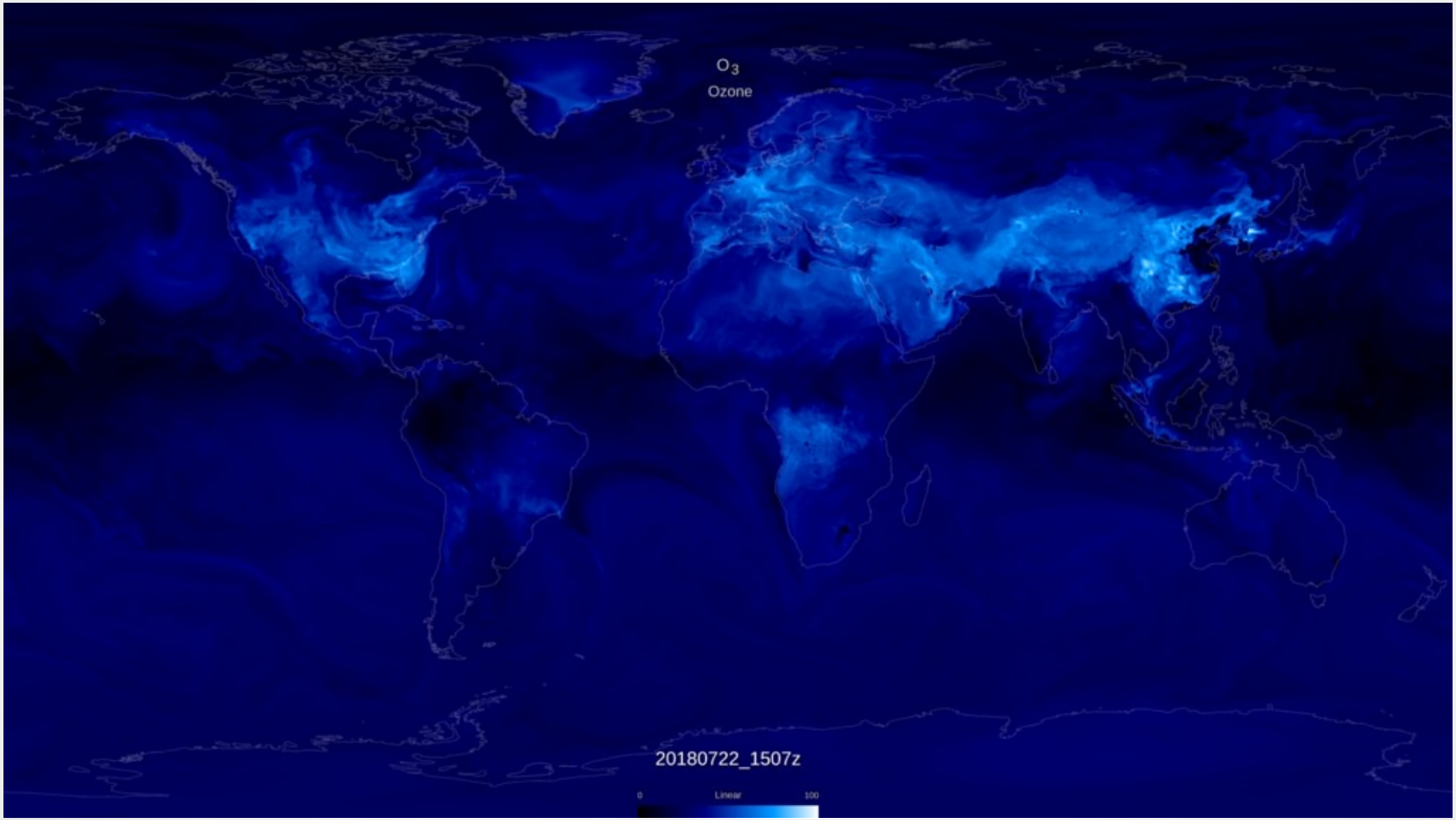
GEOS-Chem

GEOS - CF

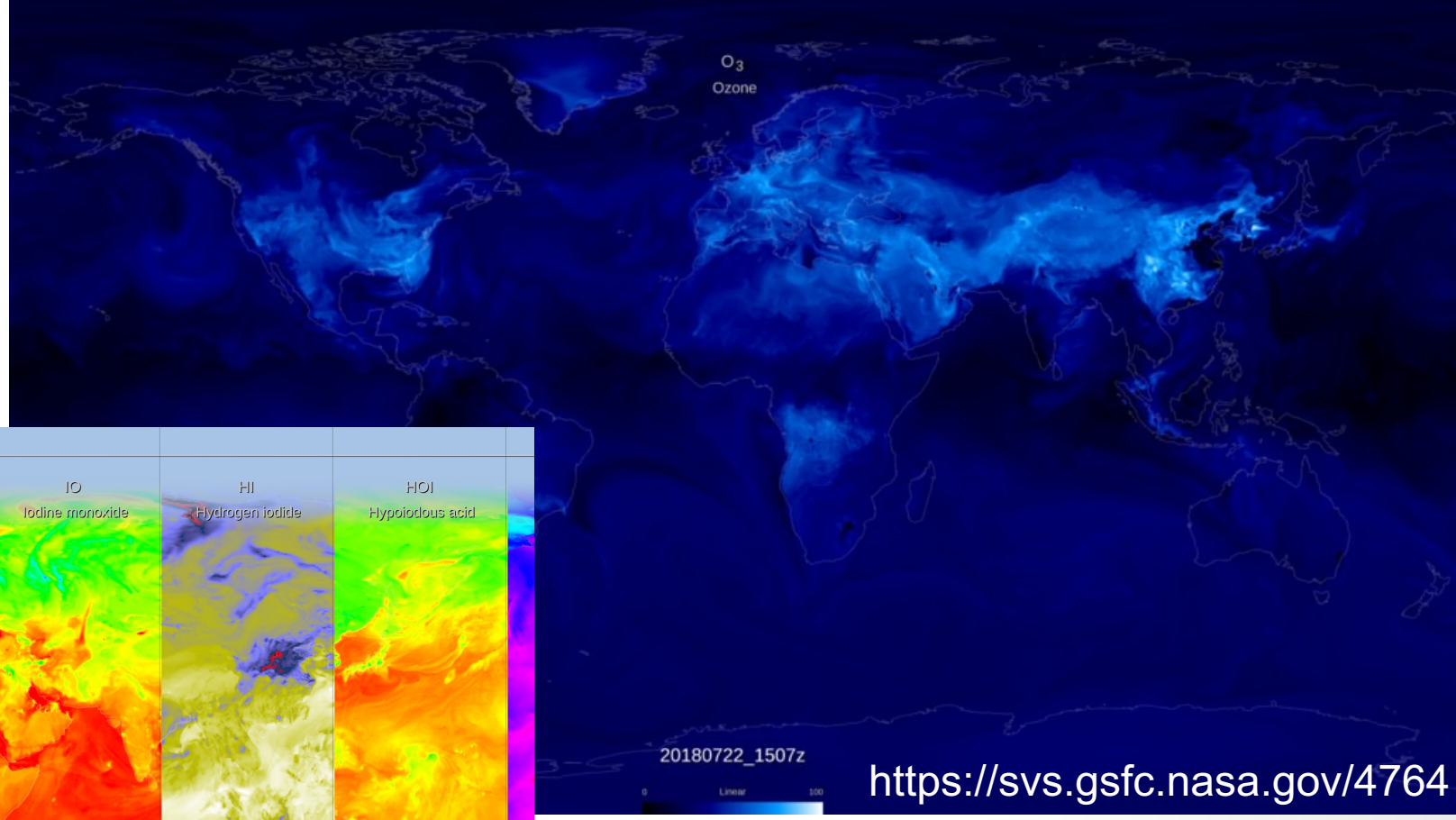
Version 12

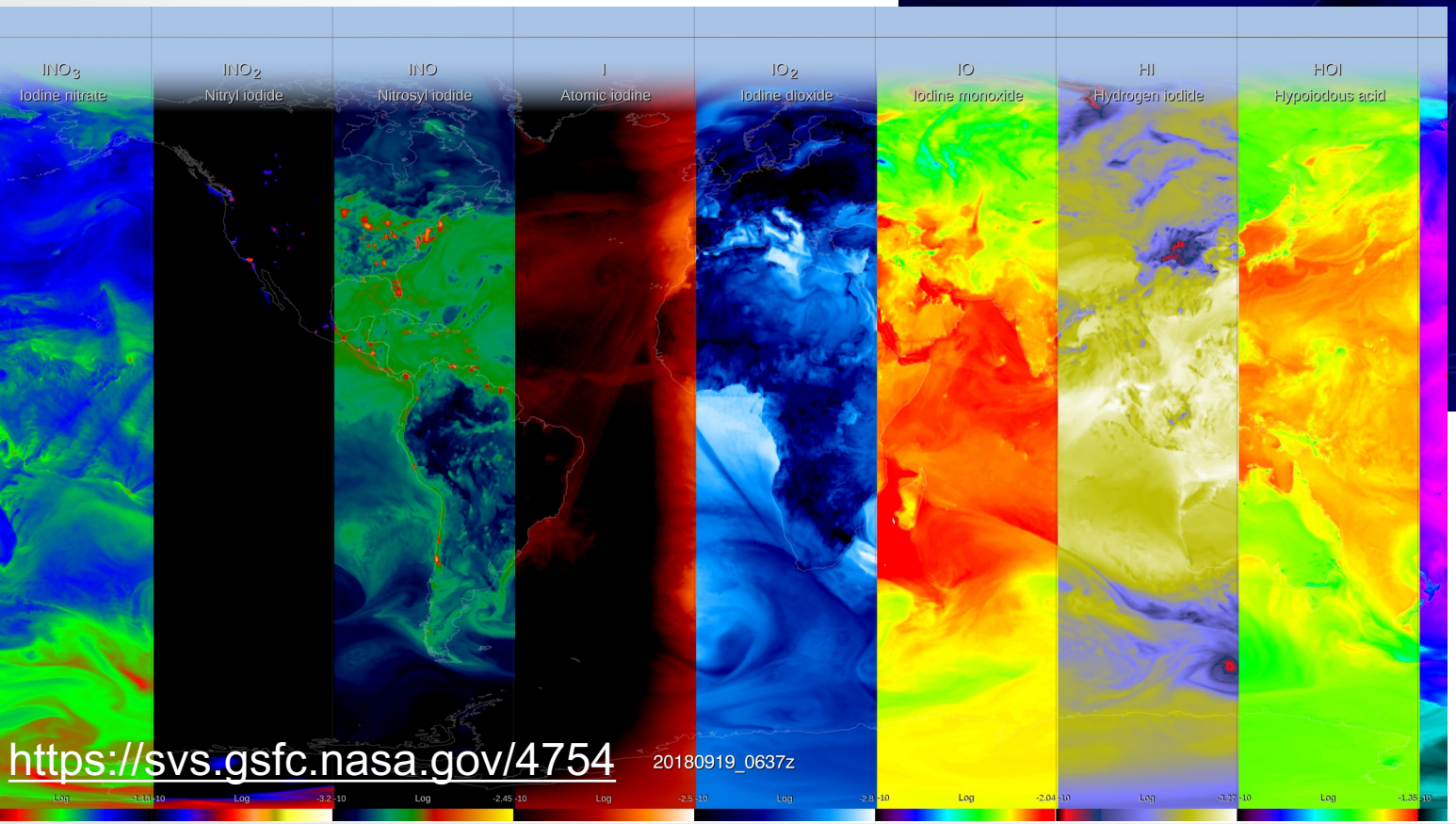
Tropospheric and Stratospheric chemistry

- 250 Chemical Species
- 725 Chemical Reactions


O_3
Ozone
20180722_1507z
0
Linear
100

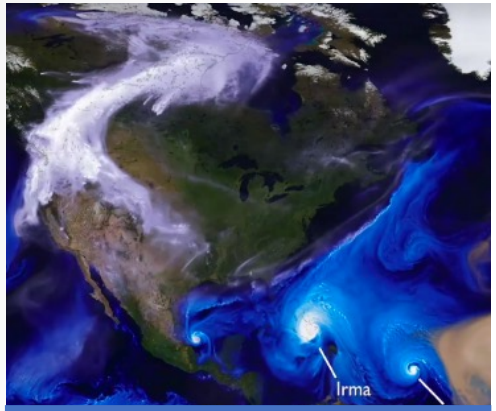
GEOS - CF



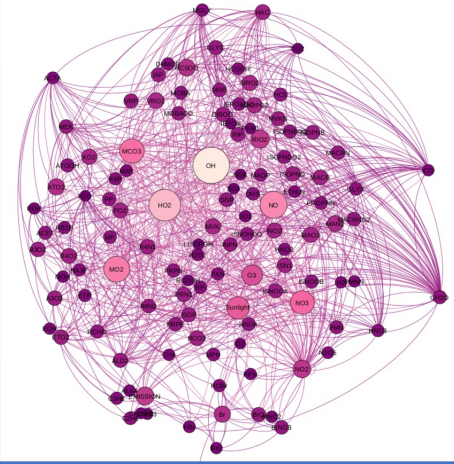


Global historical model estimates and daily 5-day forecasts of major air pollutants like Ozone & $PM_{2.5}$

Daily composition forecast



GEOS Weather



GEOS - Chem

GEOS - CF

One **5-day forecast** per day

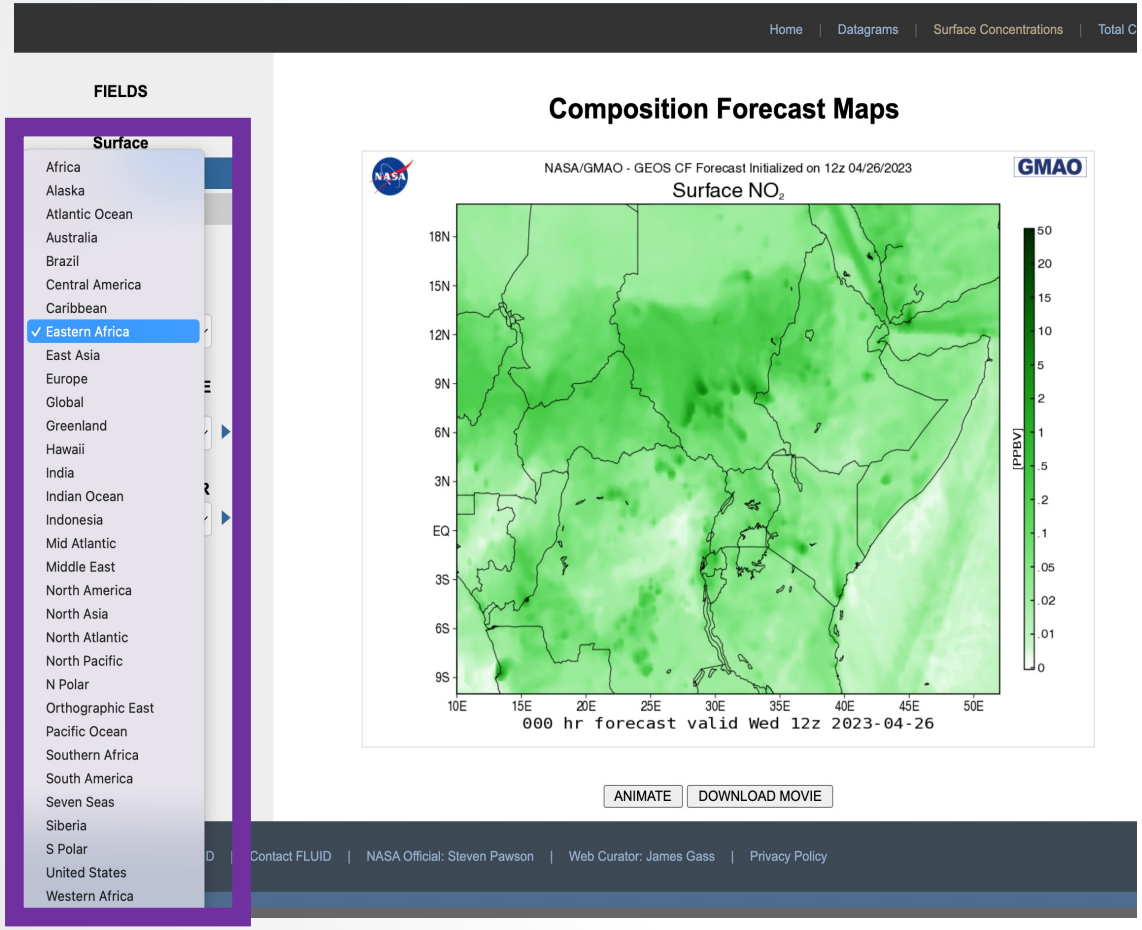
- 1-day simulation of previous 24 hours (“replay”)
- 5-day forecast
- c360 (0.25°, **~25x25 km²**) resolution, 72 model layers

- **15 minute** “surface”
- **1-hour** average and instantaneous 2D & 3D
- **Available since**
  - **1 January 2018** (replay)
  - **1 January 2019** (forecast)

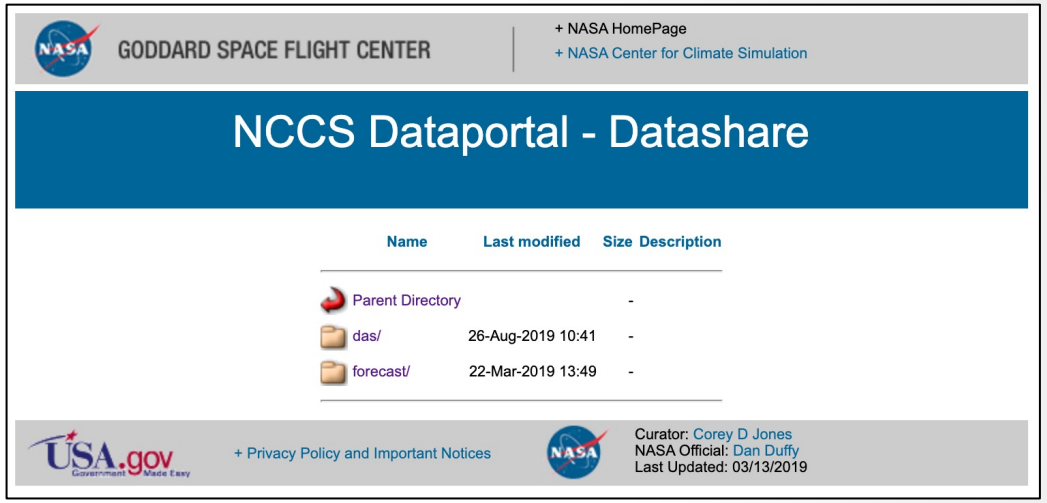
GEOS-CF output is available online in near real-time

Fluid is a mobile-friendly website

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



https://portal.nccs.nasa.gov/datashare/gmao/geos-cf/v1/



https://opendap.nccs.nasa.gov/dods/gmao/geos-cf/

GrADS Data Server - info for /gmao/geos-cf/assim/chm_tavg_1hr_g1440x721_v1 : dds das

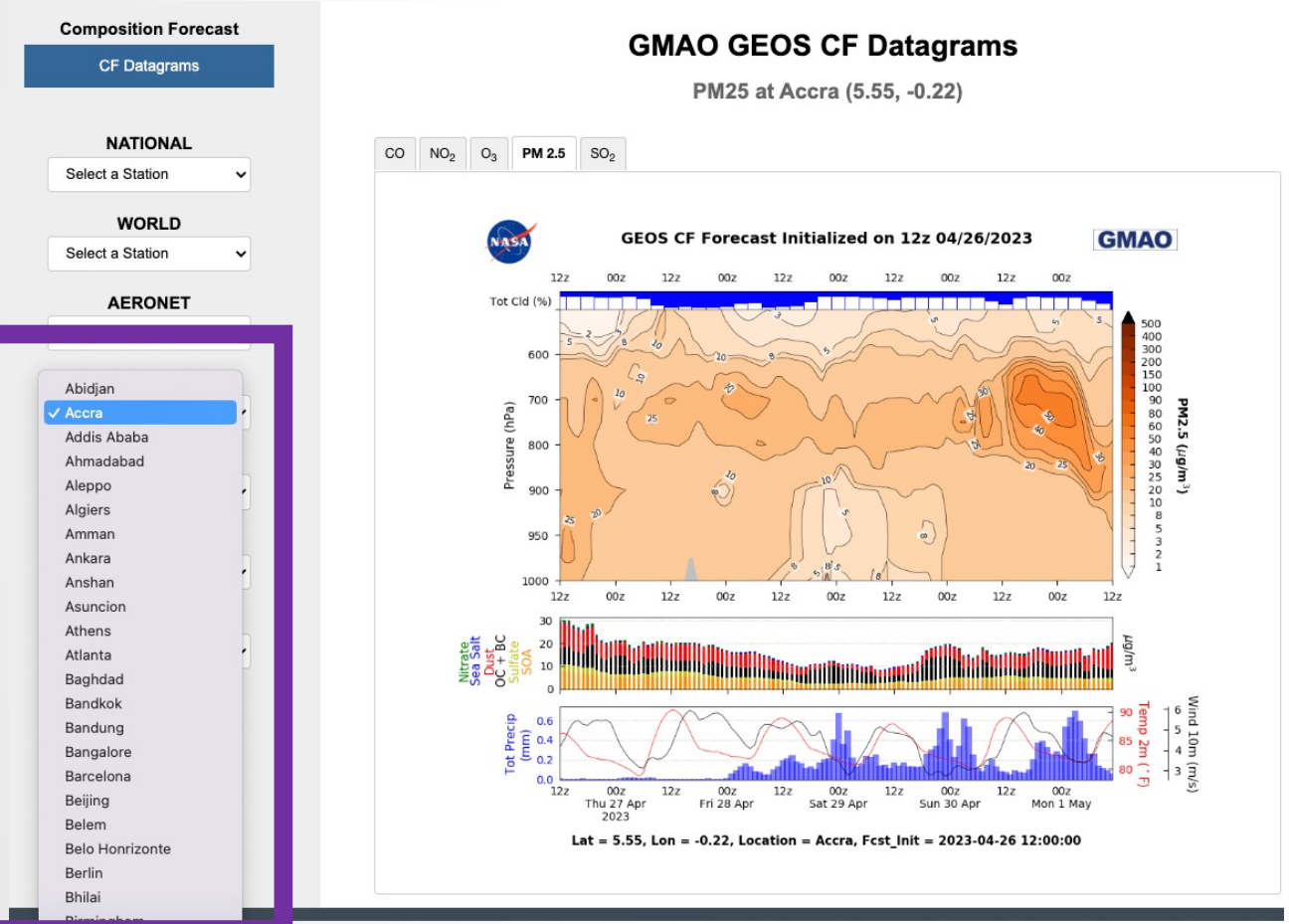
OPeNDAP/DODS Data URL: https://opendap.nccs.nasa.gov/dods/gmao/geos-cf/assim/chm_tavg_1hr_g1440x721_v1

Description:	GEOS CF (Composition Forecast)	
Documentation:	(none provided)	
Longitude:	-180.00000000000°E to 179.75000000000°E	(1440 points, avg. res. 0.25°)
Latitude:	-90.00000000000°N to 90.00000000000°N	(721 points, avg. res. 0.25°)
Altitude:	72.00000000000 to 72.00000000000	(1 points)
Time:	00:30Z01JAN2018 to 11:30Z31OCT2019	(16044 points, avg. res. 0.042 days)
Variables:	**(total of 52)**	
xyle	xylene (c8h10, mw = 106.16 g mol-1) volume mixing ratio dry air	
dst2	dust aerosol, reff = 1.4 microns (mw = 29.00 g mol-1) volume mixing ratio dry air	
hno4	peroxynitric acid (hno4, mw = 79.00 g mol-1) volume mixing ratio dry air	
pm25su_rh35_gcc	sulfate_particulate_matter_with_diameter_below_2.5_um_rh_35	

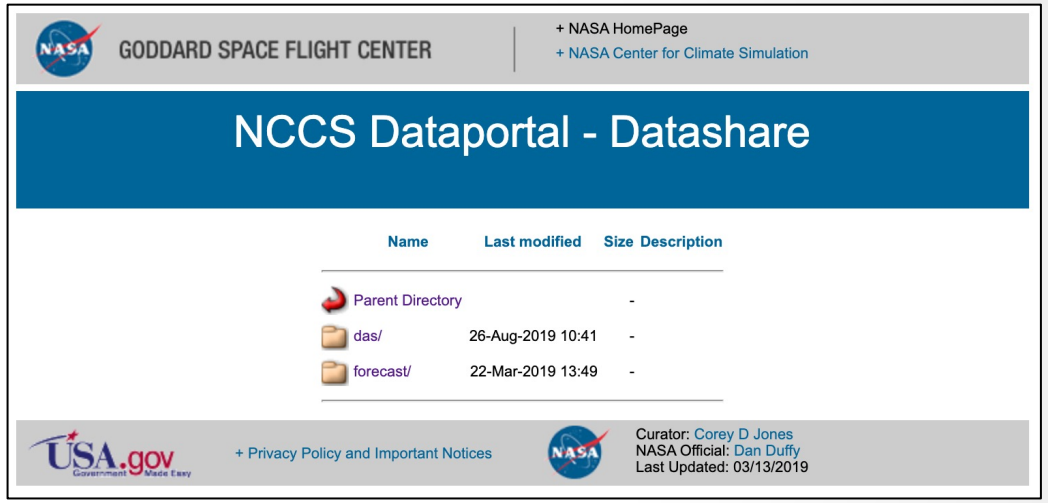
GEOS-CF output is available online in near real-time

Fluid is a mobile-friendly website

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



https://portal.nccs.nasa.gov/datashare/gmao/geos-cf/v1/



https://opendap.nccs.nasa.gov/dods/gmao/geos-cf/

GrADS Data Server - info for /gmao/geos-cf/assim/chm_tavg_1hr_g1440x721_v1 : dds das

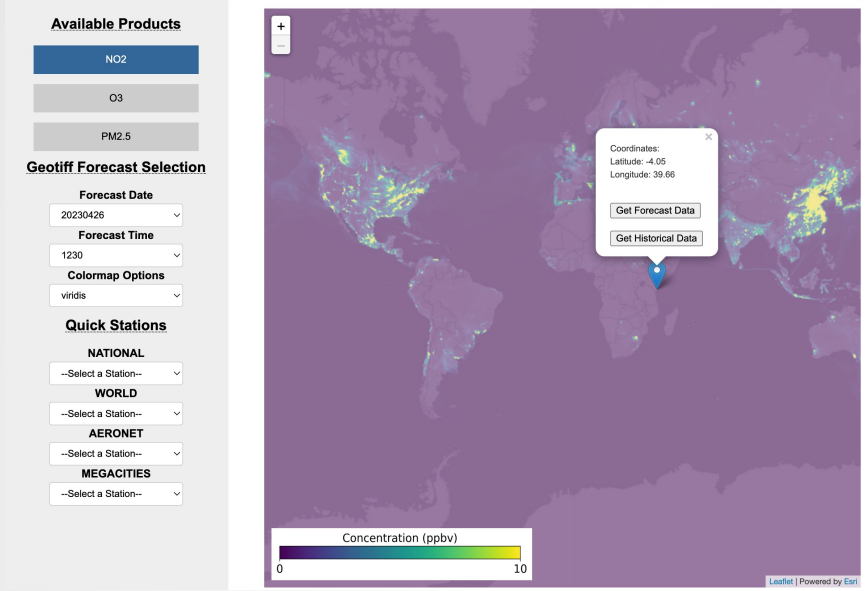
OPeNDAP/DODS Data URL: https://opendap.nccs.nasa.gov/dods/gmao/geos-cf/assim/chm_tavg_1hr_g1440x721_v1

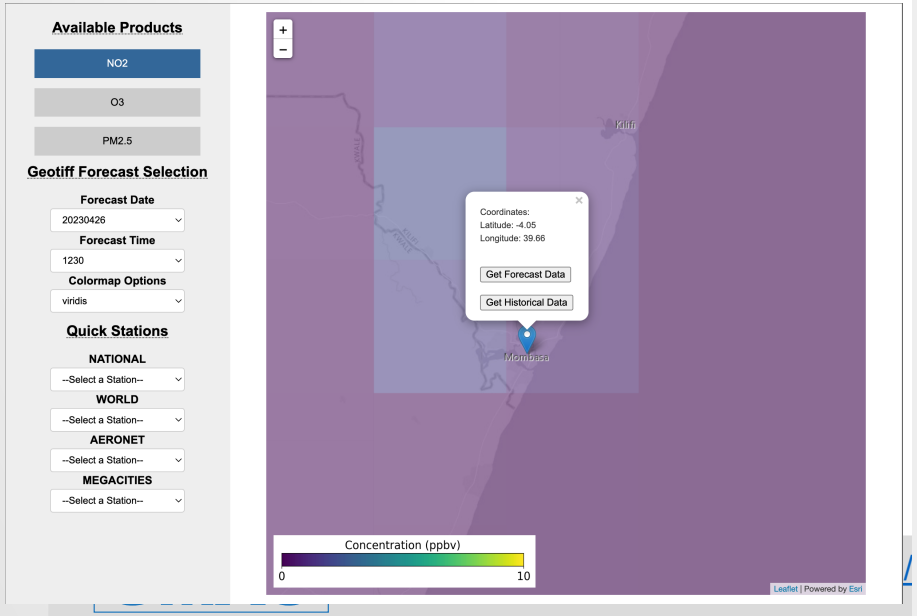
Description:	GEOS CF (Composition Forecast)	
Documentation:	(none provided)	
Longitude:	-180.00000000000°E to 179.75000000000°E	(1440 points, avg. res. 0.25°)
Latitude:	-90.00000000000°N to 90.00000000000°N	(721 points, avg. res. 0.25°)
Altitude:	72.00000000000 to 72.00000000000	(1 points)
Time:	00:30Z01JAN2018 to 11:30Z31OCT2019	(16044 points, avg. res. 0.042 days)
Variables:	**(total of 52)**	
xyle	xylene (c8h10, mw = 106.16 g mol-1) volume mixing ratio dry air	
dst2	dust aerosol, reff = 1.4 microns (mw = 29.00 g mol-1) volume mixing ratio dry air	
hno4	peroxynitric acid (hno4, mw = 79.00 g mol-1) volume mixing ratio dry air	
pm25su_rh35_gcc	sulfate_particulate_matter_with_diameter_below_2.5_um_rh_35	

GEOS-CF forecast imagery is available on-demand

National Aeronautics and Space Administration

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





Downloadable Data

Chemistry Data

NO_2 O_3 PM2.5

--Select File Format--

--Select Dataset--

Download Chem Data

Meteorology Data

--Select File Format--

Download Met Data

Other Data Sets

Pressure Level Plot

Historical CF Plot

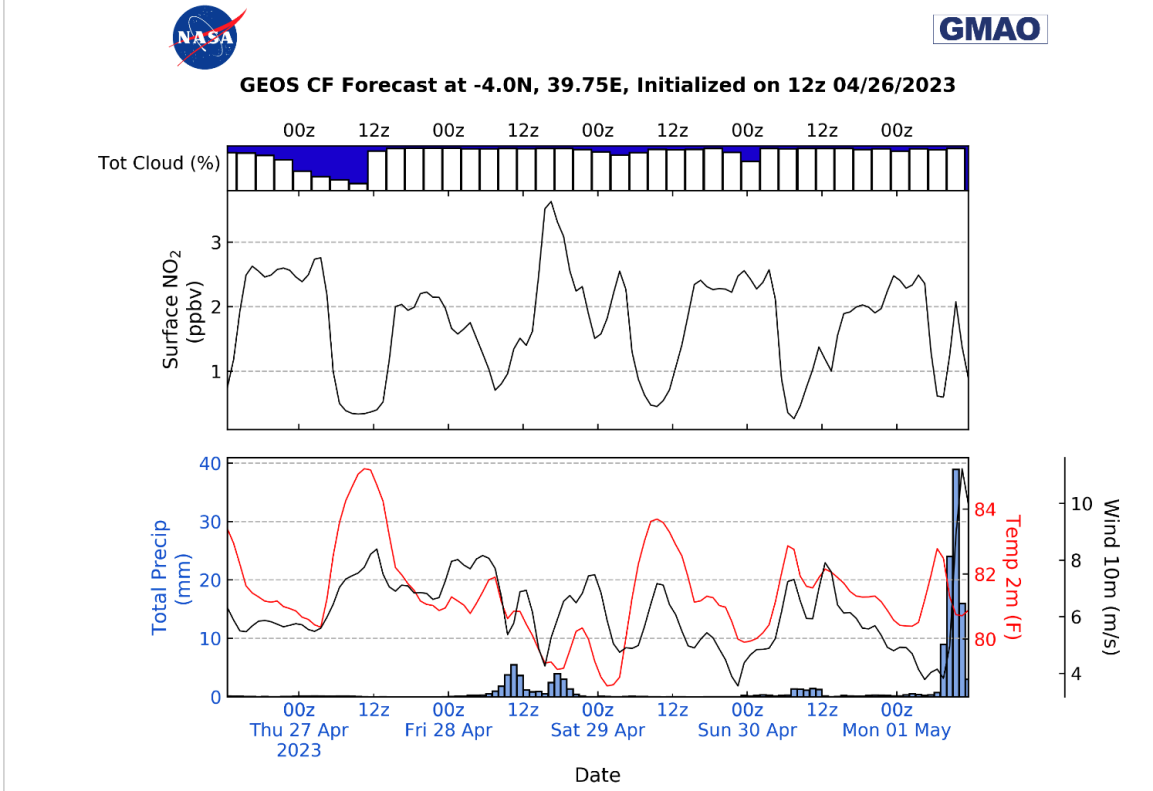
NO2

O3

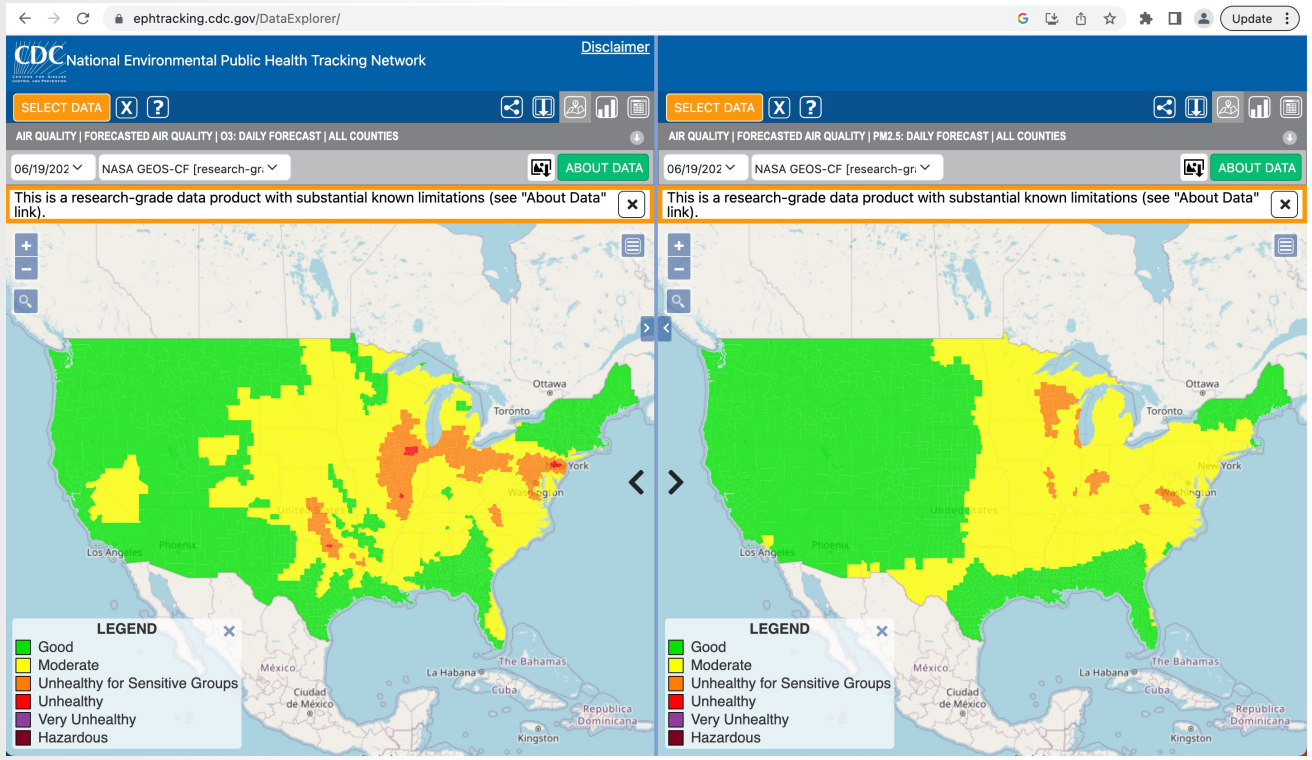
PM 2.5

GMAO GEOS CF Datagrams

NO_2 at (-4.0N, 39.75E)

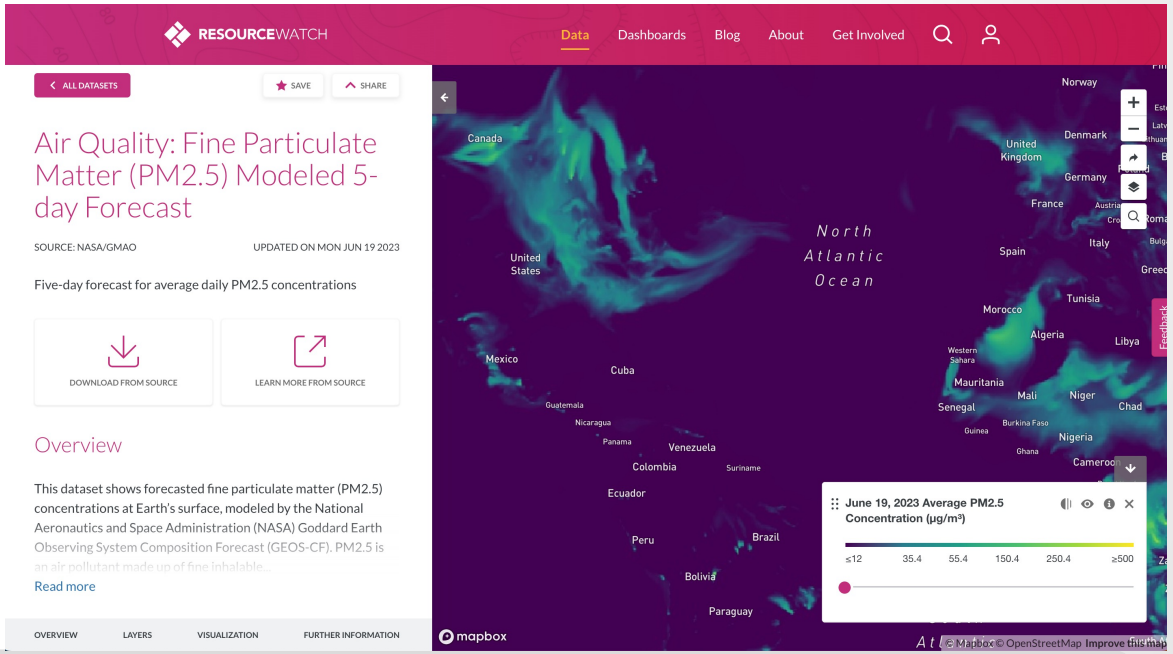


Examples of other ways to access GEOS-CF forecasts



https://ephtracking.cdc.gov/DataExplorer/

https://resourcewatch.org/data/explore

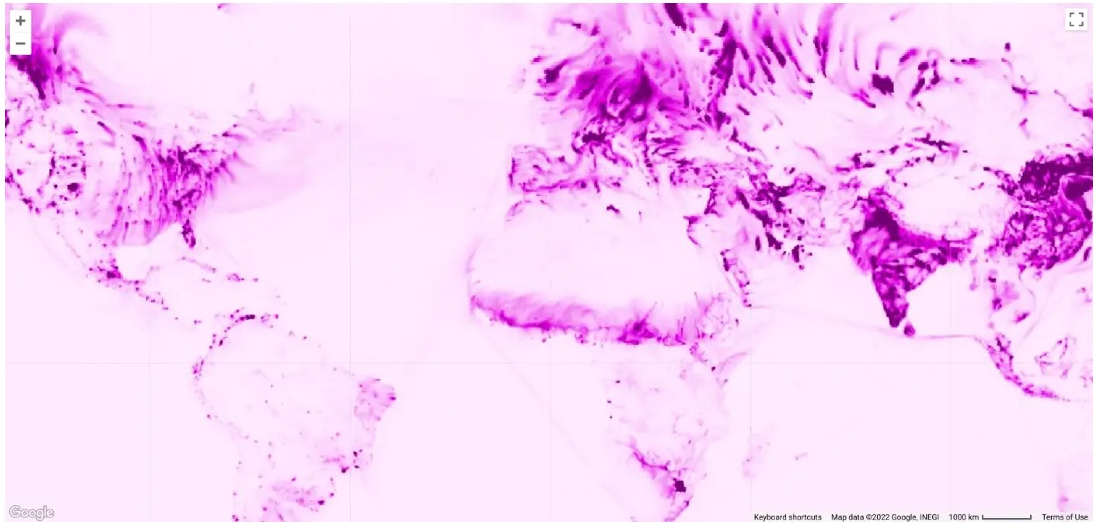


Global -> local scale estimates

Google Earth
Sep 16, 2022 · 3 min read · Listen

How NASA and Google are teaming up to understand and analyze air quality around the world

By Karin Tuxen-Bettman, PhD, Program Manager, Google Earth Outreach
Nicholas Clinton, PhD, Developer Advocate, Google Earth Engine
Argyro Kavvada, PhD, Program Lead, Sustainable Development Goals, Applied Sciences Program, NASA Earth Science Division

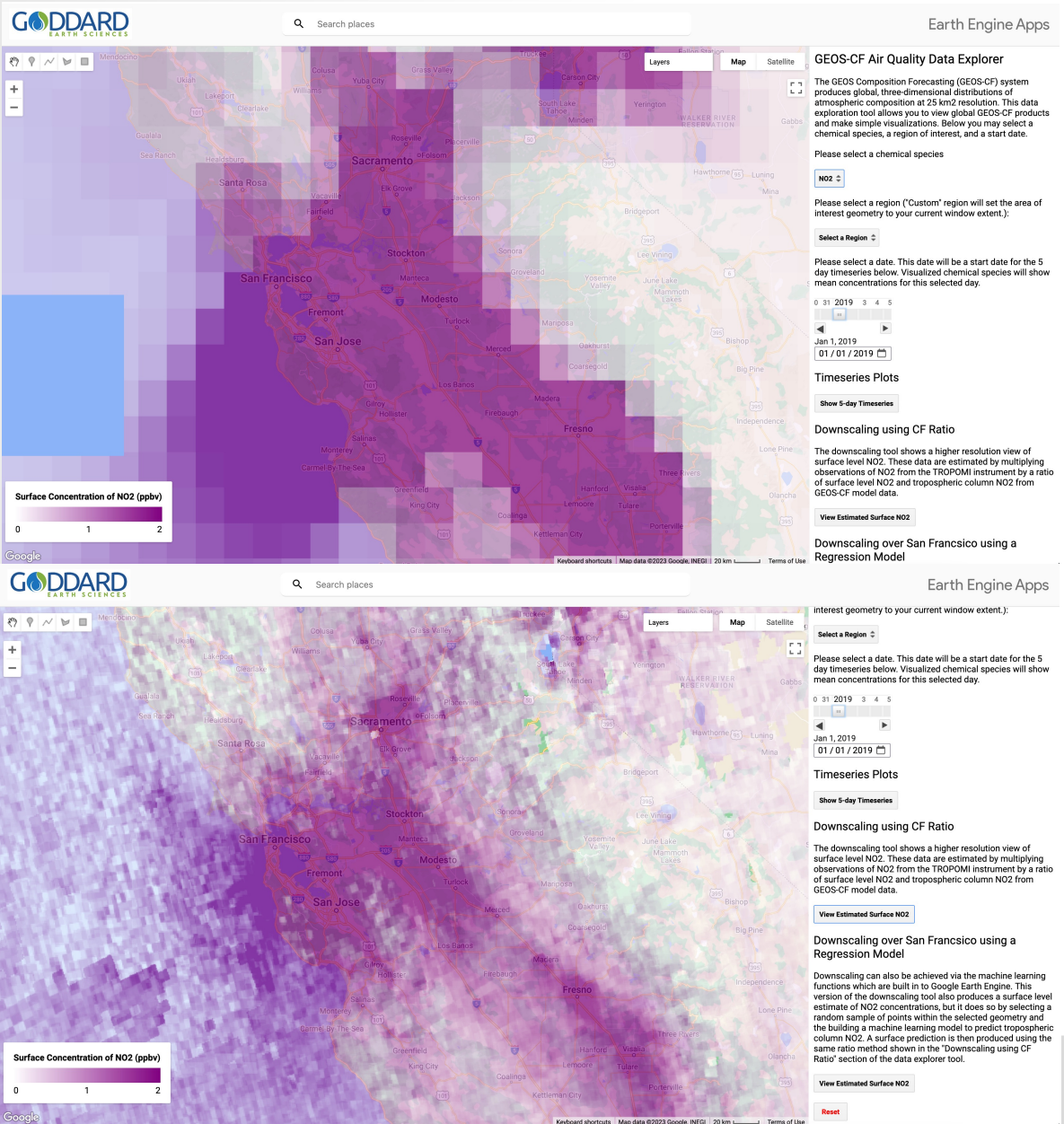


Part of the NASA-Google Partnership, select GMAO products are now being ingested into GEE:

- **GEOS-CF** hourly-average surface fields (chemistry and meteorology)
- **MERRA-2** hourly-average aerosol fields

https://medium.com/google-earth/how-nasa-and-google-are-teaming-up-to-understand-and-analyze-air-quality-around-the-world-7c89f6efad3d

Global -> local scale estimates



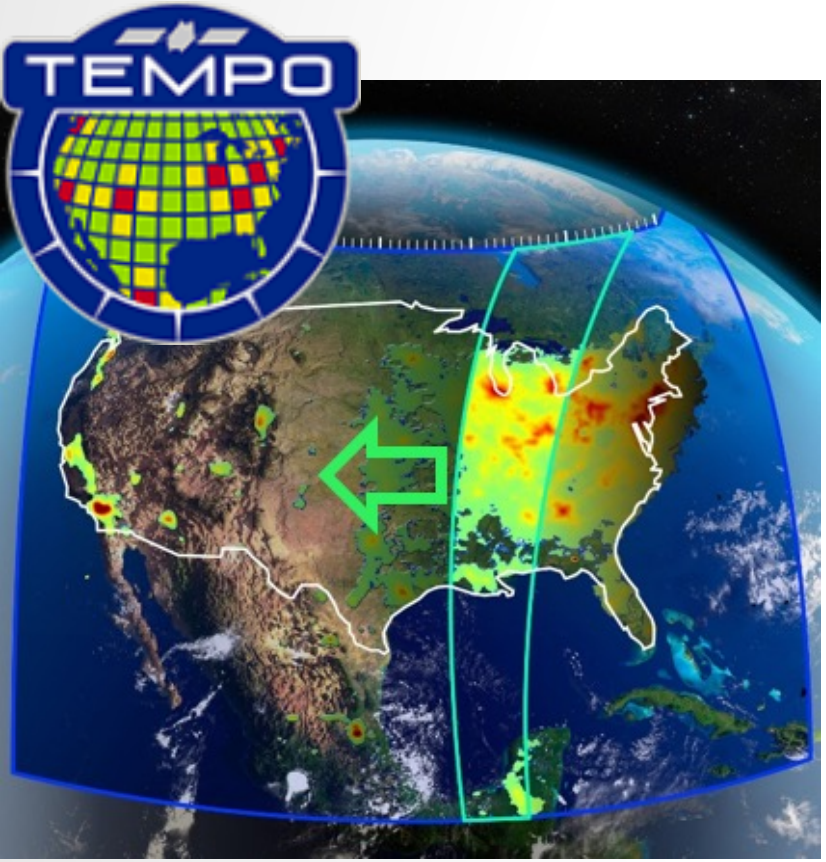
Part of the NASA-Google Partnership, select GMAO products are now being ingested into GEE:

- **GEOS-CF** hourly-average surface fields (chemistry and meteorology)
- **MERRA-2** hourly-average aerosol fields

Google Earth Engine and Google Cloud tools provide researchers the option to layer together different datasets in GEE and perform statistical methods and machine learning techniques which can downscale the global model reanalysis and forecasts to help improve on the biases at the sub-grid scale.

https://gmao.gsfc.nasa.gov/research/science_snapshots/2023/new-generation-gmao-apps.php

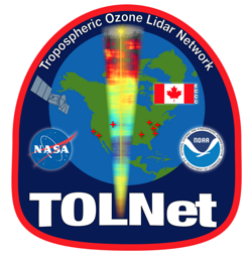
Daily *atmospheric composition* forecast

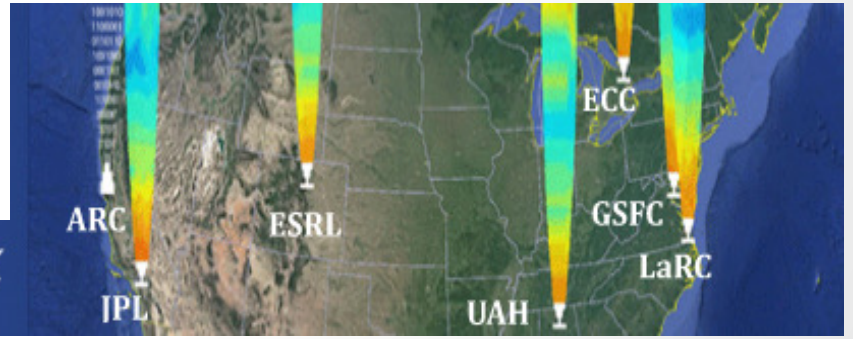


A realistic stratosphere in GEOS-CF is essential to support a broad range of NASA applications, including:

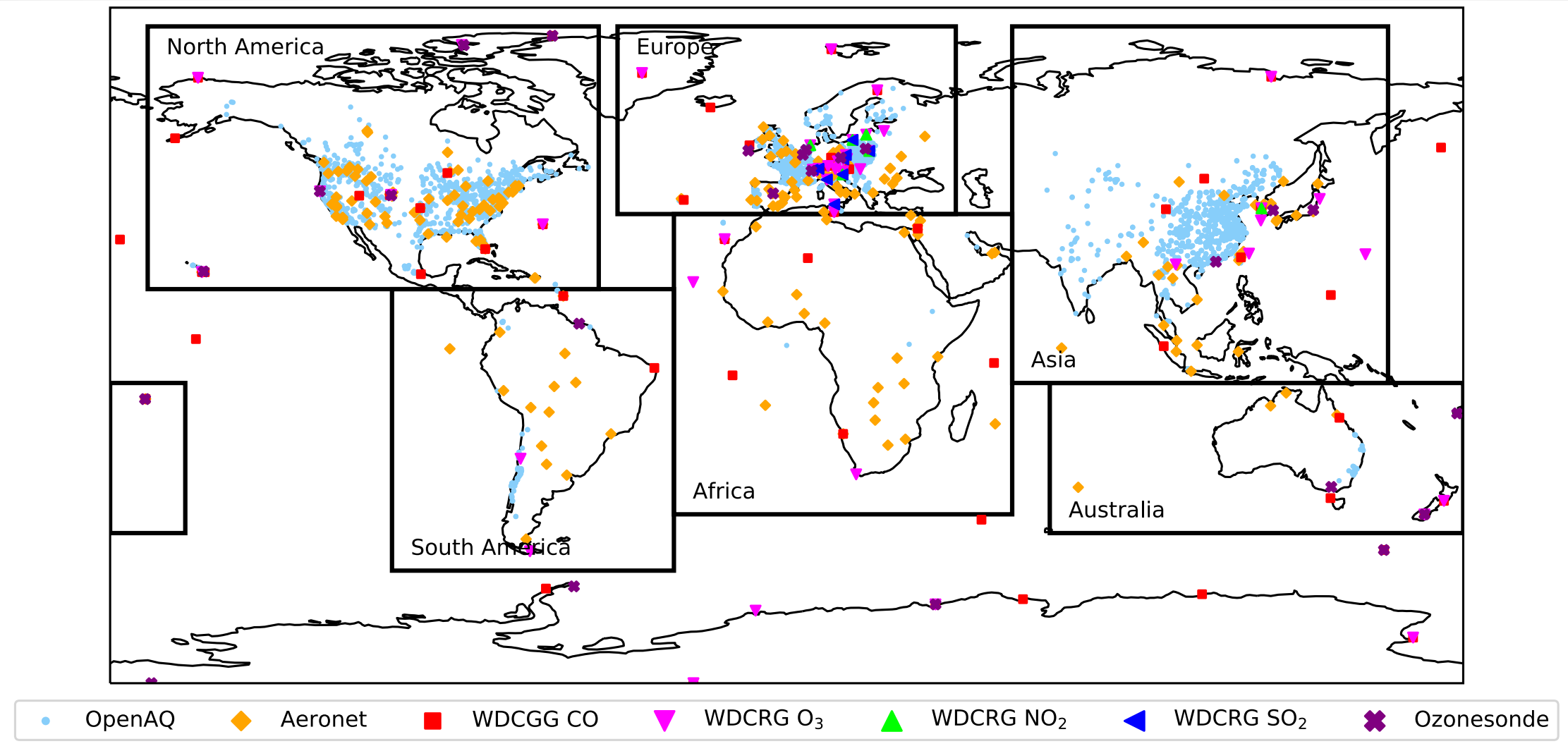
- Satellite retrievals of trace gases
- Airborne campaigns
- Stratosphere-troposphere exchange

TOLNET
Tropospheric Ozone LIDAR Network





Observations for evaluation



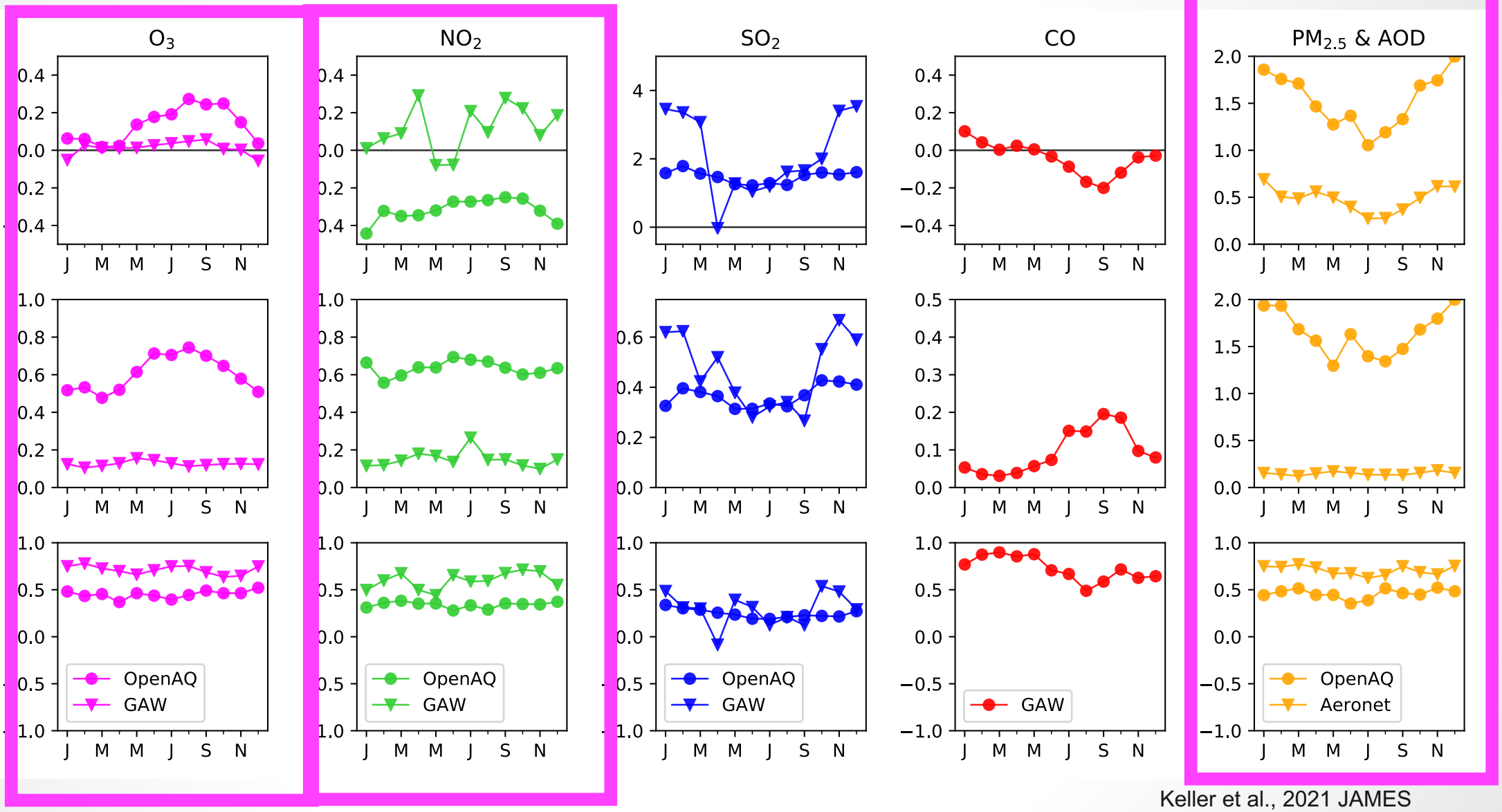
Keller et al., 2021 JAMES

Global surface comparisons - monthly

Normalized Mean Bias (NMB)

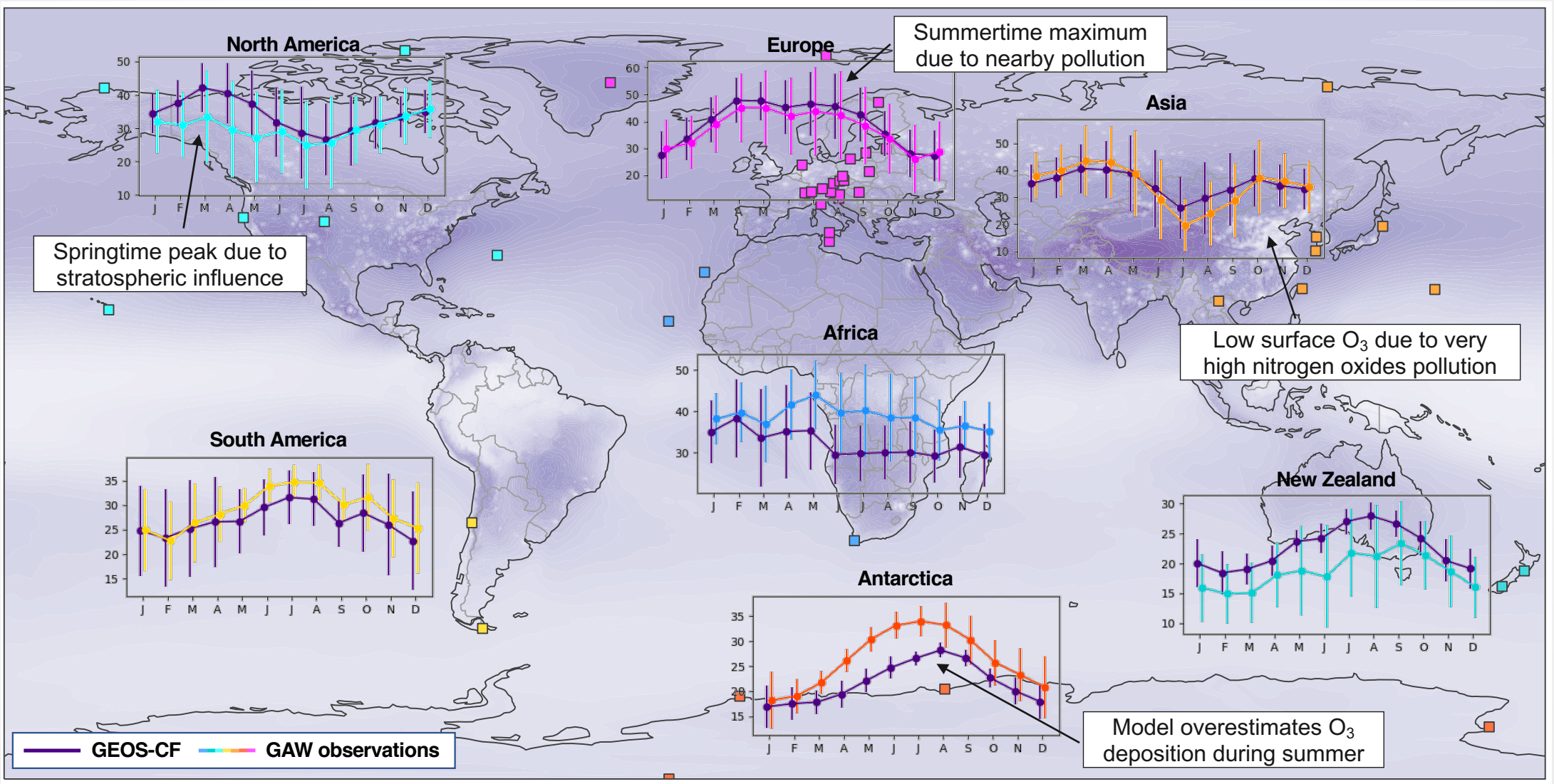
Normalized Root Mean Square Error (NRMSE)

Pearson Correlation Coefficient (R)

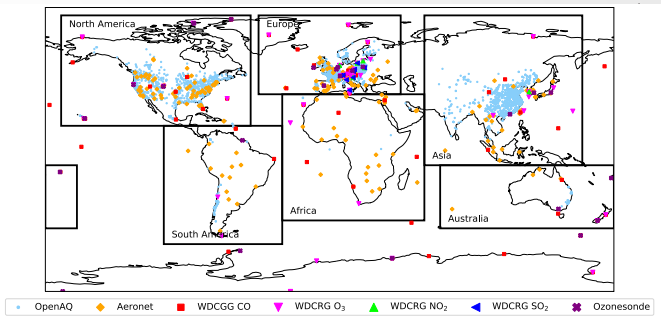


Keller et al., 2021 JAMES

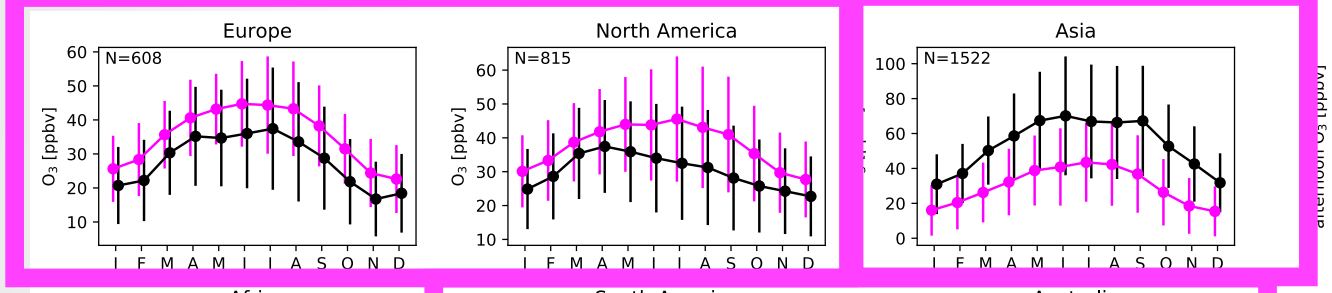
GEOS-CF surface ozone compares well against background observations from the Global Atmospheric Watch (GAW) network



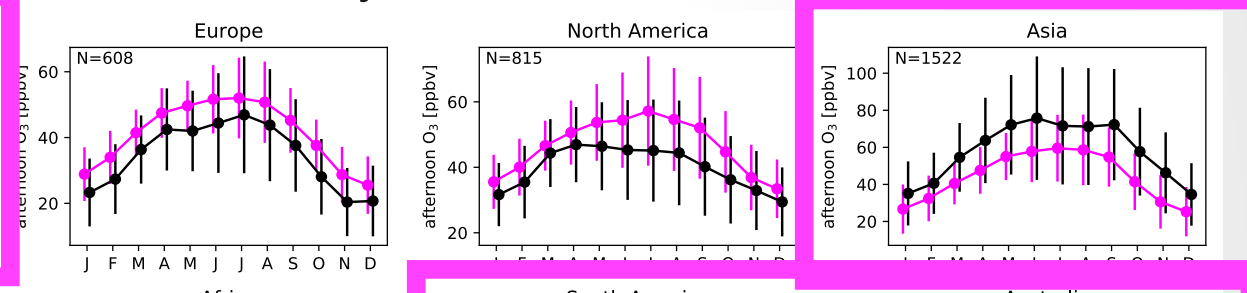
Surface O_3 (OpenAQ sites only)



Daily



Afternoon only

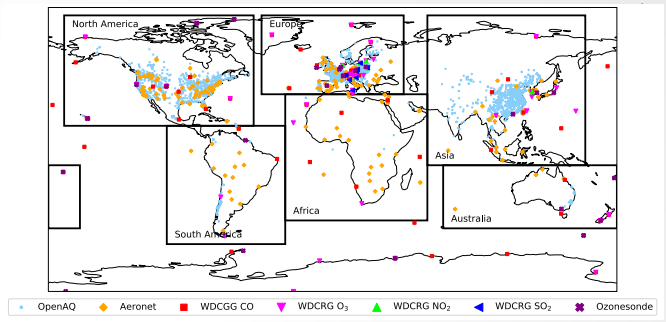


- GEOS-CF captures the overall seasonal cycle in the six regions, but generally overestimates in Europe, North & South America and Australia, while underestimating in Asia and Africa.
- In polluted regions, such as Asia, the bias is reduced when focusing on peak photochemical production period. GEOS-CF still has a bias over US during summer and fall, a known GEOS-Chem issue (Travis et al., 2016;2019; Hu et al., 2018).

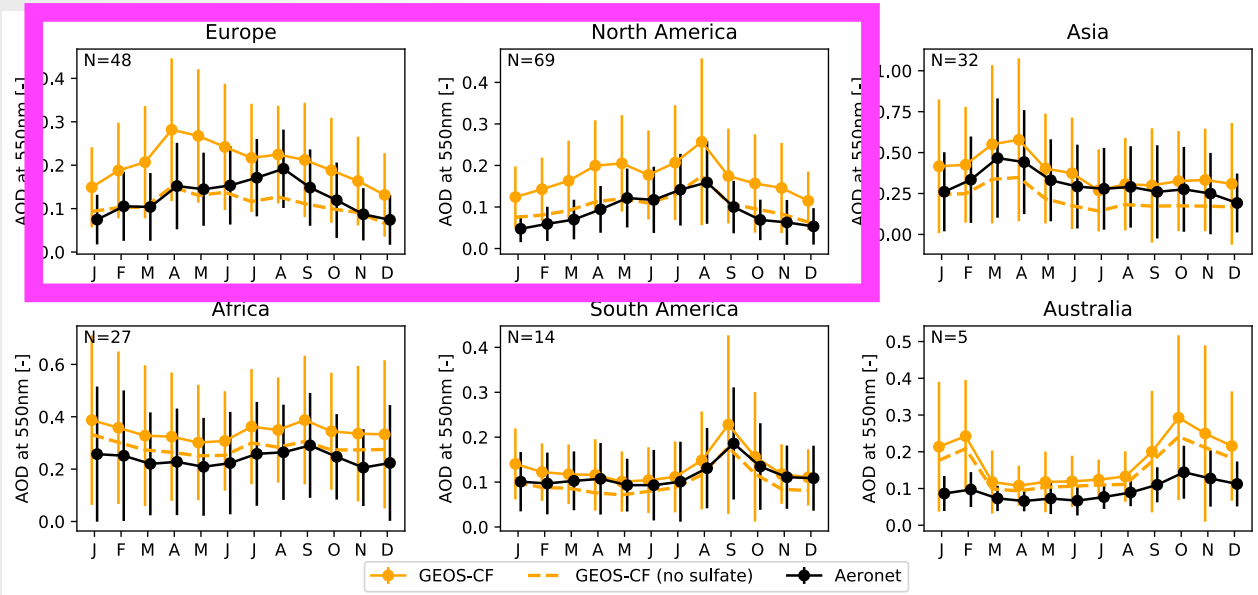
Observations GEOS-CF

Keller et al., 2021 JAMES

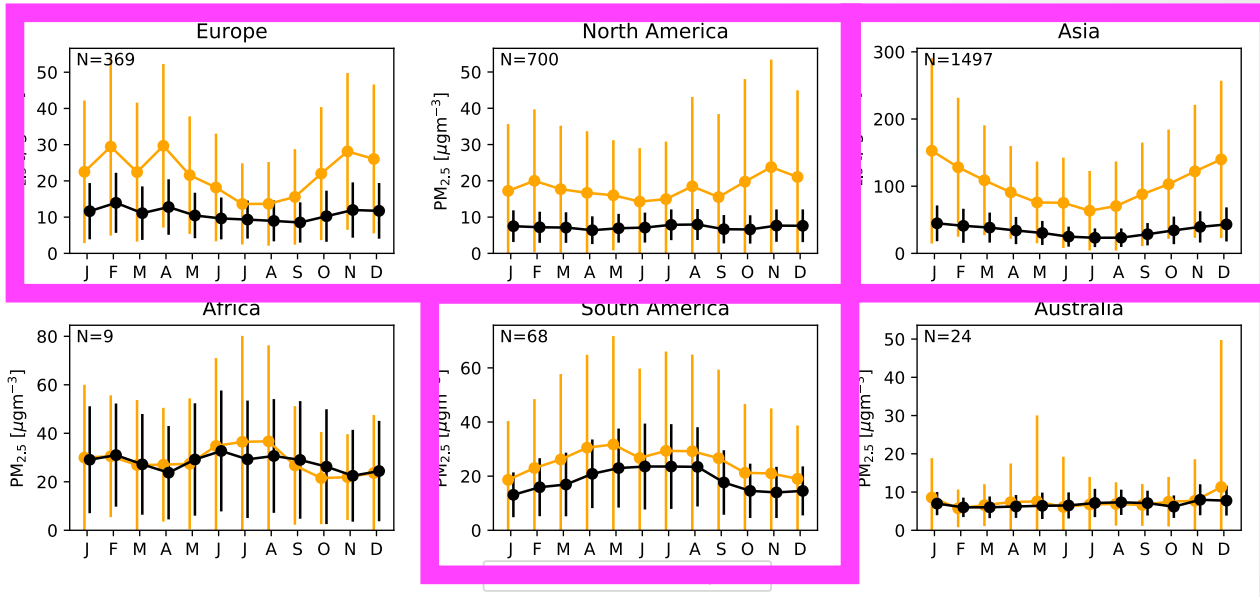
GEOS-CF generally overestimates aerosols (AOD and $PM_{2.5}$)



AOD from Aeronet network



PM2.5 from OpenAQ database

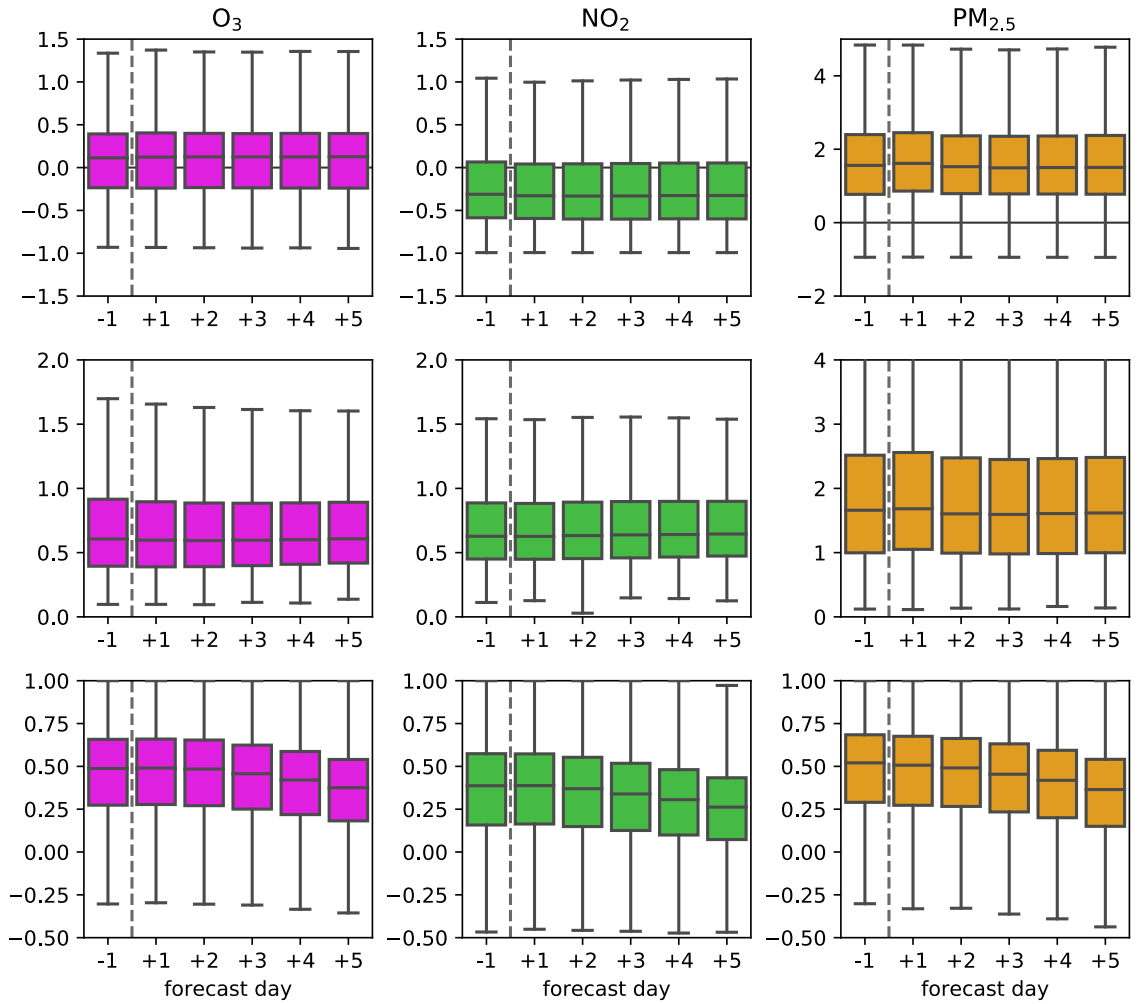


- GEOS-CF overestimates AOD 550 nm at most Aeronet sites. This is likely due to the overestimation of sulfates in the model.
- Similarly, GEOS-CF $PM_{2.5}$ is overestimated in Europe, North America, Asia and South America.

Observations **GEOS-CF**

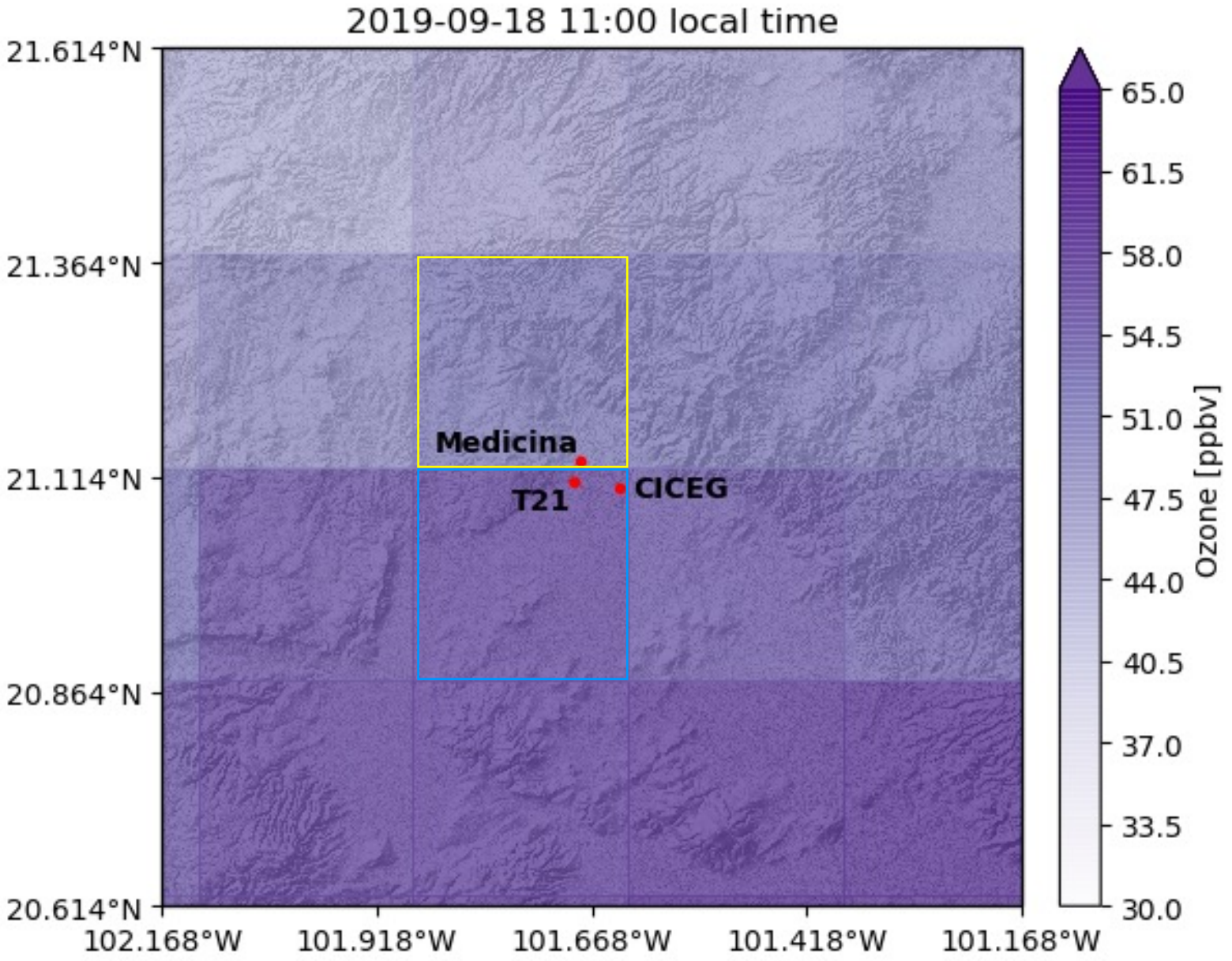
Keller et al., 2021 JAMES

GEOS CF Forecast skill (GAW and OpenAQ)



- Little variation in the skill scores out to 5 days
- Correlation tends to decrease after day 2 and this is likely due to changes in the meteorological forecast and biomass burning emissions.

Keller, et al. (2021). Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0. *JAMES* https://doi.org/10.1029/2020MS002413

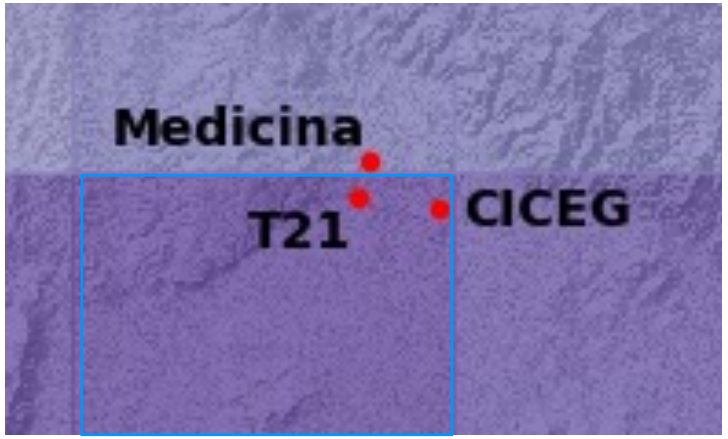


Improve local forecasts using statistical bias correction

3 monitoring stations in Leon, Mexico

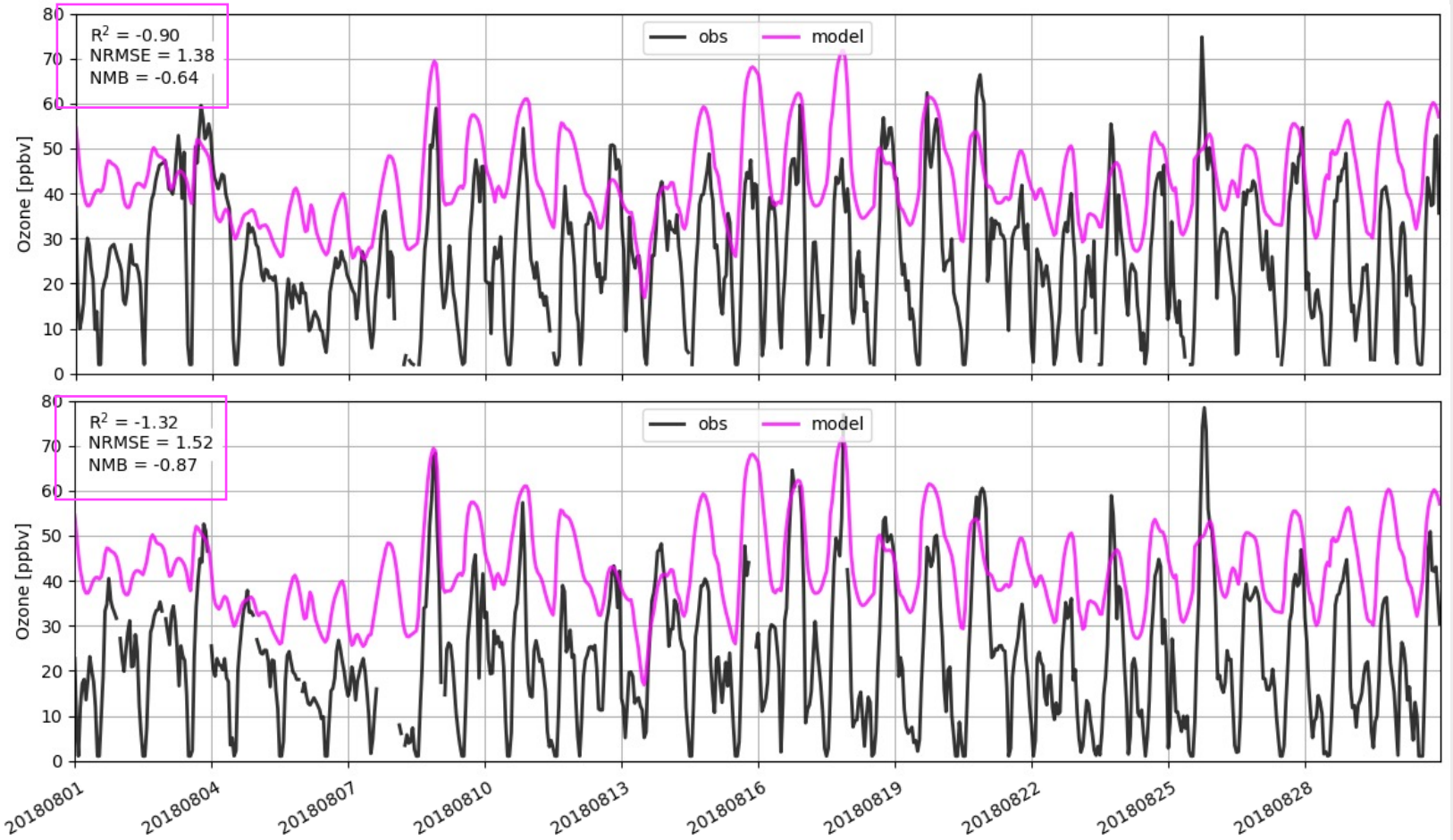
- 1 in one grid box
- 2 share a grid box
- Difficult terrain within each grid box

Improve local forecasts using statistical bias correction



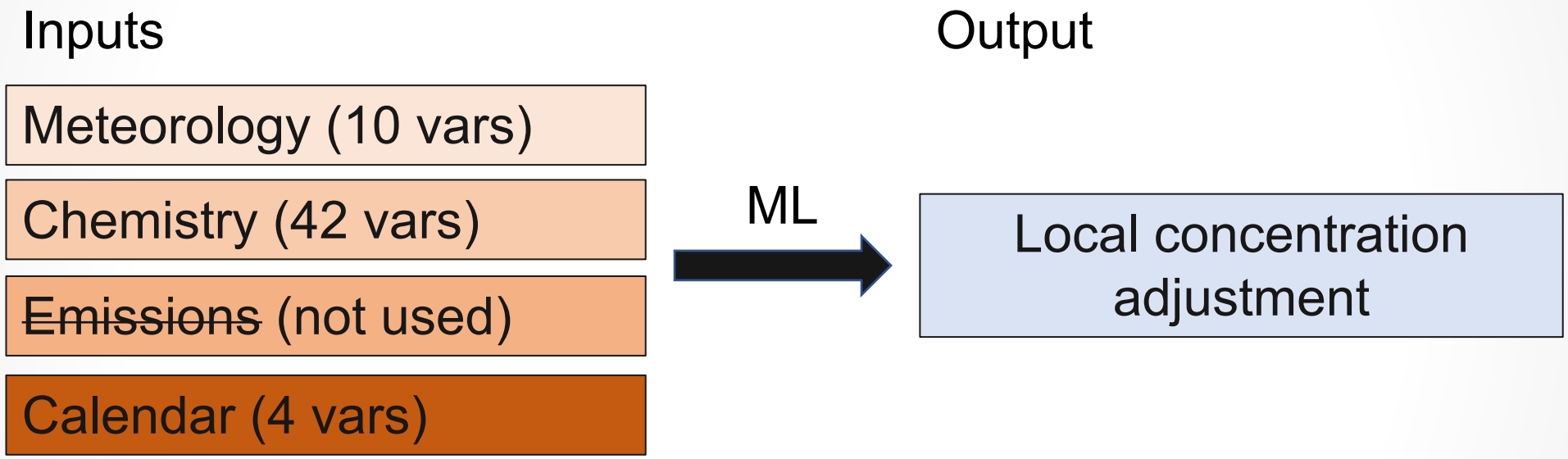
Two observation sites in the same grid box

- GEOS-CF generally over-estimates



Observations Model

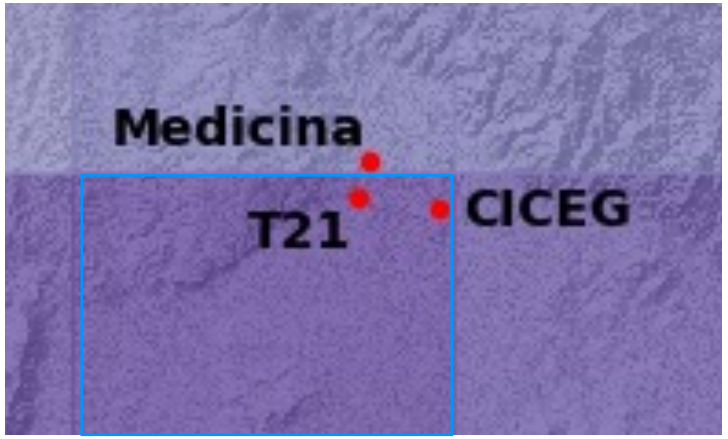
Use machine learning to correct for small scale variability and/or model biases



- Algorithm: gradient boosted decision trees (XGBoost)
- Train separate algorithm for each site

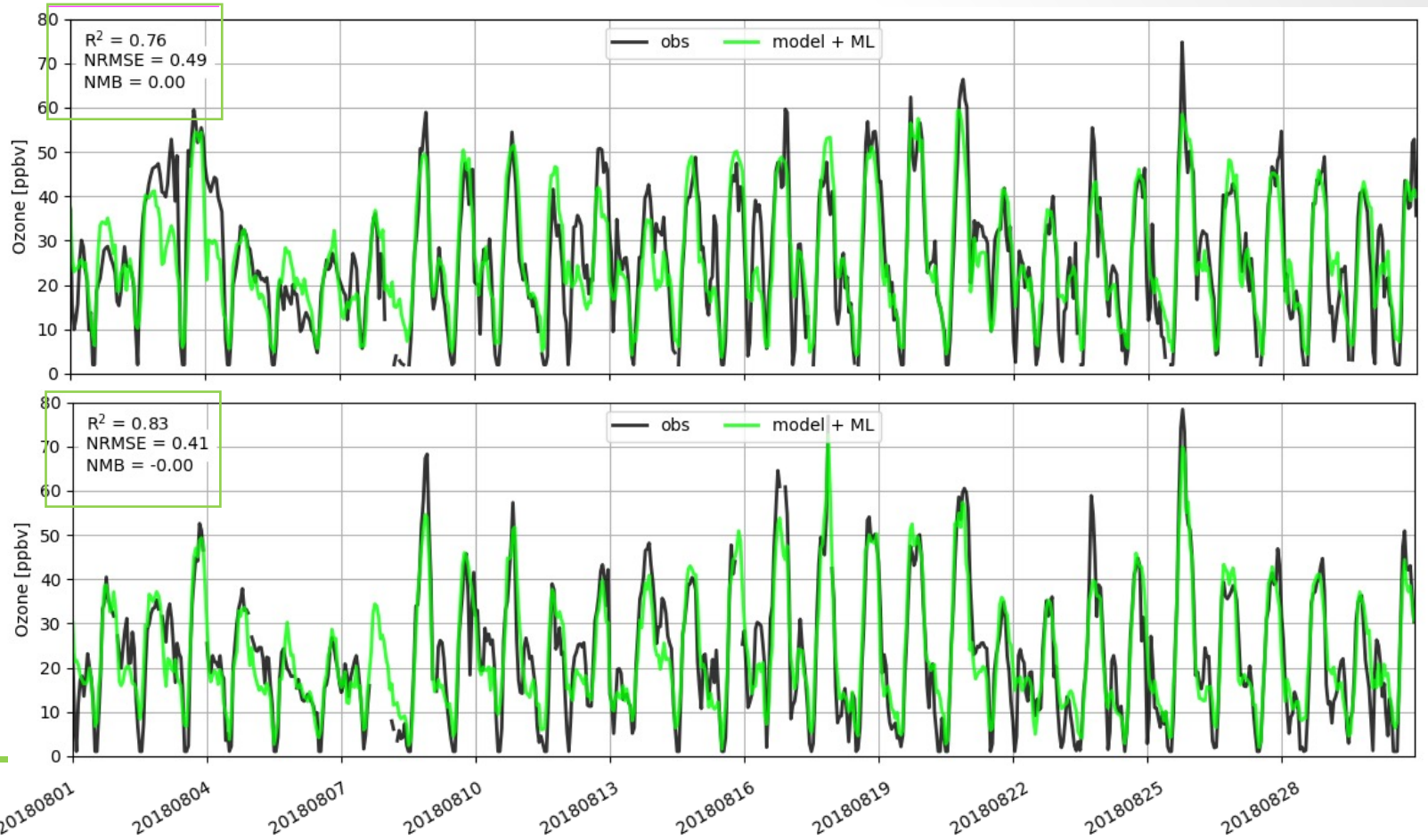
Keller et al., 2021 ACP

Improve local forecasts using statistical bias correction



Two observation sites in the same grid box

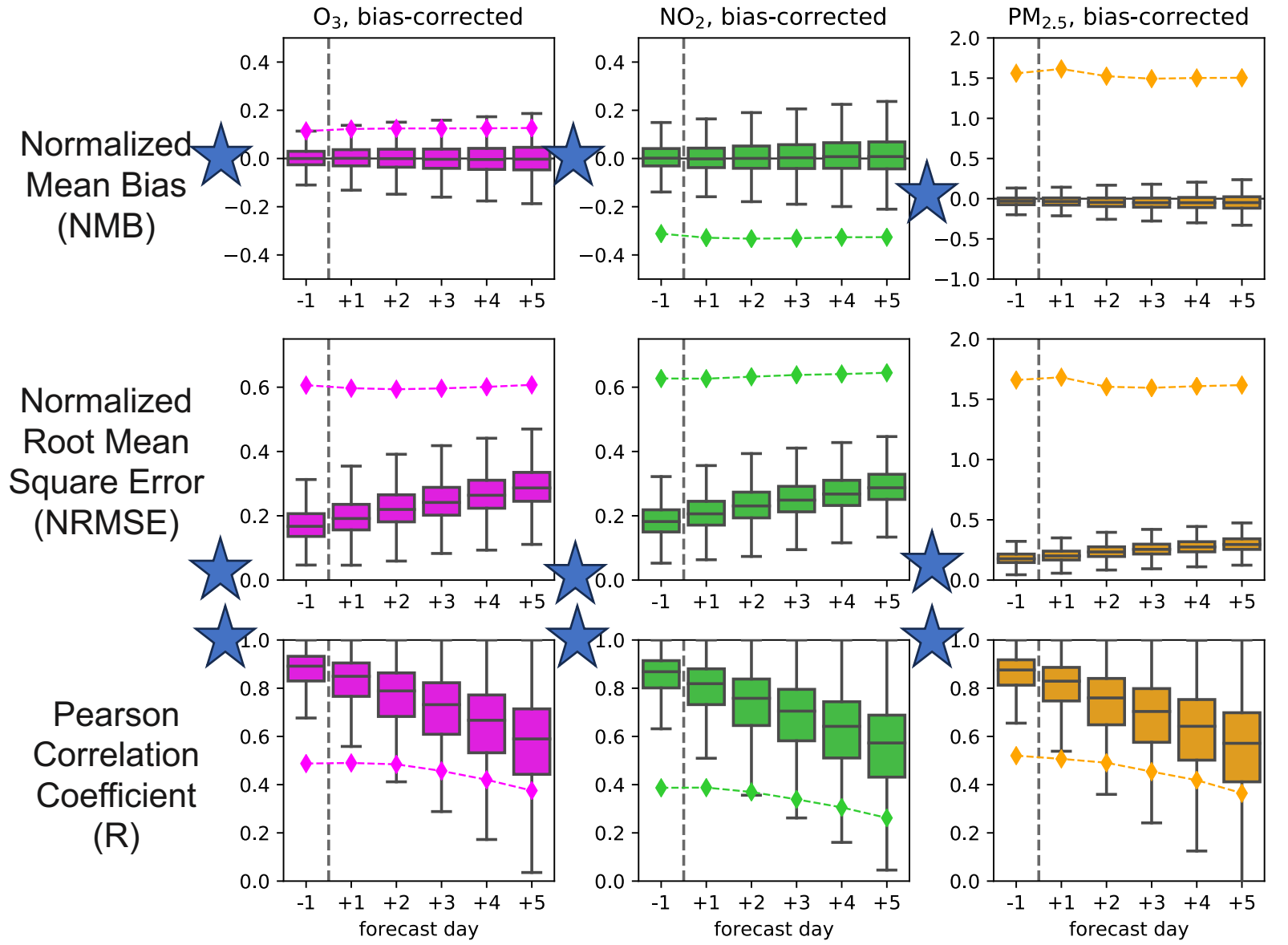
- GEOS-CF+ML captures diurnal variability at sub-grid scale



Observations Model + ML

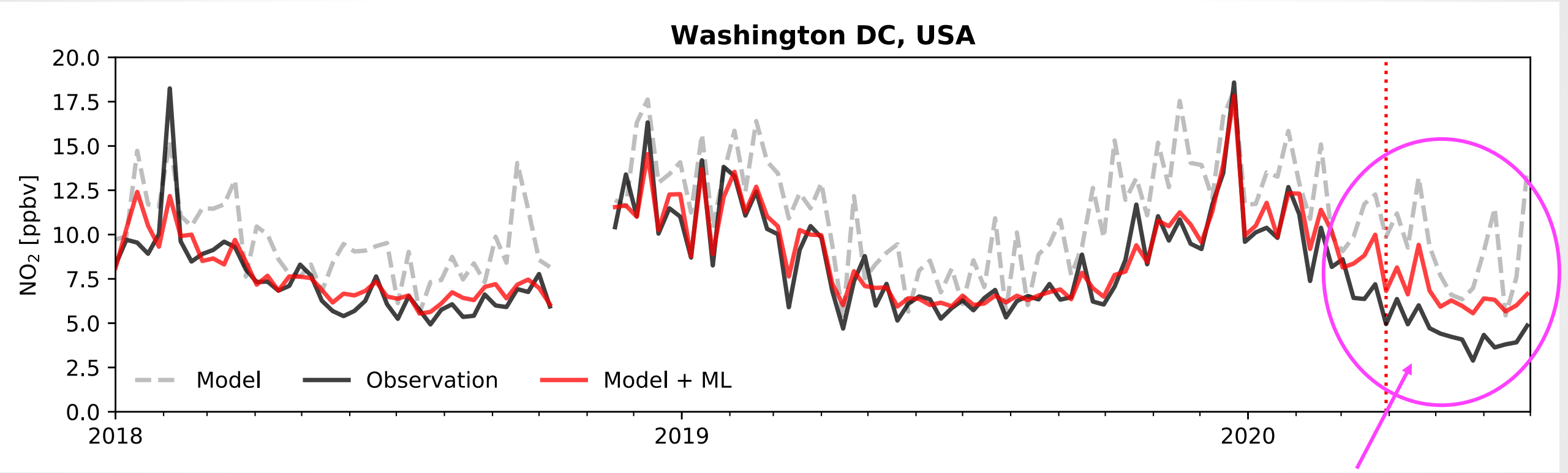
GEOS CF Forecast skill

- ✓ Using a Machine Learning (ML) algorithm to calculate bias-correction term for each monitoring site can drastically improve the forecast skill at the individual locations



Keller et al., 2021 JAMES

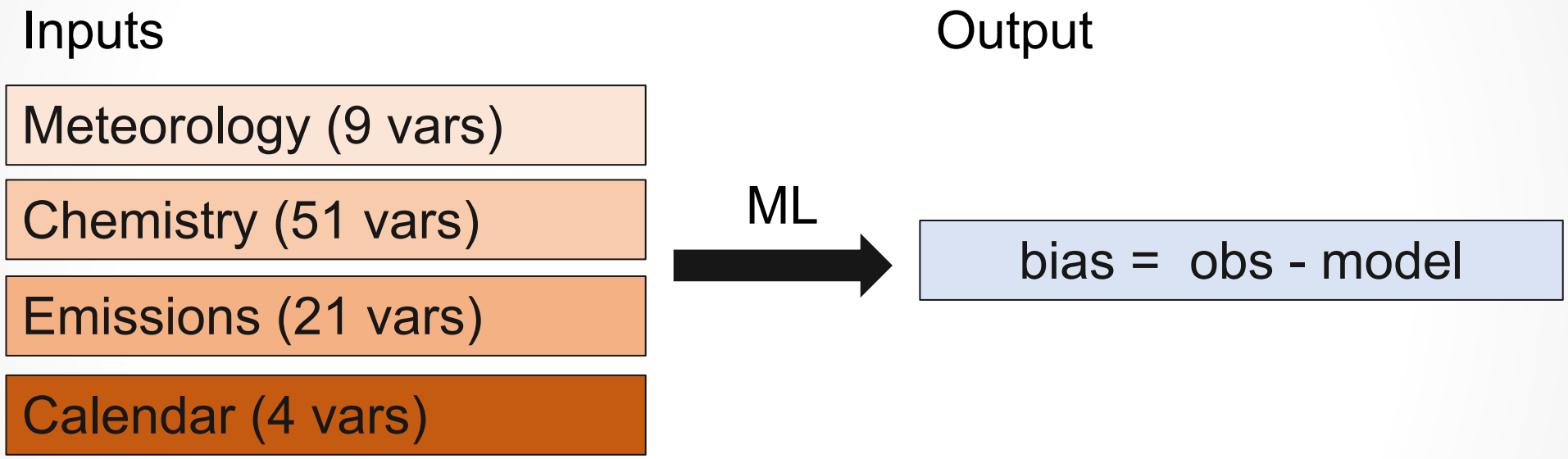
New application of the GEOS-CF ML algorithm



Impact of COVID-19 restricti

Keller et al., 2021 ACP

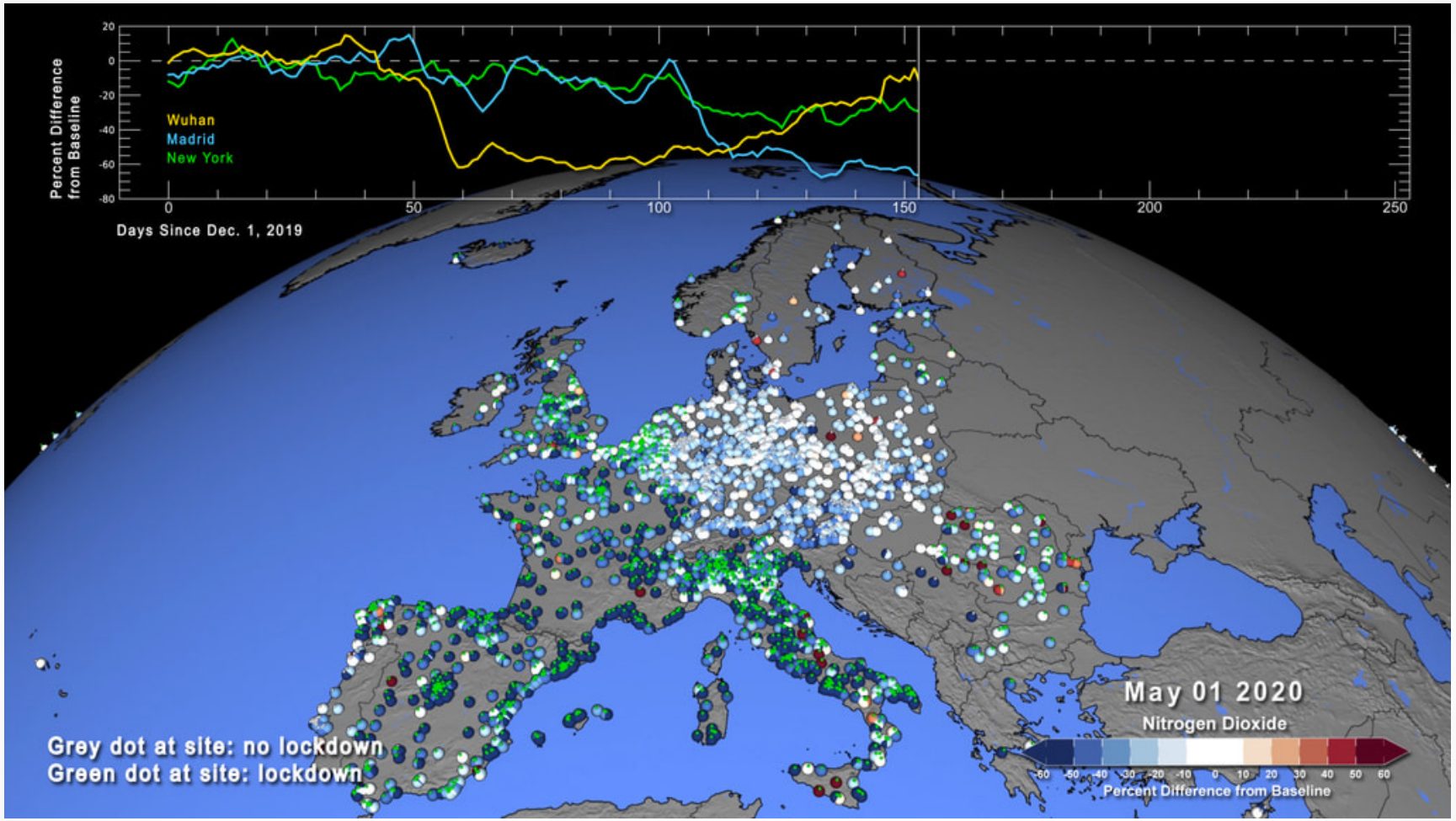
Apply bias-correction to model output using machine learning (using historical observation-model comparisons)



- Algorithm: gradient boosted decision trees (XGBoost)
- Training: 2018-2019 (8-fold cross validation)

Keller et al., 2021 ACP

Apply analysis to 5756 sites worldwide



openaq

European Environment Agency

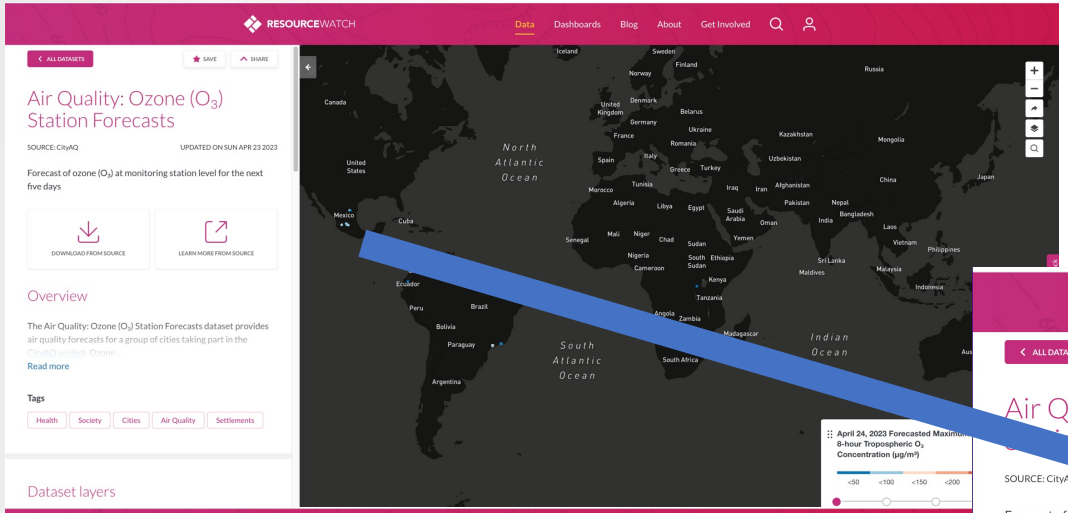
RIO PREFEITURA

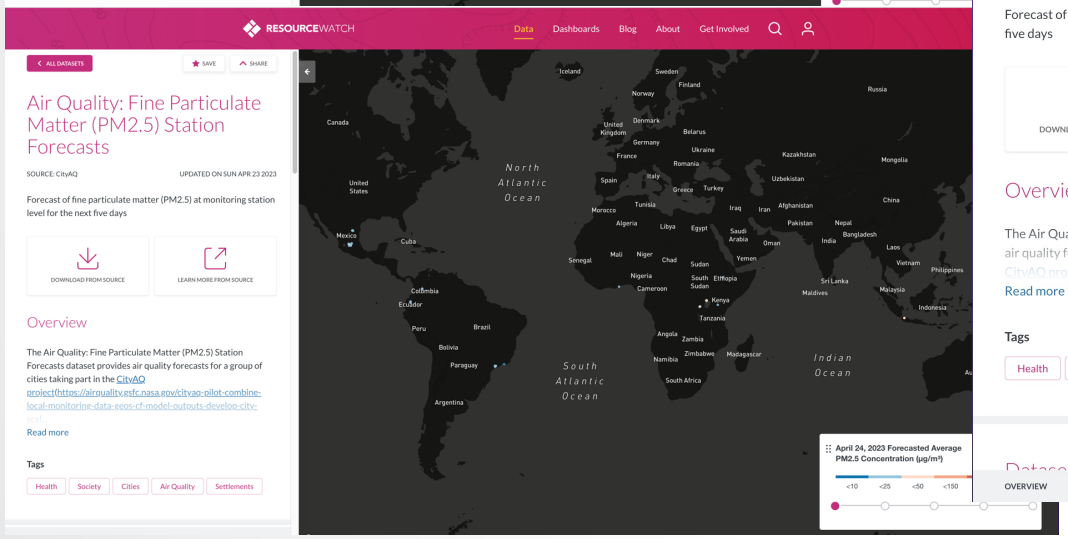
QUITO ALCALDÍA

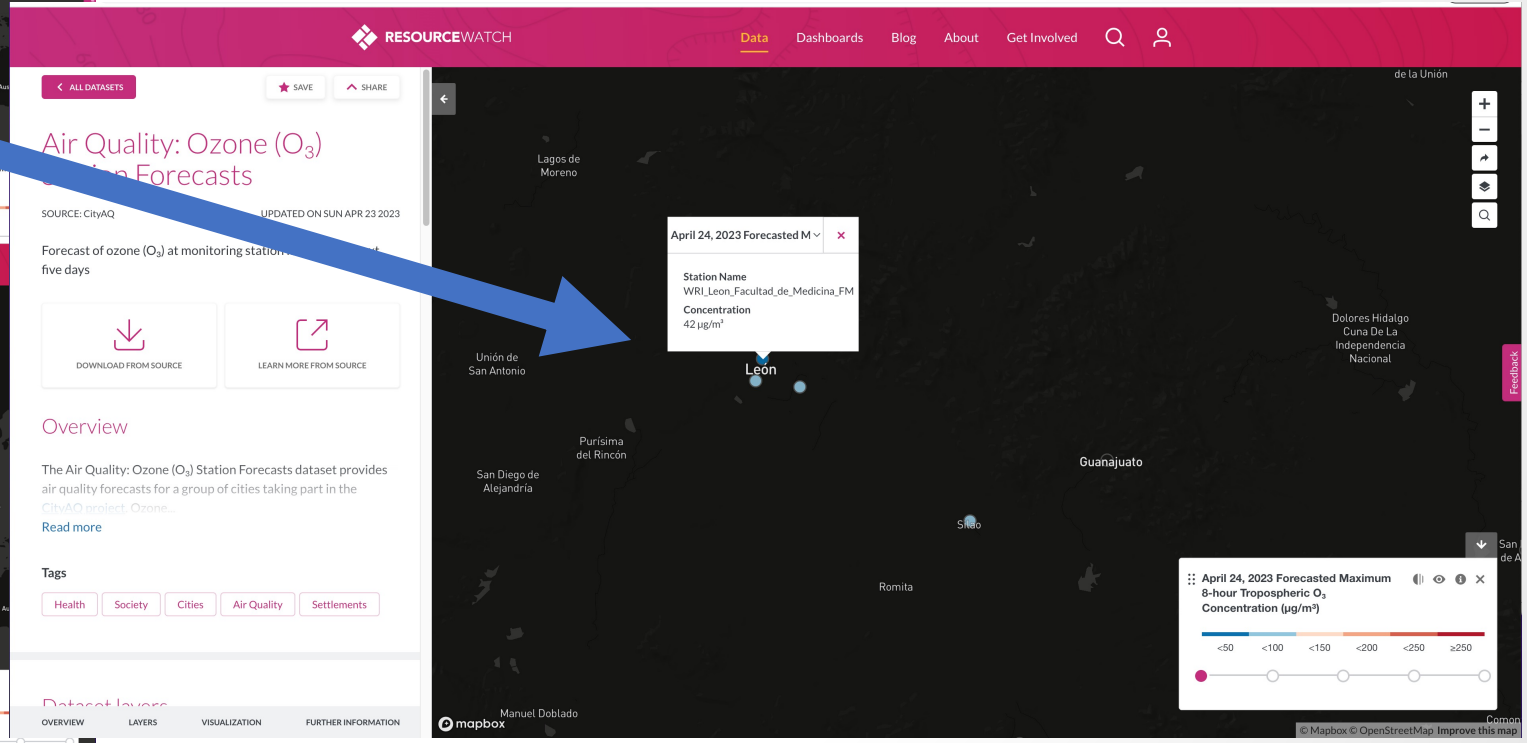
https://svs.gsfc.nasa.gov/13753 https://svs.gsfc.nasa.gov/4872

https://www.nasa.gov/feature/goddard/2020/nasa-model-reveals-how-much-covid-related-pollution-levels-deviated-from-the-norm

Share and expand localized forecast capabilities and other downscaling techniques is a priority

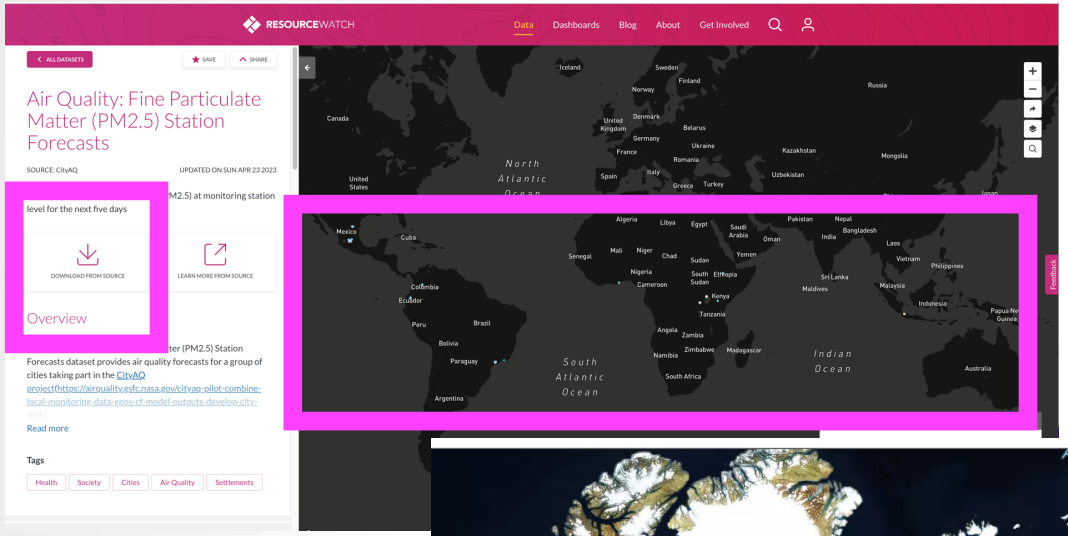


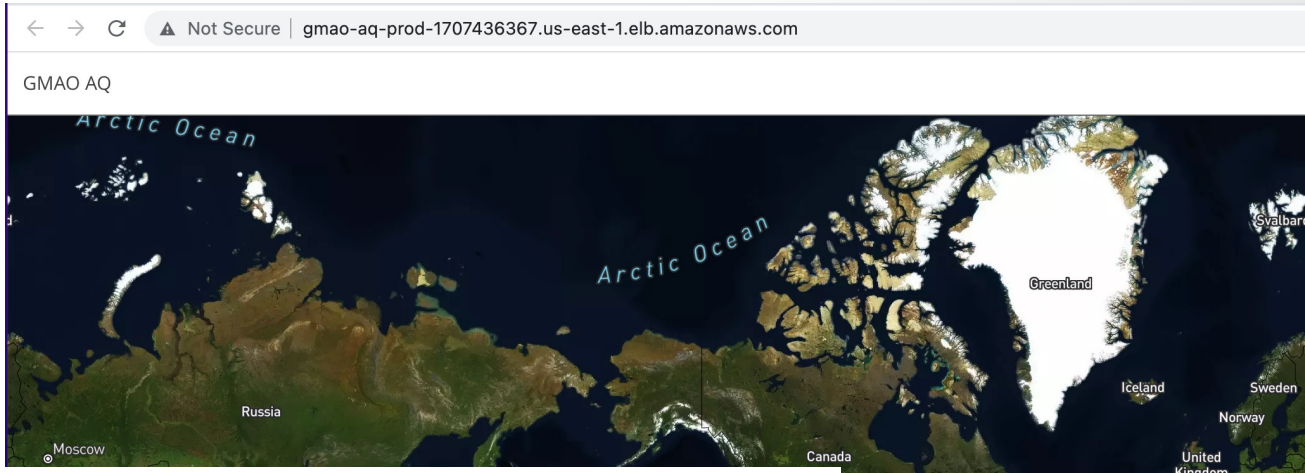




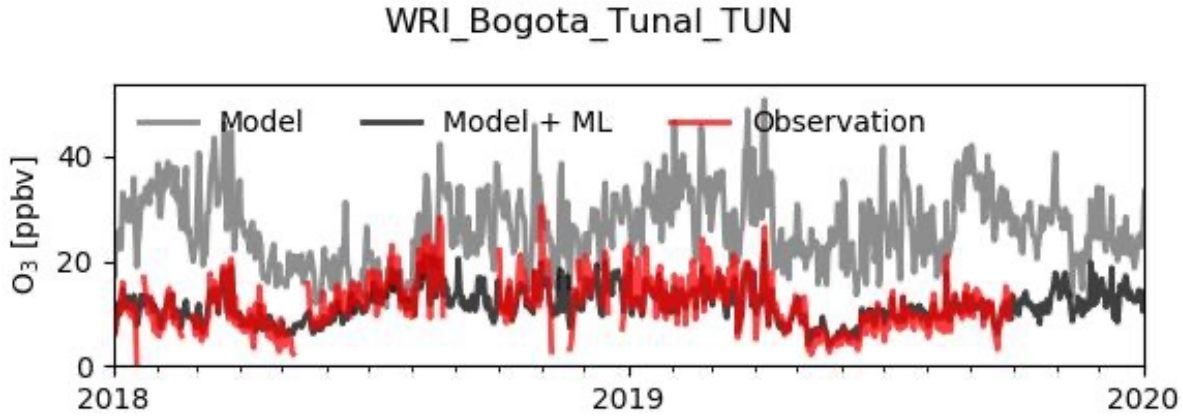
CityAQ: https://www.wri.org/initiatives/cityaq
CanAIRy: https://www.wri.org/initiatives/canairy-alert

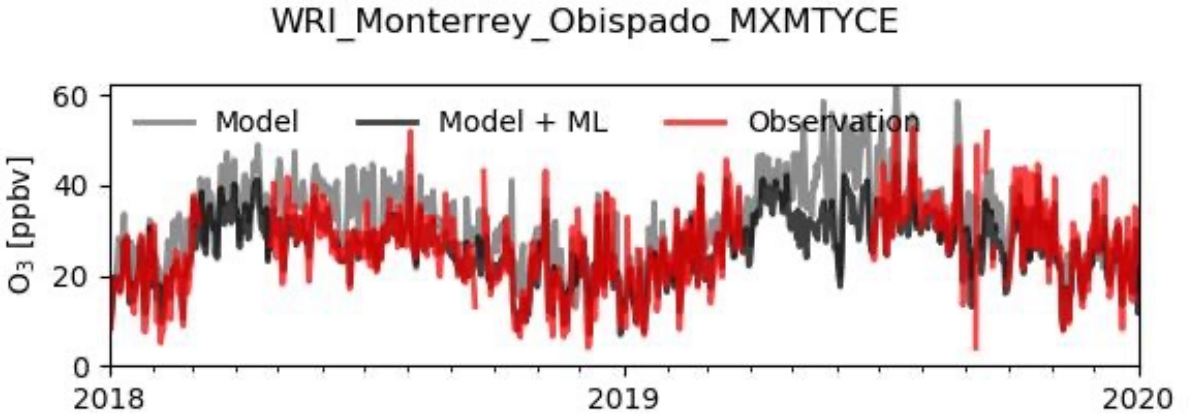
Share and expand localized forecast capabilities and other downscaling techniques is a priority

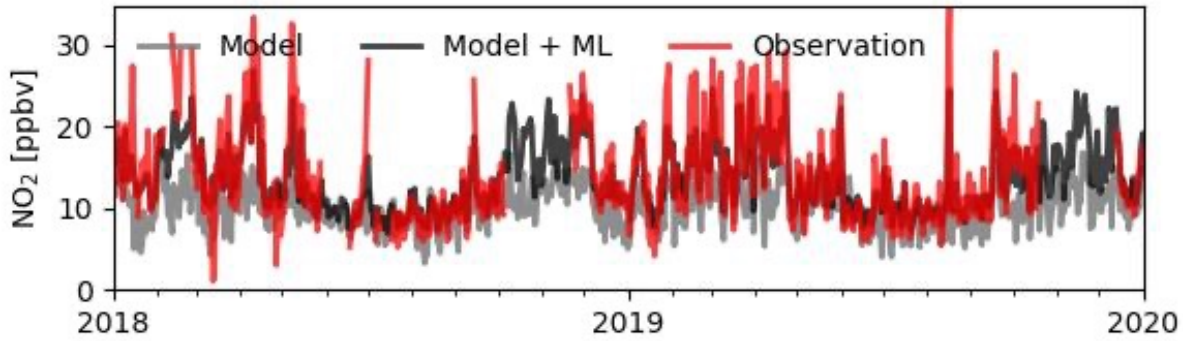


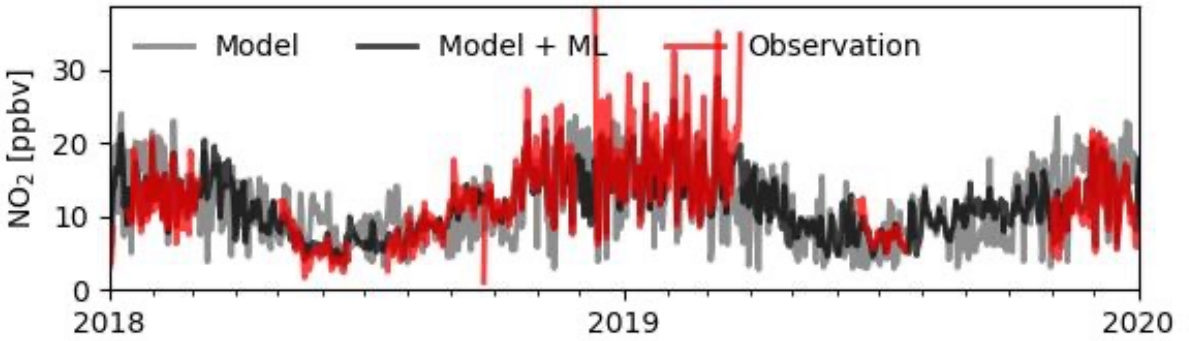


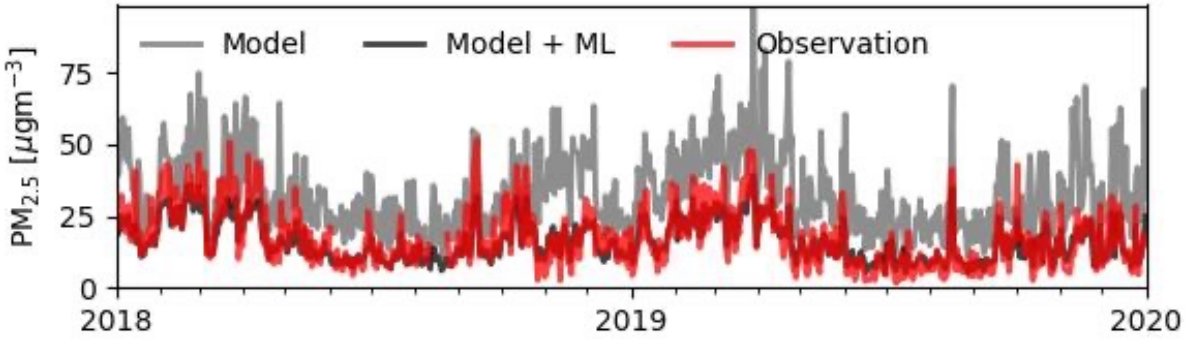
Training data (we need at least 1 year)

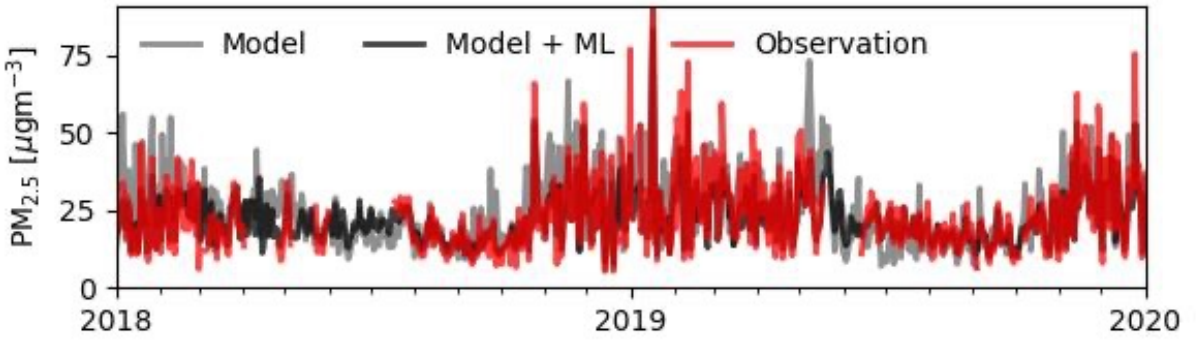




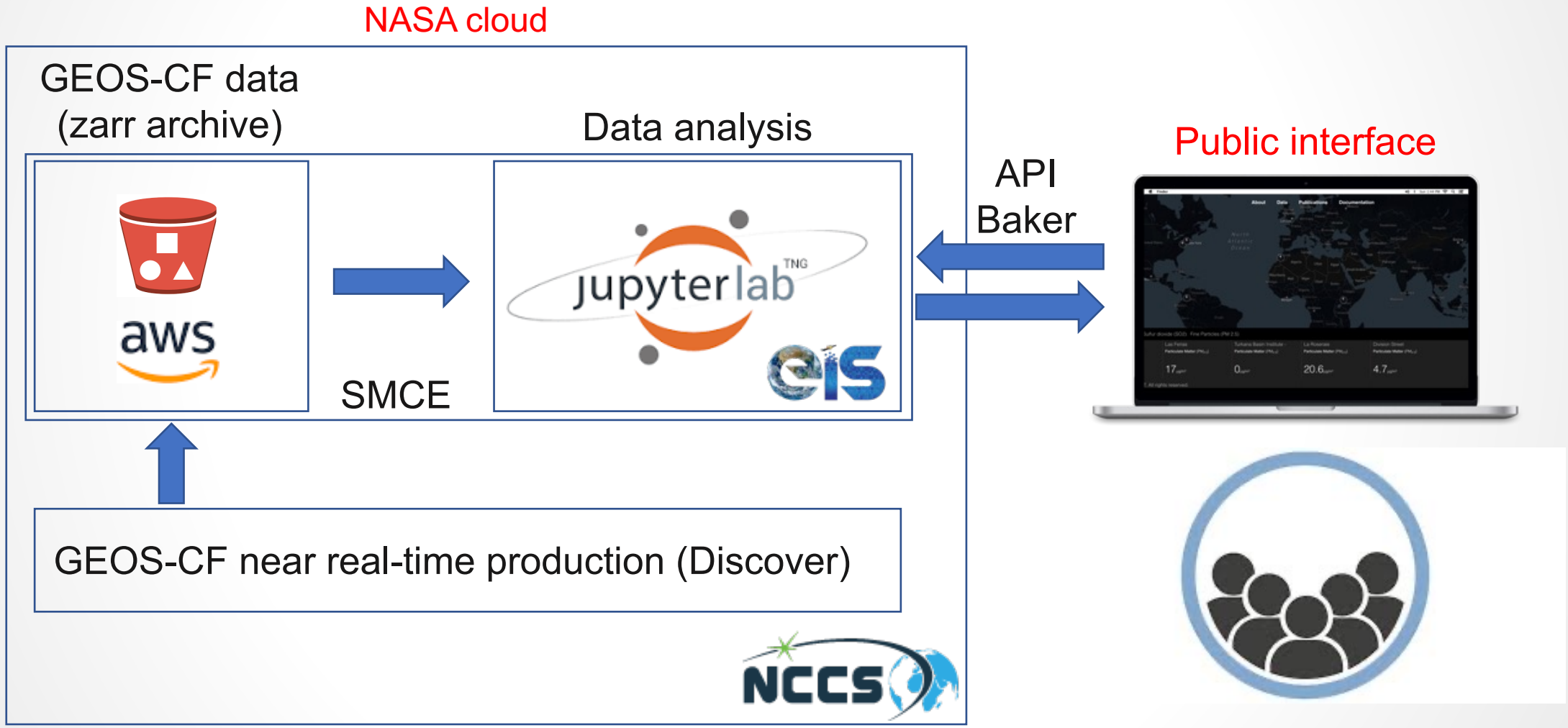






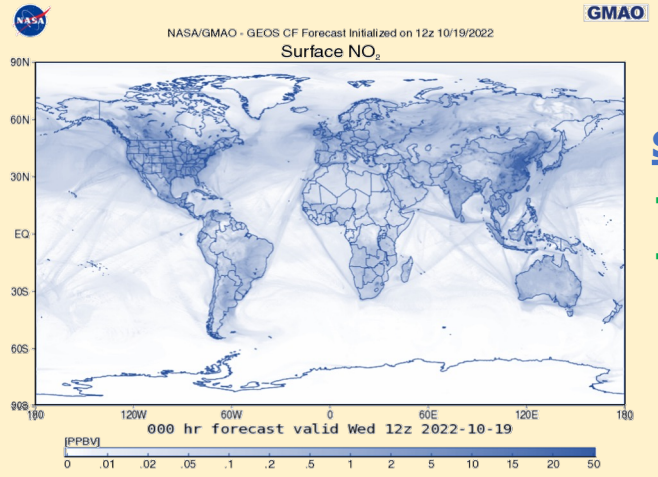


New interface to enable on-demand access to localized forecasts



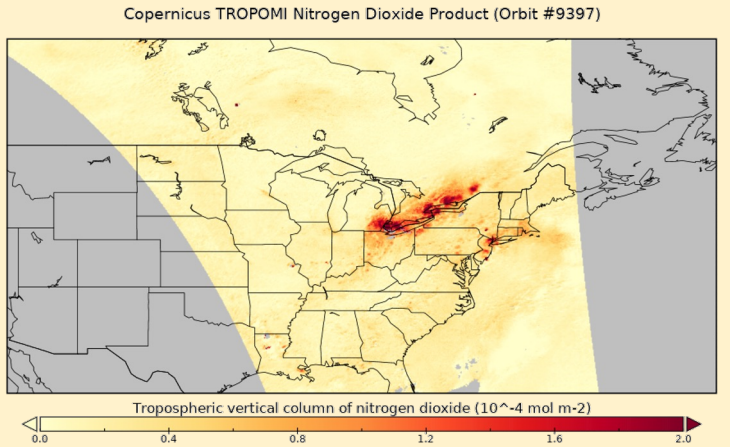
There are many different air quality data sources

Global Data Sources



simulation models
+ global coverage
+ forecasting
- limited resolution
? updated emissions

satellite retrievals
+ global coverage
- low time resolution
- column-integrated
? relevant to surface



regulatory monitoring
+ accurate
- expensive
? representative

low-cost monitoring
+ relatively inexpensive
+ dense/remote deployment
- greater noise and bias
? locally calibrated

Local Data Sources

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

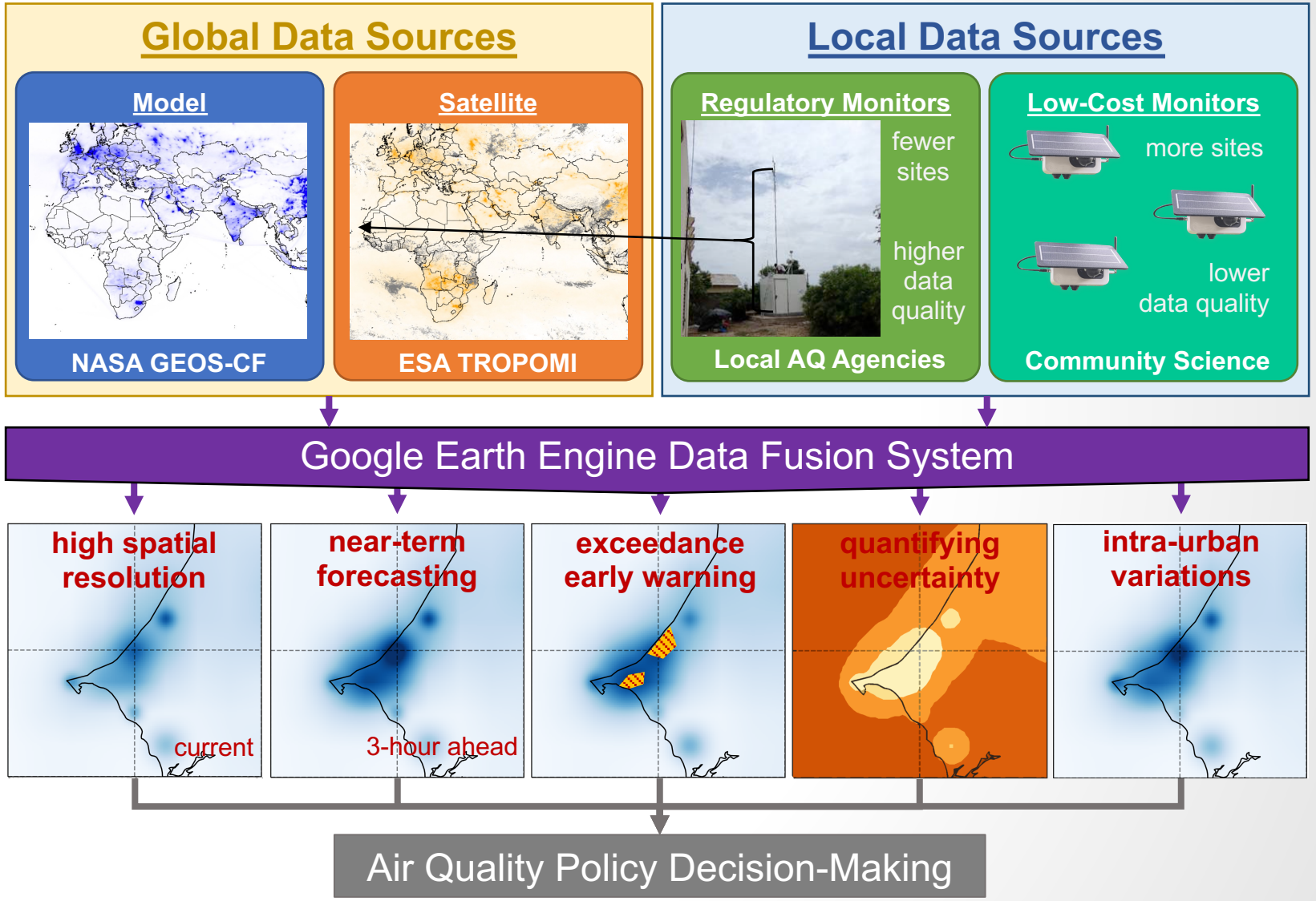
Our project's objective is to…

…integrate diverse **global** and **local** air quality data sources…

…using the cloud computing platform of **Google Earth Engine**…

…to provide synthesized **estimates** and **forecasts** of air quality at a **local scale** but with a **global scope**…

…which will be freely accessible by air quality managers worldwide, facilitating their **decision-making** processes.



PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Project Scheme and Partners

NASA GMAO: basic algorithm development & refinement

Clarity: low-cost sensor integration

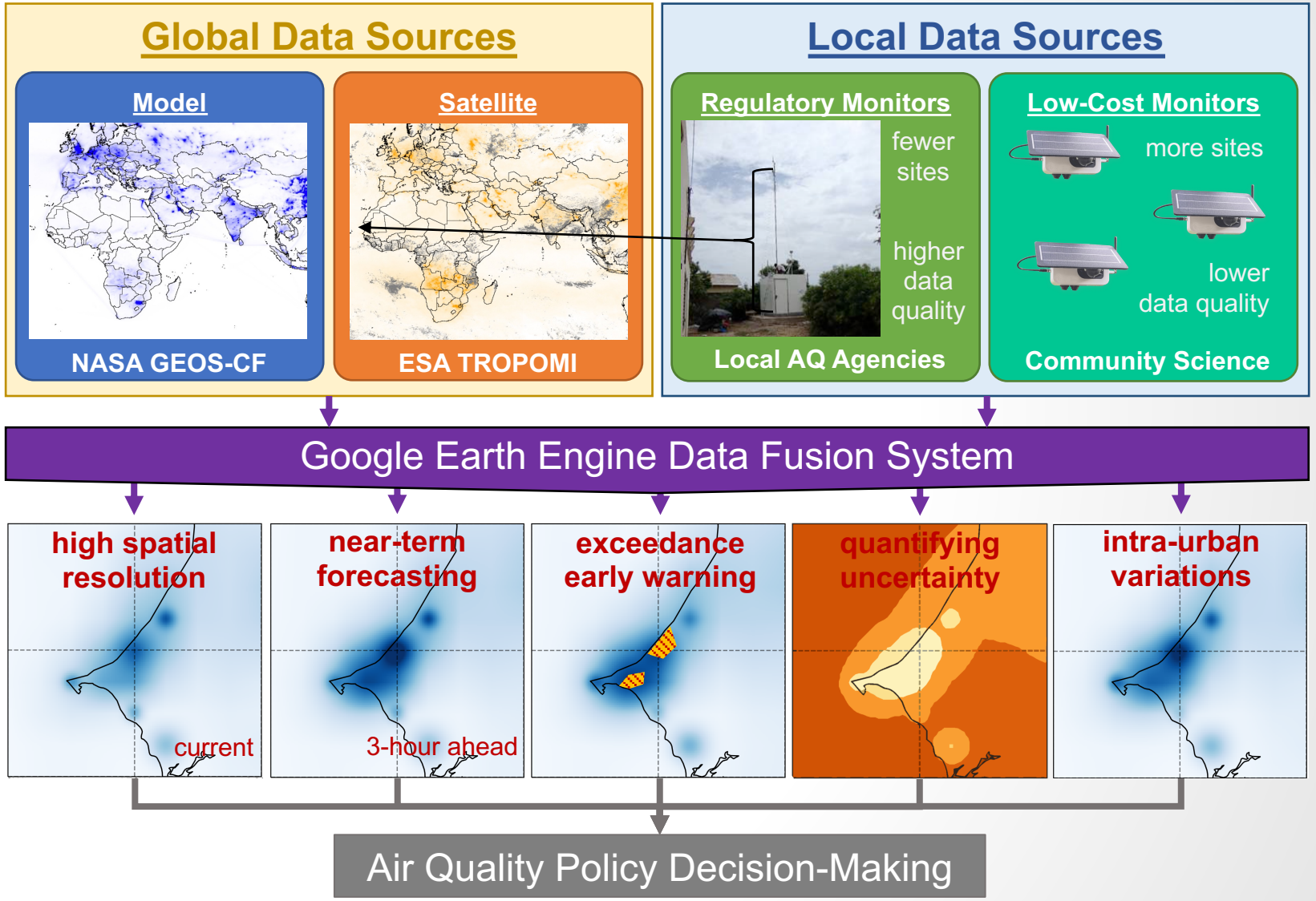
Sonoma Technologies: data fusion system implementation & user interface

WUSTL: air quality data integration expertise (monthly/annual timescales)

Columbia LDEO: experience training end-users in AQ data interpretation

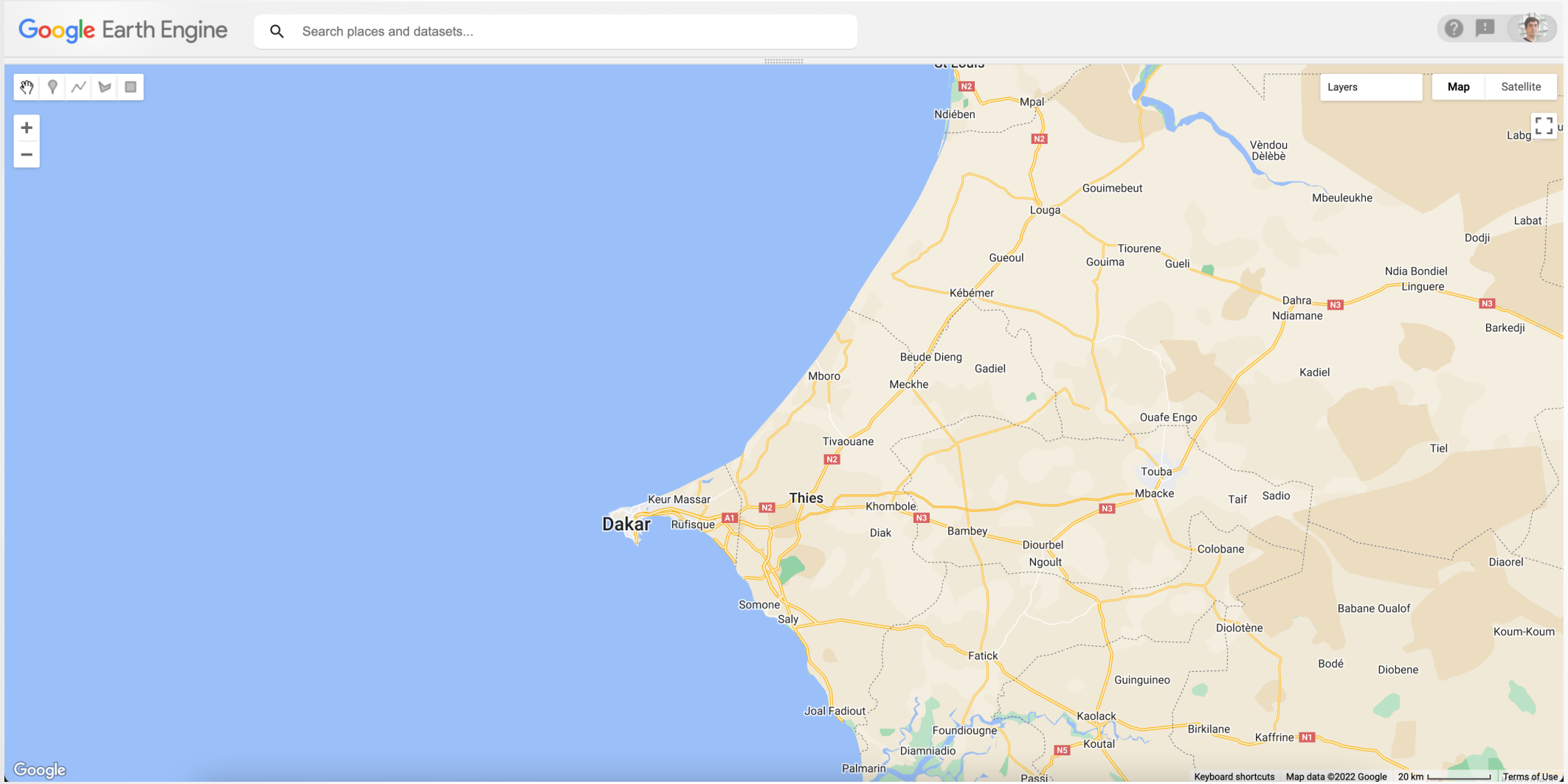
UNEP: integration with global end-users
Dakar, Senegal
Rio de Janeiro, Brazil

US EPA: integration with US end-users in cities TBD



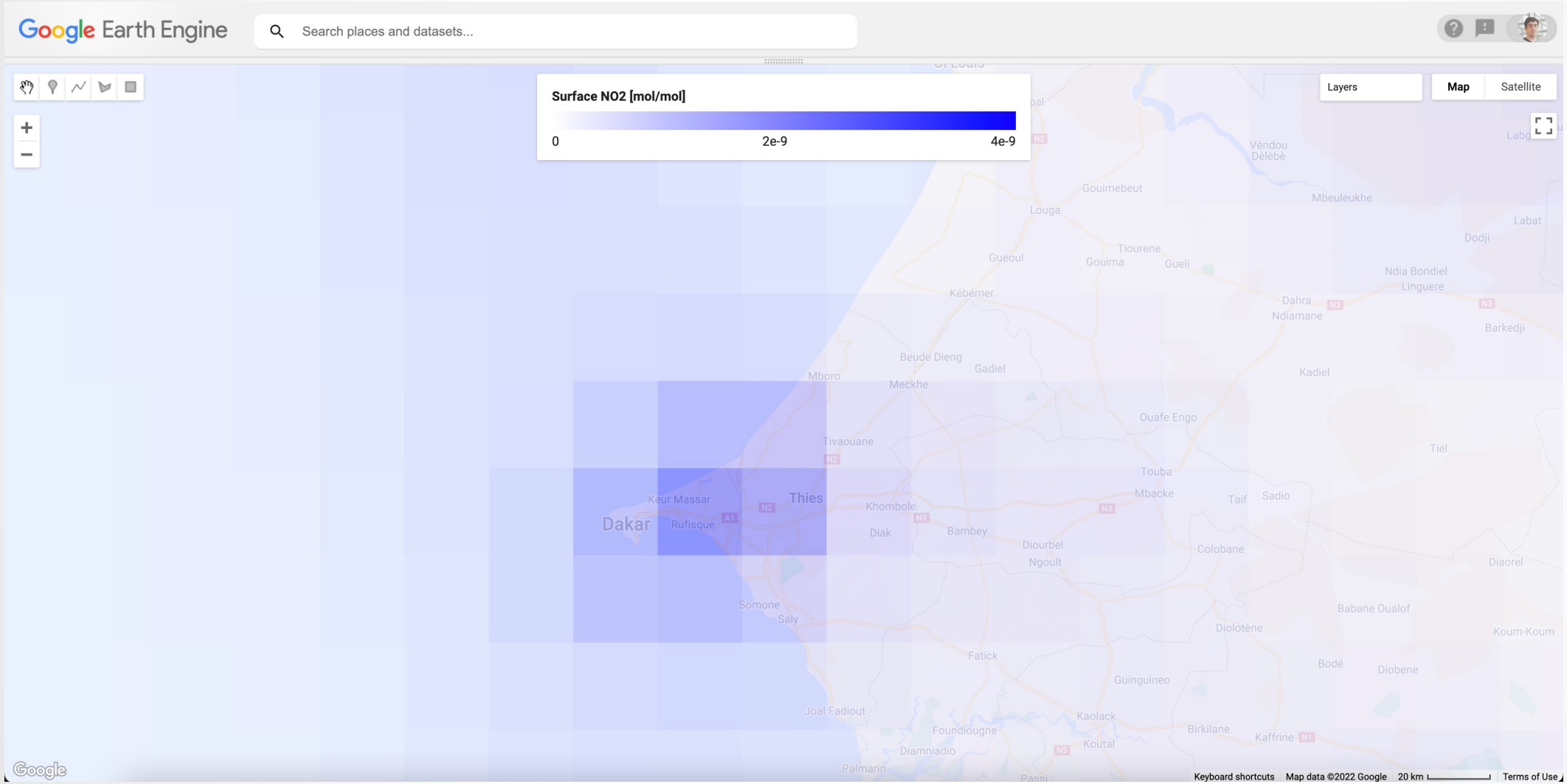
PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



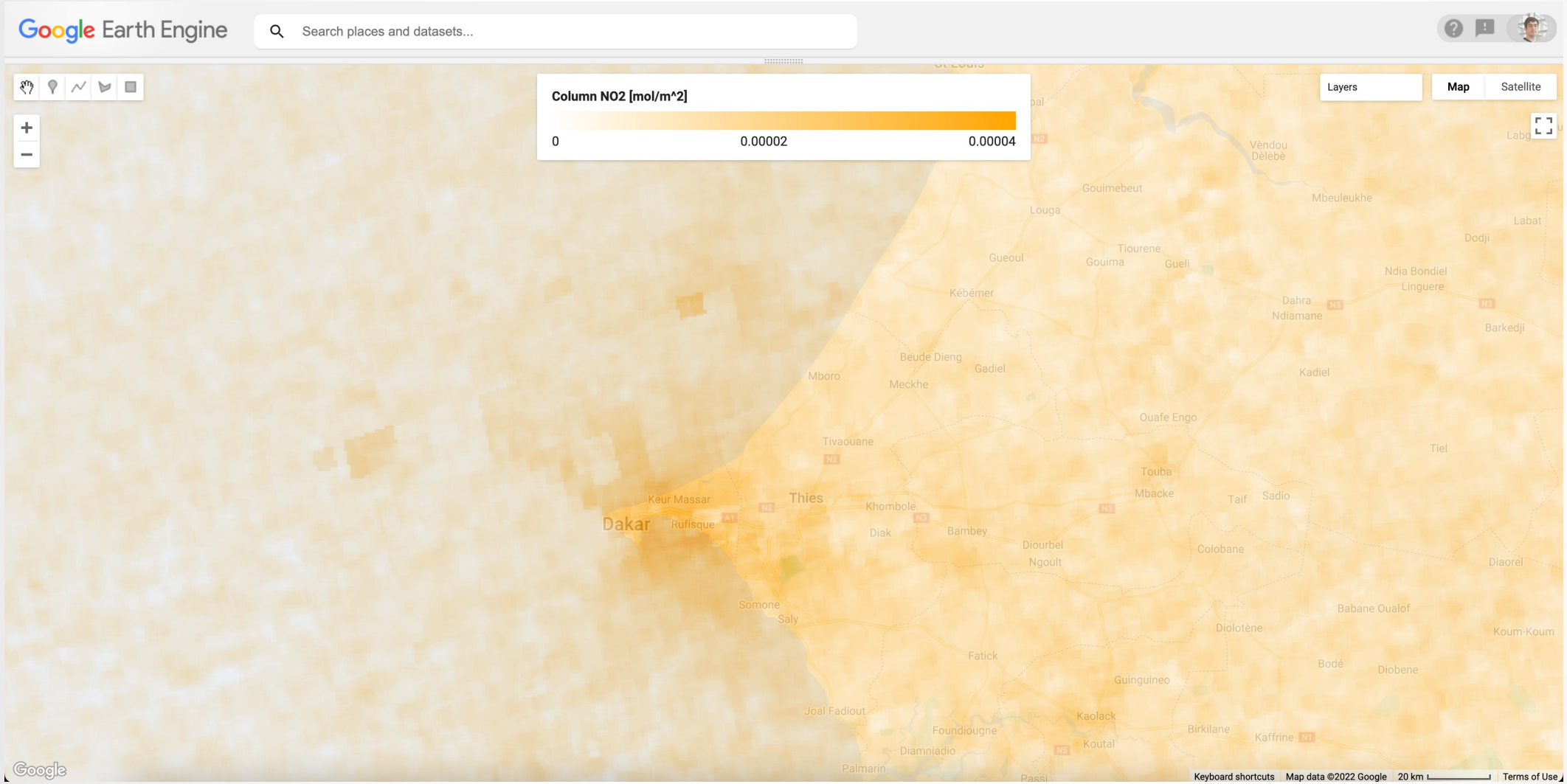
Calibration

Model

The data fusion starts from the model estimated surface concentrations. In this case the GEOS-CF model is used to provide the basic surface concentration estimates. What is shown here is a one-week average of GEOS-CF surface NO_2. This is already available through GEE. Note the spatial resolution of the model (the sizes of the squares) as they compare to Dakar.

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



Calibration

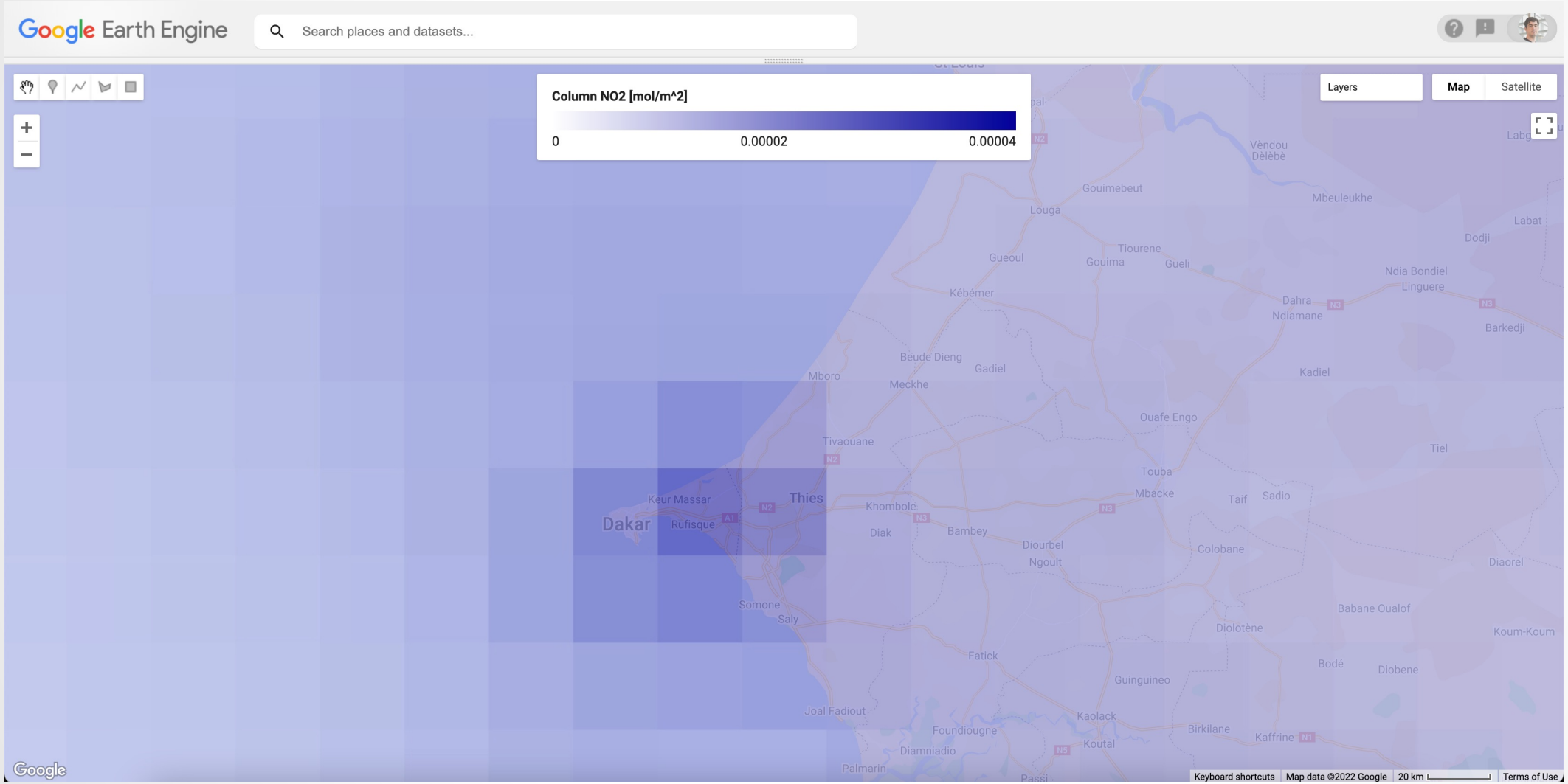
Model

Satellite

The next step is to bring in satellite data (in this case, tropospheric column NO_2 information from the ESA TROPOMI instrument, available through GEE). This is the average for the same week as before. Note the differences in spatial resolution and detail compared to what was available from the model.

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



Calibration

Model

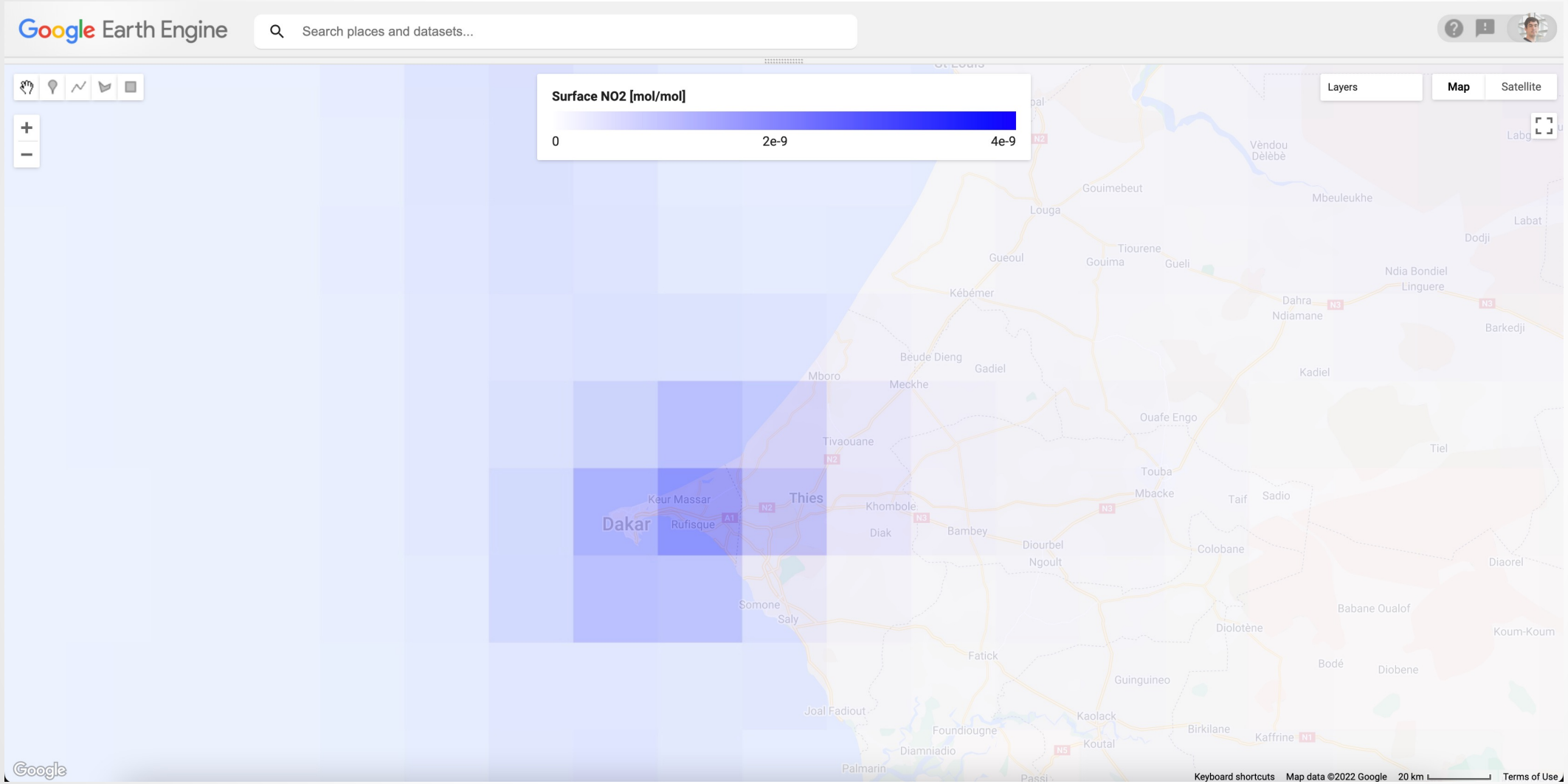
Satellite

Calibrated Relationship

Based on data from a calibration period, typical relationships are established between the satellite and model datasets, facilitated by ancillary information (e.g., the model-estimated tropospheric NO_2 columns, shown here).

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



Calibration

Model

Satellite

Calibrated Relationship

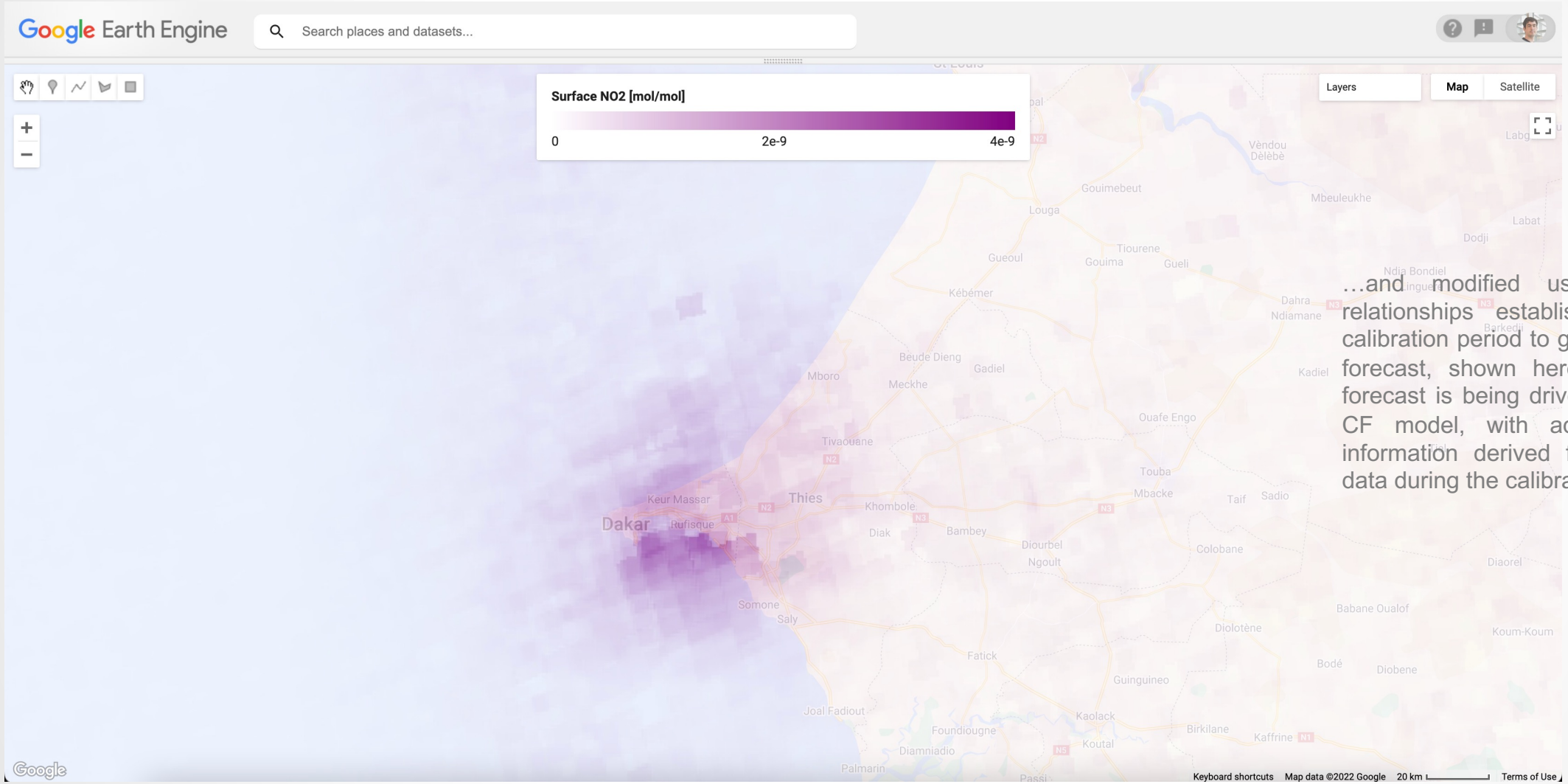
Application

Model Forecast

For forecasting, GEOS-CF forecasts are used to drive the data fusion forecast…

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



Calibration

Model

Satellite

Calibrated Relationship

…and modified using the typical relationships established during the calibration period to generate the fused forecast, shown here. Note that the forecast is being driven by the GEOS-CF model, with additional sub-grid information derived from the satellite data during the calibration period.

Application

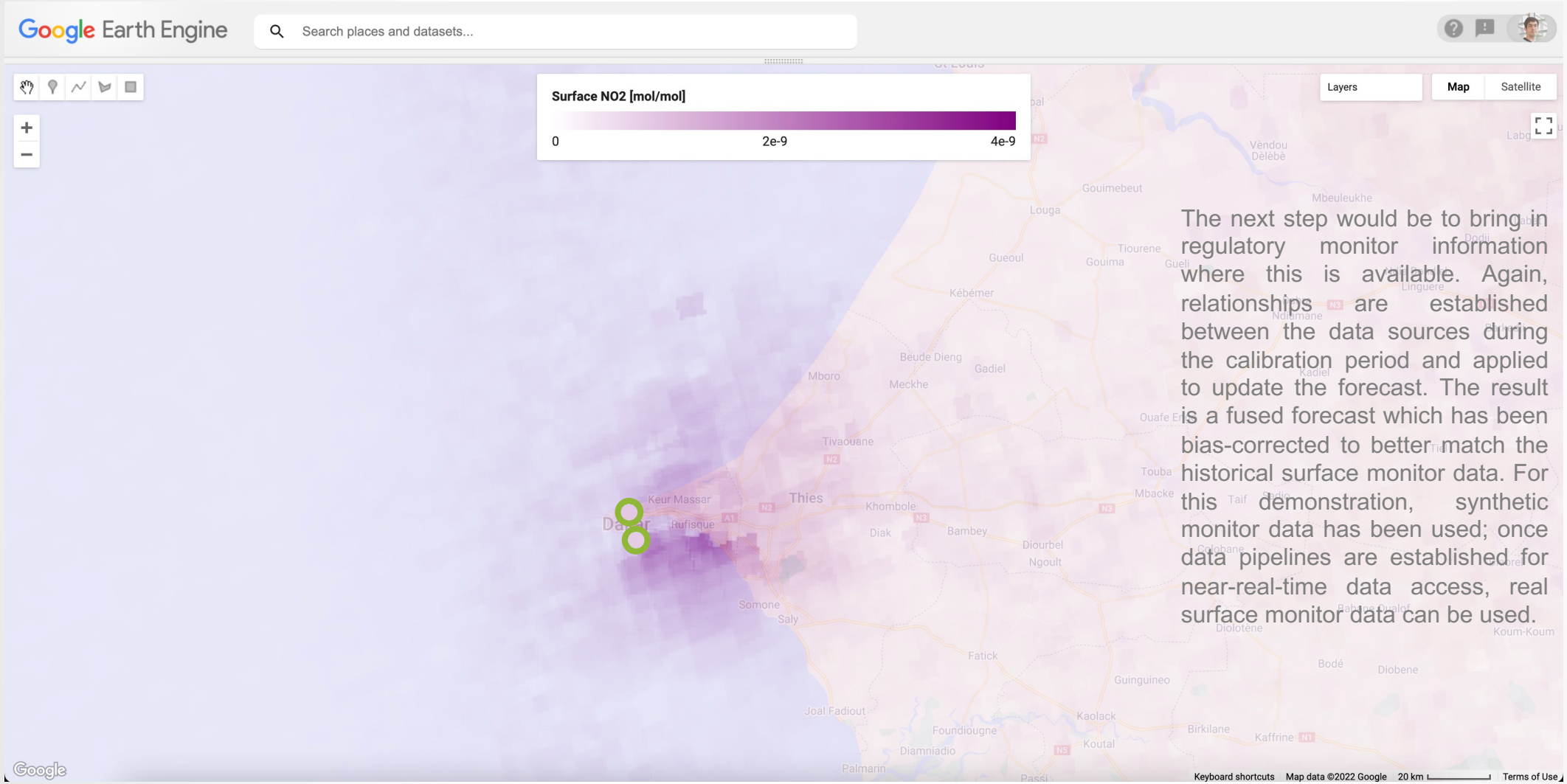
Model Forecast

★

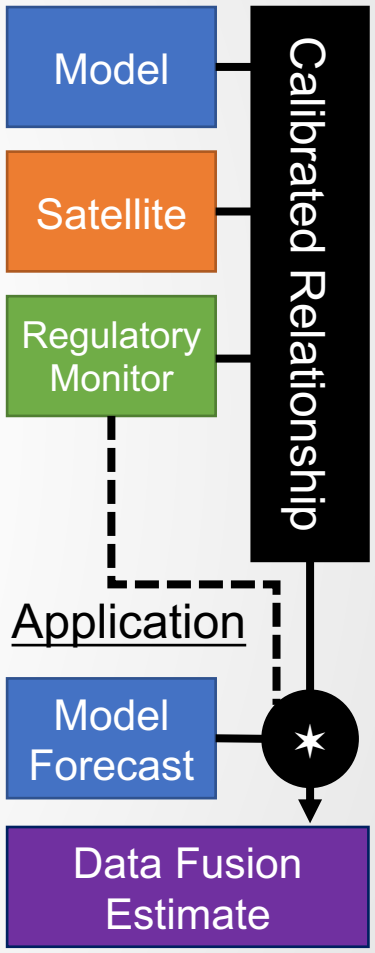
Data Fusion Estimate

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)

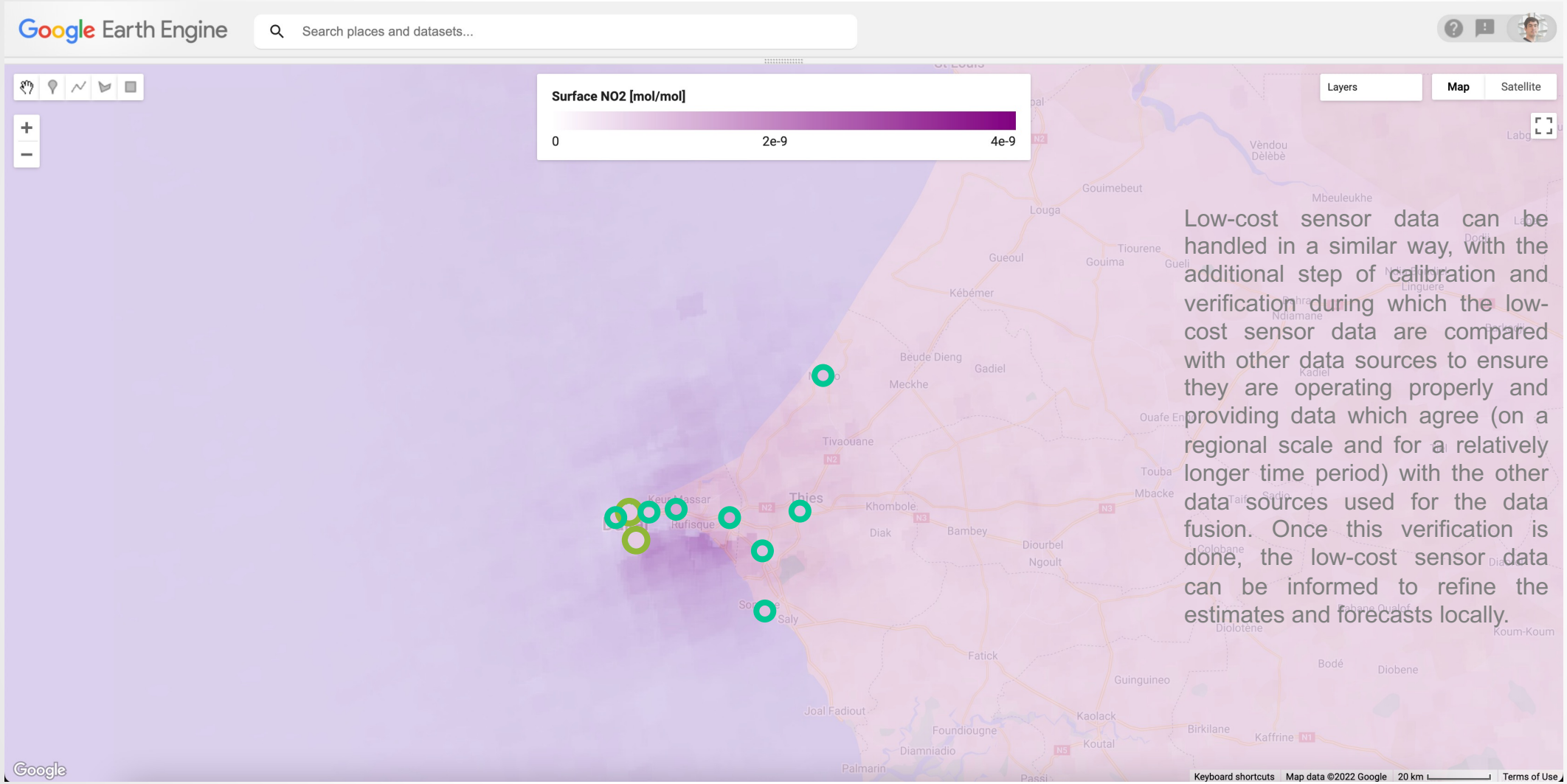


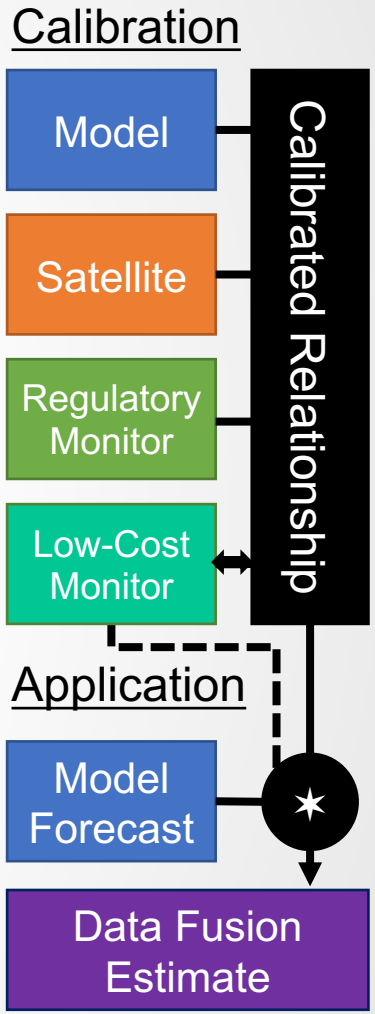
The next step would be to bring in regulatory monitor information where this is available. Again, relationships are established between the data sources during the calibration period and applied to update the forecast. The result is a fused forecast which has been bias-corrected to better match the historical surface monitor data. For this demonstration, synthetic monitor data has been used; once data pipelines are established for near-real-time data access, real surface monitor data can be used.



PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

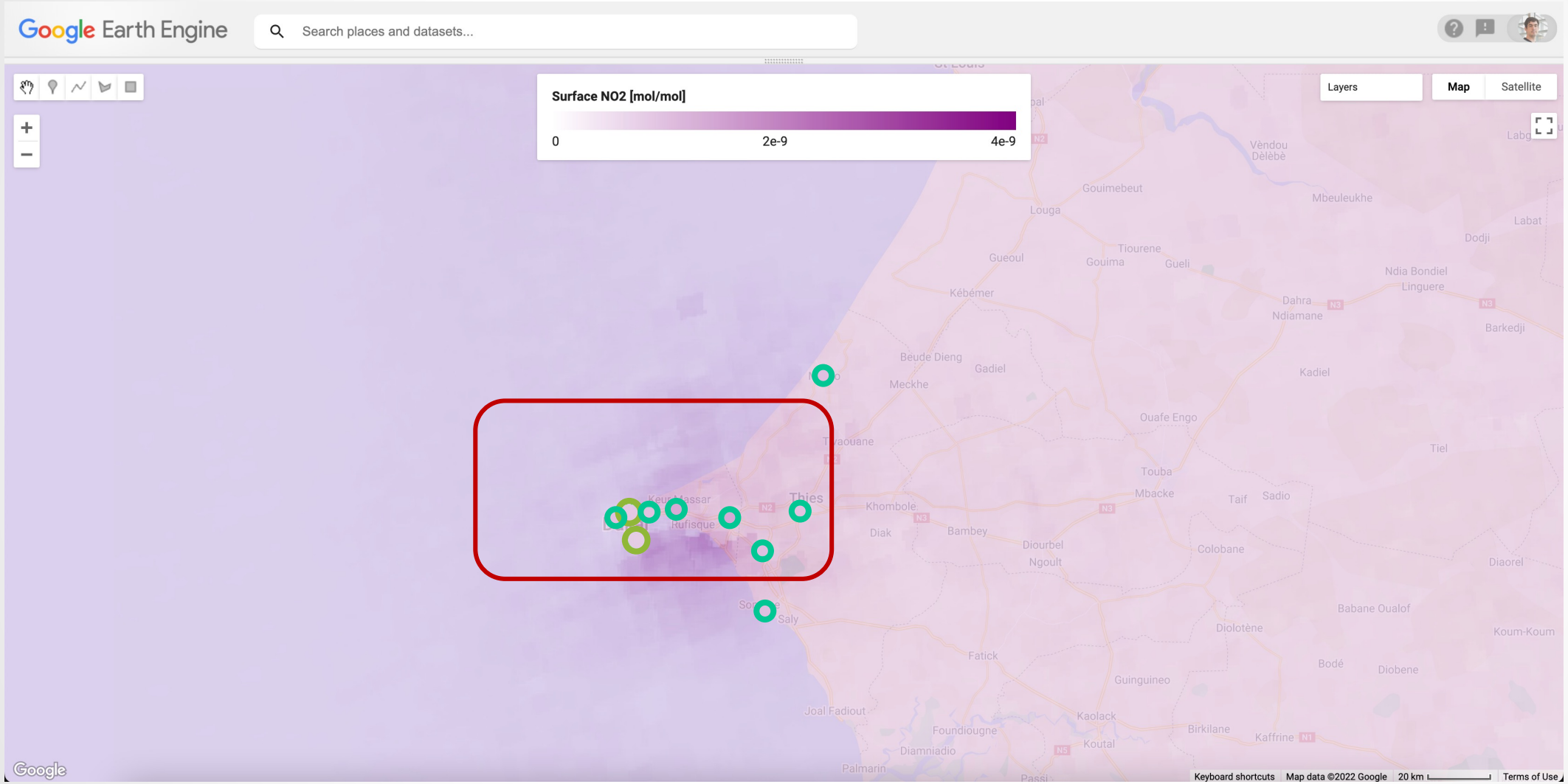
Demonstration of Data Fusion in GEE (preliminary)

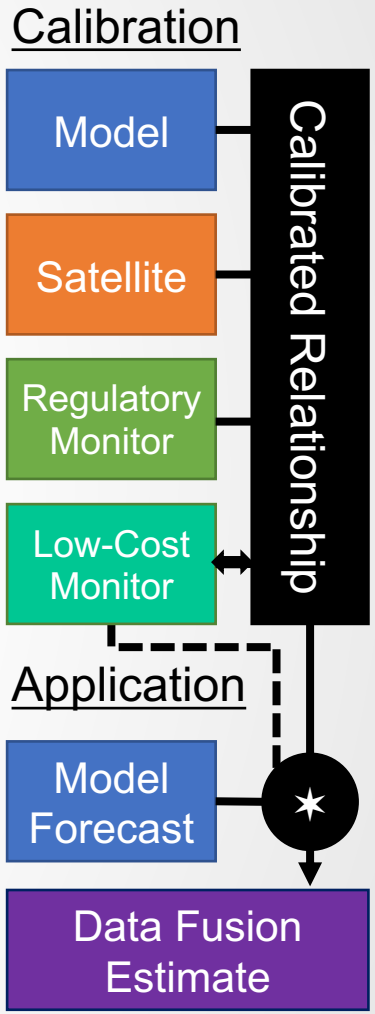




PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)



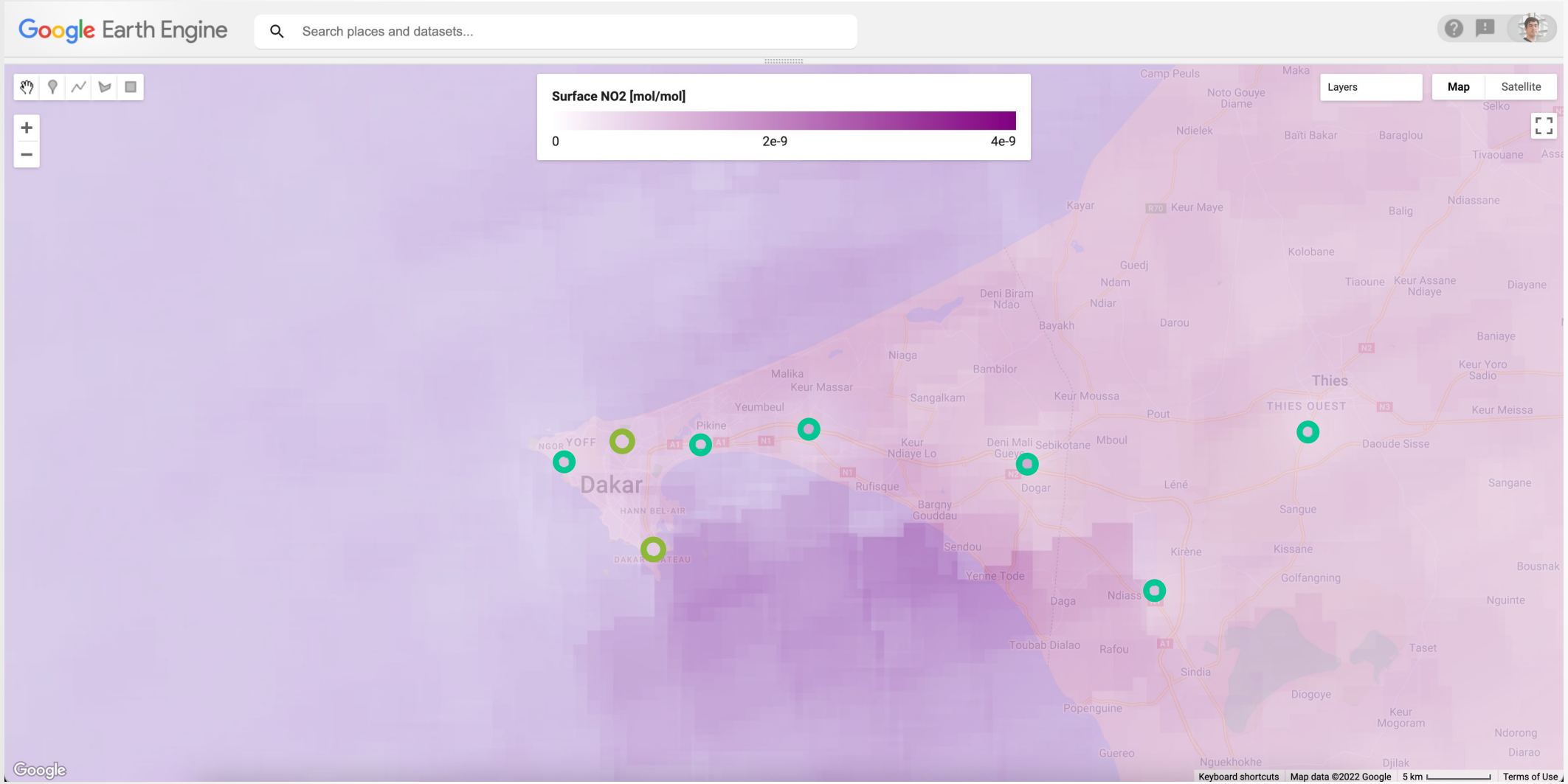


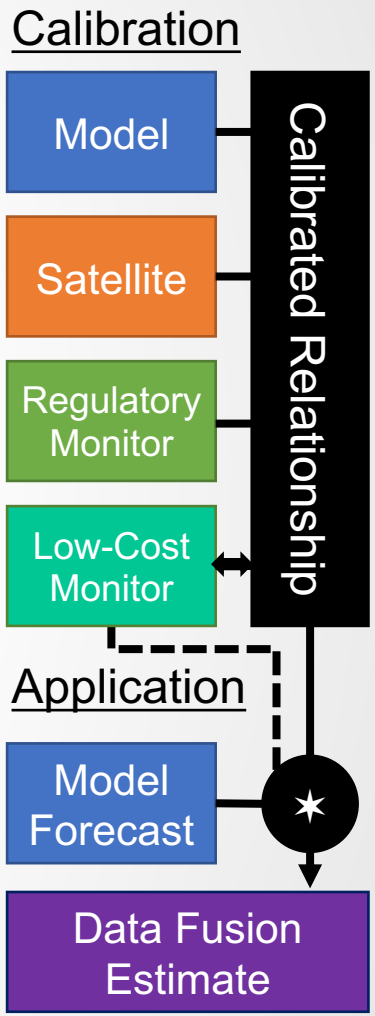
GMAO Global Modeling and Assimilation Office
gmao.gsfc.nasa.gov

PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

GODDARD EARTH SCIENCES

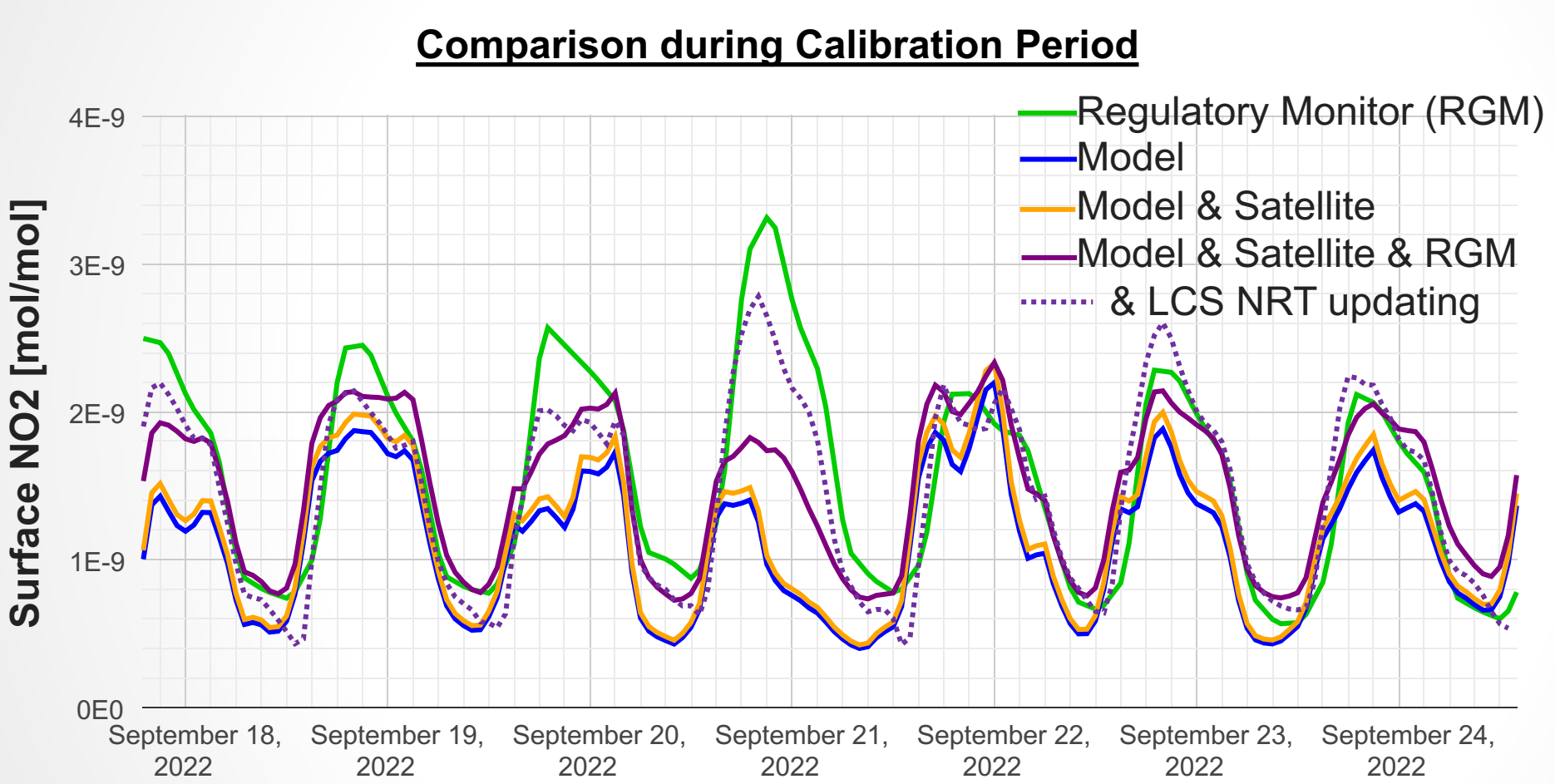
Demonstration of Data Fusion in GEE (preliminary)

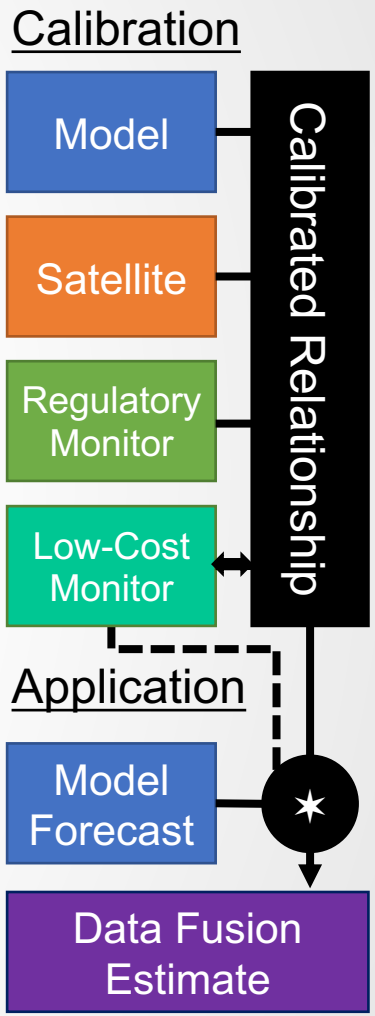




PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

Demonstration of Data Fusion in GEE (preliminary)





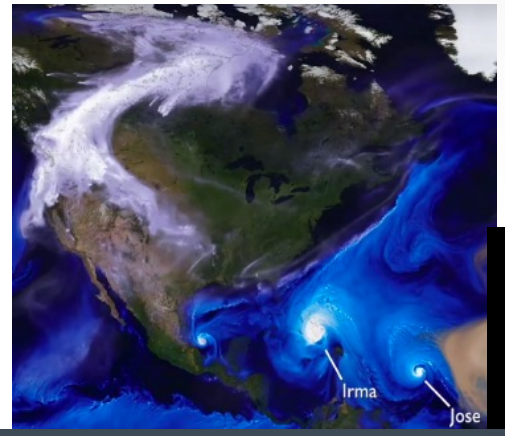
This timeseries illustrates how the data fusion approach, by successively incorporating more data sources, brings the estimates into closer agreement with the "ground truth" represented by the local regulatory monitor data. Keep in mind that in this example, these regulatory monitor data are synthetic, but nonetheless the data fusion methodology is performing as we would expect.

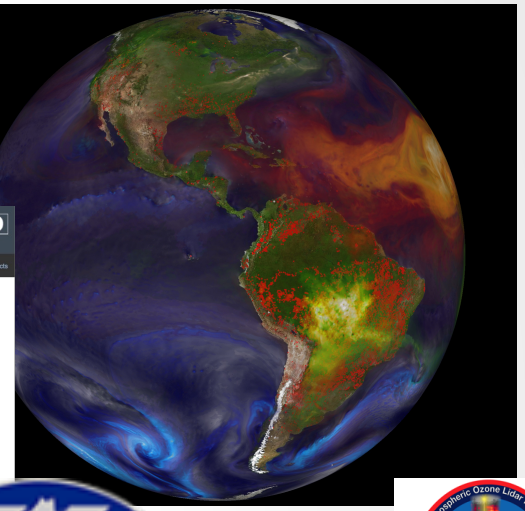
PI K. Emma Knowland: k.e.knowland@nasa.gov
Co-I Nathan Pavlovic: npavlovic@sonomatech.com
Co-I Carl Malings carl.a.malings@nasa.gov

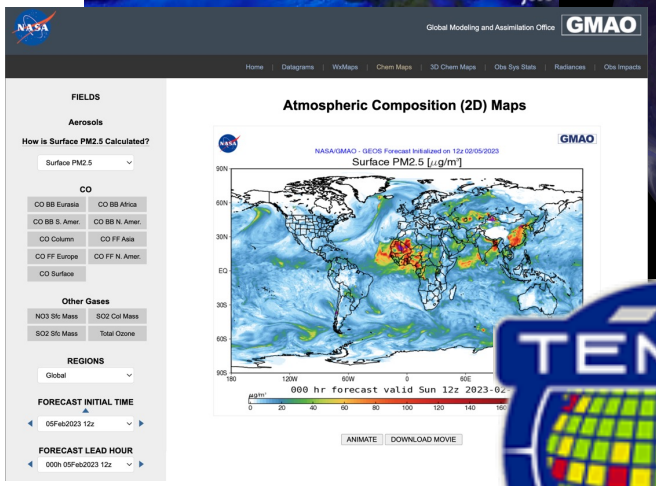
Summary

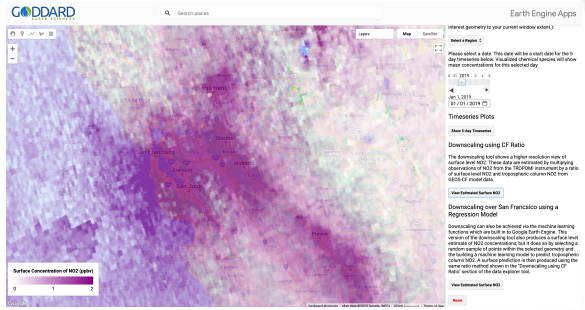
- GMAO has a state-of-the-science Earth System model and data assimilation system
- GMAO products are available to the public with file formats for both scientists and air quality managers to use.
- NASA products are research products
- Expanding our capability to provide downscaled bias-corrected and data-constrained forecasts is a priority
- We want to hear from you so we can make our products as useful as possible!

Thank you for listening!









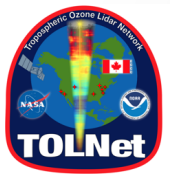
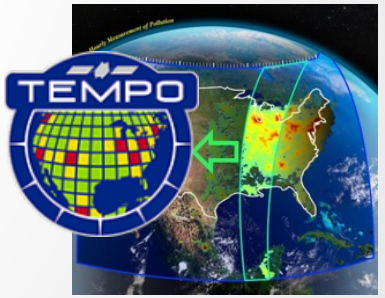
GMAO Global Modeling and Assimilation Office
gmao.gsfc.nasa.gov

K. Emma Knowland
k.e.knowland@nasa.gov

GEOS-CF v1 Status

- Daily GEOS-CF global 5-day composition forecasts at 0.25° (25km) resolution are generated in near-real time:
  - High-resolution historical estimates for fields are available since January 2018
  - Forecast visualizations and links to data available at fluid.nccs.nasa.gov/cf and /cf_map

Keller, C. A., et al. (2021). **Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0**. *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002413. https://doi.org/10.1029/2020MS002413

Knowland, K. E., et al. (2022). **NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0: Stratospheric Composition.** *JAMES* https://doi.org/10.1029/2021MS002852

- Applications include:
  - NASA field missions (e.g., SCOAPE, FIREX-AQ, ACT-America, TRACER-AQ)
  - Daily alerts sent to NASA TOLNet lidar teams (Matt Johnson, NASA Ames)
  - TEMPO a priori for trace gas retrieval
  - Cloud platforms, e.g., Google Earth Engine, WRI Resource Watch, CDC Tracker