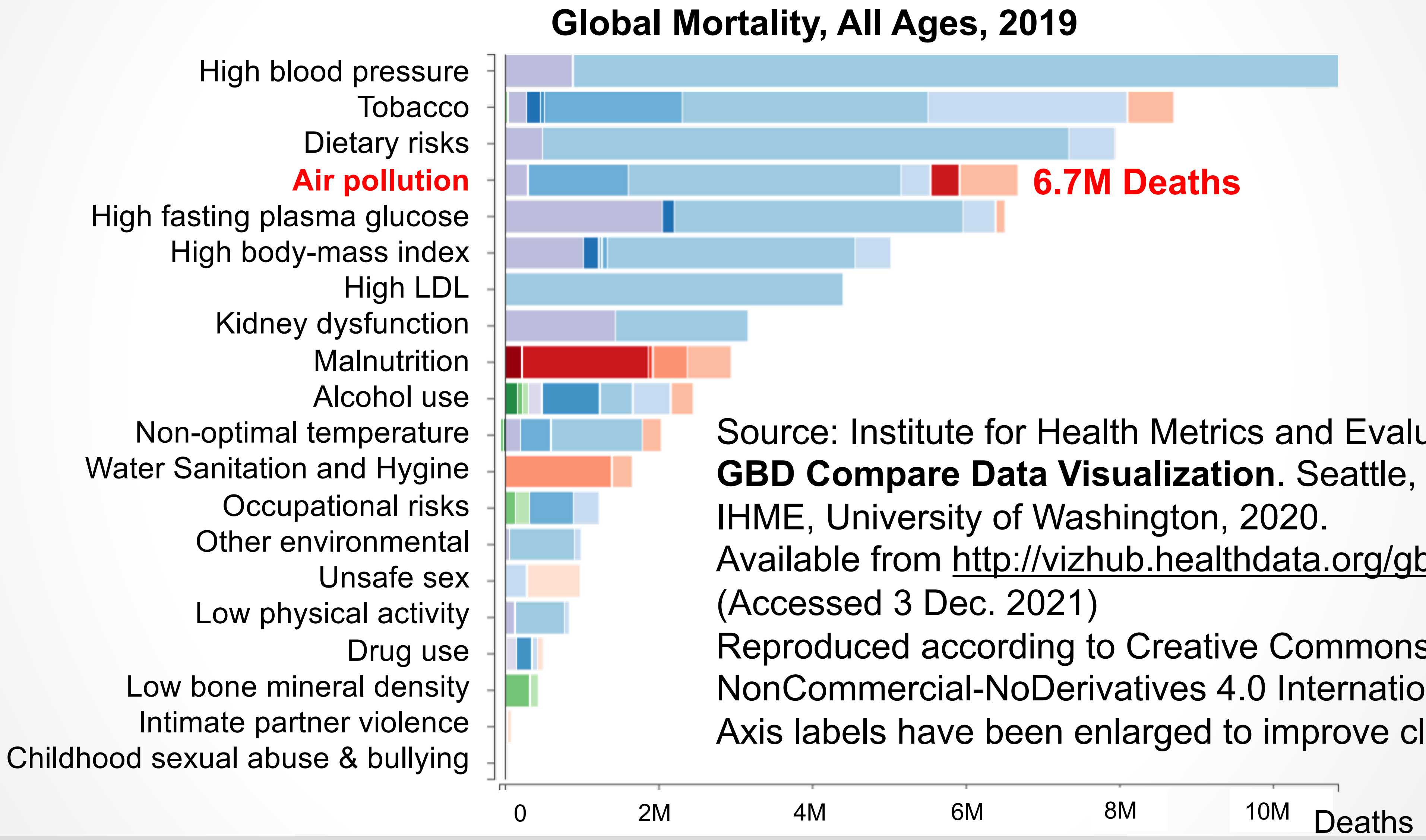
Air Quality Data Fusion with Sensors, Satellites, and Models

Carl Malings

Assistant Research Scientist at Morgan State University (MSU)
working in the Global Modelling and Assimilation Office (GMAO)
part of the Earth Sciences Division (ESD)
at the NASA Goddard Space Flight Center (GSFC)

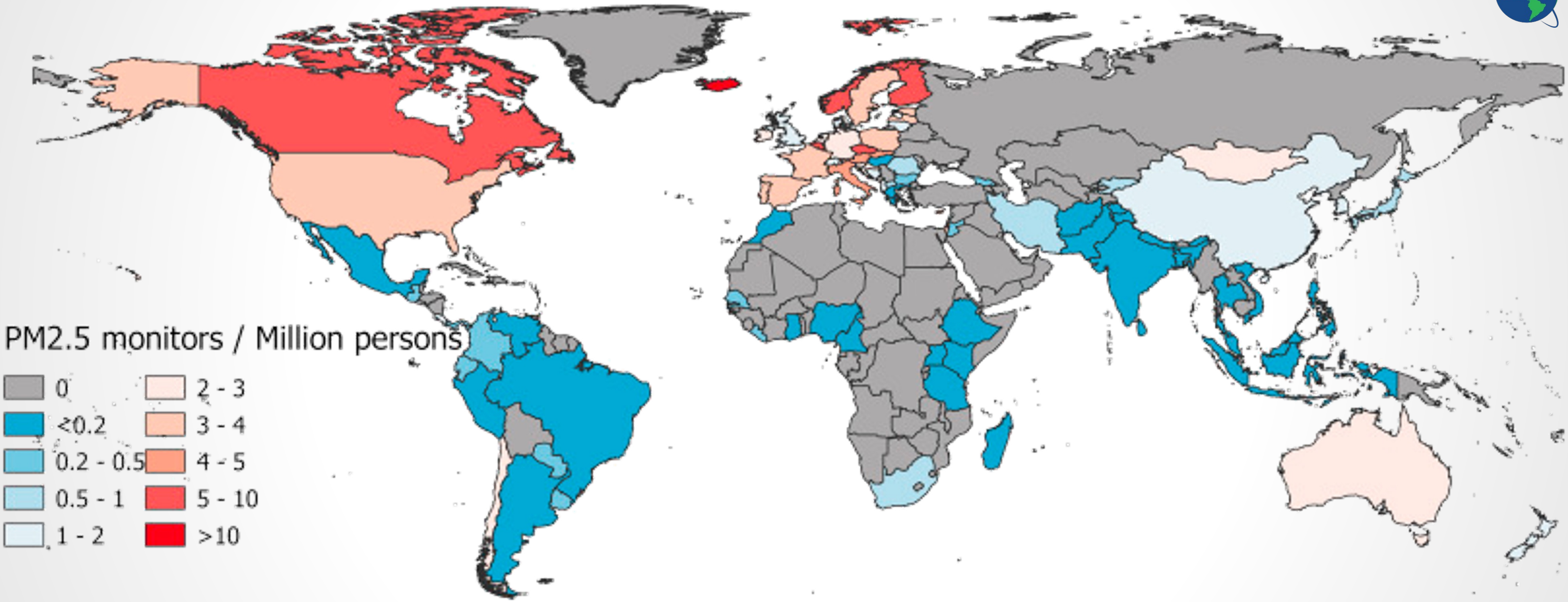
The global risks of poor air quality are severe…



Source: Institute for Health Metrics and Evaluation (IHME). **GBD Compare Data Visualization**. Seattle, WA: IHME, University of Washington, 2020. Available from http://vizhub.healthdata.org/gbd-compare. (Accessed 3 Dec. 2021)

Reproduced according to Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. Axis labels have been enlarged to improve clarity.

…but our knowledge of air quality is incomplete!



Source: Martin et al. (2019), "**No one knows which city has the highest concentration of fine particulate matter**" *Atmospheric Environment*. https://doi.org/10.1016/j.aeaoa.2019.100040

Outline

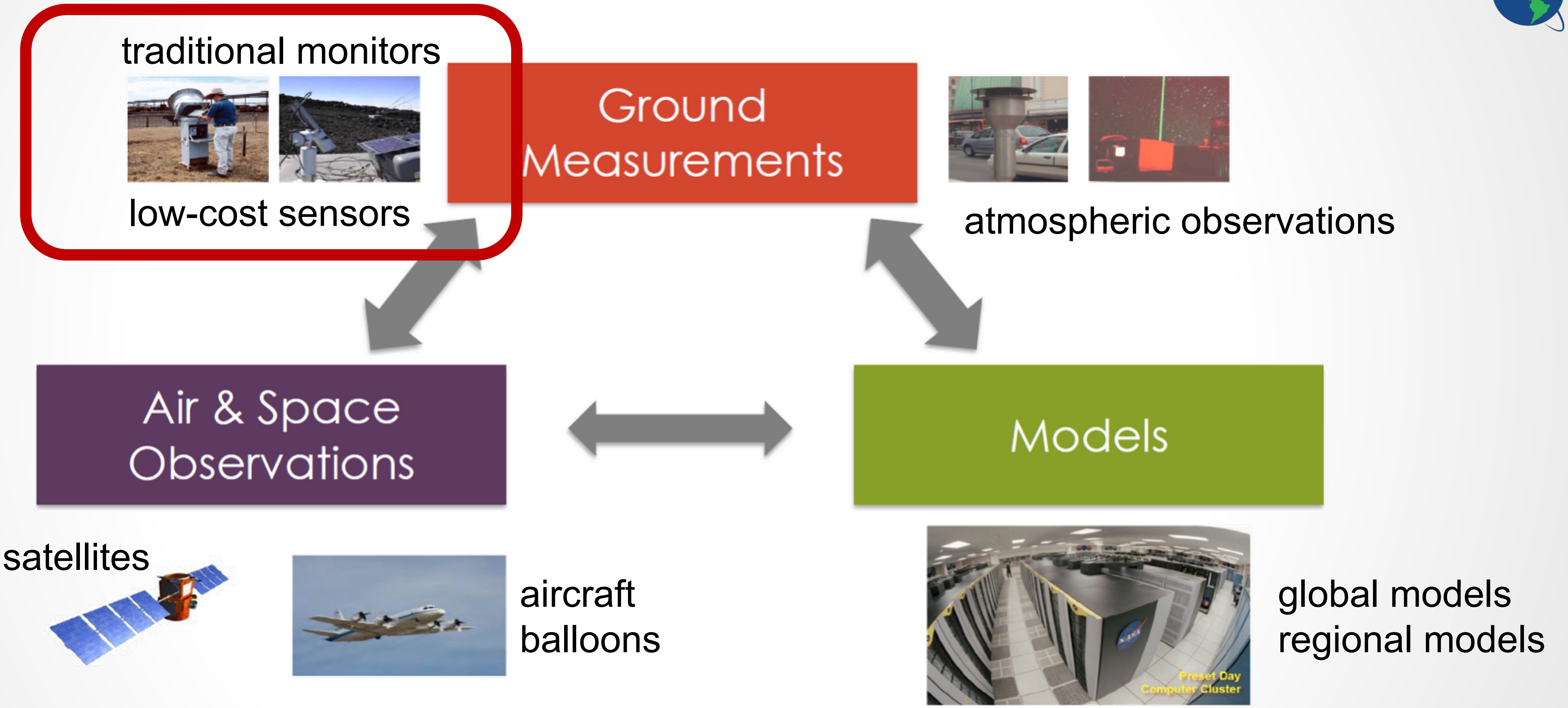
Part 1: Data Sources

- Regulatory Monitors
- Low-Cost Sensors
- Satellite Remote Sensing
- Atmospheric Chemistry and Transport Models
- Details on NASA GEOS-CF

Part 2: Data Fusion

- General Technique
- Example application in Google Earth Engine
- Ongoing Work: Uncertainty Quantification

How do we measure and understand air quality?



Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

Traditional Regulatory Monitors

air quality monitoring trailer of the Houston Health Department, www.houstontx.gov/health/Environmental

MetOne BAM-1020 for Particulate Matter, metone.com

2BTech Model 405 for NOx, twobtech.com

+ accurate
- expensive
? representativity

These form the "backbone" of the monitoring system, but are relatively sparse (especially when taking a global view).

Low-Cost Air Quality Sensors

SENSIT RAMP multi-pollutant sensor, gasleaksensors.com

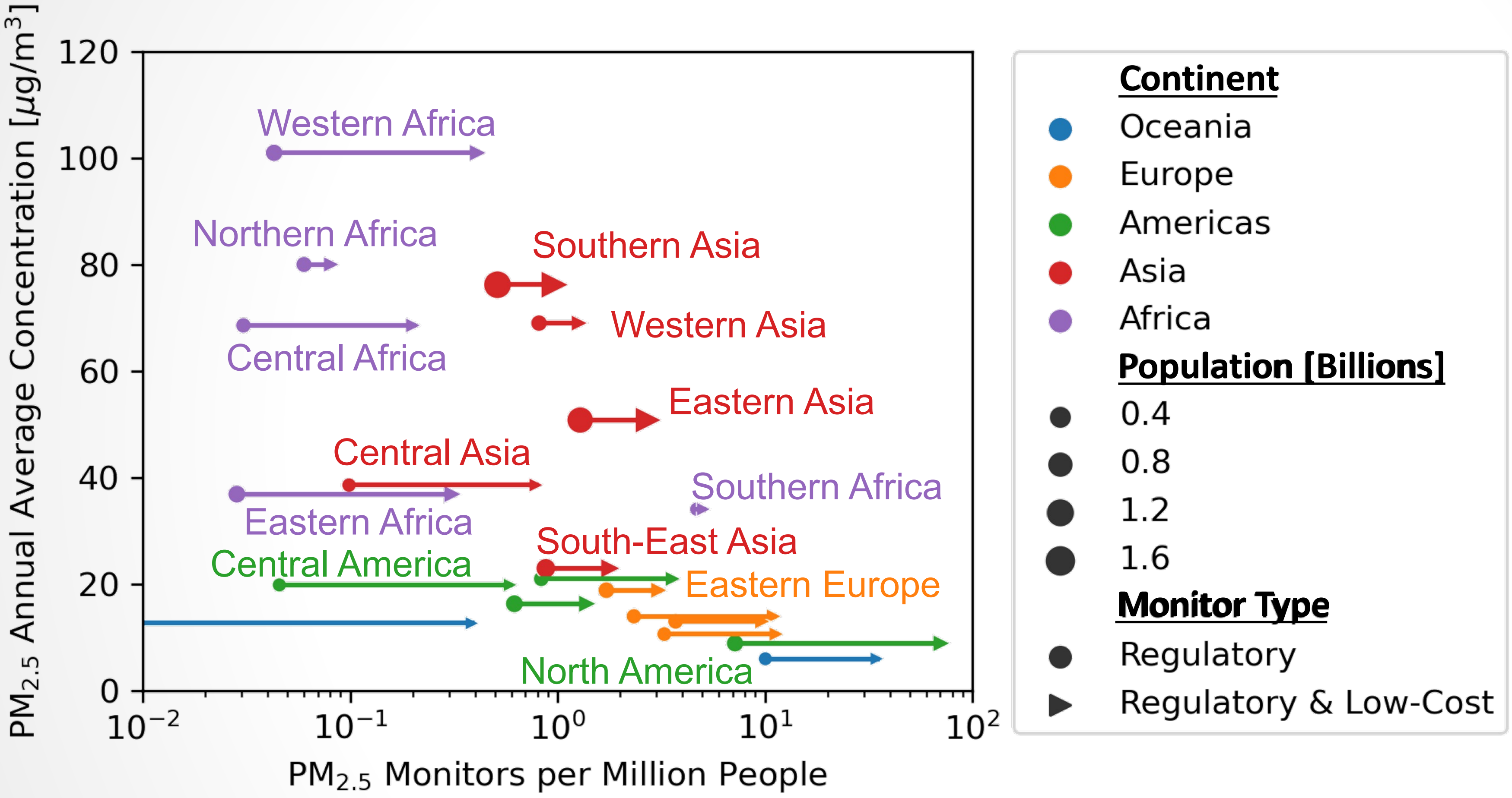
PurpleAir for Particulate Matter, purpleair.com

+ relatively inexpensive
+ dense and/or remote deployment
- greater noise and bias

Clarity Node for PM and NOx, clarity.io

calibration is an open issue, but leveraging network density can offset some of these shortcomings, and allows greater access to air quality monitoring technologies

Global Monitoring of Air Quality (surface monitors)



Source: Malings et al. (2020). "Application of low-cost fine particulate mass monitors to convert satellite AOD to surface concentrations in North America and Africa." *Atmospheric Measurement Techniques.* DOI: 10.5194/amt-13-3873-2020.
Updated analysis based on open air quality data available from openAQ.org

Many regions (especially Africa & Asia) feature high $PM_{2.5}$ concentration but low per-capita $PM_{2.5}$ monitor density, leading to **poor AQ data coverage**.

Including low-cost sensors **increases per-capita AQ monitor density by up to an order of magnitude** in several major regions.

This highlights the need for **openly accessible AQ data** and the need to **integrate traditional & low-cost measurements**

How do we measure and understand air quality?

traditional monitors

low-cost sensors

Ground Measurements

atmospheric observations

Air & Space Observations

Models

satellites

aircraft
balloons

global models
regional models

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

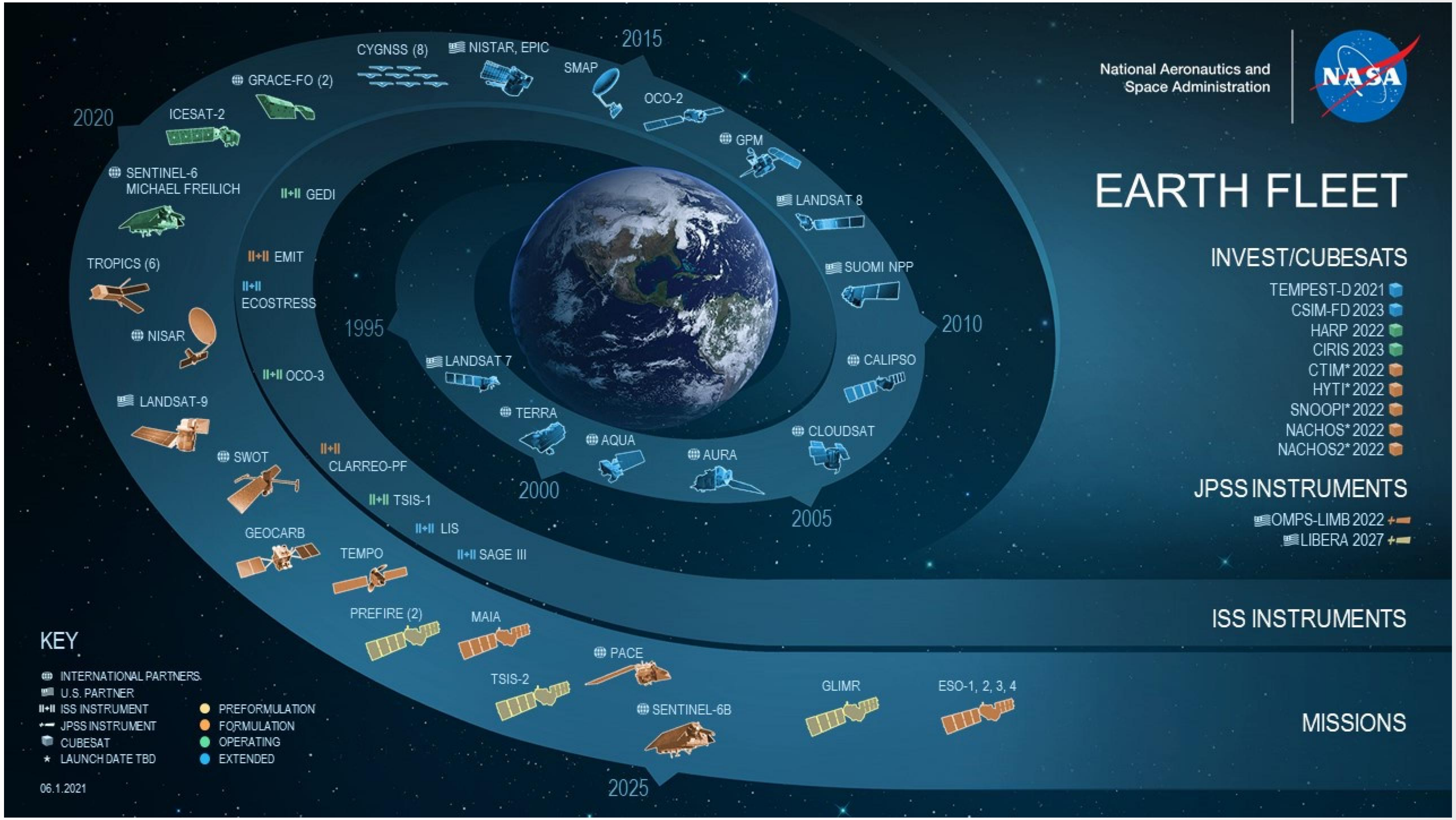
Satellite instruments and retrievals

Geophysical Parameters (useful data product)

Retrieval Algorithm (data processing)

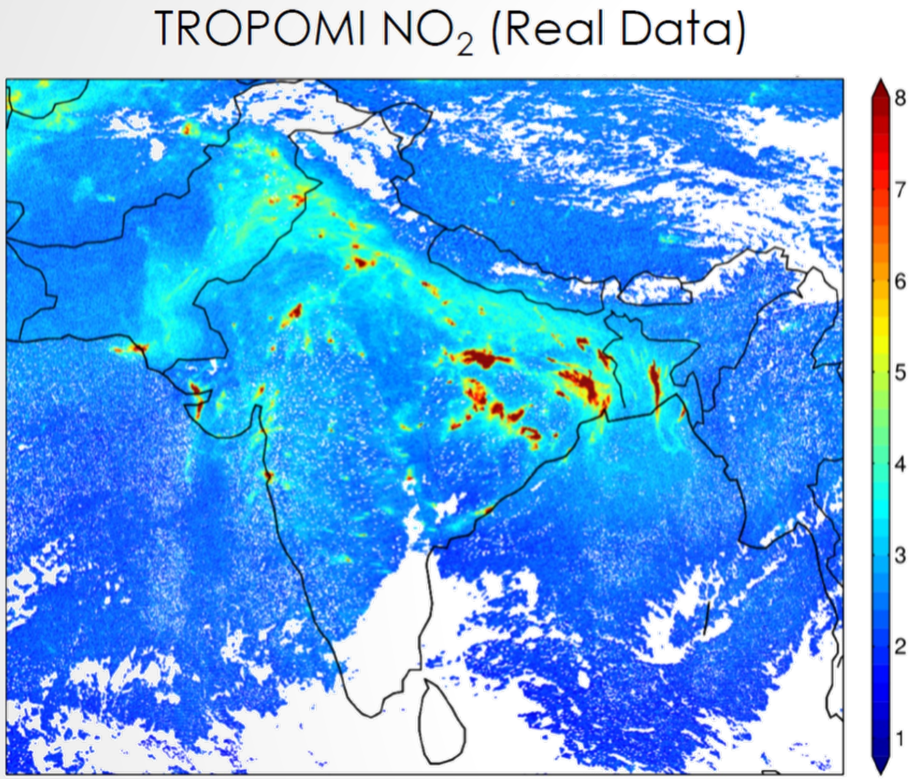
Radiative Transfer Theory (physics)

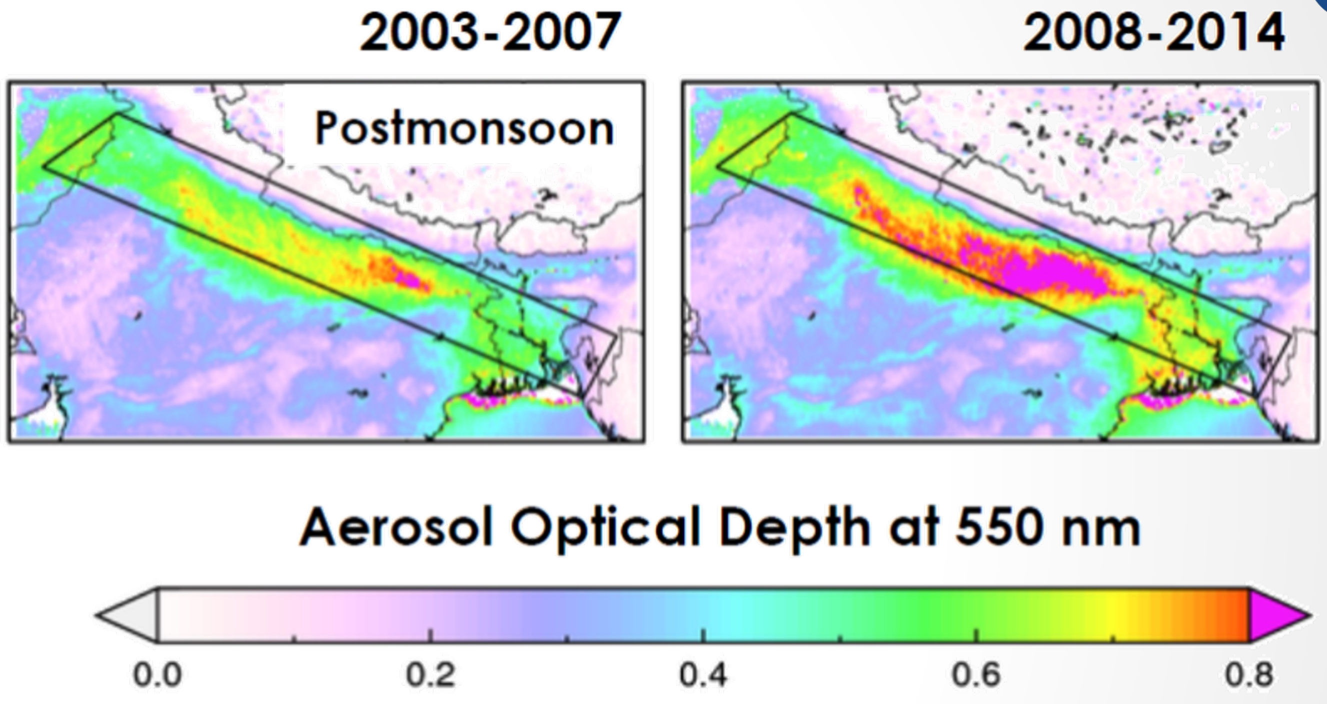
Spectral Radiance (instrument)



Source: NASA Earth Science https://science.nasa.gov/earth-science

What a satellite CAN do for air quality

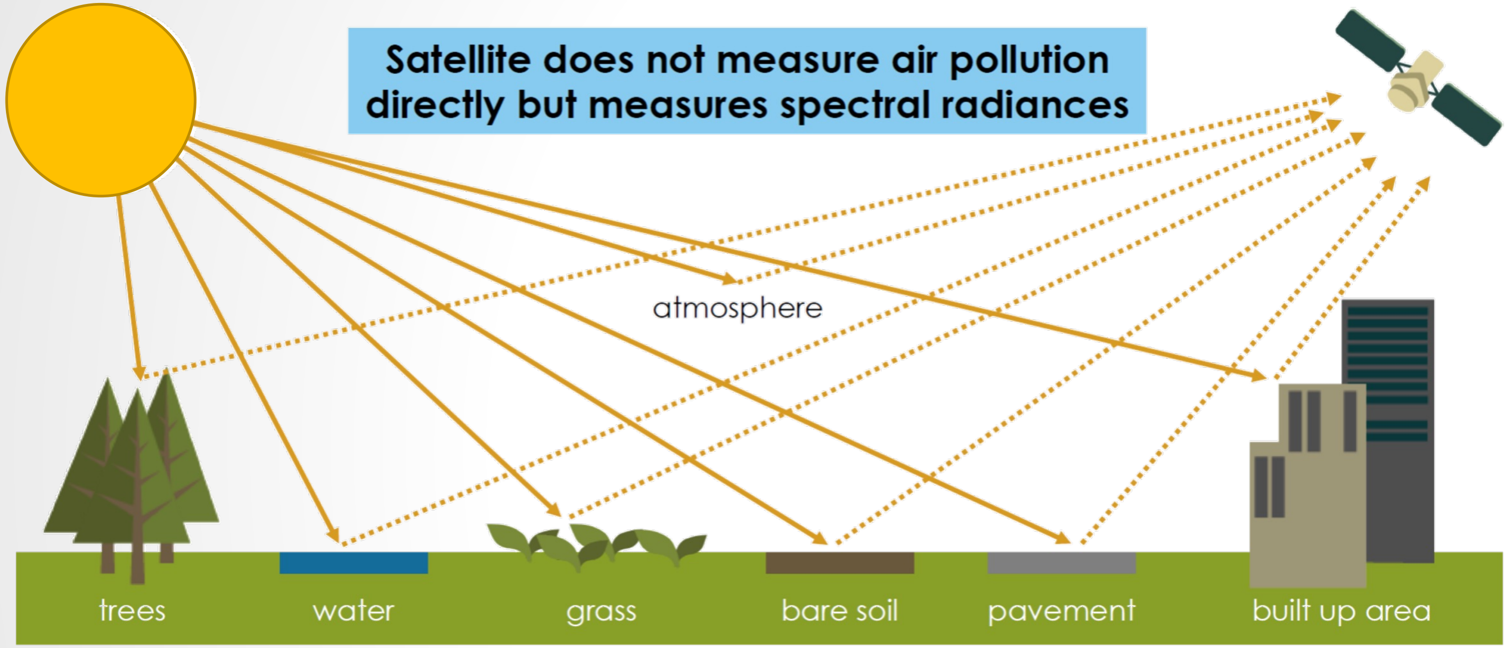


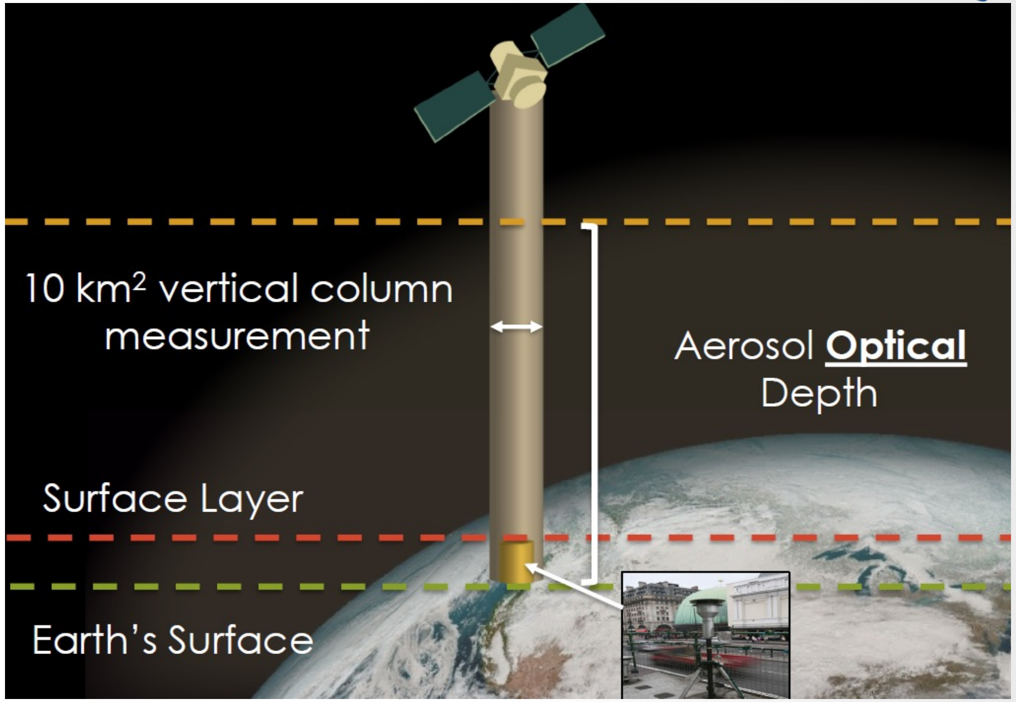


- **Examine a large area**: where are the hotspots? how is long-range transport happening?
- **Track changes over time:** how much has the average concentration over an area changed over time?
- **A picture is worth a million datapoints:** Anyone can understand a satellite photo of a smoke plume.

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

What a satellite CANNOT do for air quality

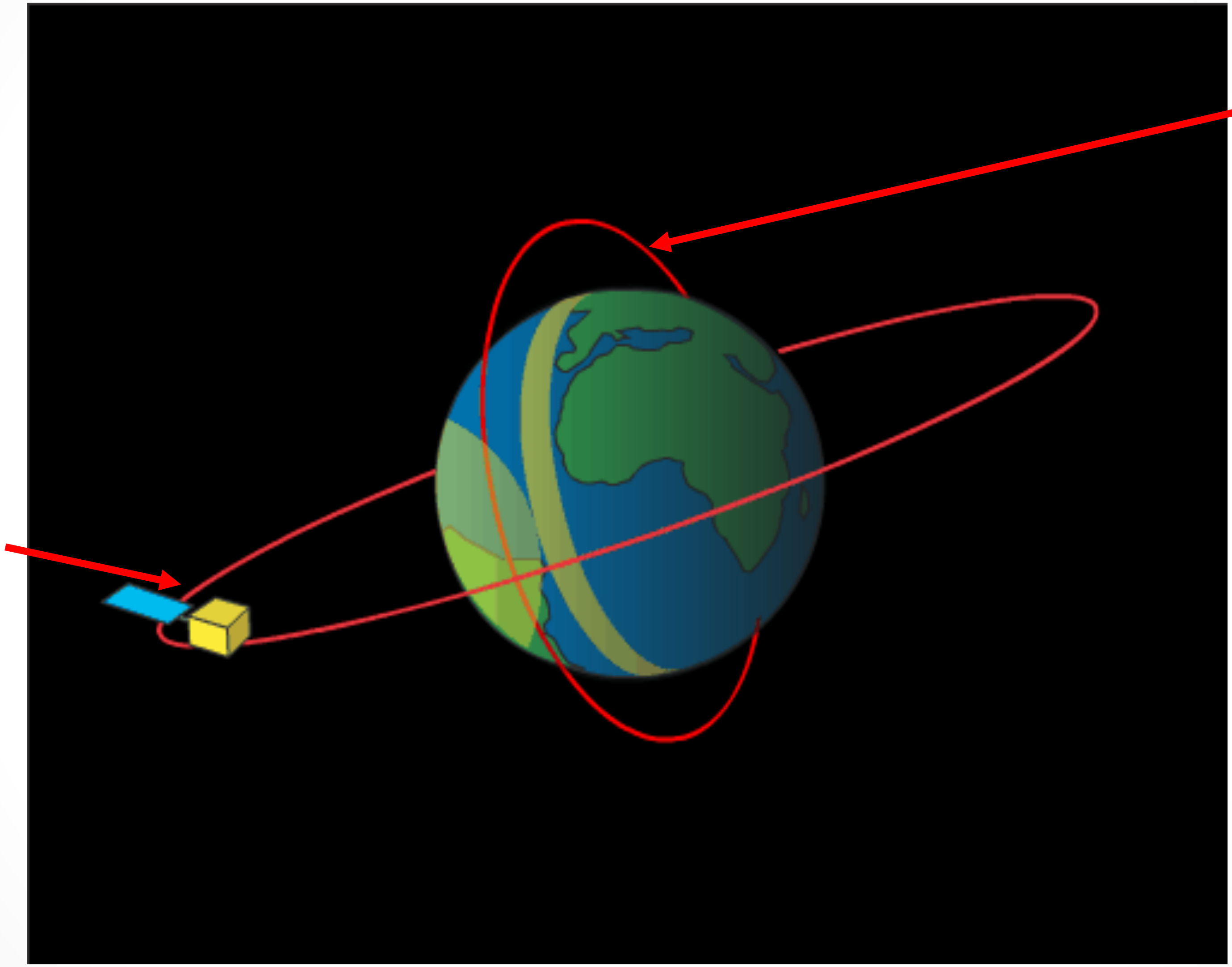




- **See at night:** satellites measure the properties of reflected sunlight passing through the atmosphere.
- **See through clouds:** most satellite measurements are blocked by cloud cover.
- **See what is happening at "nose level":** satellites measure quantities in the whole atmosphere.
- **See at different times of day:** polar-orbiting satellites will observe a location once per day.

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

Common types of orbits for air quality satellites



Polar Orbit

Observes a location about once a day (weather permitting)

Observes at about the same time of day (sun-synchronous)

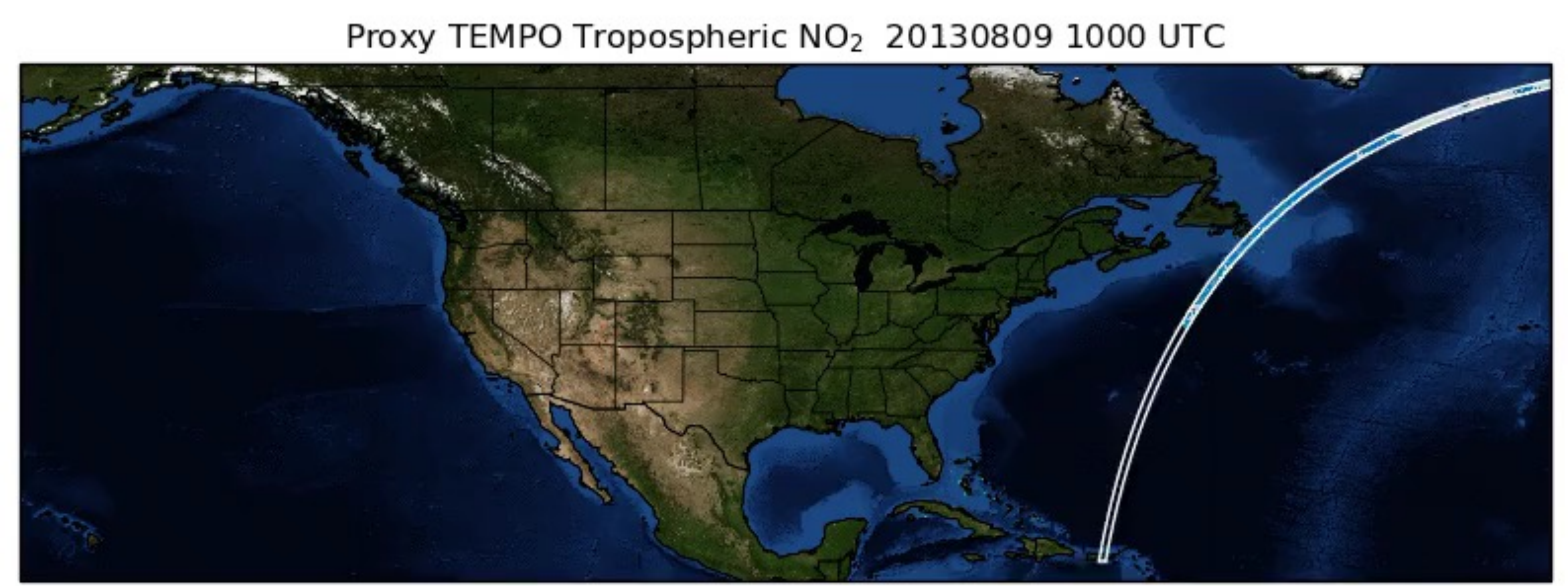
Geostationary Orbit

Observes the same area all the time

Observes throughout the day (weather and light permitting)

source: NOAA
https://scijinks.gov/orbit/

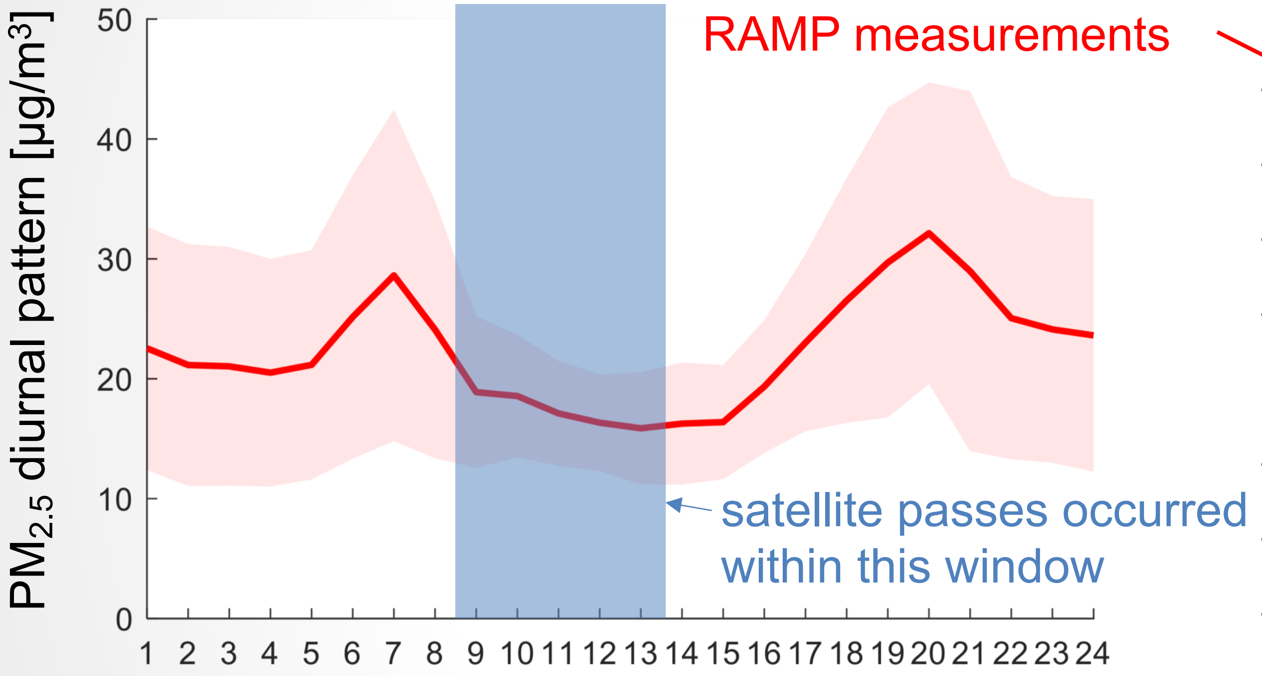
Tropospheric Emissions: Monitoring of POllution (TEMPO)



New geostationary instrument providing hourly trace gas products over North America
Launched April 7, 2023; [“first look” images released August 24, 2023], full data expected mid-2024.

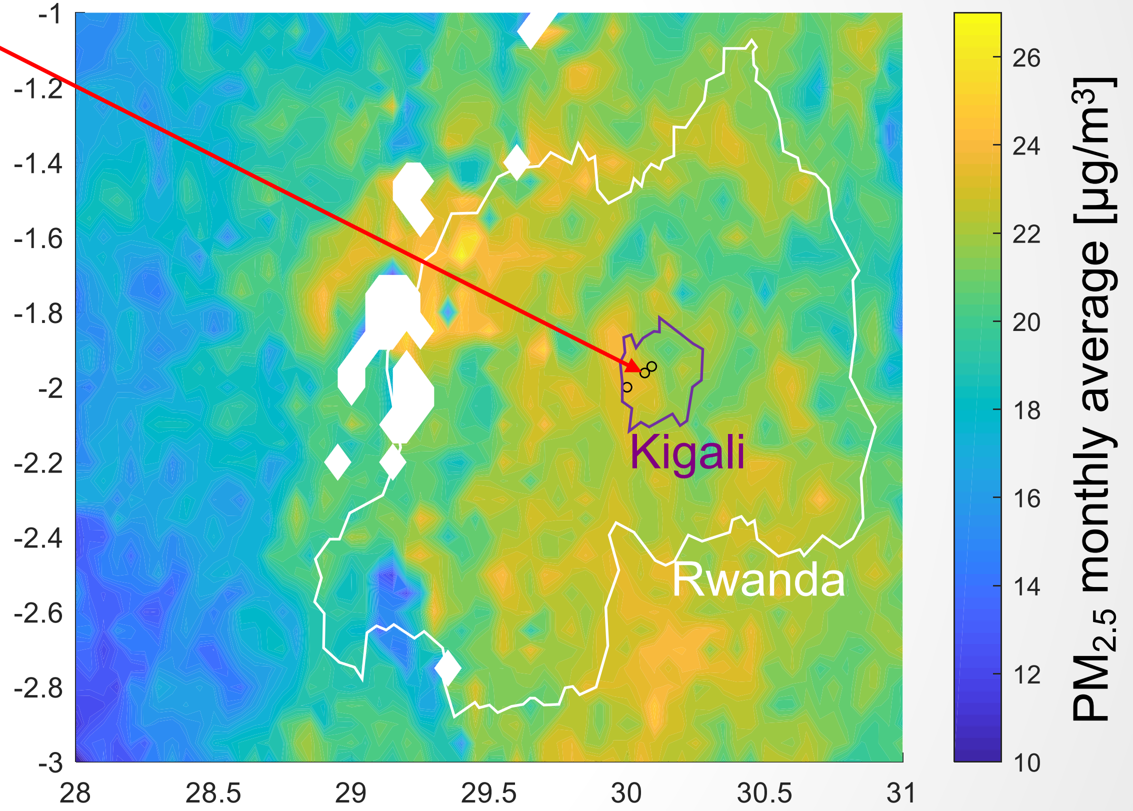
Satellites and surface sensors are complementary

Temporal Coverage (local measurements)



passes during low-concentration
periods may bias long-term averaging

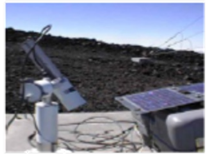
Spatial Coverage (satellite)



Source: Malings et al. (2020), "Application of low-cost fine particulate mass monitors to convert satellite AOD to surface concentrations in North America and Africa" *Atmospheric Measurement Techniques.* DOI: 10.5194/amt-13-3873-2020

How do we measure and understand air quality?

traditional monitors



low-cost sensors

Ground Measurements

atmospheric observations

Air & Space Observations

Models

satellites

aircraft
balloons

global models
regional models

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

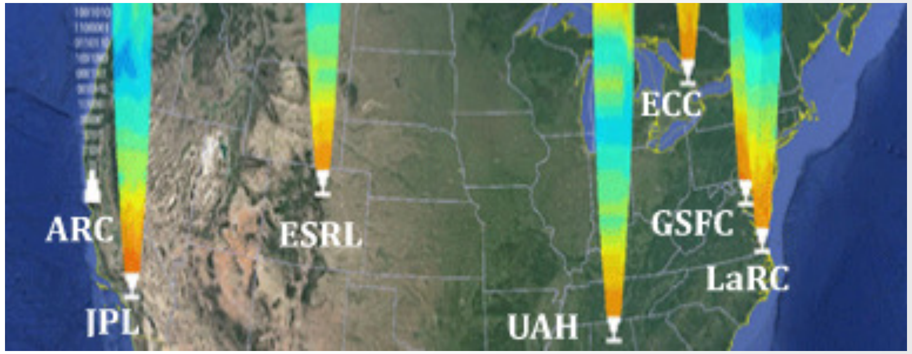
Ground-based atmospheric column observations

Source: https://pandora.gsfc.nasa.gov/

PANDORA
atmospheric gases
(NO_2, Ozone)

TOLNET
LIDAR measurements
of Ozone profiles

AERONET
aerosol optical depth
(relevant to PM)



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

Source: https://www-air.larc.nasa.gov/missions/TOLNet/index.html

How do we measure and understand air quality?

traditional monitors

low-cost sensors

Ground Measurements

atmospheric observations

Air & Space Observations

Models

satellites

aircraft
balloons

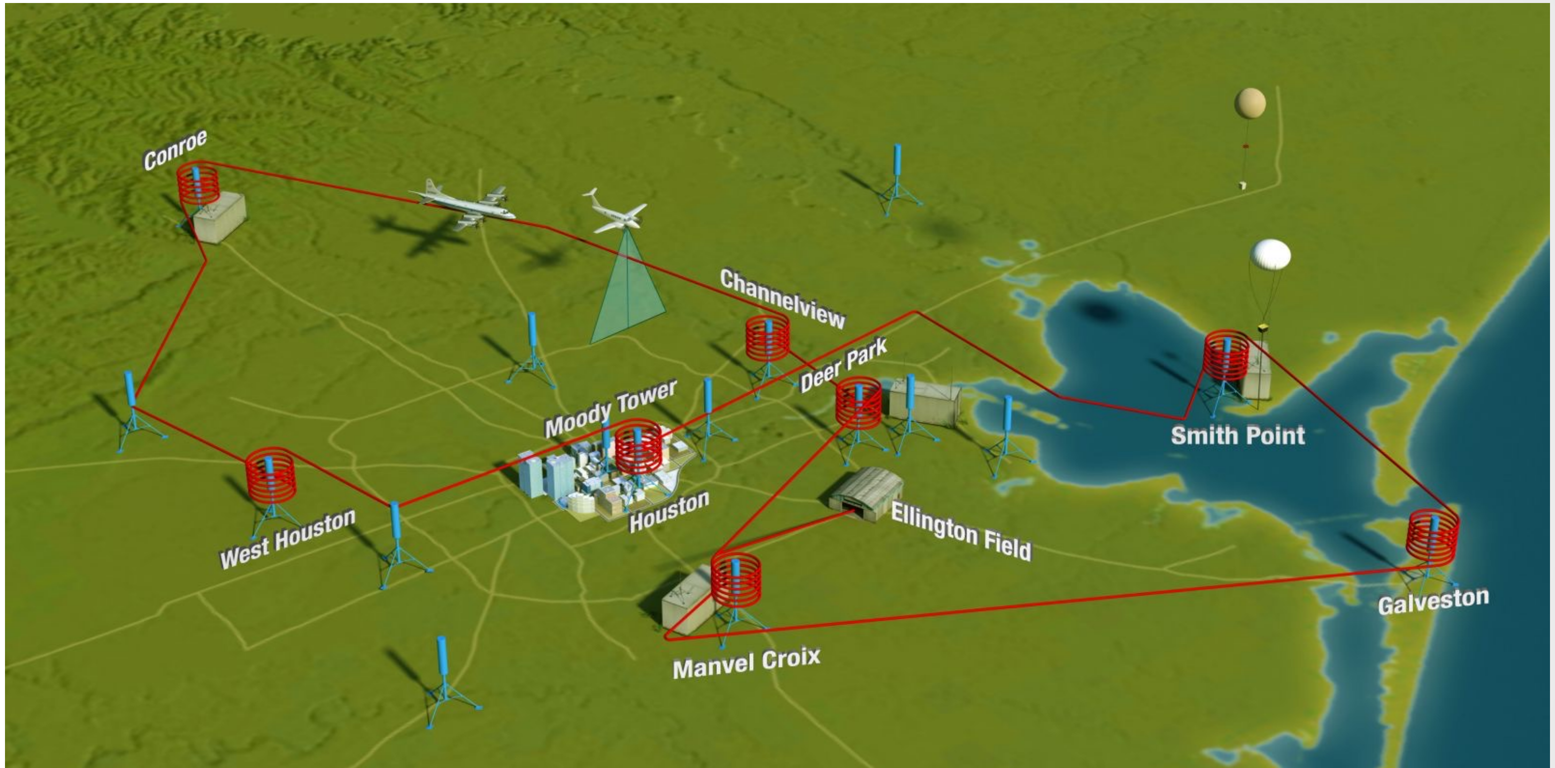
global models
regional models

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

A few airborne air quality campaigns

Typically, these campaigns gather data to improve satellite retrieval algorithms and models.

Instruments aboard NASA DC-8 Aircraft (photo credit: Pedro Campuzano-Jost)



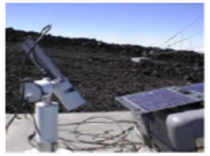
Flight patterns over Houston during DISCOVER-AQ

Source: https://espo.nasa.gov/firex-aq/content/FIREX-AQ

Source: https://www.nasa.gov/mission_pages/discover-aq/index.html

How do we measure and understand air quality?

traditional monitors



low-cost sensors

Ground Measurements

atmospheric observations

Air & Space Observations

satellites

aircraft
balloons

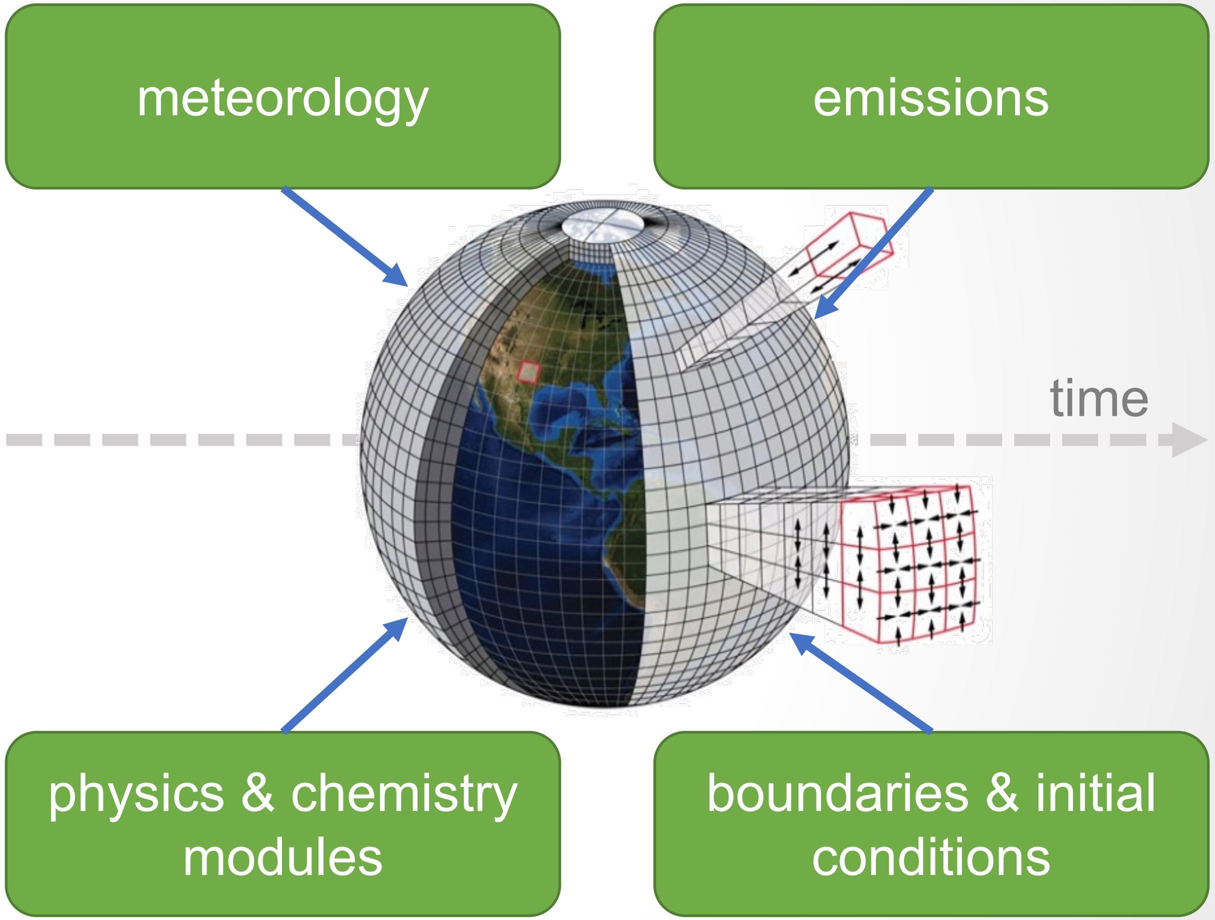
Models

global models
regional models

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

Atmospheric transport & chemistry models

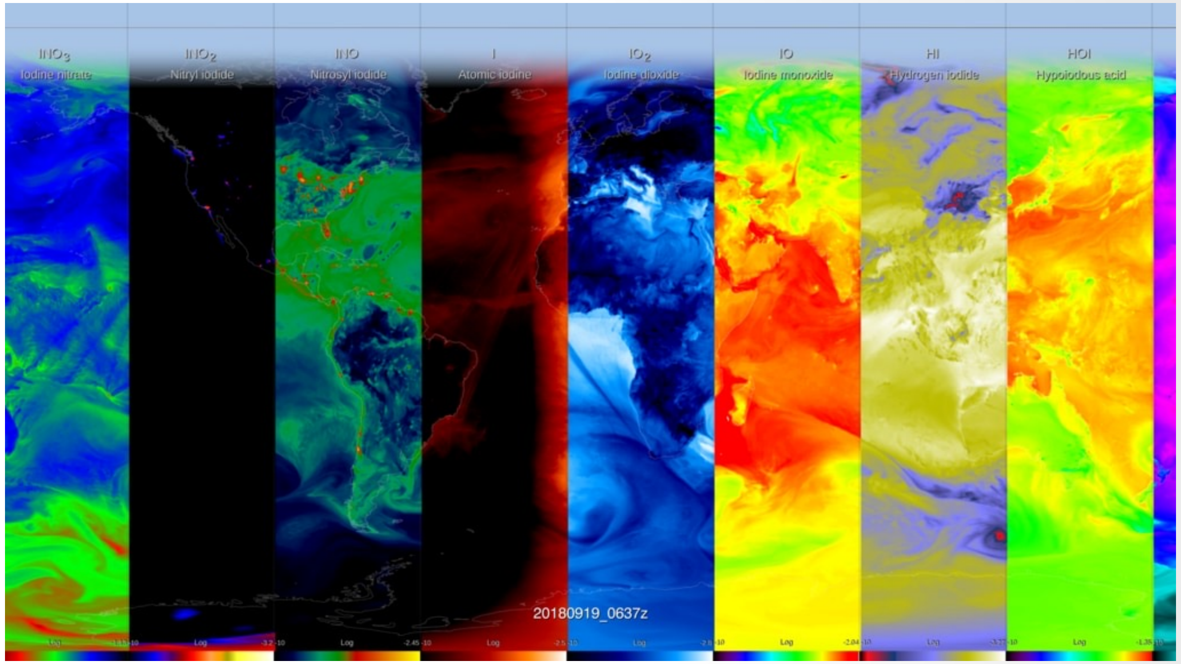
- **Mathematically represent the processes which influence air quality** (emission, transport, diffusion, transformation, removal)
- Operate on a **4D grid** with a specified resolution (horizontal, vertical, temporal)
- Can be part of an **Earth Systems Model** simulating the atmosphere, hydrosphere, geosphere, biosphere, etc.
- Models require decades of research and development; **updates integrate the latest science**, but make it harder to compare between different versions of the model
- **Different models use different approaches**, and so give different results.



Source: Gupta, P.; Follette-Cook, M.; Parrington, M.; Stewart, C. (2021). *Introduction and Access to Global Air Quality Forecasting Data and Tools*. NASA ARSET. https://appliedsciences.nasa.gov/join-mission/training/english/arset-introduction-and-access-global-air-quality-forecasting-data-and

Advantages of air quality models

- Simulate physical and chemical processes which affect air quality; **data are self-consistent**.
- Applicable to **many space & time scales**.
- **No “missing data”**. The model grid is complete.
- Can forecast **future conditions**.
- Can utilize measurements (through data assimilation), but do not require measurements; this makes models **useful in data-sparse regions**.
- **Freely available** global models are run by various groups (including NASA), whose outputs can be used by anyone around the world for free.
- Simulations can **identify sources** of pollution and their relative importance to local air quality.



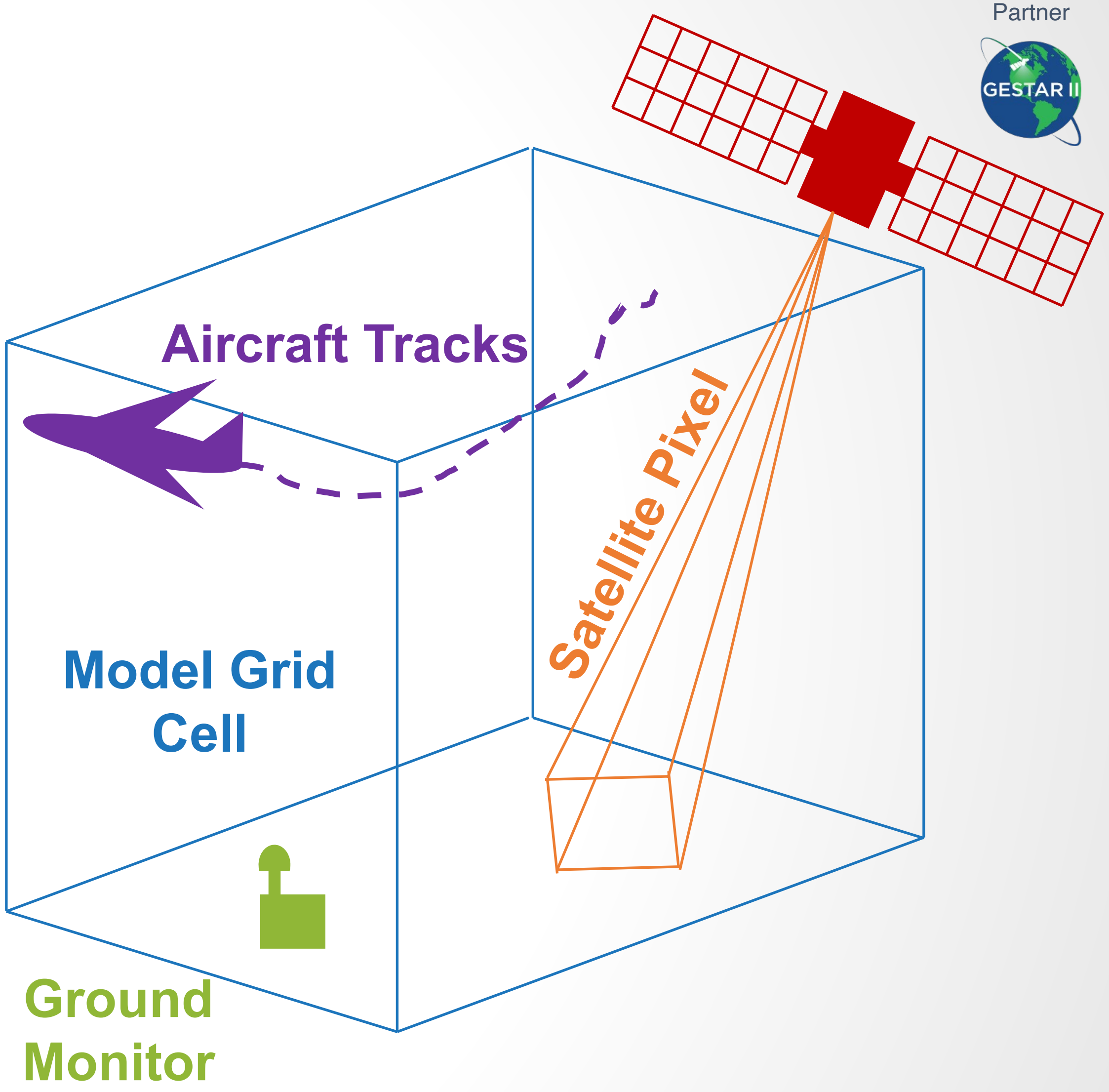
Visualization of several outputs from the NASA GEOS atmospheric chemistry model system.

Source: https://svs.gsfc.nasa.gov/4754

Source: Gupta, P.; Follette-Cook, M.; Parrington, M.; Stewart, C. (2021). *Introduction and Access to Global Air Quality Forecasting Data and Tools*. NASA ARSET. https://appliedsciences.nasa.gov/join-mission/training/english/arset-introduction-and-access-global-air-quality-forecasting-data-and

Weaknesses of air quality models

- **"garbage in, garbage out"**; model outputs are only as good as the emissions data, model assumptions, and initial conditions that are used.
- **Out of date** and/or **coarse resolution emissions inventories** cause uncertainty.
- **Model outputs are not directly comparable to ground or other data sources** due to the scale mis-match; the model estimates average concentrations across its grid, which are not the same as measurements at specific locations.
- Models are difficult and **computationally intensive** to run.
- **Large amount of data** requires expertise & software to interpret and visualize.



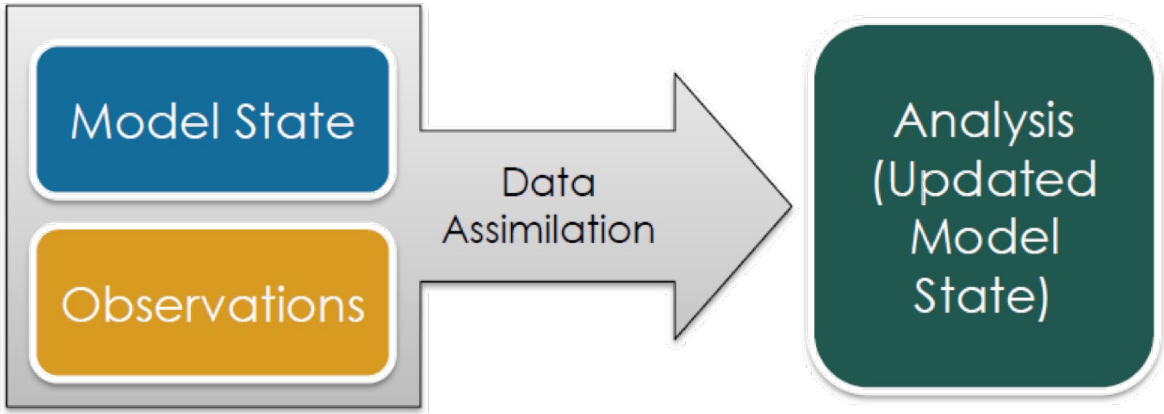
Source: Gupta, P.; Follette-Cook, M.; Parrington, M.; Stewart, C. (2021). *Introduction and Access to Global Air Quality Forecasting Data and Tools*. NASA ARSET. https://appliedsciences.nasa.gov/join-mission/training/english/arset-introduction-and-access-global-air-quality-forecasting-data-and

How do we measure and understand air quality?

Source: Gupta, P.; Follette-Cook, M. (2018). Satellite Remote Sensing of Air Quality. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-remote-sensing-air-quality

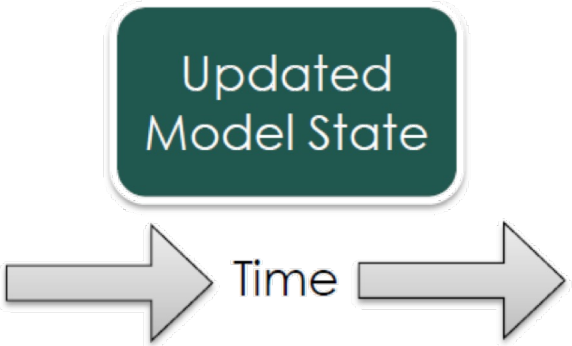
Assimilation, analysis, reanalysis & forecasting

Data assimilation describes the process of assimilating, or incorporating, observations into a model state to produce the best estimate of the atmosphere, land, and ocean conditions.



An **analysis** is the blend of the model and observations.

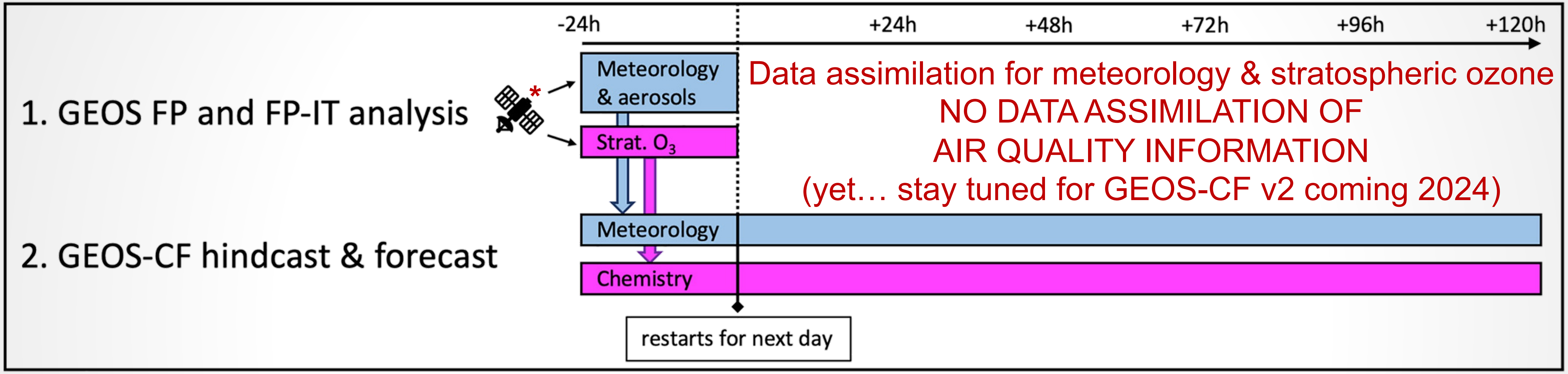
A **reanalysis** blends observations with model simulations of the past using a single model version.

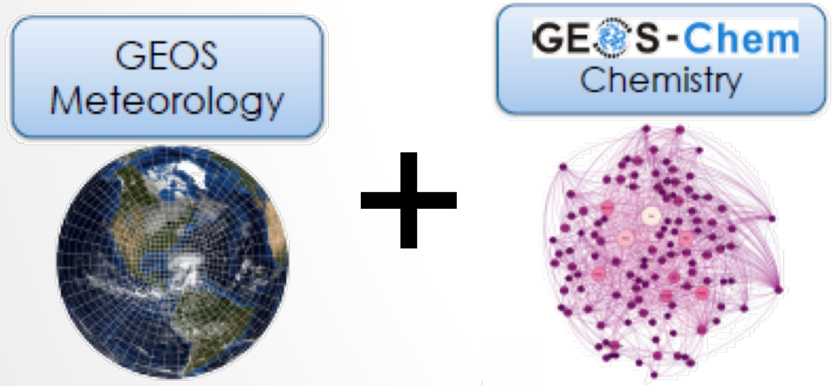


A **forecast** is a model simulation run forward in time to predict a future state.

Source: Gupta, P.; Follette-Cook, M.; Parrington, M.; Stewart, C. (2021). *Introduction and Access to Global Air Quality Forecasting Data and Tools*. NASA ARSET. https://appliedsciences.nasa.gov/join-mission/training/english/arset-introduction-and-access-global-air-quality-forecasting-data-and

GEOS Composition Forecast (GEOS-CF)



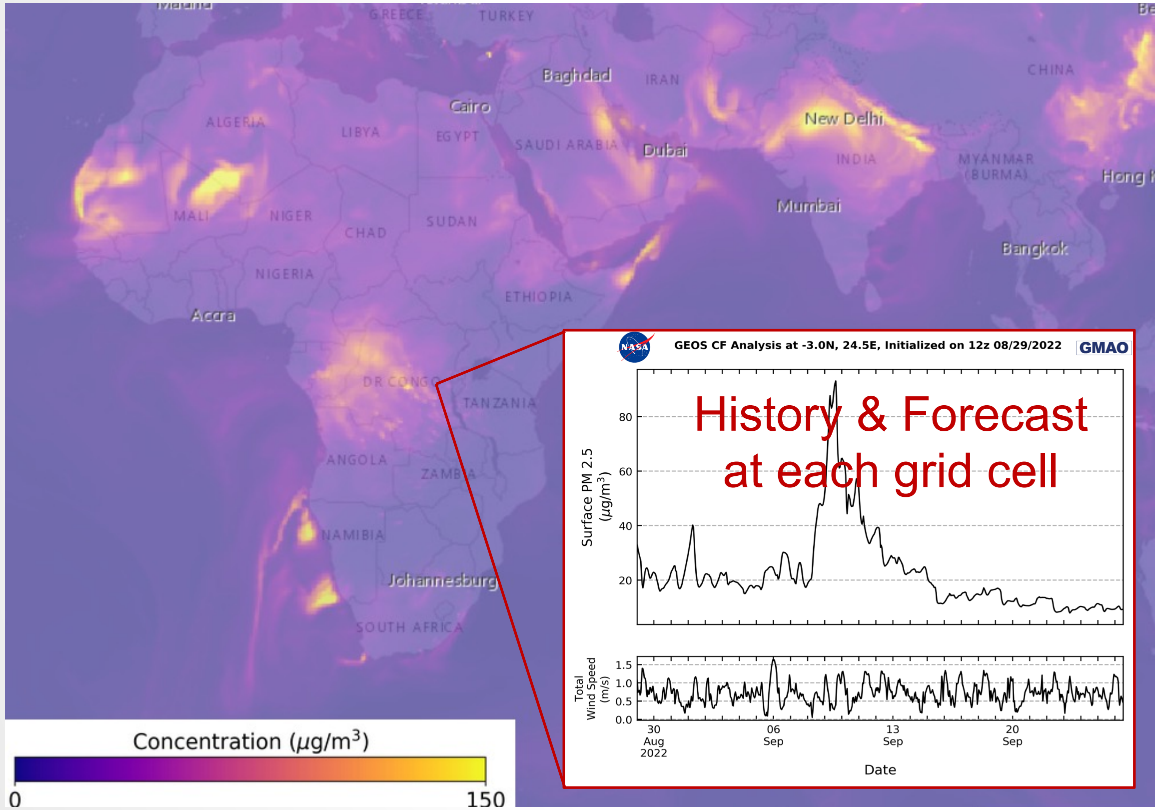


- Combine GEOS meteorology with GEOS-Chem chemistry
- 250 chemical species
- Hourly temporal resolution
- 0.25 degree (25 km) spatial resolution
- Global coverage
- Daily 1-day hindcast and 5-day forecast

Source: Keller, C., et al. (2021) "Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0". *Journal of Advances in Modeling Earth Systems*, 13:4. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020MS002413

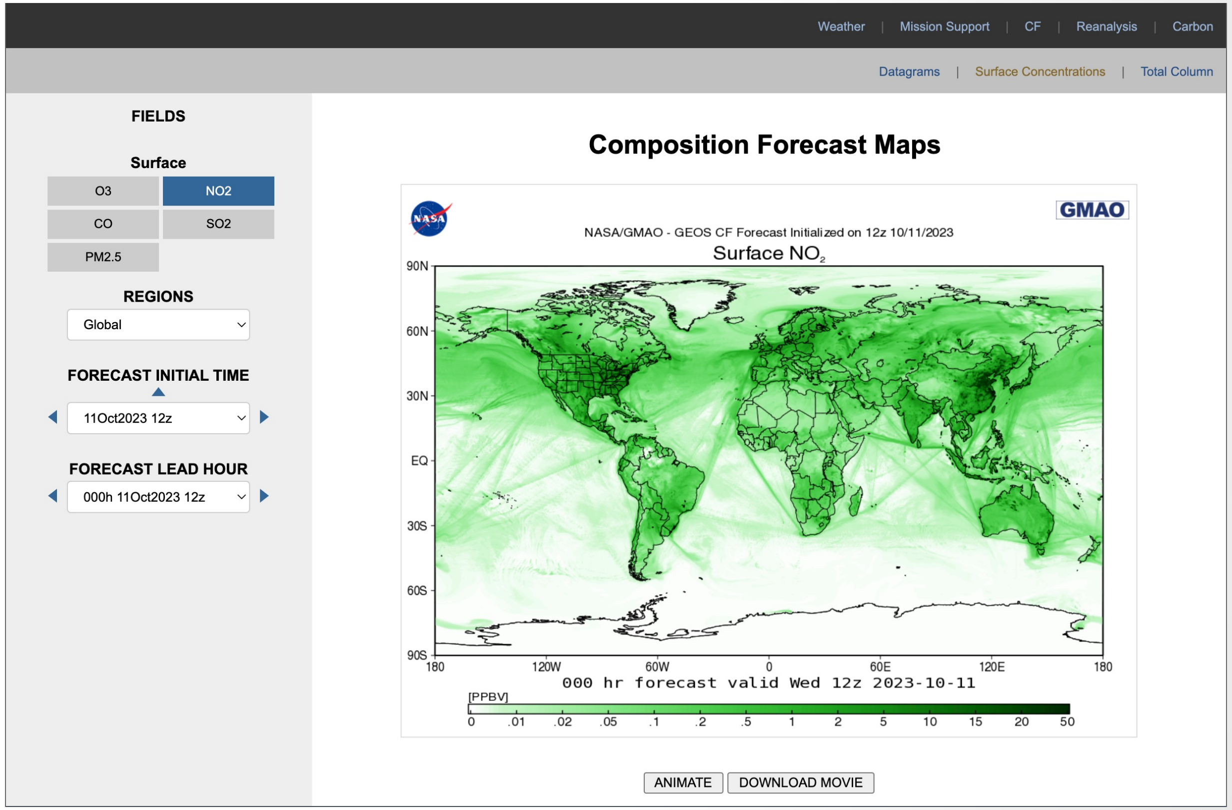
Interacting with NASA GEOS-CF data products

Example: interactive FLUID map for GEOS-CF $PM_{2.5}$



GEOS-CF FLUID interactive map https://fluid.nccs.nasa.gov/cf_map/index

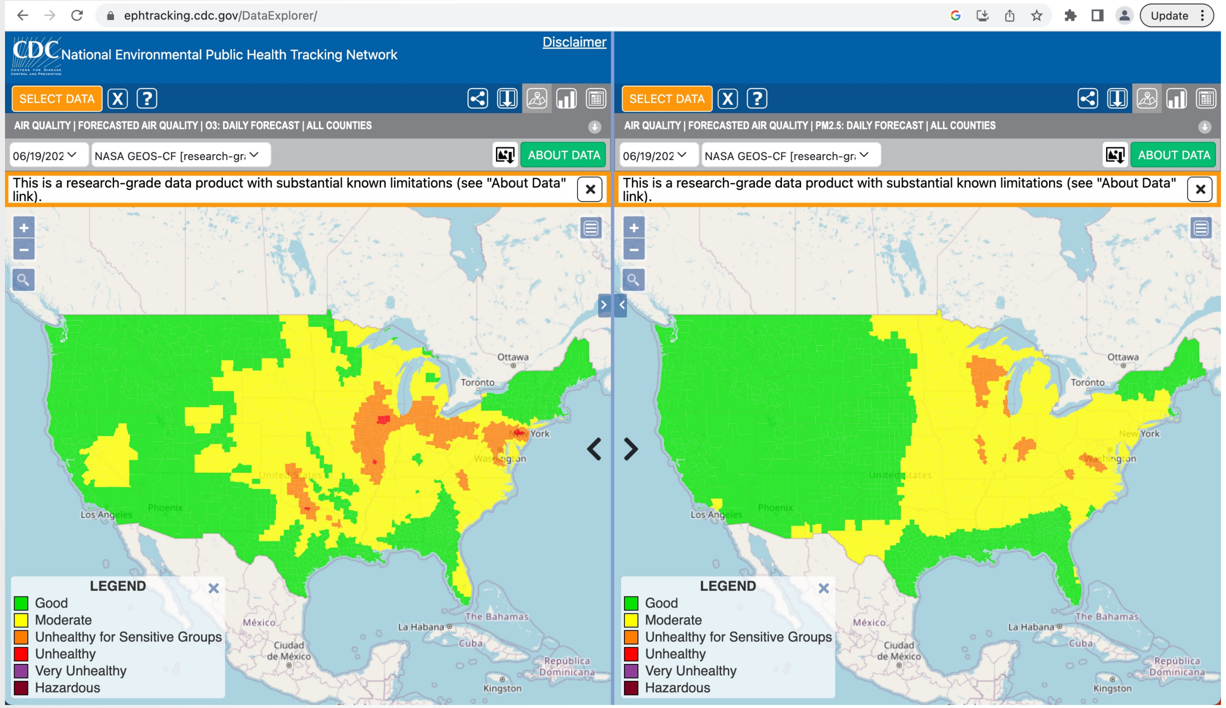
Example: FLUID animated map for GEOS-CF NO_2



GEOS-CF FLUID animated maps https://fluid.nccs.nasa.gov/cf/classic_geos_cf/

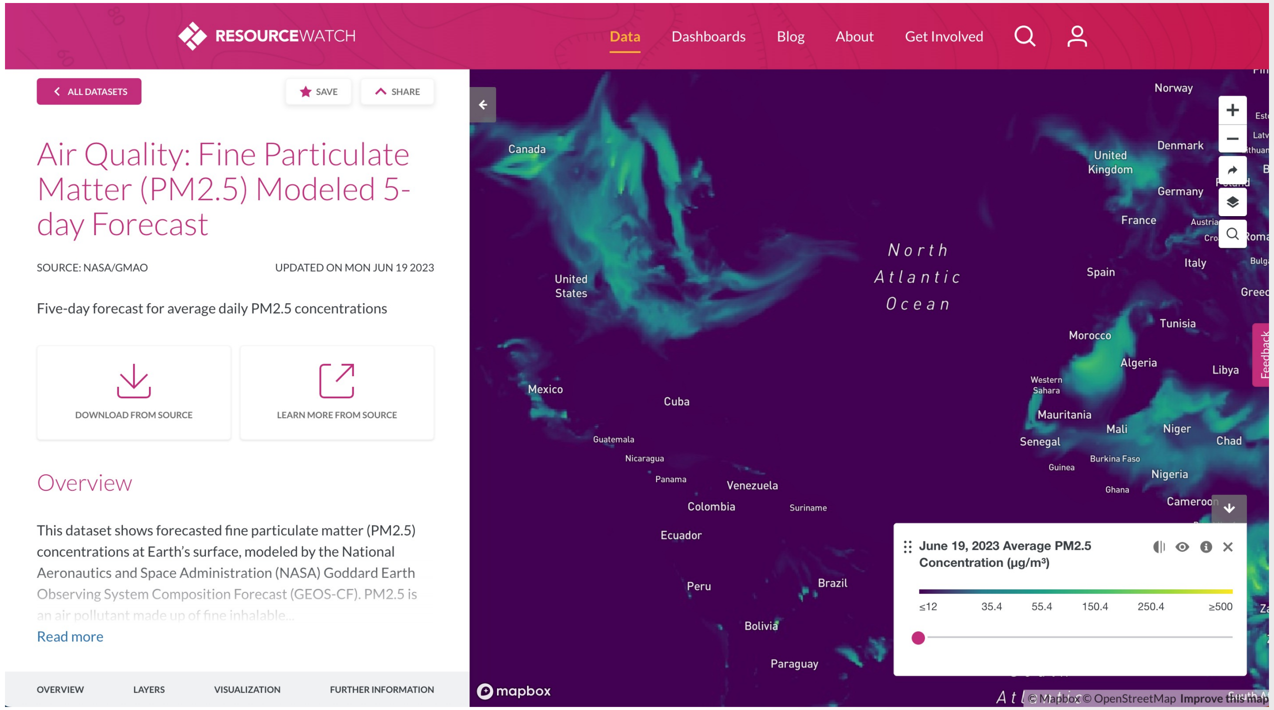
Interacting with NASA GEOS-CF data products

Example: Derived Air Quality Index in CDC Tracker



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

Example: Forecasts in WRI CityAQ ResourceWatch



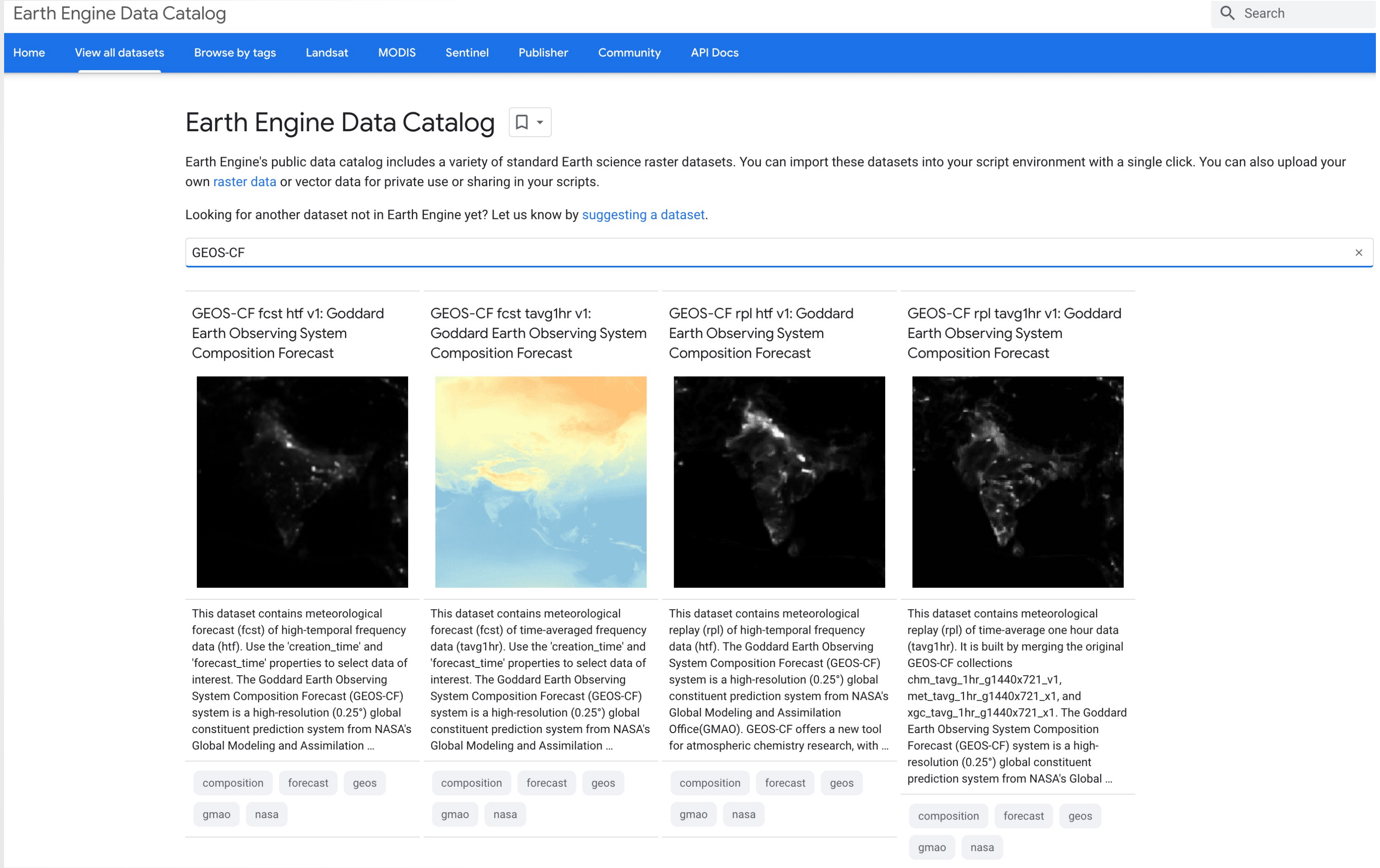
Source: https://www.wri.org/initiatives/cityaq

National Aeronautics and Space Administration

Partner

GESTAR II

NASA GEOS-CF Data in Google Earth Engine



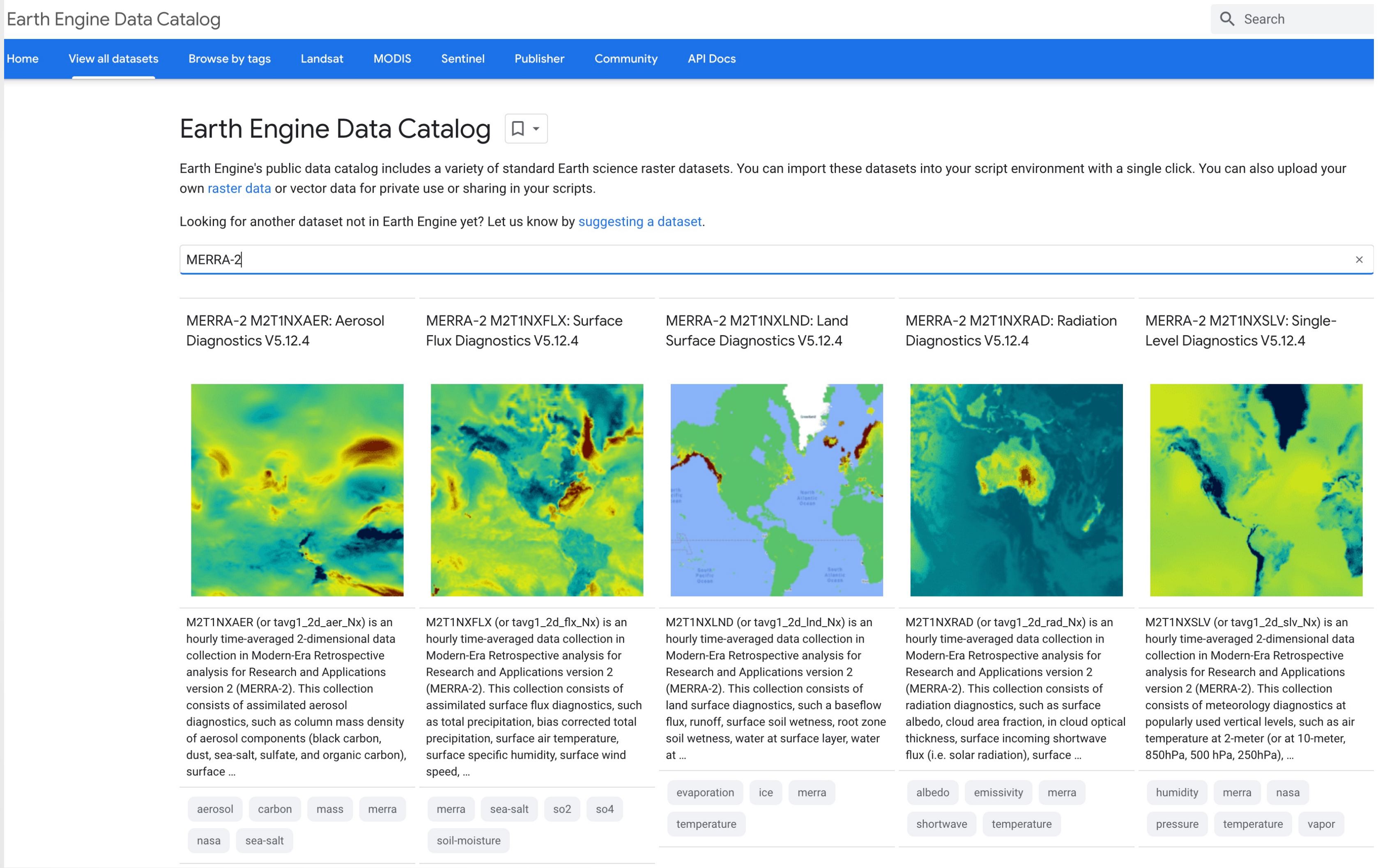
NASA/GEOS-CF/v1/rpl/tavg1hr
Replay (best estimate of past conditions)
1-hour time-averaged
Surface air quality bands:
PM25_RH35_GCC, NO2, CO, SO2, O3

NASA/GEOS-CF/v1/rpl/htf
Replay (best estimate of past conditions)
High time frequency (15 minute instantaneous)
Surface air quality bands:
PM25_RH35_GCC, NO2, CO, SO2, O3

NASA/GEOS-CF/v1/fcst/tavg1hr
Forecast (meteorological forecast with chemisty)
1-hour time-averaged
Surface air quality bands:
PM25_RH35_GCC, NO2, CO, SO2, O3

NASA/GEOS-CF/v1/fcst/htf
Forecast (meteorological forecast with chemisty)
High time frequency (15 minute instantaneous)
Surface air quality bands:
PM25_RH35_GCC, NO2, CO, SO2, O3

NASA MERRA-2 Data in Google Earth Engine



NASA/GSFC/MERRA/aer/2
Reanalysis (assimilation of satellite data)
1-hour time-averaged
Surface aerosol bands:
BCSMASS, DMSMASS, DUSMASS25, OCSMASS, SSSMASS25, SO4SMASS

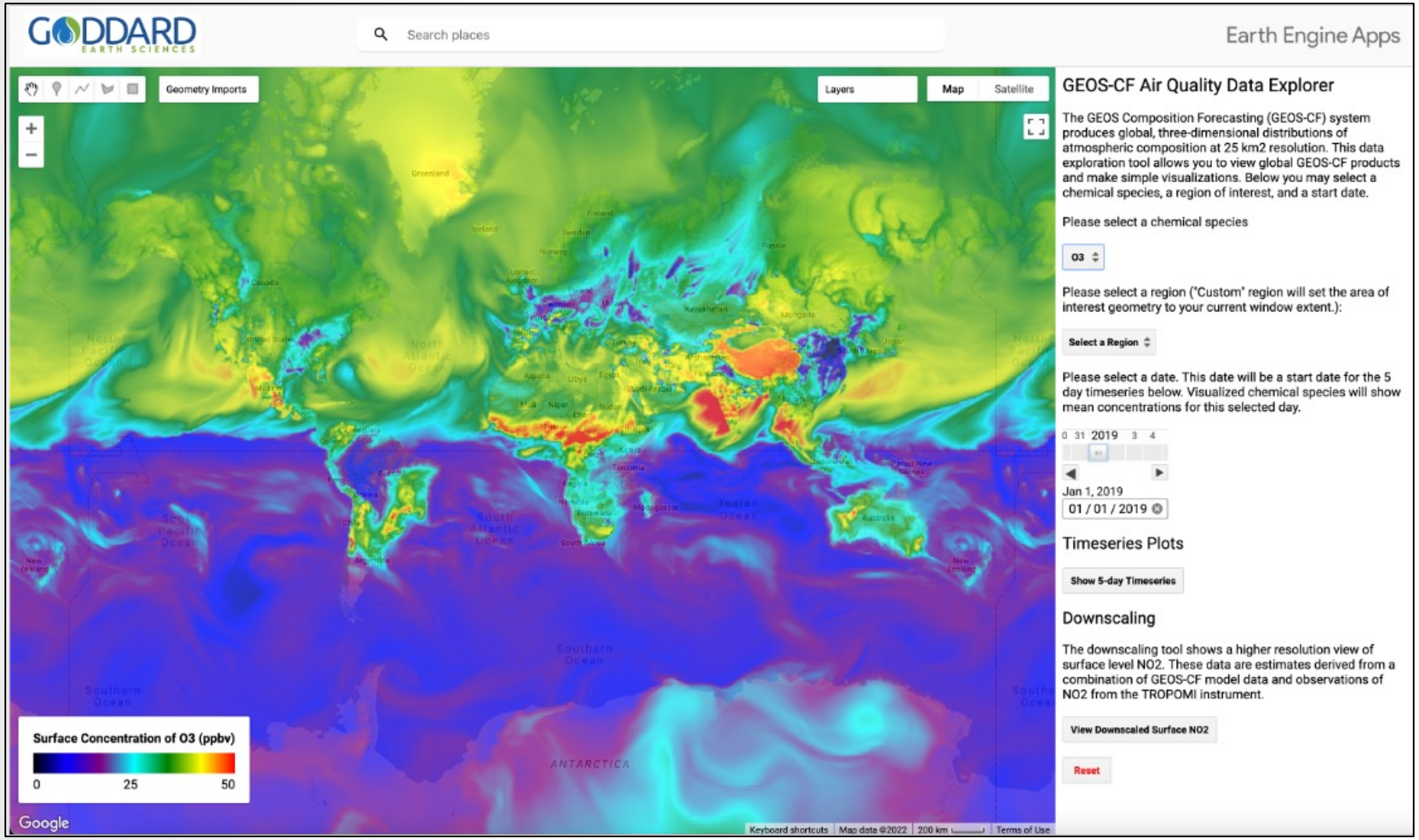
Aerosol Optical Depth Bands:
BCEXTTAU, BCSCATAU, DUEXTTAU, DUSCATAU, OCEXTTAU, OCSCATAU, SSEXTTAU, SSSCATAU, SUEXTTAU, SUSCATAU

Surface $PM_{2.5}$ can be calculated from aerosol constituents:

$PM_{2.5}$ = DUSMASS25 + OCSMASS+ BCSMASS + SSSMASS25 + SO4SMASS $\times$ (132.14/96.06)

Source:
gmao.gsfc.nasa.gov/reanalysis/MERRA-2/FAQ/

NASA Applications in Google Earth Engine



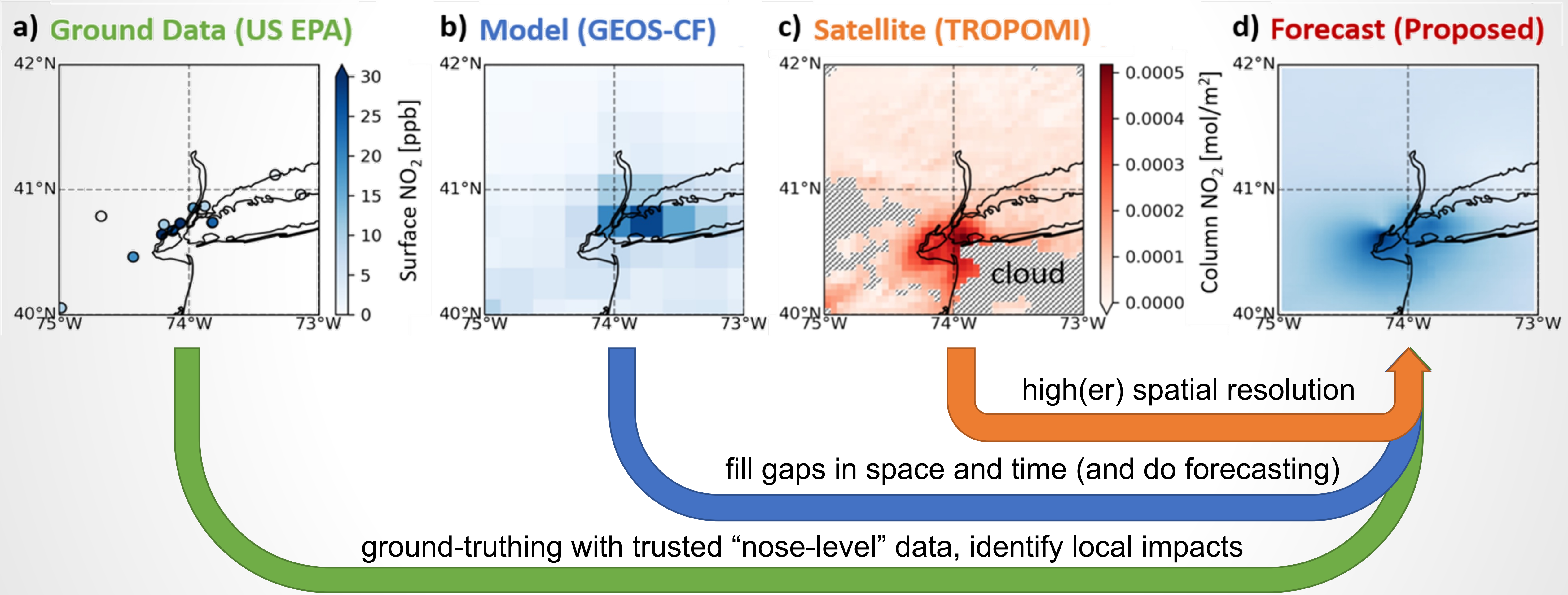
Source: https://callumwayman.users.earthengine.app/view/geoscfexplorer

Through a NASA-Google Partnership Agreement, Callum Wayman (NASA GMAO & SSAI) has developed applications for viewing and manipulating NASA GEOS-CF and MERRA-2 data in Google Earth Engine, alongside other NASA & ESA data sources (e.g., TROPOMI).

Examples:

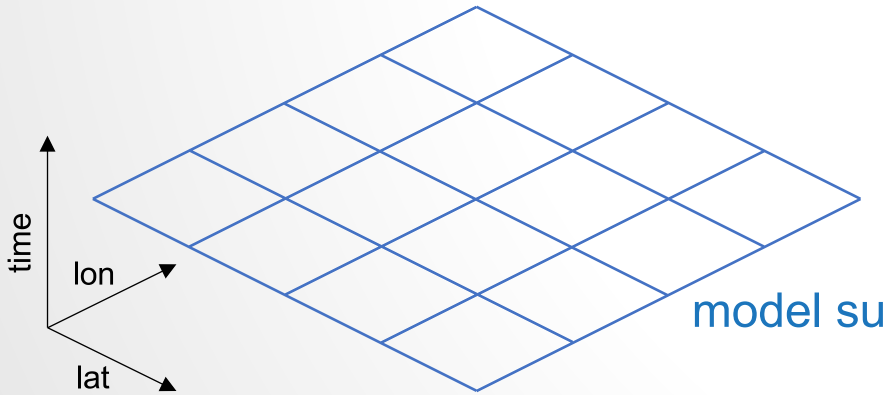
- Air quality data explorer
- GEOS-CF machine learning downscaling for surface NO_2
- Exploring Google Street View air quality data alongside GEOS-CF model outputs

Data Fusion



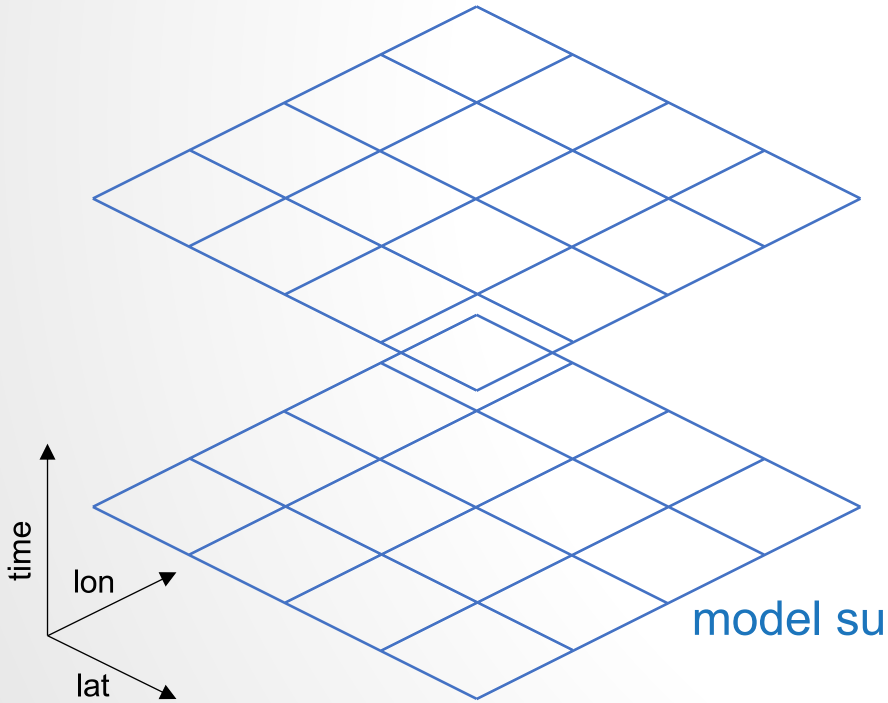
Source: Malings et al. (2021), "Sub-City Scale Hourly Air Quality Forecasting by Combining Models, Satellite Observations, and Ground Measurements" *Earth & Space Science.* DOI: 10.1029/2021EA001743

Phase 1: Model Only



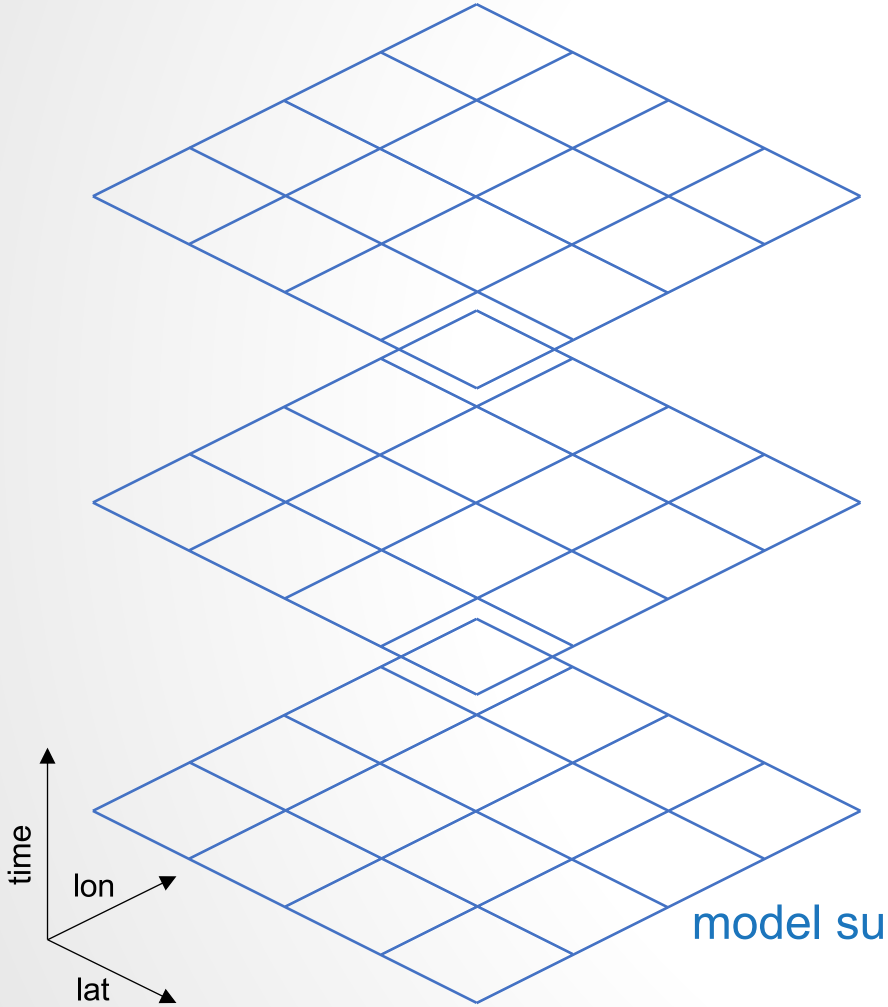
model surface concentration $M(x,t)$

Phase 1: Model Only



model surface concentration $M(x,t)$

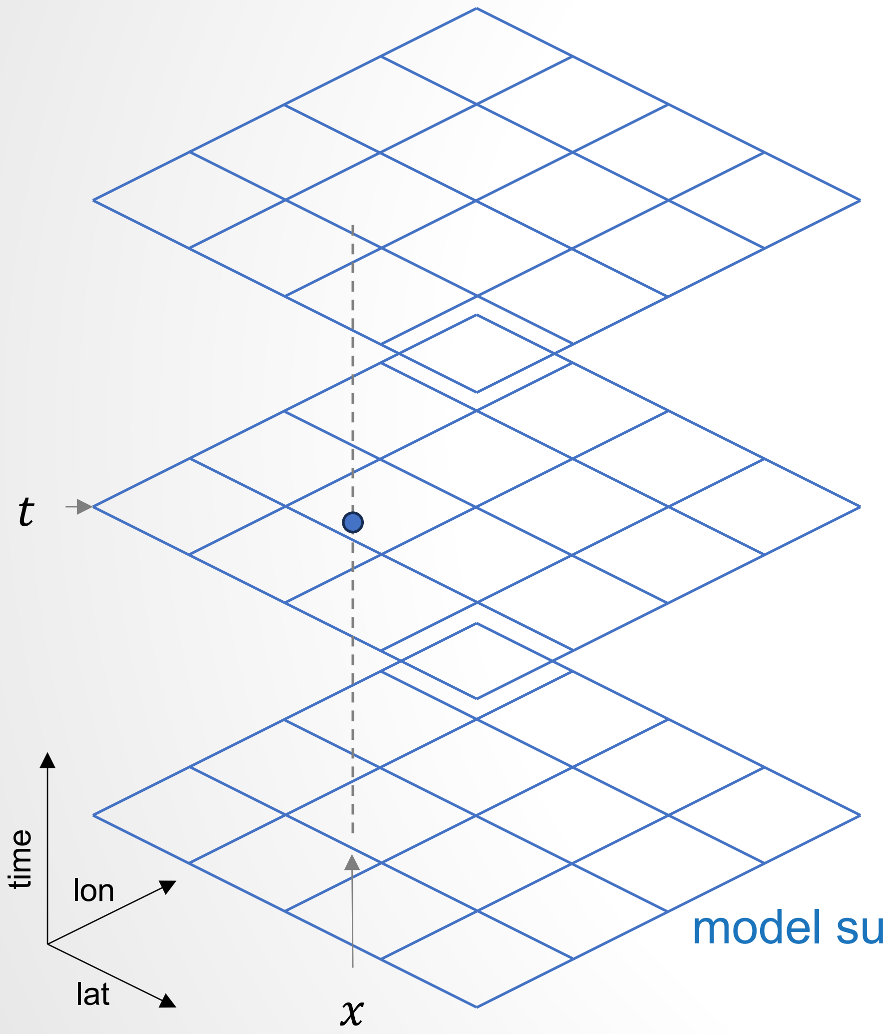
Phase 1: Model Only



model surface concentration $M(x,t)$

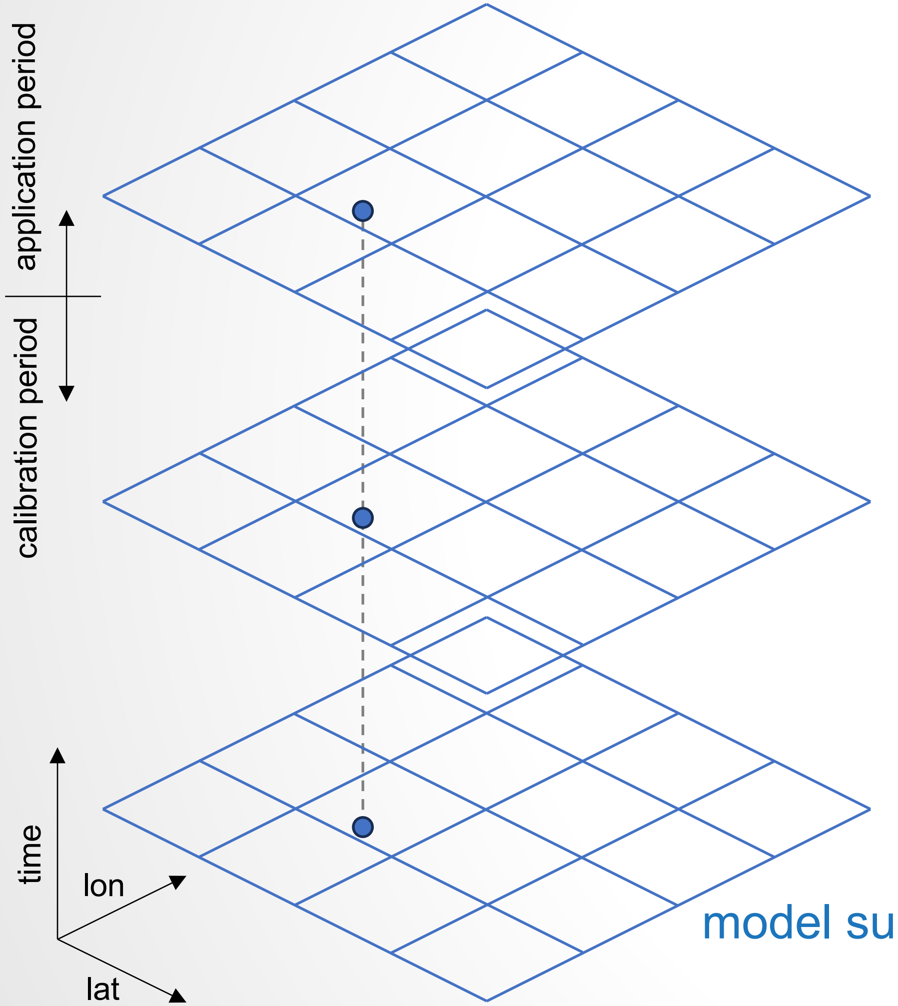
Phase 1: Model Only

$$F_1(x,t) = M(x,t)$$



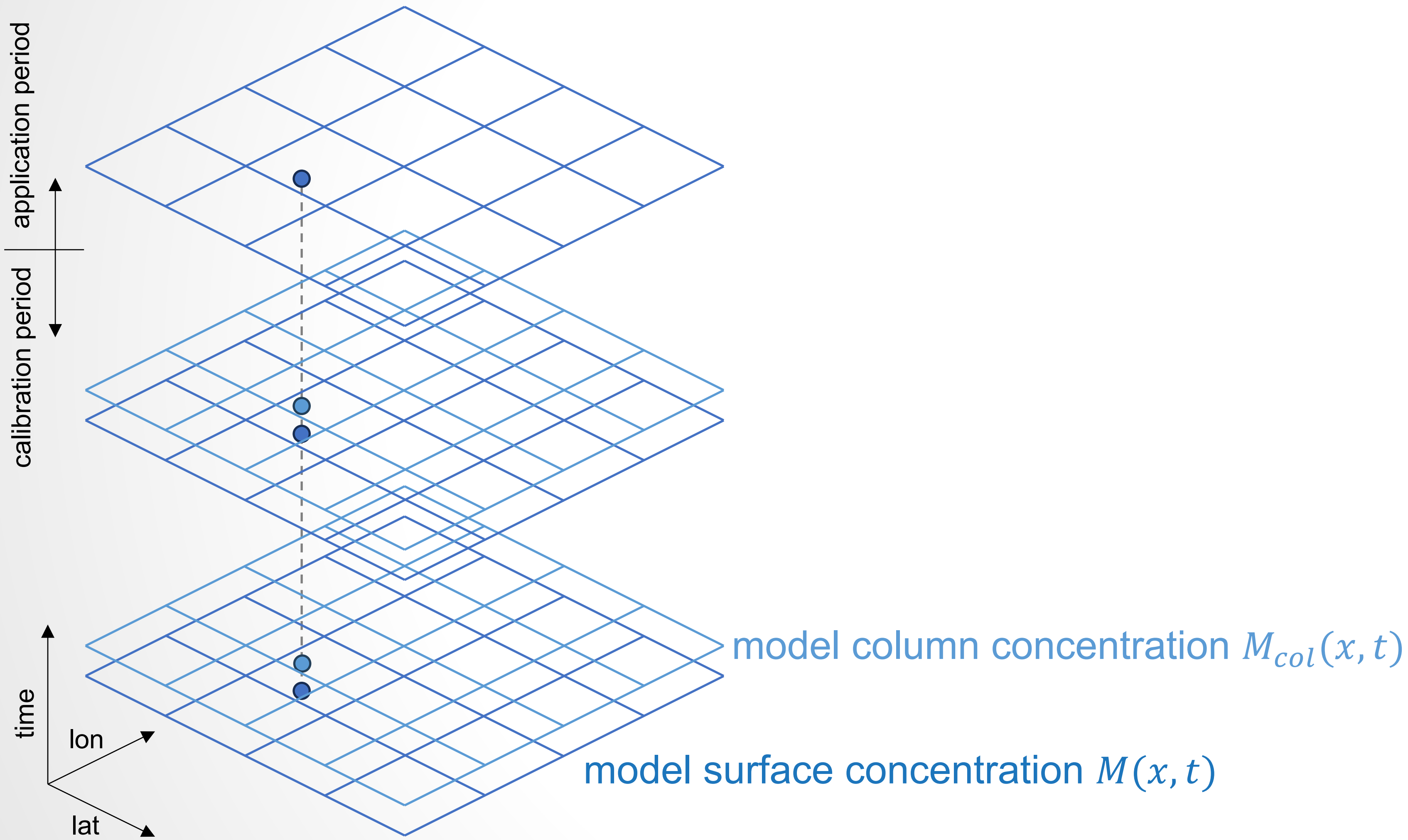
model surface concentration $M(x,t)$

Phase 2: Model & Satellite

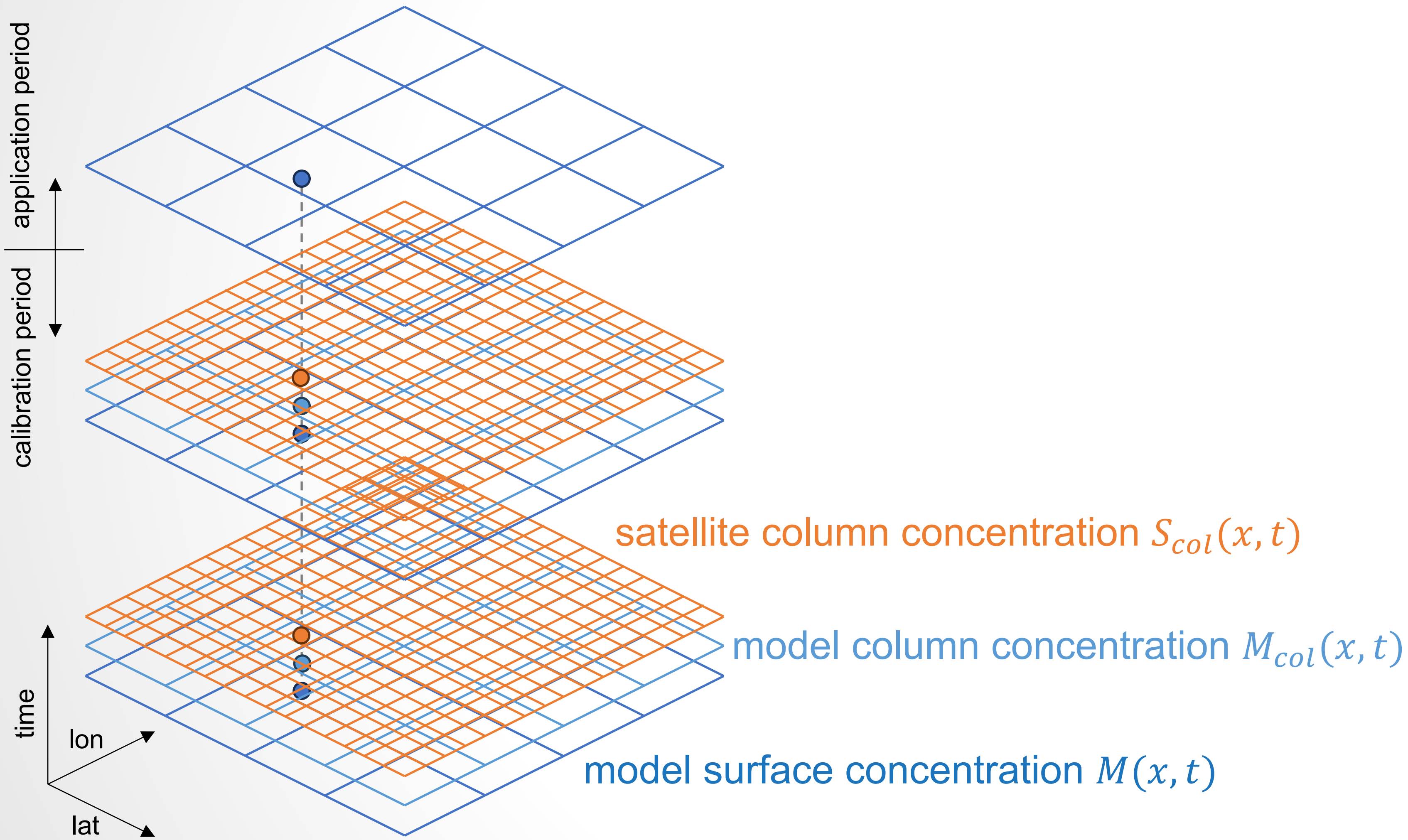


model surface concentration $M(x,t)$

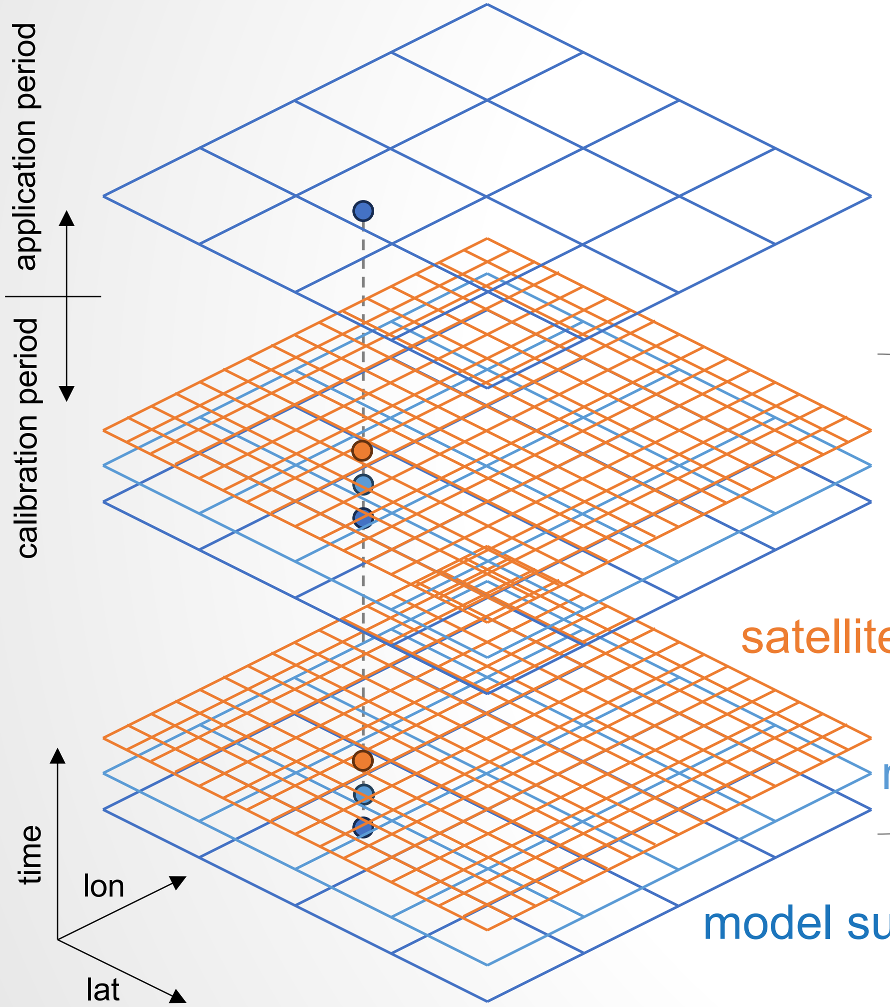
Phase 2: Model & Satellite



Phase 2: Model & Satellite



Phase 2: Model & Satellite



$$F_2(x,t) = F_1(x,t) + D(x,t)$$

$$D(x,t) = \mathbb{E}_{t' \in T_{c,overpass}(t)}\left[\left(S_{col}(x,t') - M_{col}(x,t')\right)\phi(x,t')\,\psi(x,t,t')\right]$$

$$\phi(x,t') \approx \frac{M(x,t')}{M_{col}(x,t')}$$

surface-to-column relationship

$$\psi(x,t,t') \approx \frac{M(x,t)}{M(x,t')}$$

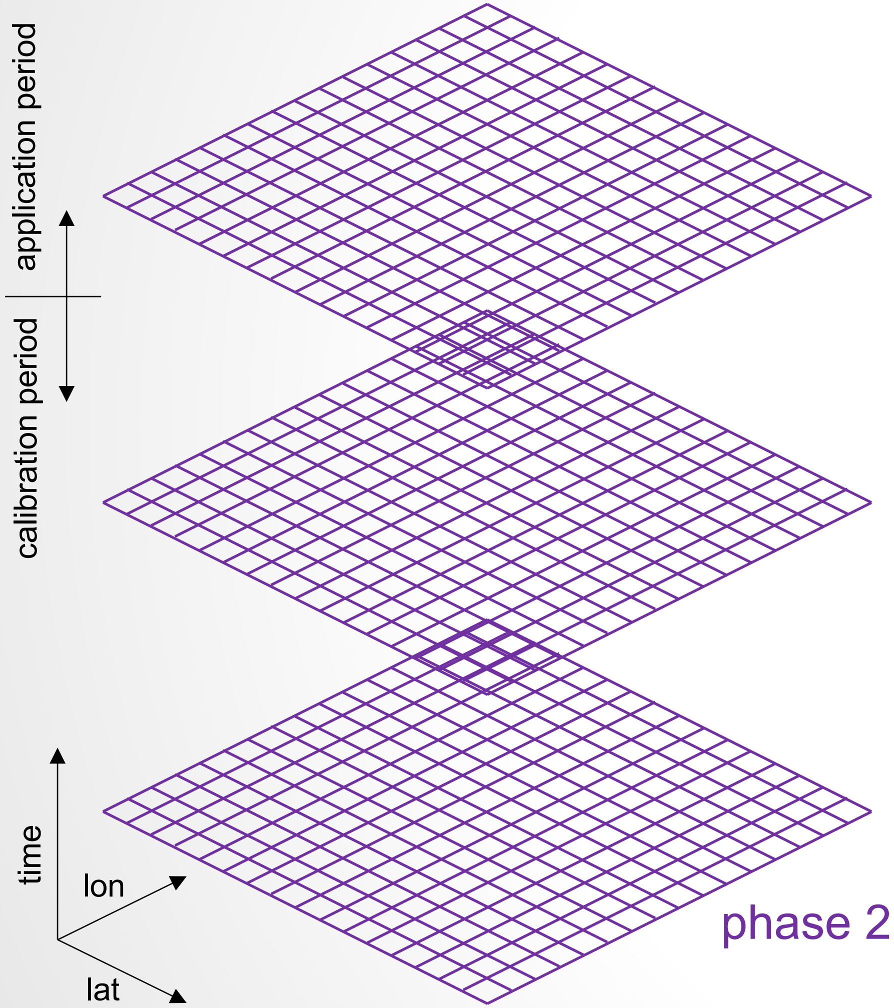
target-time-to-overpass-time relationship

satellite column concentration $S_{col}(x,t)$

model column concentration $M_{col}(x,t)$

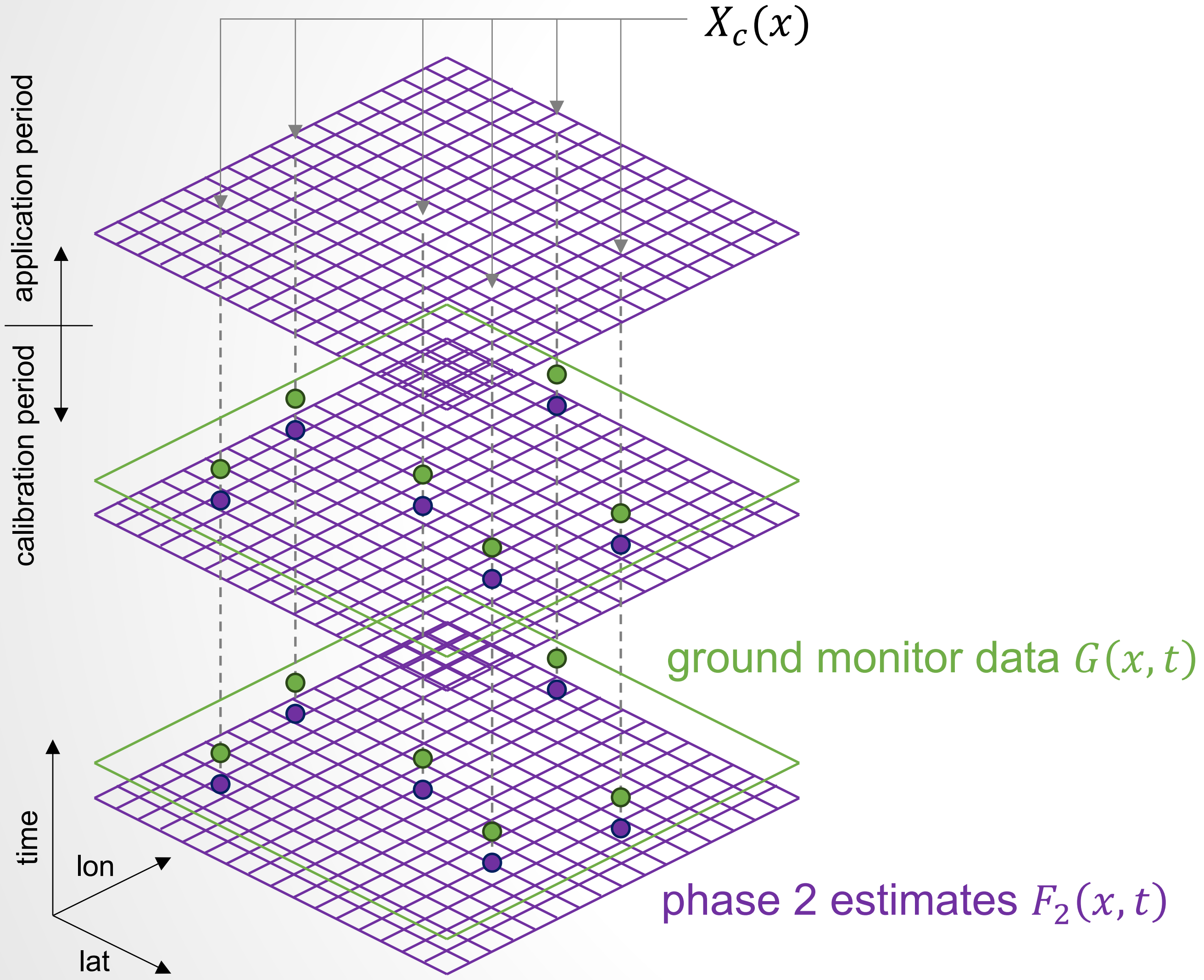
model surface concentration $M(x,t)$

Phase 3: Model & Satellite & Ground



phase 2 estimates $F_2(x, t)$

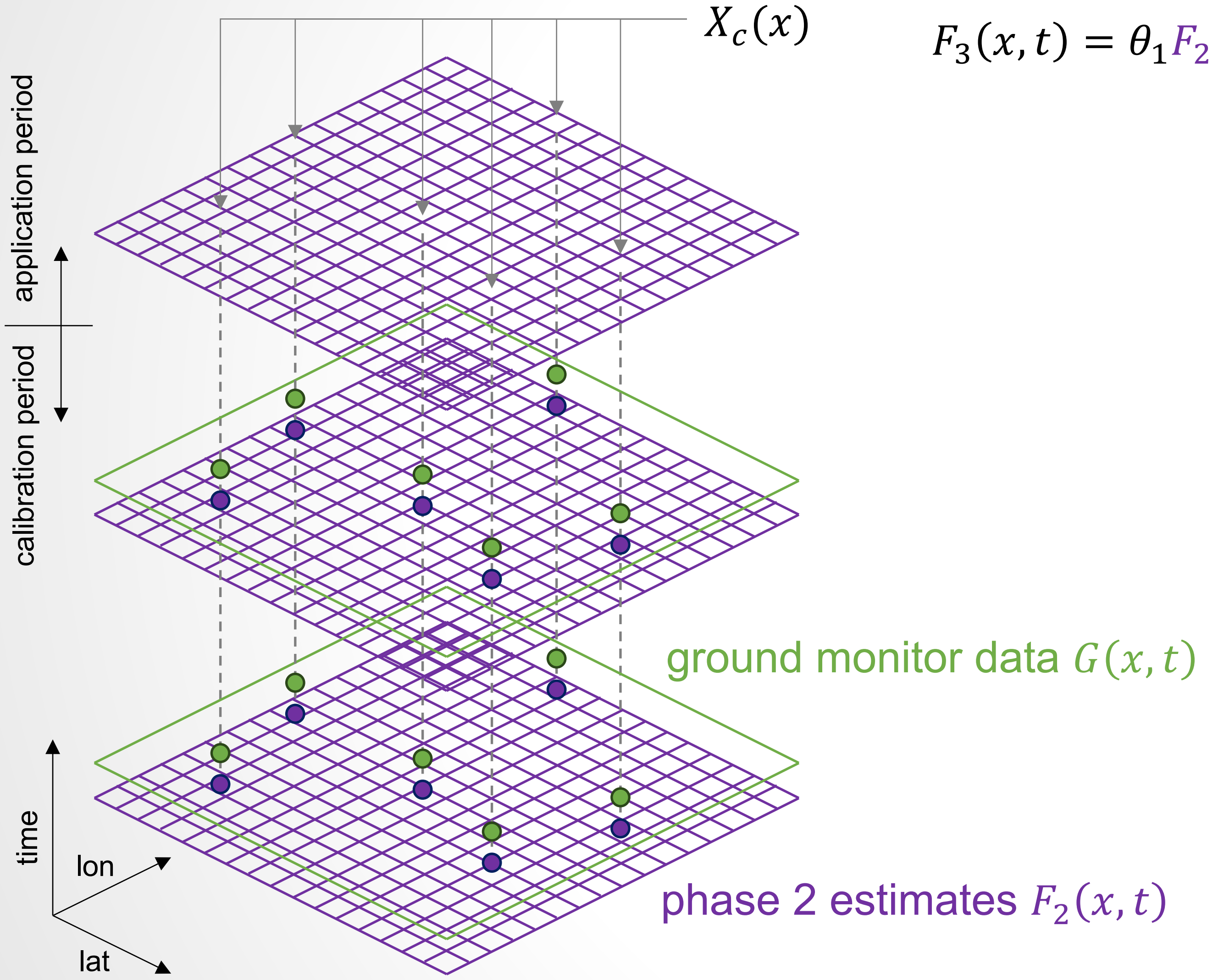
Phase 3: Model & Satellite & Ground

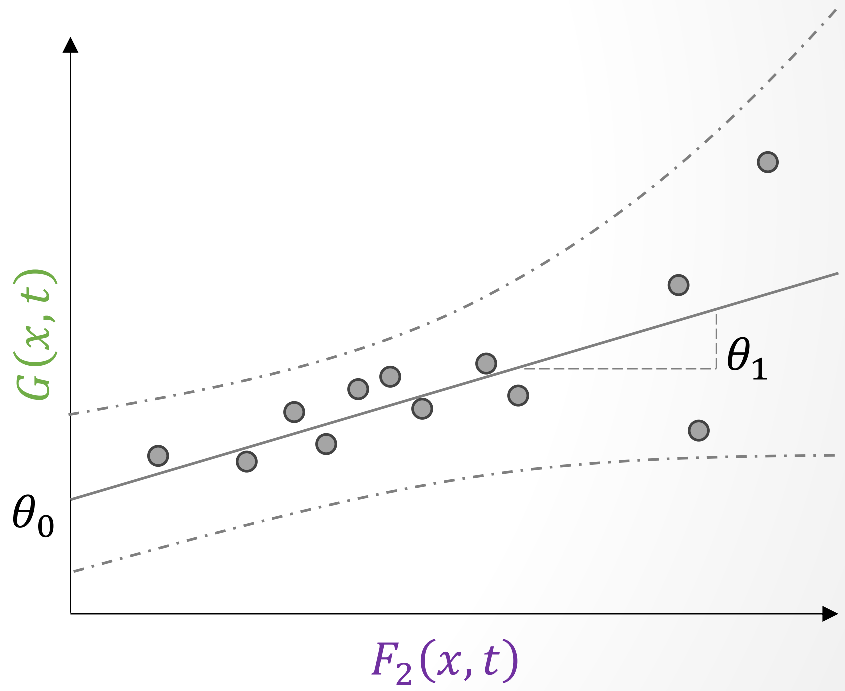


Phase 3: Model & Satellite & Ground

$$F_3(x,t) = \theta_1 F_2(x,t) + \theta_0$$

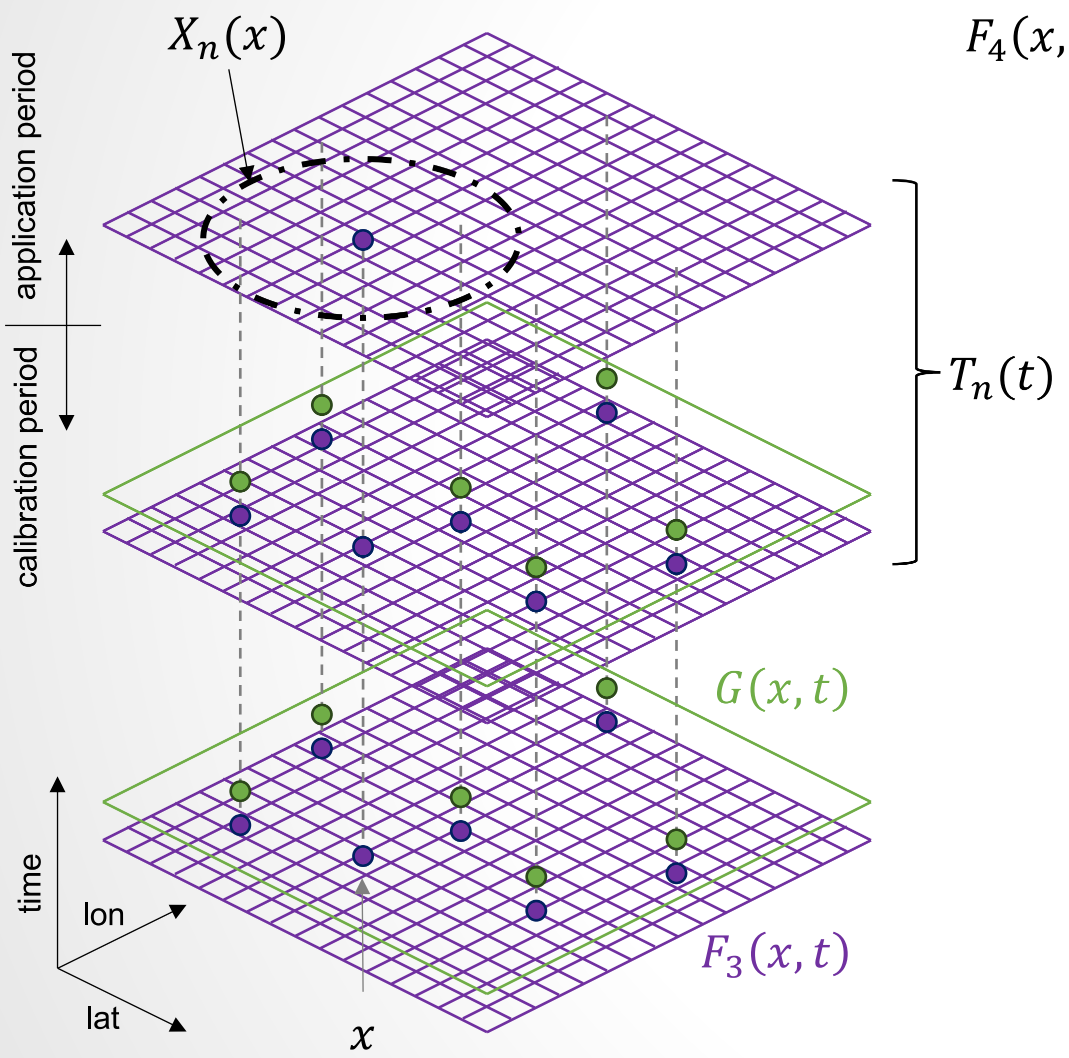
$$\theta_0, \theta_1 = \mathbb{LR}_{t' \in T_c(t), x' \in X_c(x)}[G(x',t') \sim F_2(x',t')]$$



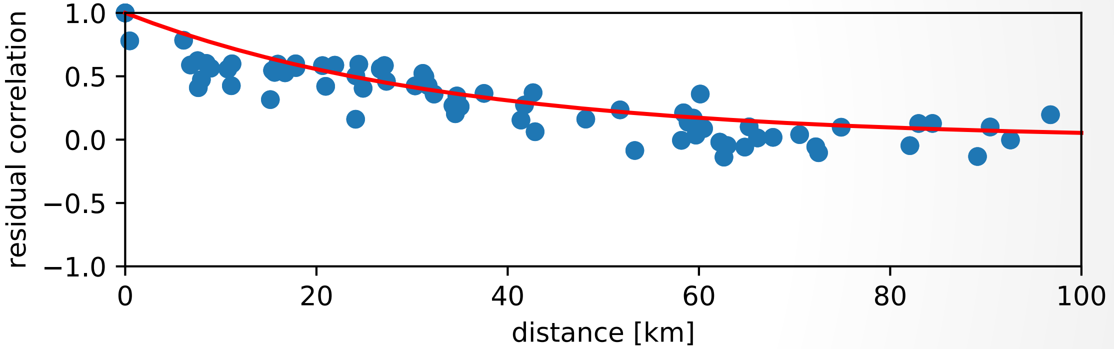


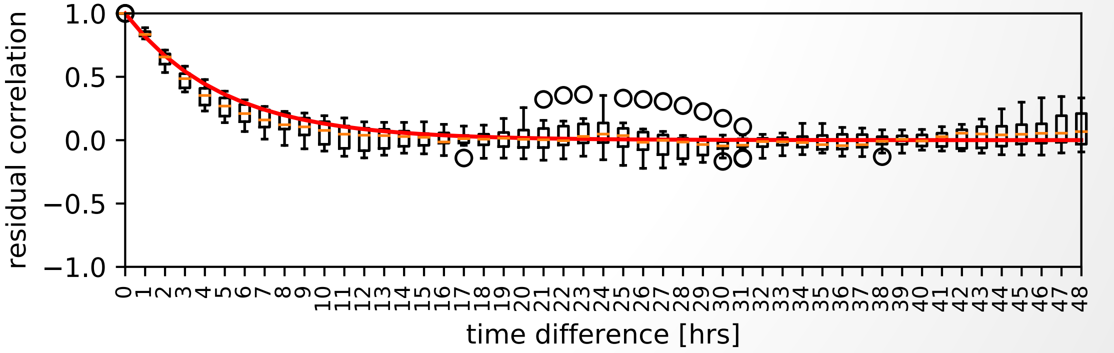
Phase 4: Residual Kriging

$$F_4(x,t) = F_3(x,t) + \sum_{x' \in X_n(x),\, t' \in T_n(t)} K(x,x',t,t')\,[G(x',t') - F_3(x',t')]$$

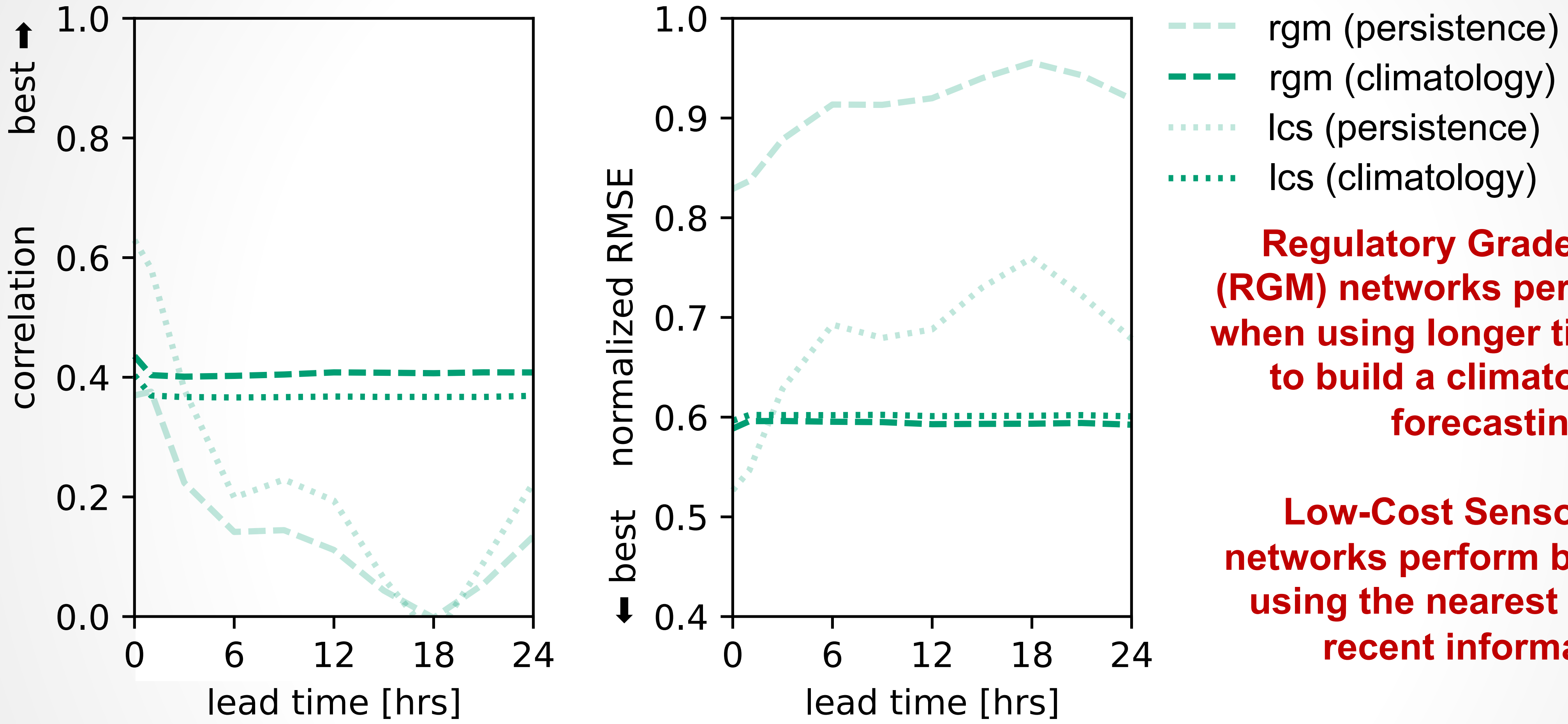


Kriging update term (function of assumed spatio-temporal correlation structure, variance & measurement uncertainty)





Case Study in London: Using local data only performs okay

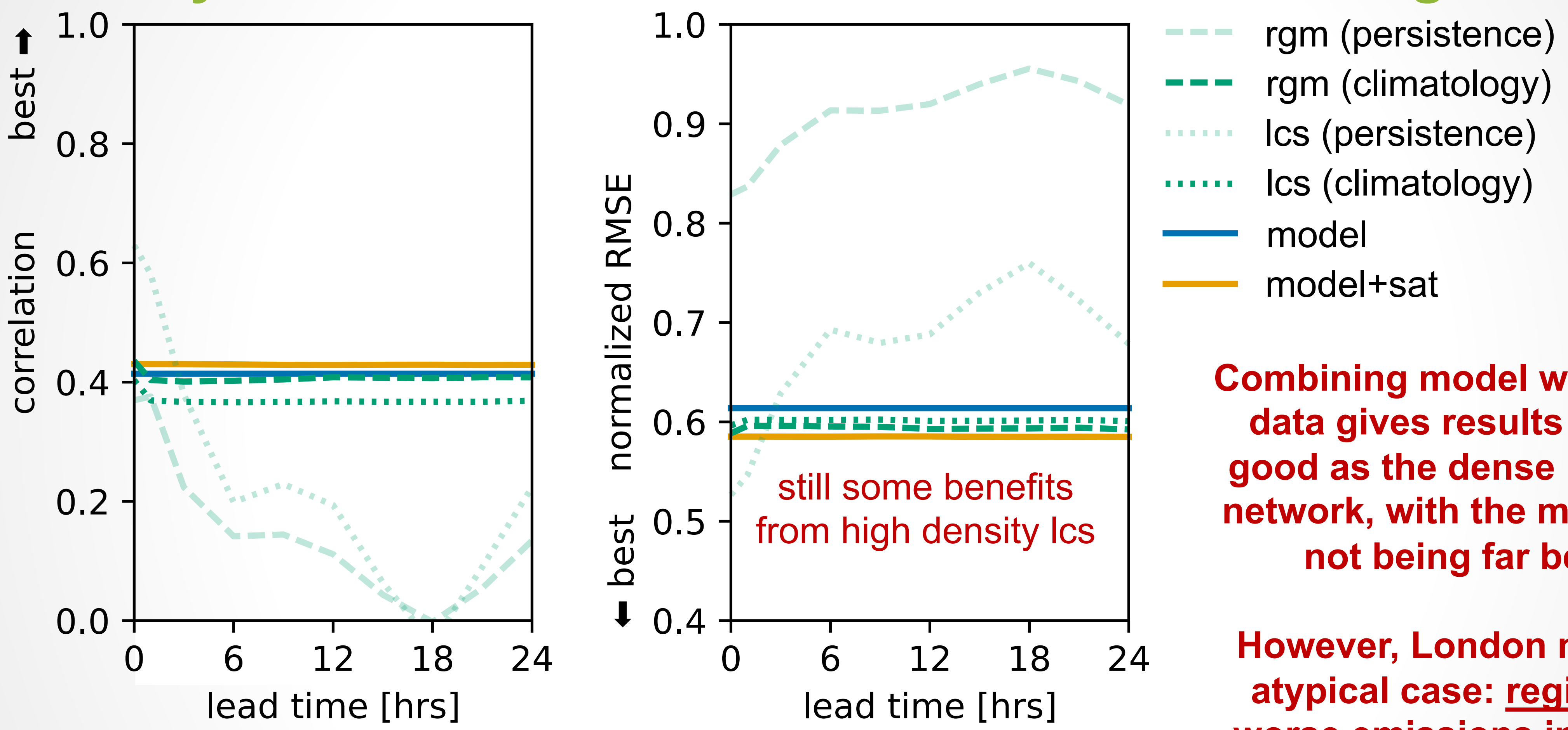


Regulatory Grade Monitor (RGM) networks perform better when using longer time periods to build a climatology for forecasting

Low-Cost Sensor (LCS) networks perform better when using the nearest and most recent information.

forecasting results for **NO_2** in ***London, October & November 2019***
cross-validation: leave-one-site-out, considering only regulatory sites
plotted results represent ***average metrics*** across validation sites

Case Study in London: Model and satellite is about as good

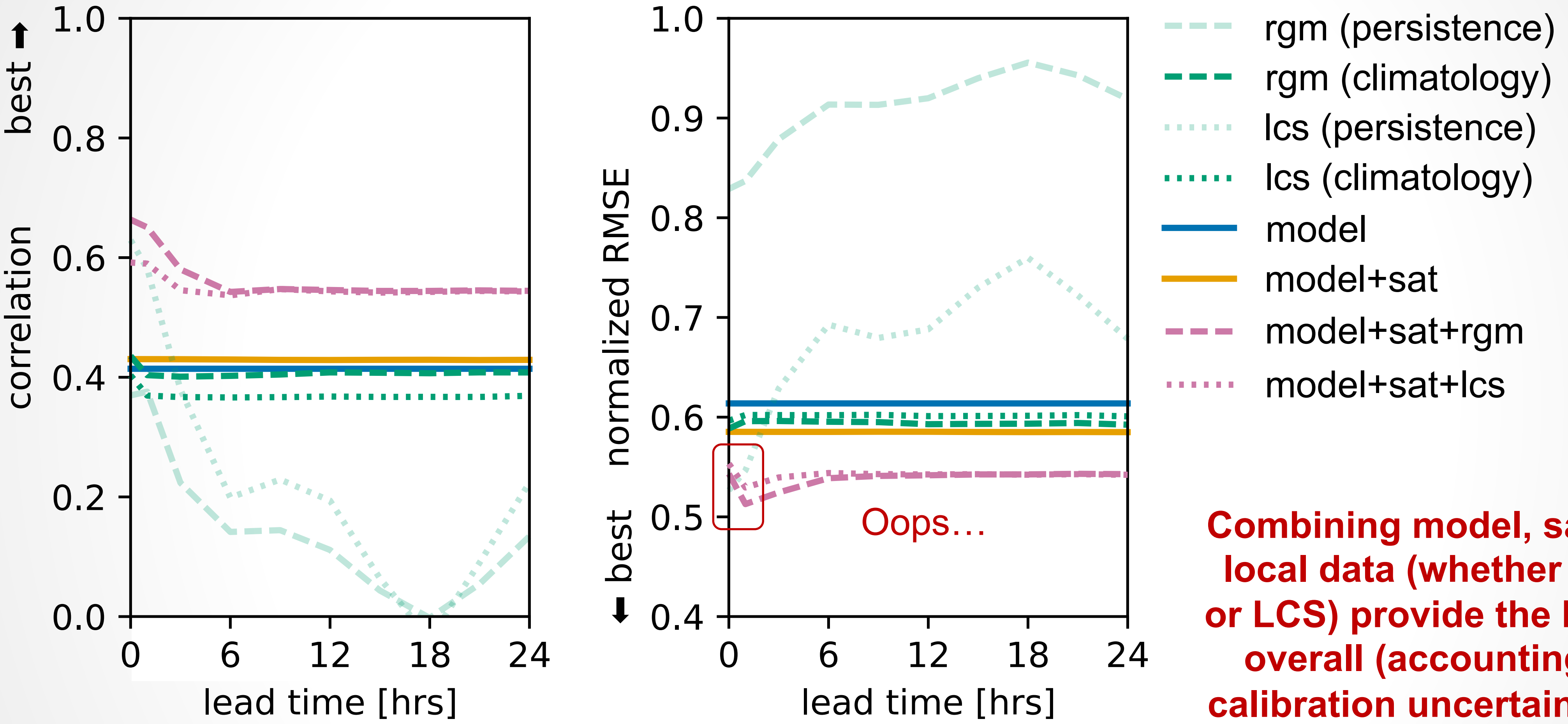


forecasting results for **NO_2** in ***London, October & November 2019***
cross-validation: leave-one-site-out, considering only regulatory sites
plotted results represent ***average metrics*** across validation sites

Combining model with satellite data gives results about as good as the dense local RGM network, with the model alone not being far behind.

However, London may be an atypical case: <u>regions with worse emissions inventories may be more poorly represented by the model!</u>

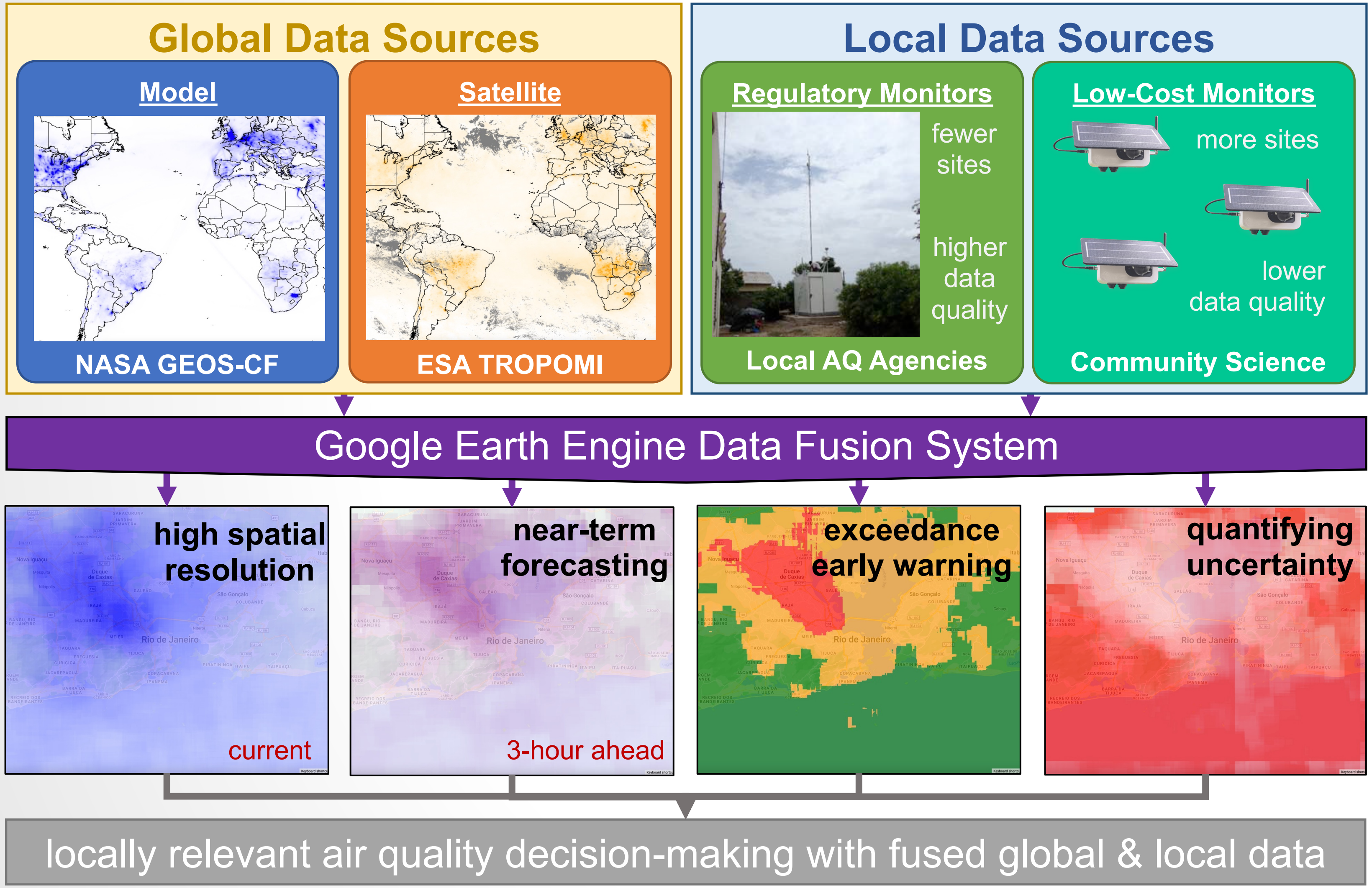
Case Study in London: Model, satellite, and local data are best



forecasting results for **NO_2** in ***London, October & November 2019***
cross-validation: leave-one-site-out, considering only regulatory sites
plotted results represent ***average metrics*** across validation sites

Combining model, satellite, and local data (whether from RGM or LCS) provide the best results overall (accounting for LCS calibration uncertainty properly is still work-in-progress).

Our ongoing NASA-funded project's objective is to…



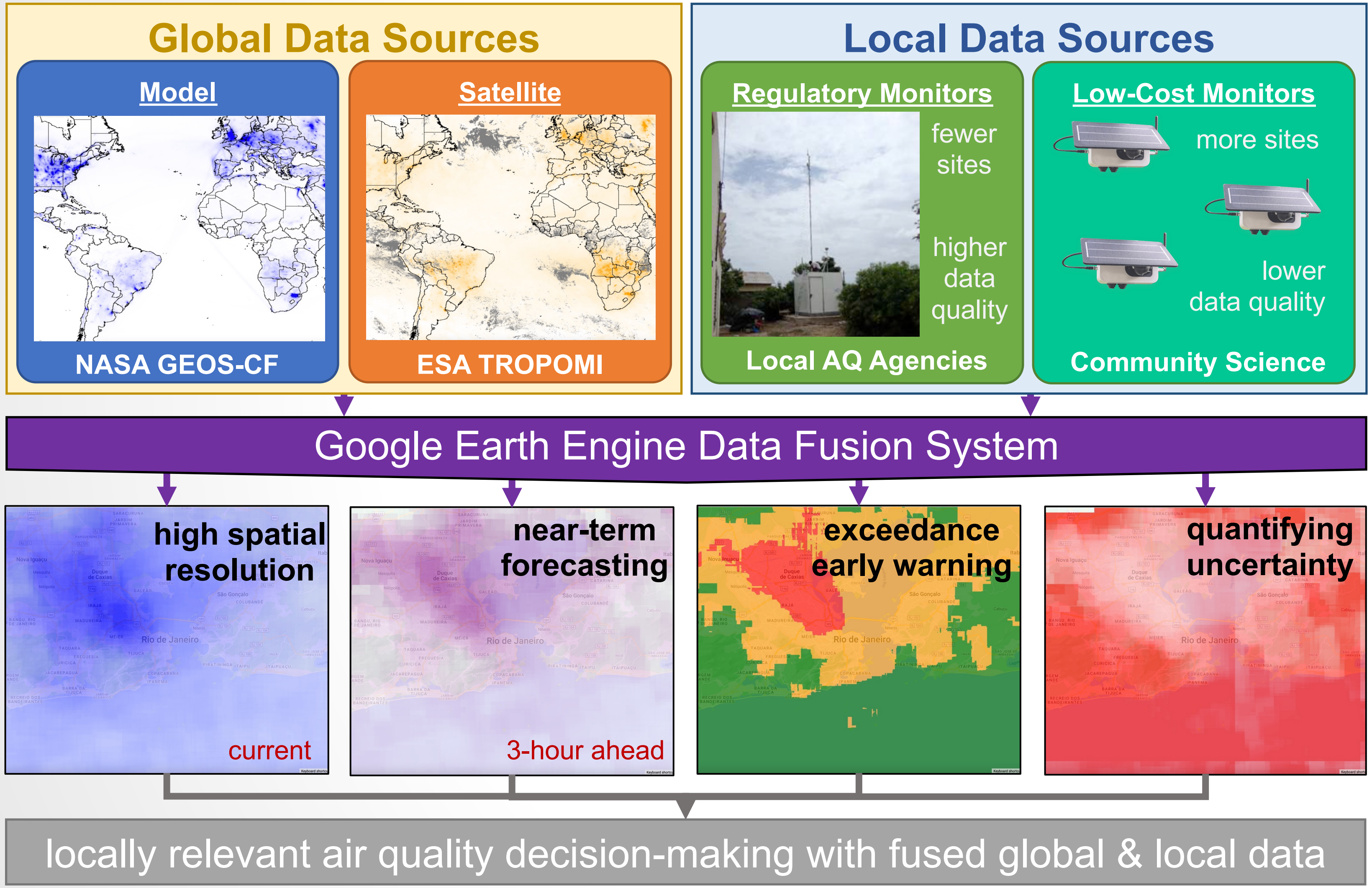
…integrate diverse **global** and **local** air quality data sources…

…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**.

Our ongoing NASA-funded project's objective is to…



NASA GMAO: basic algorithm development & refinement

Clarity: low-cost sensor integration

Sonoma Technologies: data fusion system implementation & user interface

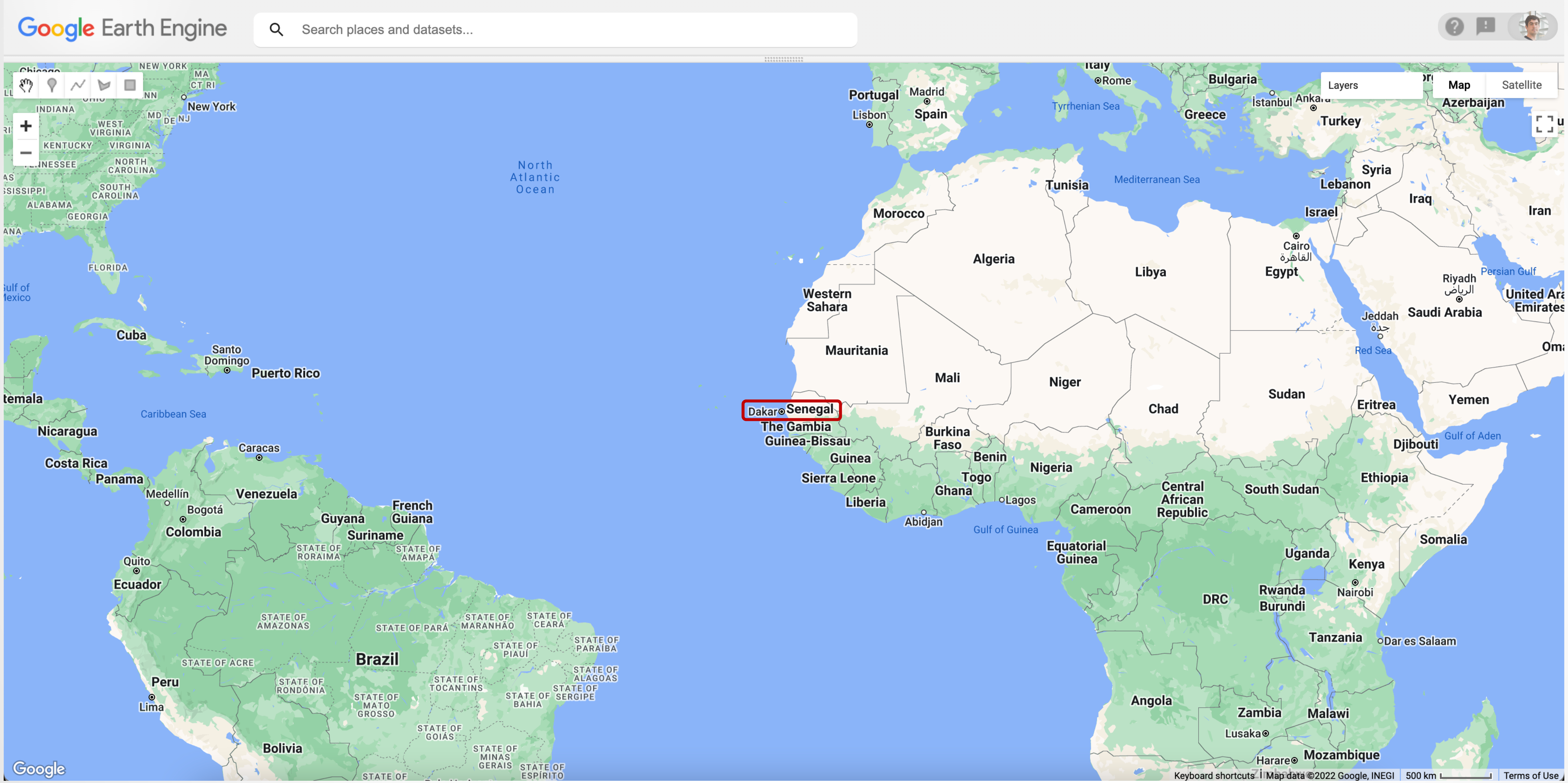
WUSTL: air quality data integration expertise (monthly/annual timescales)

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

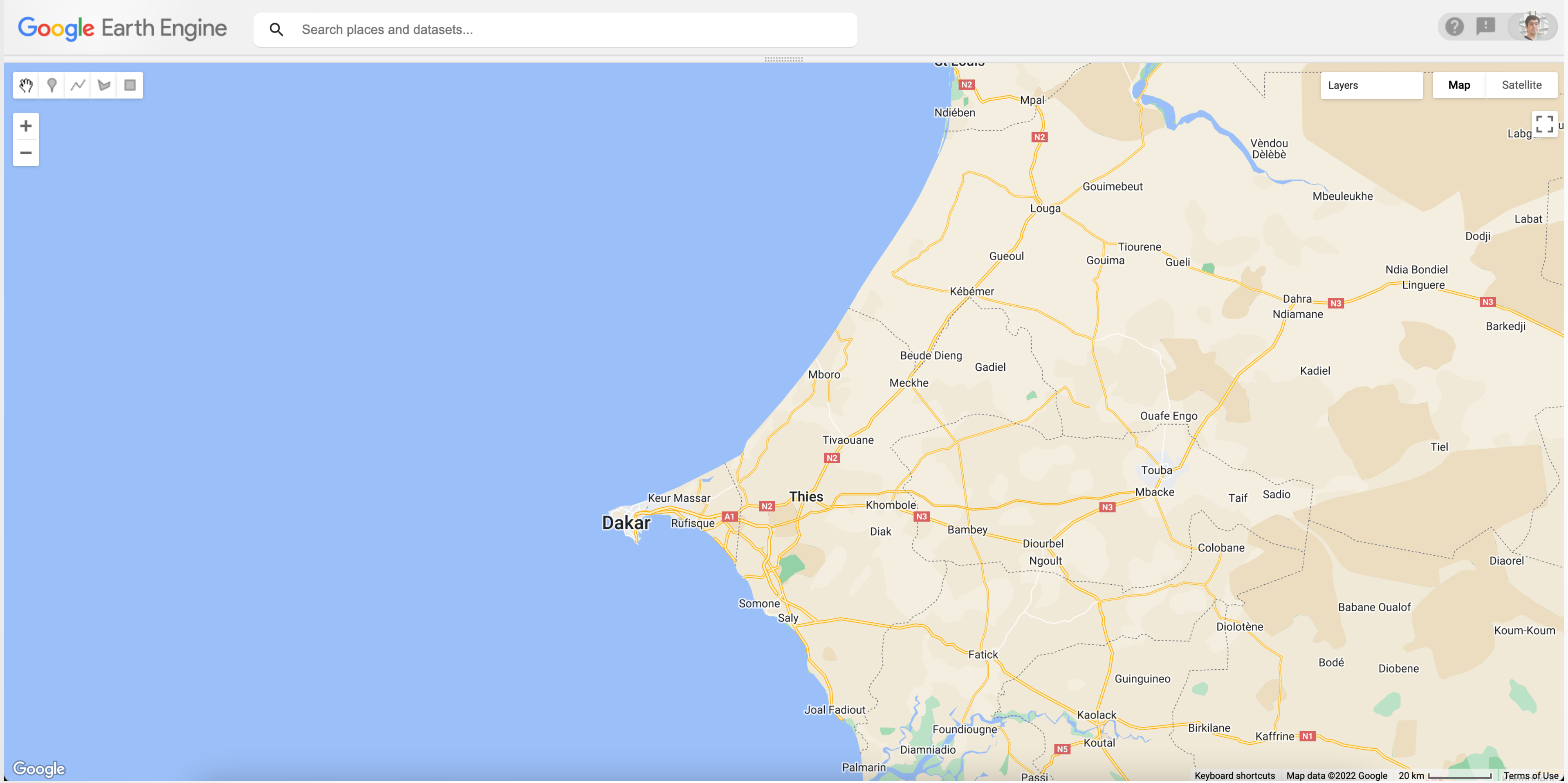
UNEP: integration with global users
Dakar, Senegal
Rio de Janeiro, Brazil

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

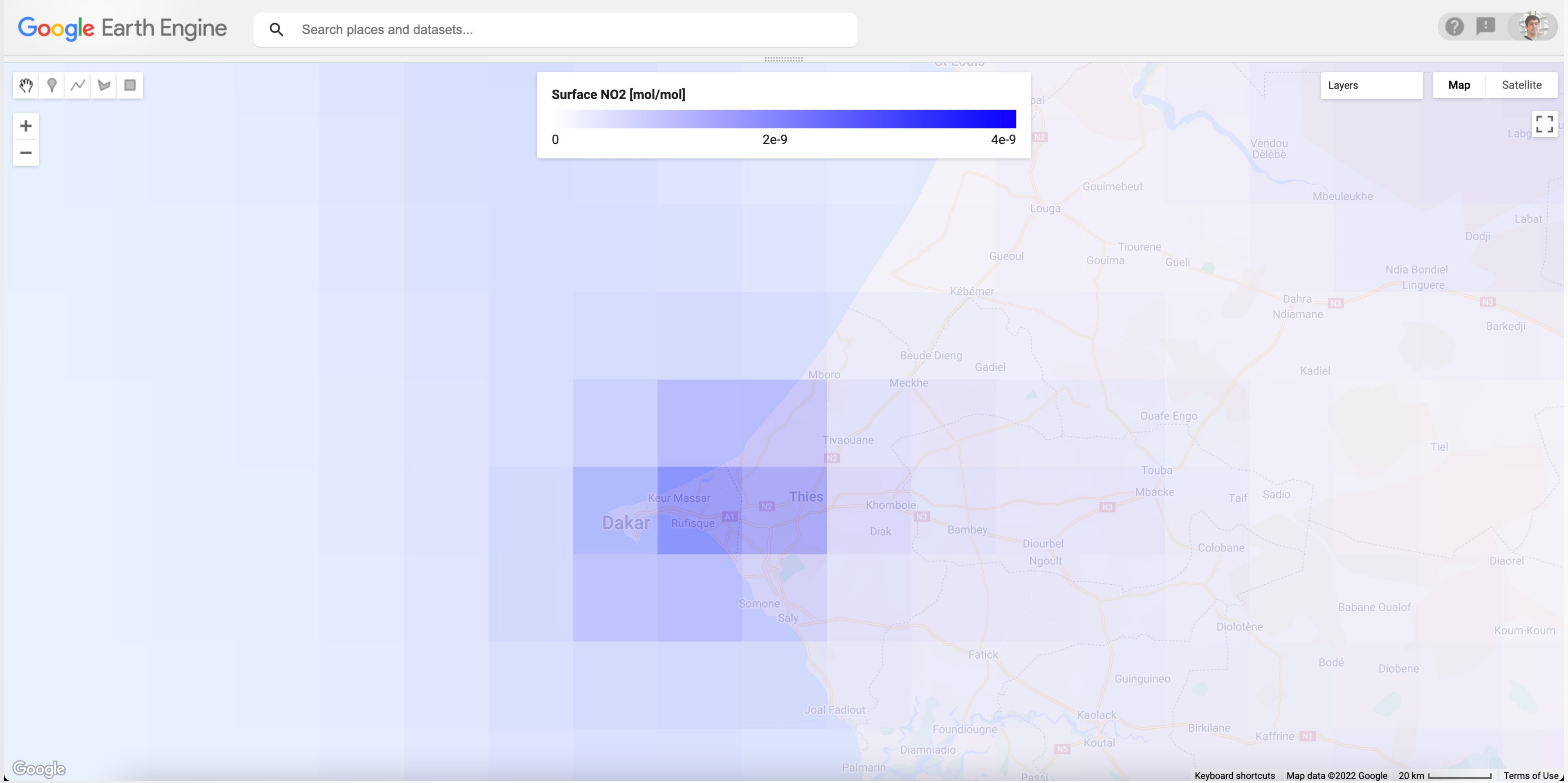
Demonstration of Data Fusion in GEE (preliminary)



Demonstration of Data Fusion in GEE (preliminary)



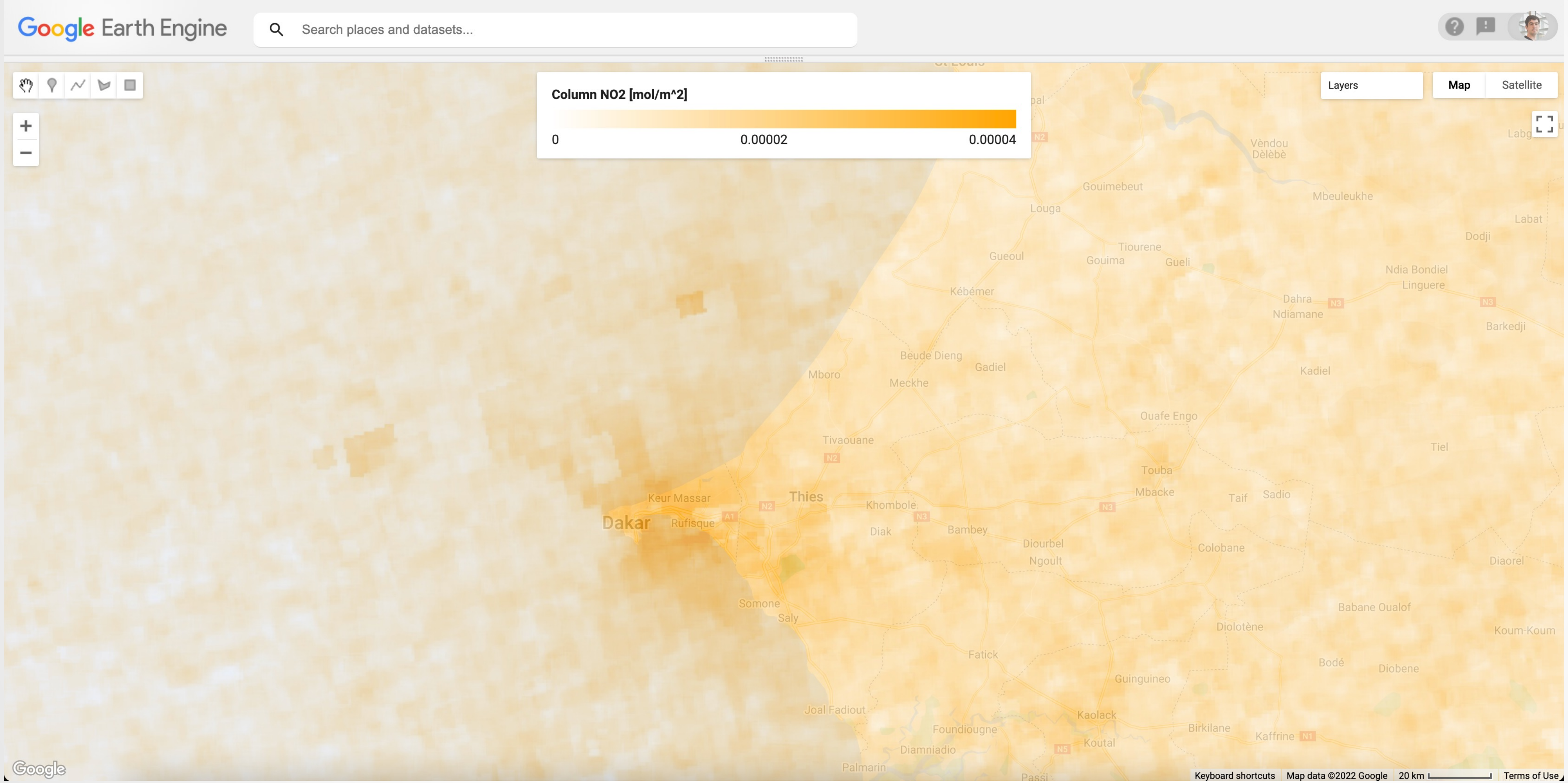
Demonstration of Data Fusion in GEE (preliminary)



Calibration

Model

Demonstration of Data Fusion in GEE (preliminary)

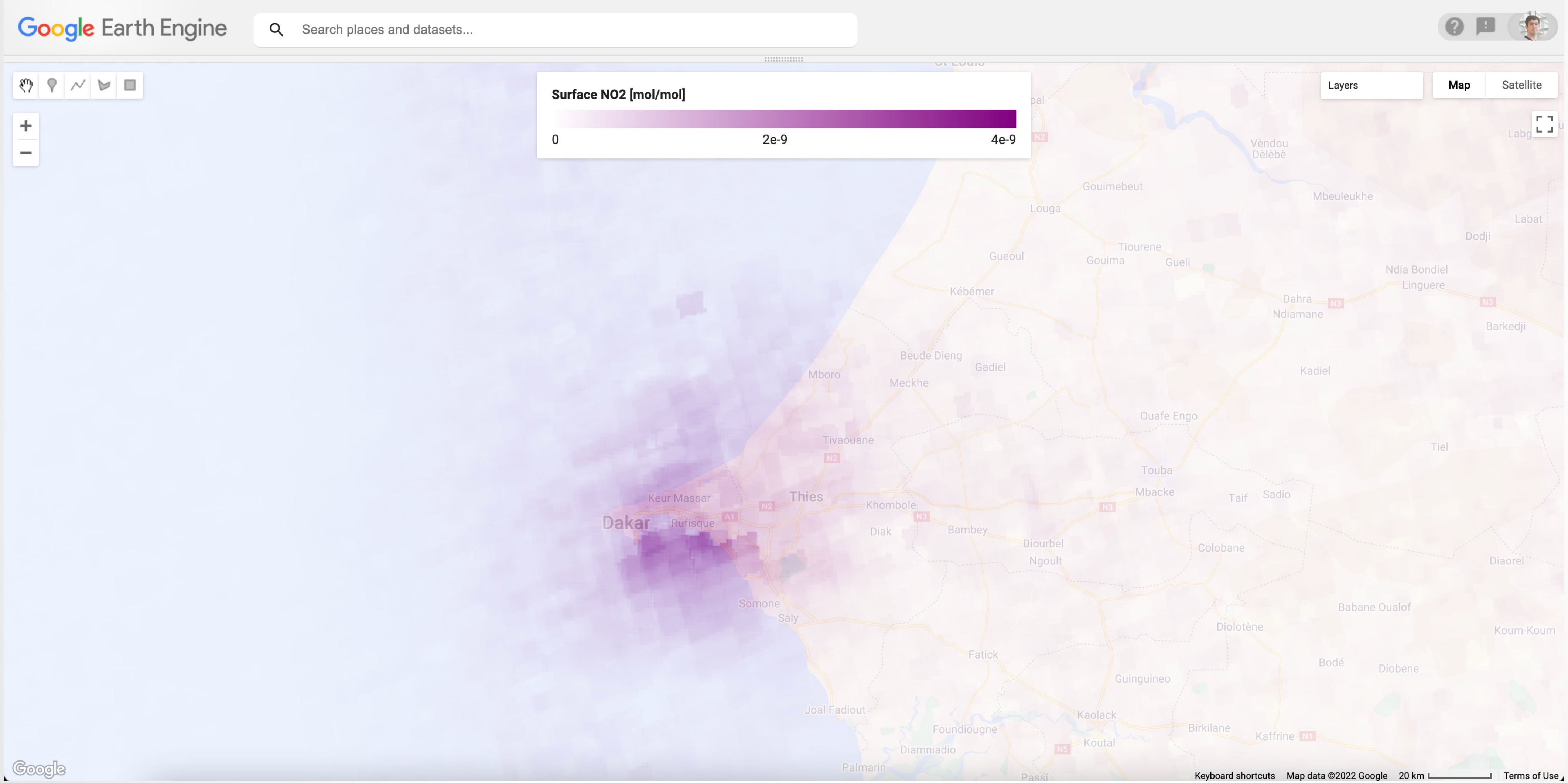


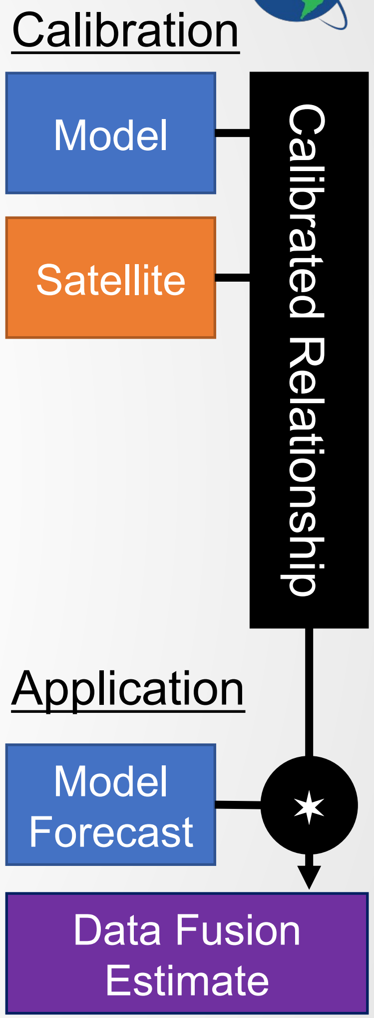
Calibration

Model

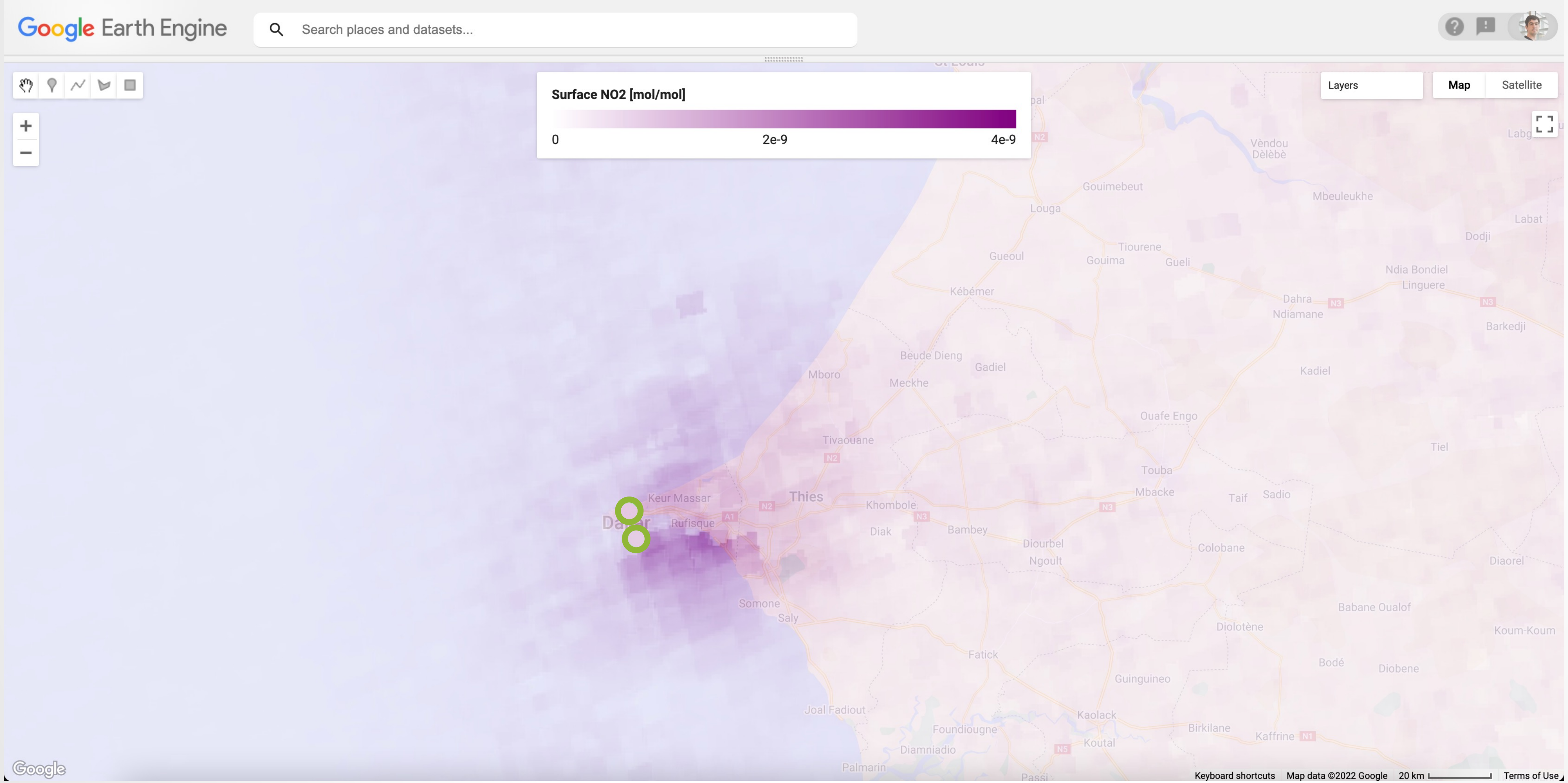
Satellite

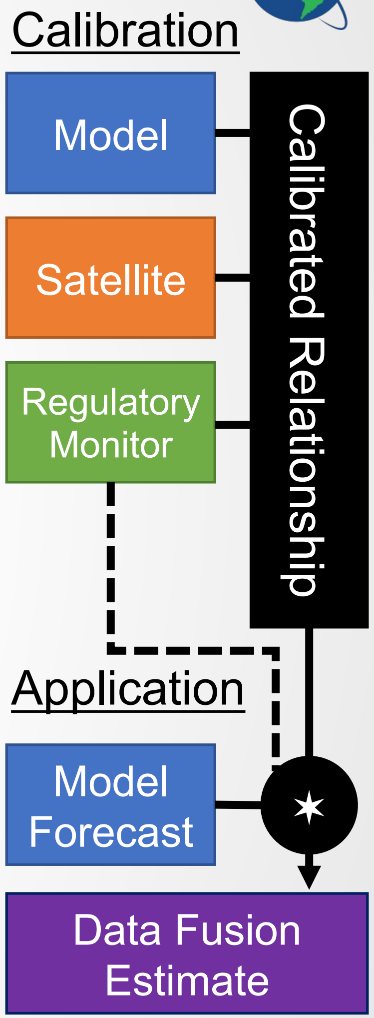
Demonstration of Data Fusion in GEE (preliminary)



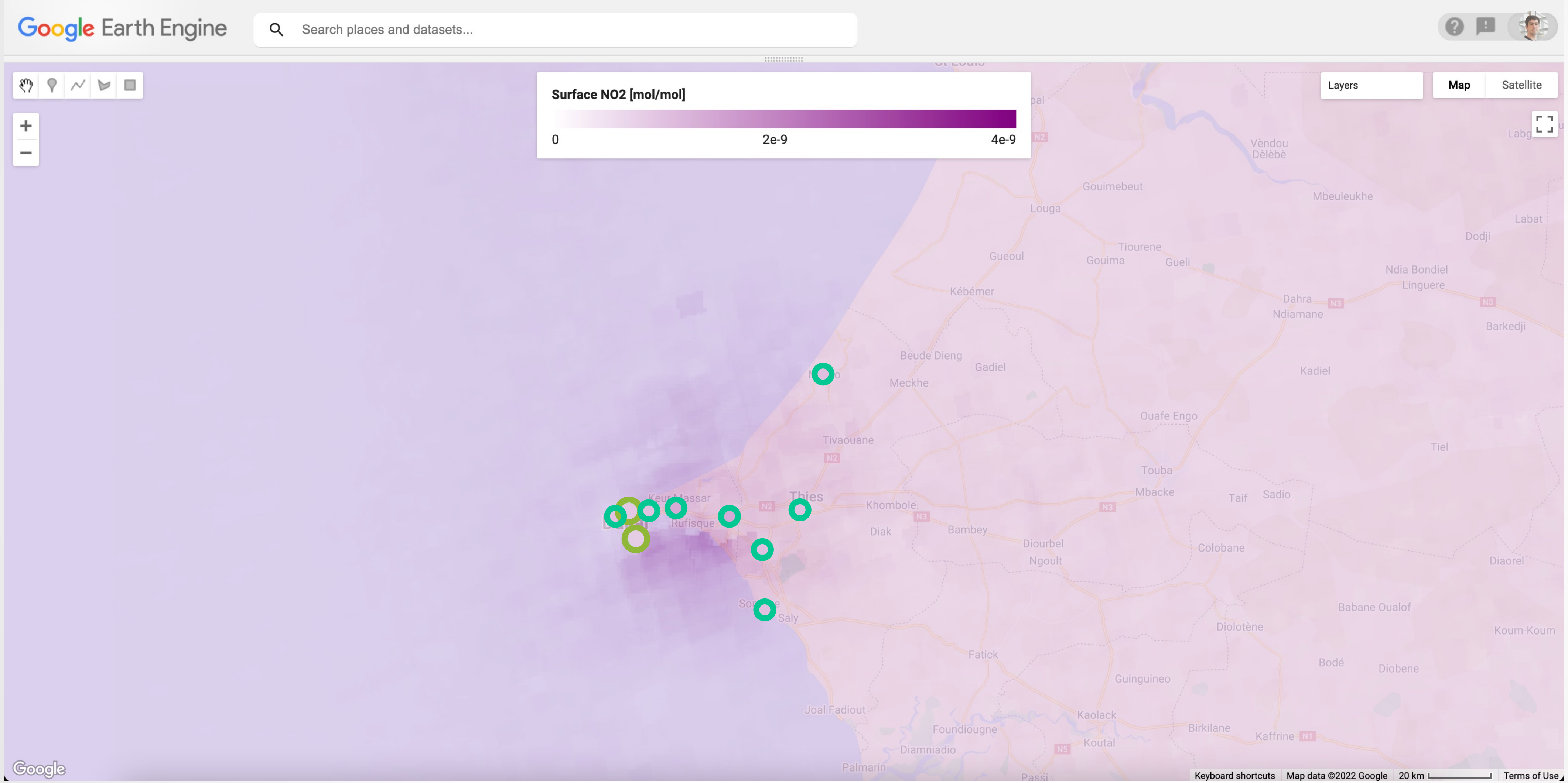


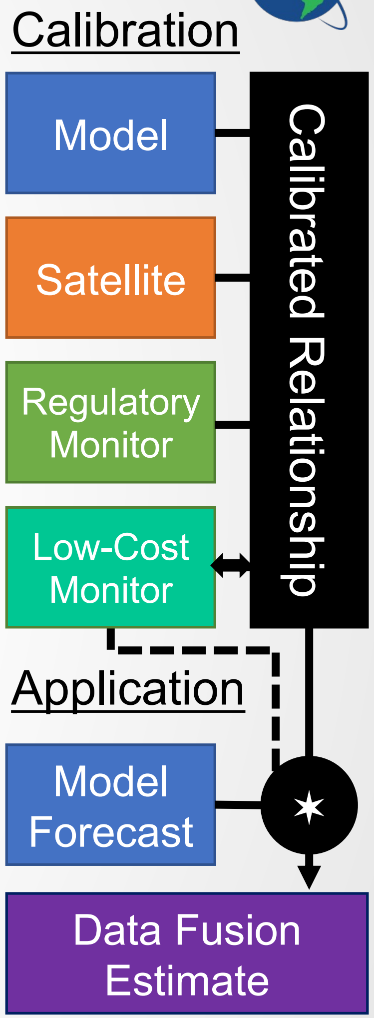
Demonstration of Data Fusion in GEE (preliminary)



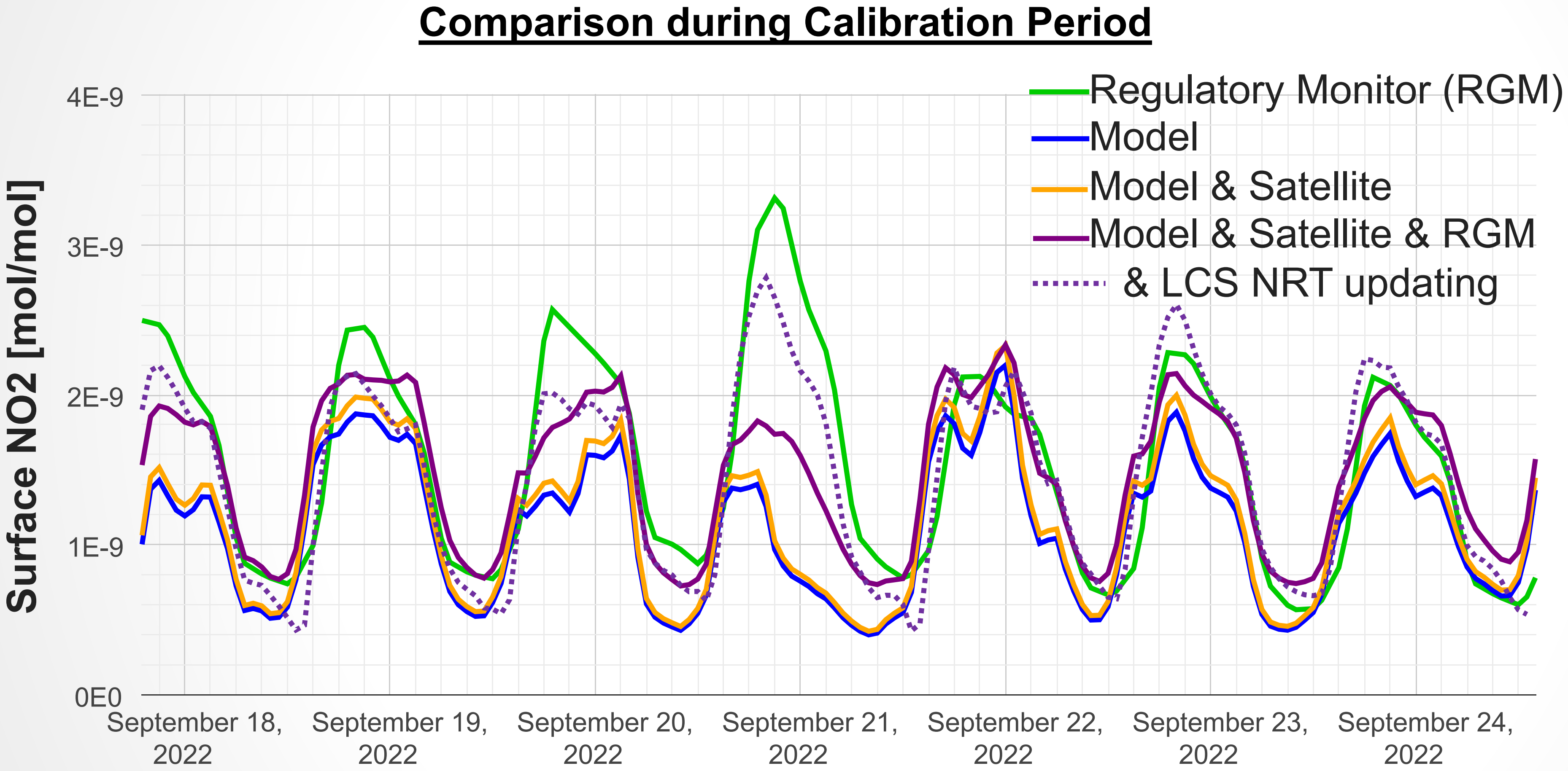


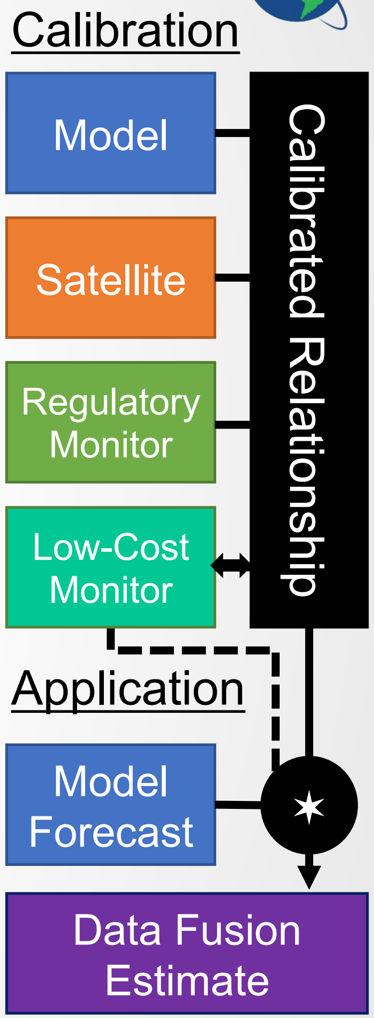
Demonstration of Data Fusion in GEE (preliminary)





Demonstration of Data Fusion in GEE (preliminary)





Prototype Data Fusion System Current Status

Earth Engine Apps

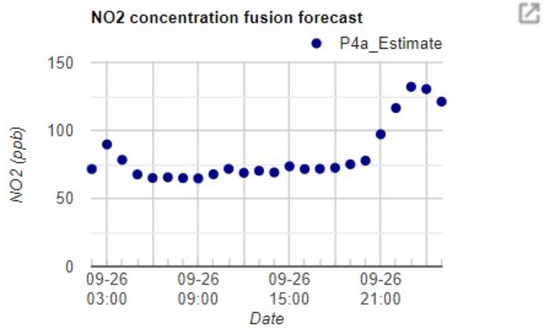
Courtesy of Nathan Pavlovic & Jonathan Coughlin, Sonoma Technologies, Inc.

Sub-city air quality forecasts

Select the region of interest to view forecasts

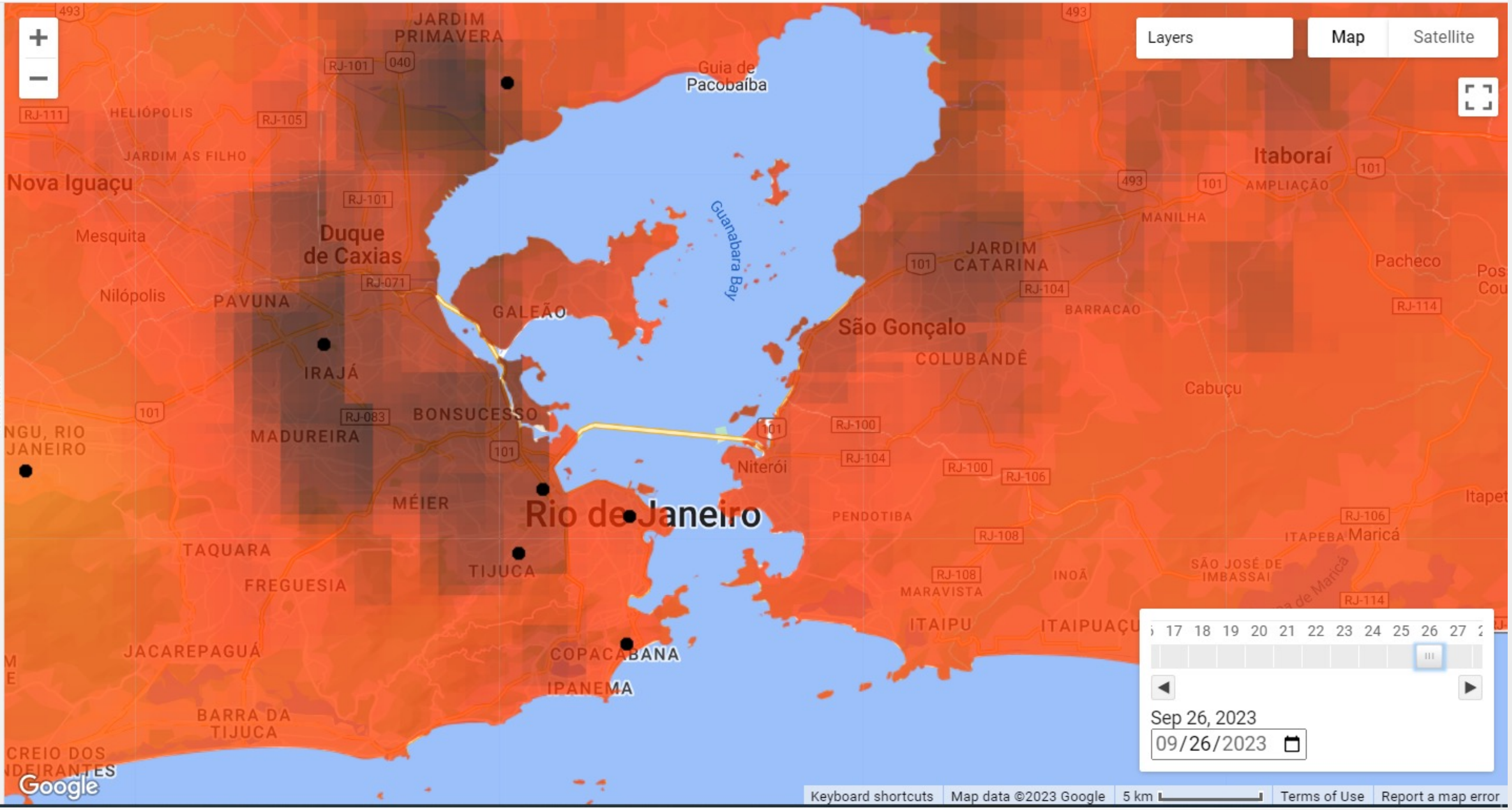
Rio de Janeiro, BR

P4

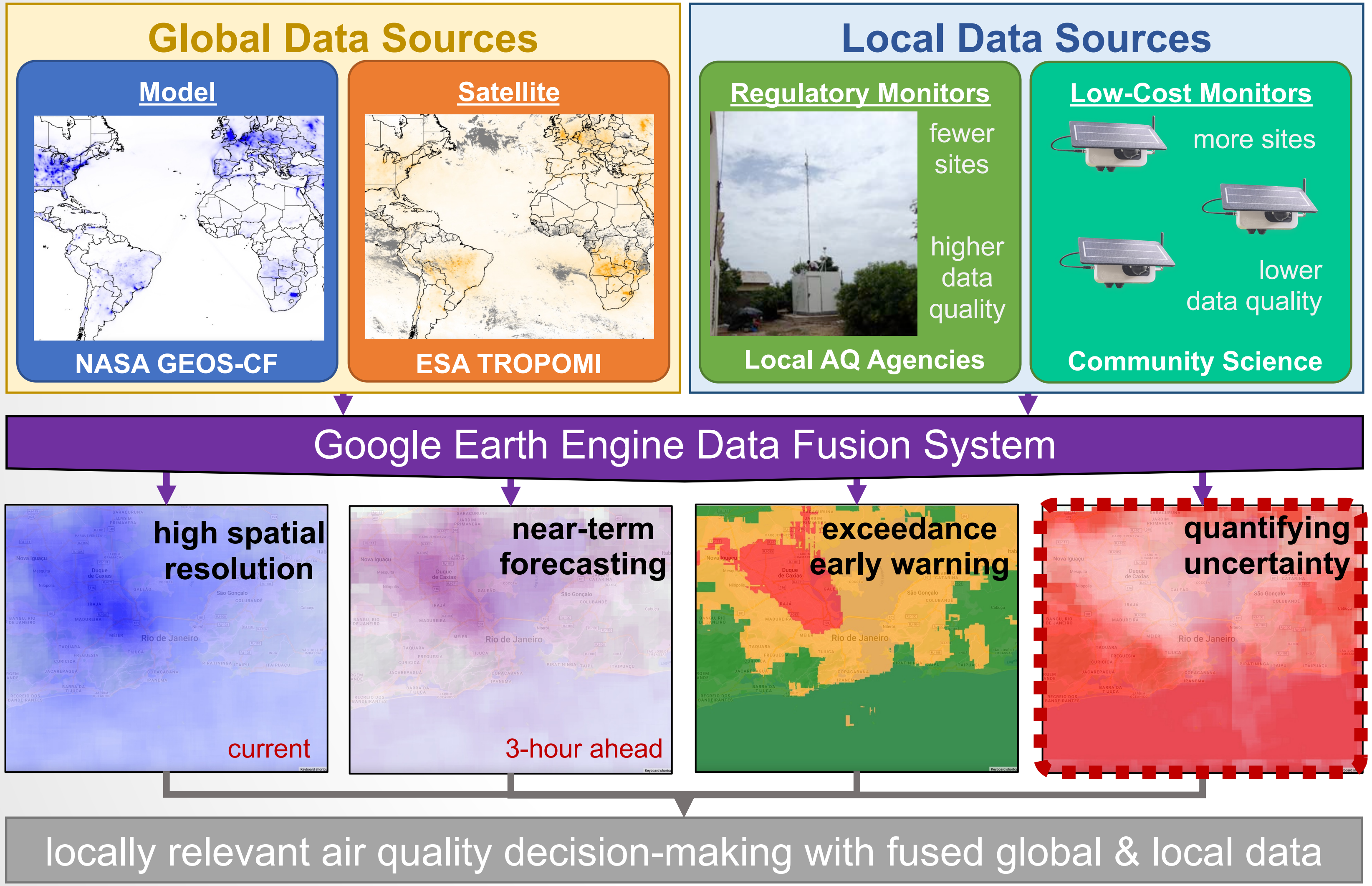


Hourly NO2 Concentration (ppb)

40 80



As part of our NASA A.37 project, we want to quantify uncertainty in our Google Earth Engine data fusion tool



Uncertainty quantification allows:

- Properly incorporating different types of measurements (e.g., low-cost sensors v. regulatory monitors)
- Establishing confidence in estimates and forecasts
- Proper communication of results to end-users
- Identifying the contributions of different data in the fusion process
- Prioritizing new data sources for end users, e.g., low-cost sensor deployments

Phase 1 Uncertainty

$$F_1(x,t) = M(x,t)$$

data fusion uncertainty (variance) at phase 1

$$V_1(x,t) = V_{F1}(x,t,\tau)$$

uncertainty due to forecasting by τ ahead (ignore this for now…)

$$+V_M(x,t)$$

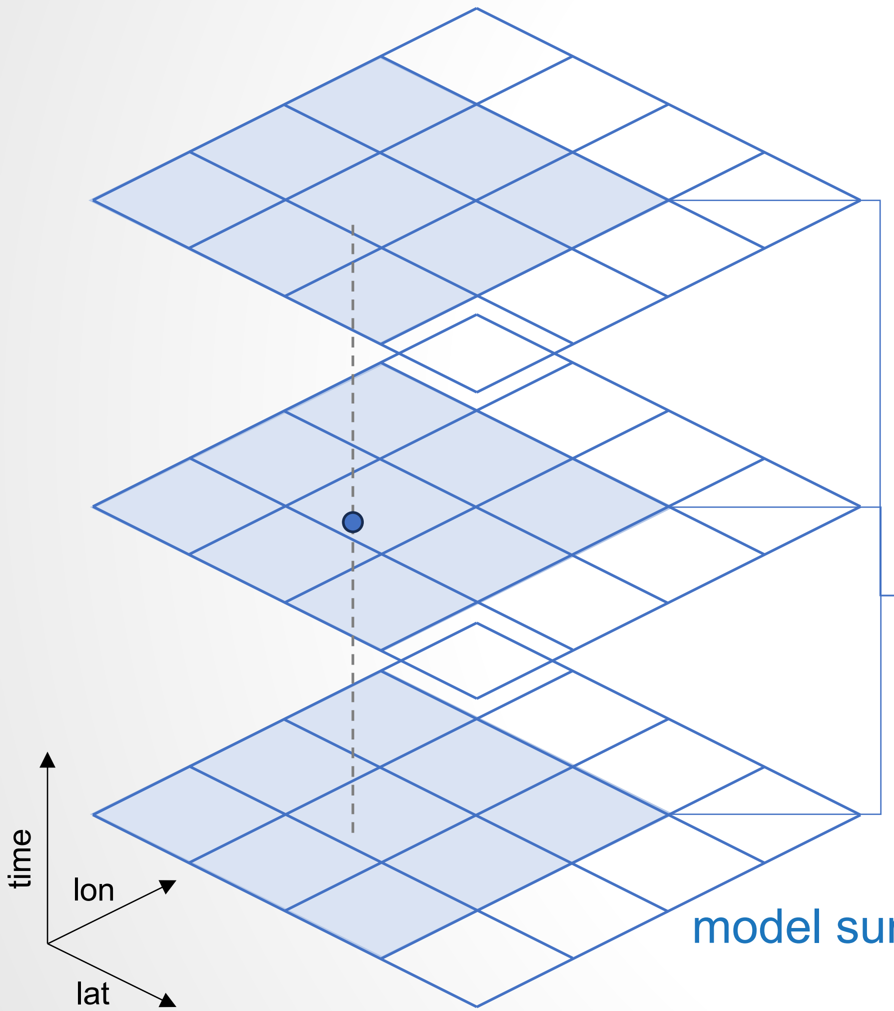
uncertainty due to model internal variability

$$+V_{B1}(x,t)$$

uncertainty due to model bias

$$+V_{R1}(x,t)$$

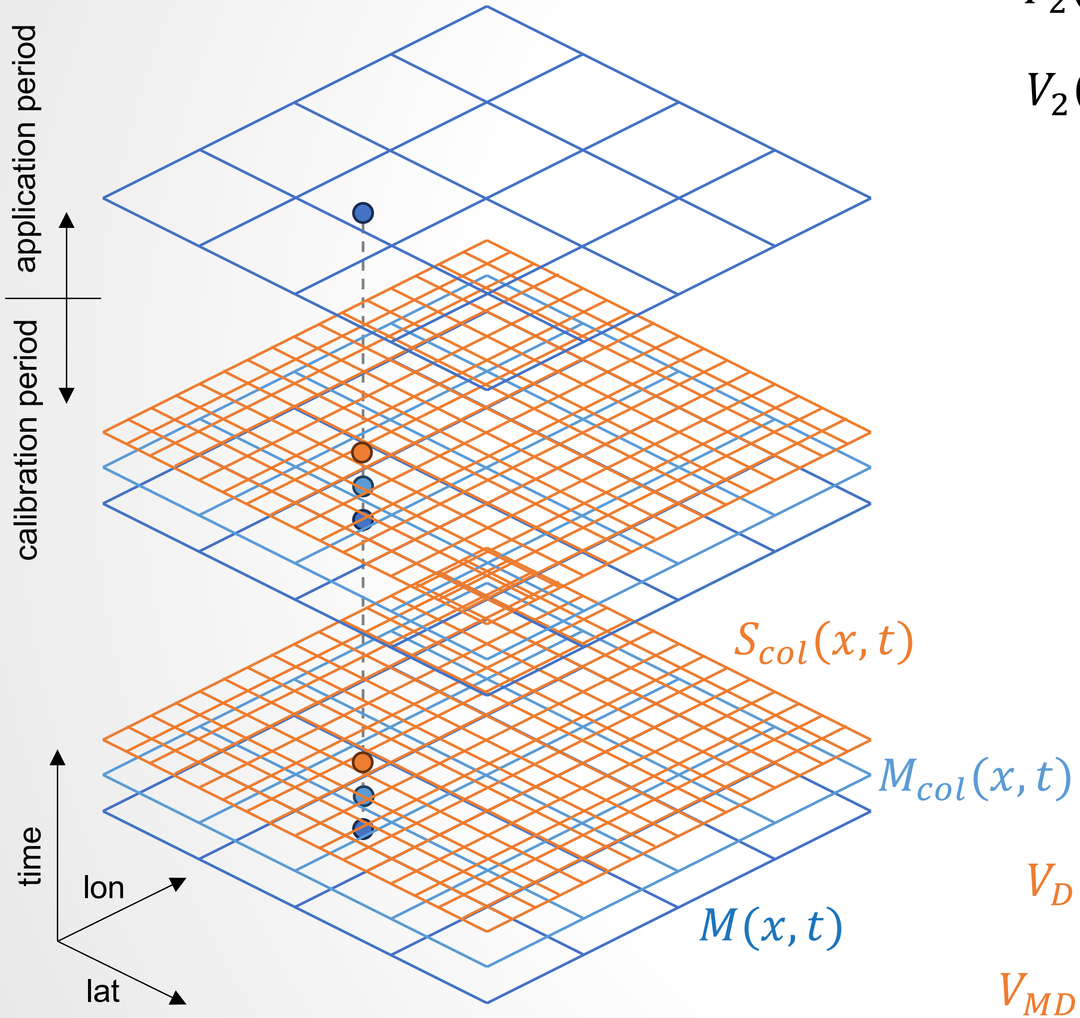
uncertainty due to spatial representativity (model scale)



$$V_M(x,t) \approx \mathbb{E}_{x' \in X_n(x), t' \in T_n(t)}\left[\left(M(x',t') - M(x,t)\right)^2\right]$$

model surface concentration $M(x,t)$

Phase 2 Uncertainty



$$F_2(x,t) = F_1(x,t) + D(x,t)$$

$V_2(x,t) = V_{F2}(x,t,\tau)$ ← uncertainty due to forecasting by τ ahead (ignore this for now…)

$+V_M(x,t)$ ← uncertainty due to model internal variability

$+V_D(x,t)$ ← uncertainty in satellite-to-model differences (estimated over the calibration period)

$+2V_{MD}(x,t)$ ← co-variance of satellite-to-model differences with model outputs (empirical estimate)

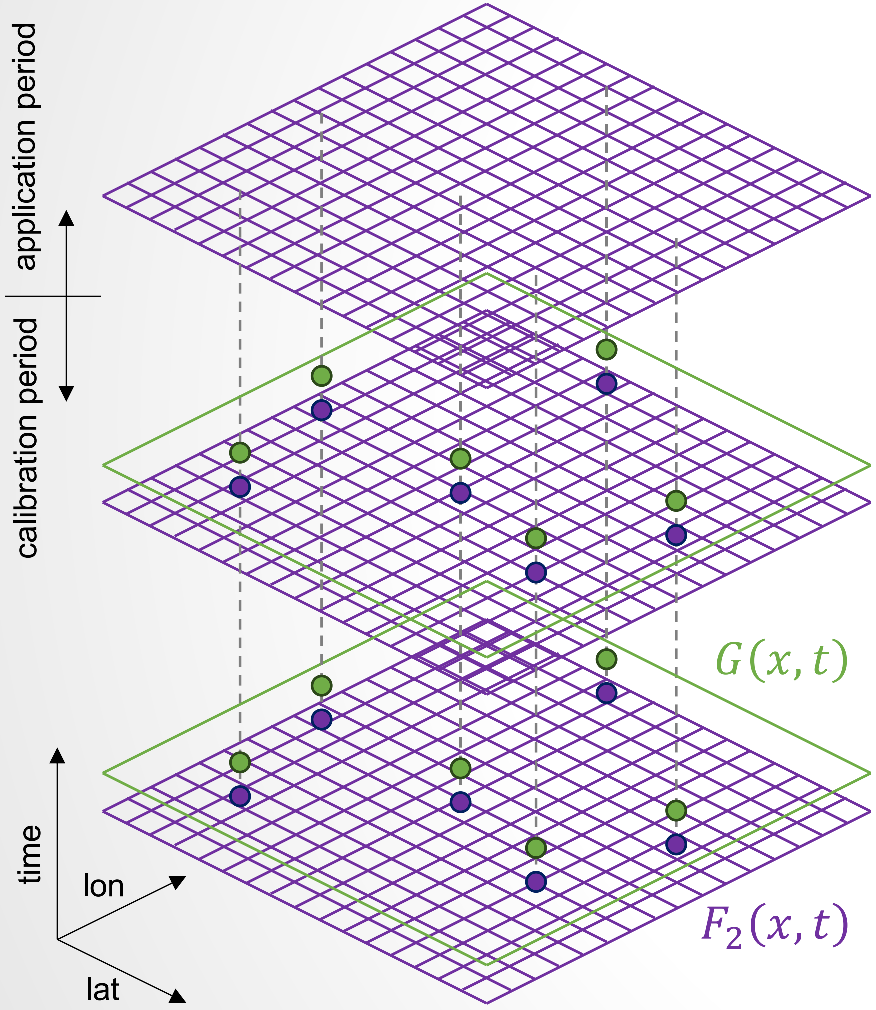
$+\mathrm{V}_{B2}(x,t)$ ← uncertainty due to model & satellite bias

$+V_{R2}(x,t)$ ← uncertainty due to spatial representativity (satellite scale)

$$V_D(x,t) \approx \mathbb{V}_{t'\in T_{c,overpass}(t)}\left[\left(S_{col}(x,t') - M_{col}(x,t')\right)\phi(x,t')\,\psi(x,t,t')\right]$$

$$V_{MD}(x,t) \approx \mathbb{E}_{x'\in X_n(x),t'\in T_n(t)}\left[\left(M(x',t') - M(x,t)\right)\left(D(x',t') - D(x,t)\right)\right]$$

Phase 3 Uncertainty



$$F_3(x,t) = \theta_1 F_2(x,t) + \theta_0$$

$$V_3(x,t) = V_{F3}(x,t,\tau) \leftarrow$$ uncertainty due to forecasting by τ ahead (ignore this for now…)

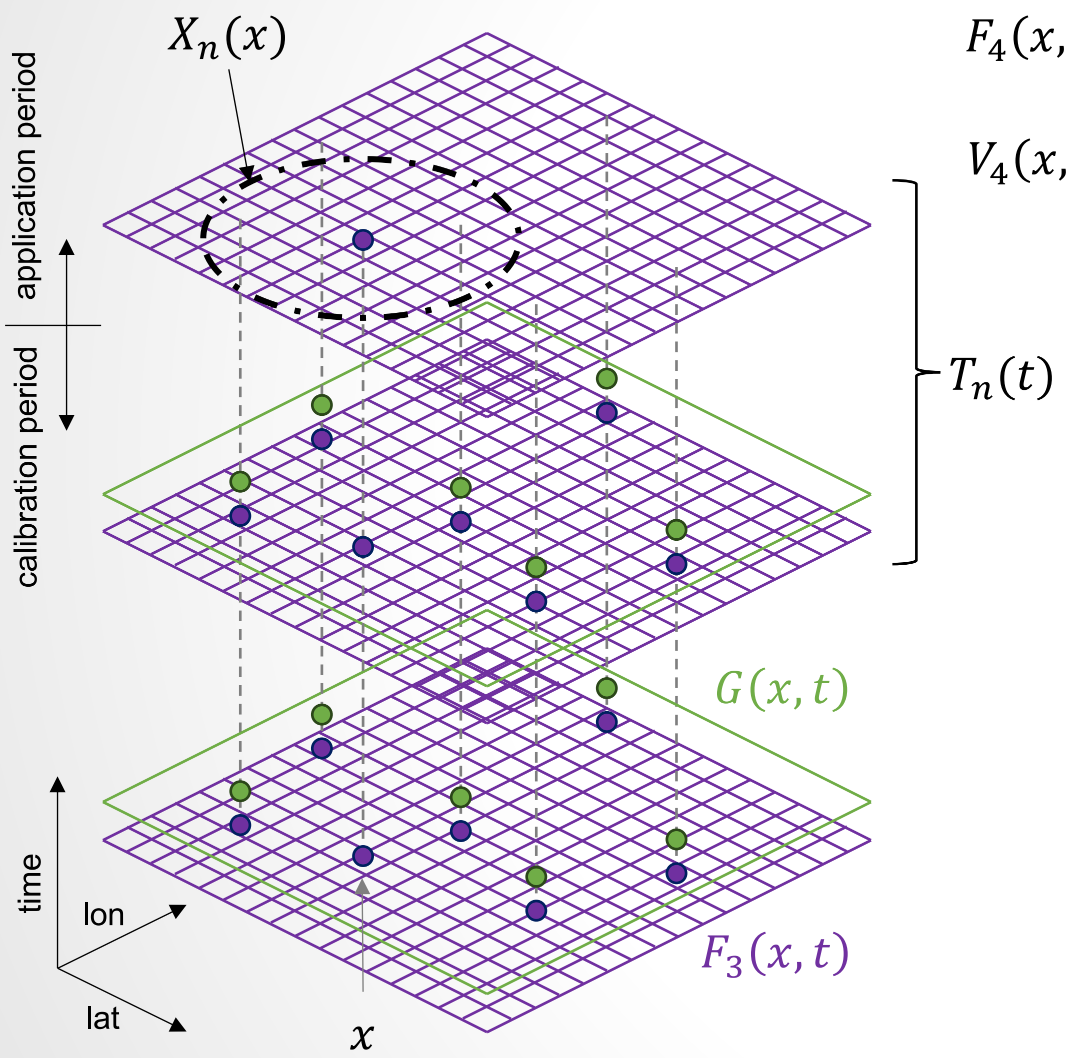
$$+\theta_1^2[V_M(x,t) + V_D(x,t) + 2V_{MD}(x,t)] \leftarrow$$ rescaled from phase 2

$$\left.\begin{array}{l} +\mathrm{var}[\theta_1]F_2(x,t)^2 \\ +2\mathrm{cov}[\theta_0,\theta_1]F_2(x,t) \\ +\mathrm{var}[\theta_0] \\ +\sigma_{residual}^2 \end{array}\right\}$$ variance and co-variance of regression parameters as well as regression residual are known

We can now quantify every term contributing to the uncertainty!

…provided that we ignore potential systematic differences between the calibration and applications periods and the fact that the ground monitors may not be representative of the domain as a whole

Phase 4 Uncertainty



$$F_4(x,t) = F_3(x,t) + \sum_{x' \in X_n(x), t' \in T_n(t)} K(x, x', t, t') \left[G(x', t') - F_3(x', t') \right]$$

$$V_4(x,t) = V_3(x,t) \underbrace{- \sum_{x' \in X_n(x), t' \in T_n(t)} K(x, x', t, t') \operatorname{cov}\left[G(x', t'), F_3(x', t') \right]}_{\text{Kriging reduction from phase 3 variance}}$$

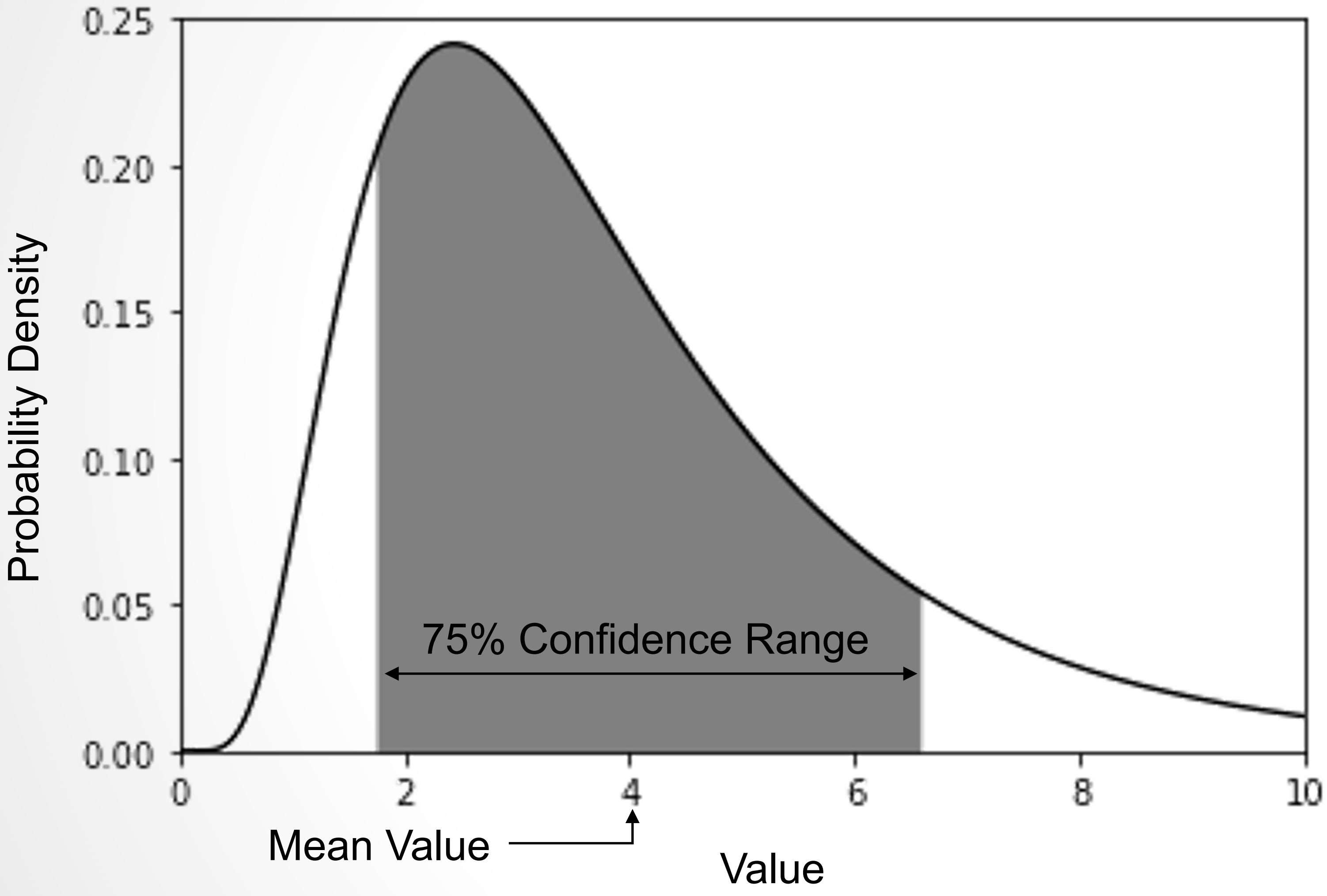
Summary of Data Fusion Estimates & Uncertainties

Phase		Estimate	Uncertainty			
			Bias	Model Variance	Model Scale Spatial Representativity	Satellite Scale Spatial Representativity
1	Model	$F_1(x,t) = M(x,t)$	$V_{B1}(x,t)$	$V_M(x,t)$	$V_{R1}(x,t)$	
2	Model & Satellite	$F_2(x,t) = \text{avg}_{t' \in T_c(t)}\left[\left(S_{col}(x,t') - M_{col}(x,t')\right)\phi(x,t')\,\psi(x,t,t')\right] + F_1(x,t)$ $= D(x,t) + F_1(x,t)$	$V_{B2}(x,t)$	$V_M(x,t)$	$V_D(x,t) + 2V_{MD}(x,t)$	$V_{R2}(x,t)$
3	Model & Satellite & Ground	$F_3(x,t) = \theta_1 F_2(x,t) + \theta_0$ $with \;\; \theta_0, \theta_1 = \mathbb{LR}_{t' \in T_c(t), x' \in X_c(x)}[G(x',t') \sim F_2(x',t')]$	0*	$\theta_1^2 V_M(x,t)$	$\theta_1^2[V_D(x,t) + 2V_{MD}(x,t)]$	$\text{var}[\theta_1]F_2(x,t)^2$ $+ 2\text{cov}[\theta_0,\theta_1]F_2(x,t)$ $+ \text{var}[\theta_0] + \sigma^2_{residual}$
4	Model & Satellite & Ground & Kriging	$F_4(x,t) = \text{F}_3(x,t) + \sum_{x' \in X_n(x), t' \in T_n(t)} K(x,x',t,t')\,[G(x',t') - F_3(x',t')]$	0*	$\theta_1^2 V_M(x,t)$	$\theta_1^2[V_D(x,t) + 2V_{MD}(x,t)]$	$\text{var}[\theta_1]F_2(x,t)^2$ $+ 2\text{cov}[\theta_0,\theta_1]F_2(x,t)$ $+ \text{var}[\theta_0] + \sigma^2_{residual}$
			$-\sum_{x' \in X_n(x), t' \in T_n(t)} K(x,x',t,t')\,\text{cov}[G(x',t'), F_3(x,t)]$			

*Ground measurements during the calibration period are assumed to be an unbiased representation of concentrations across the domain of interest during the application period

Using outputs from later phases, informed estimates can be made for uncertainty terms in earlier phases
Ad hoc estimates for these terms might be crafted later based on observed relationships with model outputs

Defining Confidence Intervals



Assuming a distribution for the values being estimated (a lognormal distribution is assumed in this case), confidence intervals can be estimated.

$$\mu(x,t) = \log\left[\frac{F(x,t)}{\sqrt{1+\frac{V(x,t)}{F(x,t)^2}}}\right]$$

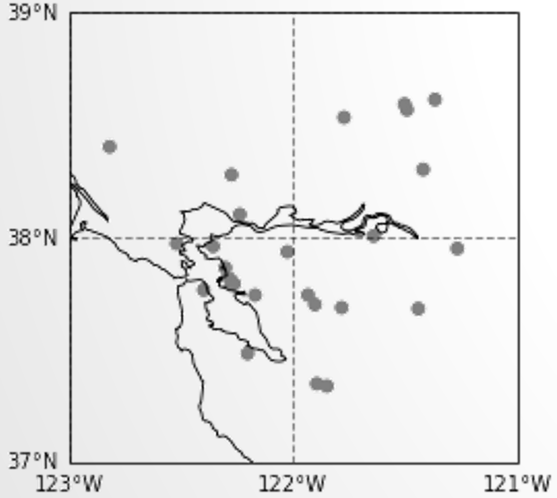
$$\sigma(x,t) = \sqrt{\log\left[1+\frac{V(x,t)}{F(x,t)^2}\right]}$$

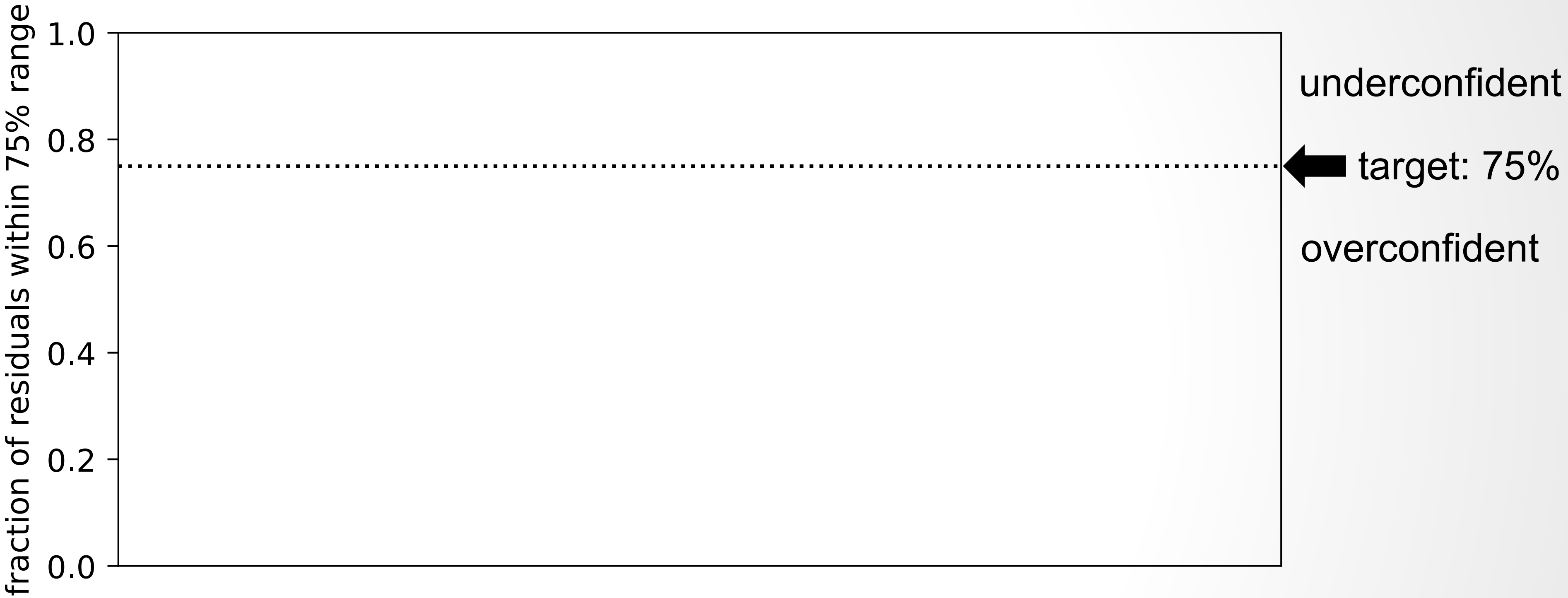
$$f(x,t) \sim LN\big(\mu(x,t), \sigma(x,t)\big)$$

Case Study Results

Case Study Details
San Francisco
September 2019
Surface NO_2
Lognormal distribution
Cross-validation test
25 ground monitors

Ground Sites

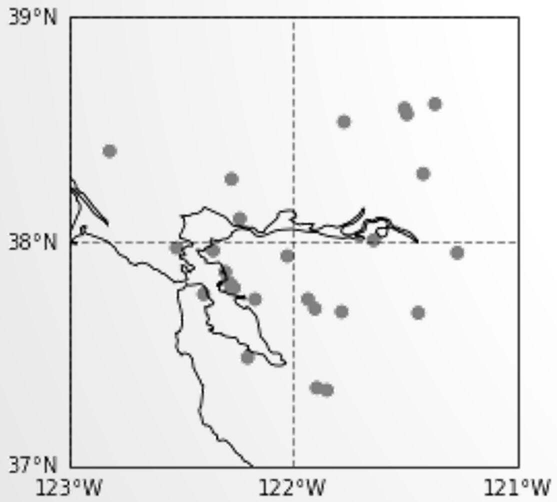


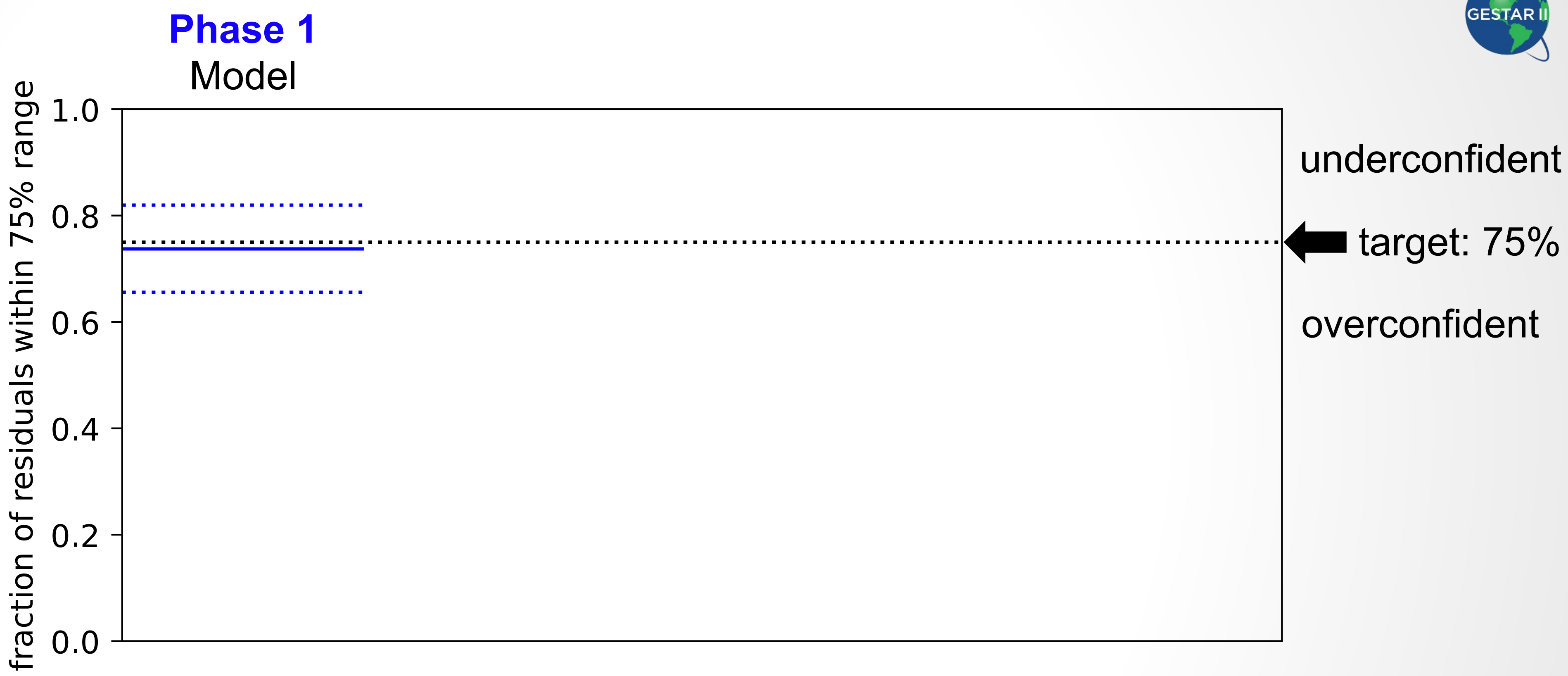


Case Study Results

Case Study Details
San Francisco
September 2019
Surface NO_2
Lognormal distribution
Cross-validation test
25 ground monitors

Ground Sites

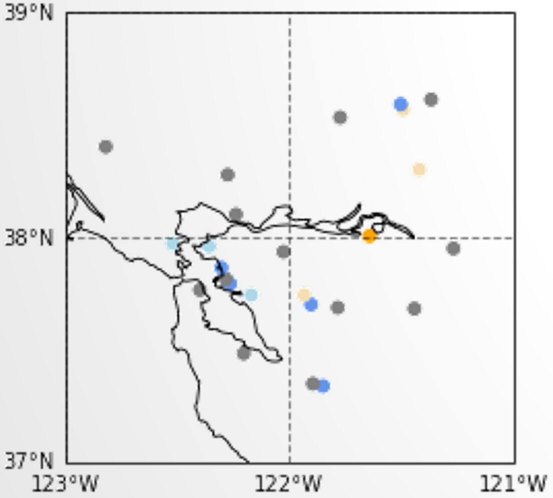


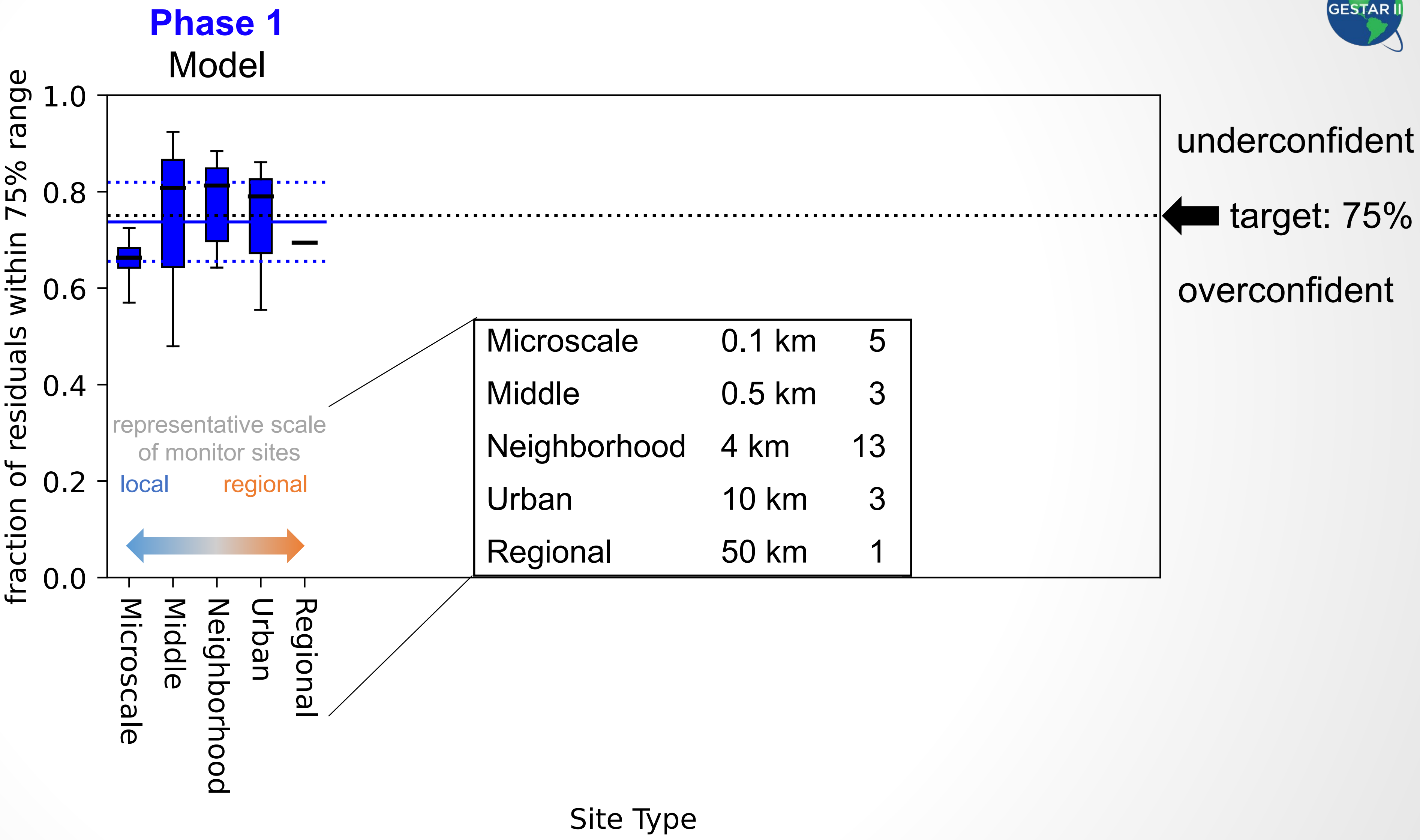


Case Study Results

Case Study Details
San Francisco
September 2019
Surface NO_2
Lognormal distribution
Cross-validation test
25 ground monitors

Ground Sites





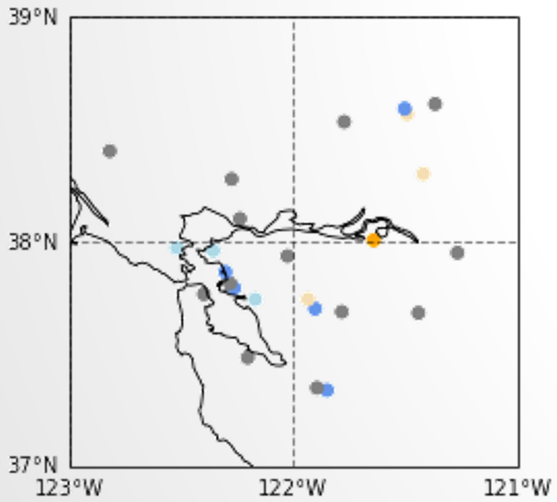
Microscale	0.1 km	5
Middle	0.5 km	3
Neighborhood	4 km	13
Urban	10 km	3
Regional	50 km	1

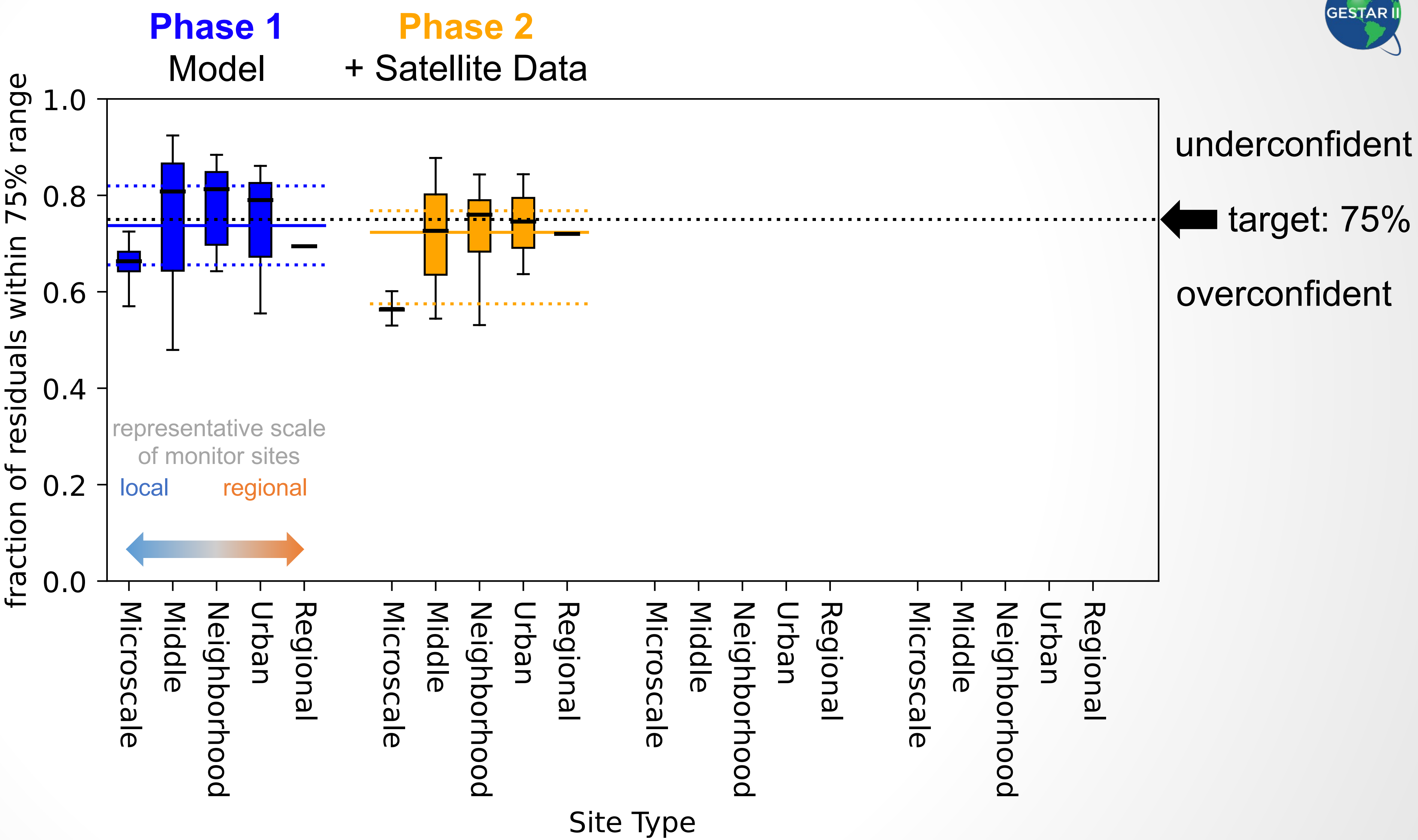
Case Study Results

Case Study Details

San Francisco
September 2019
Surface NO_2
Lognormal distribution
Cross-validation test
25 ground monitors

Ground Sites

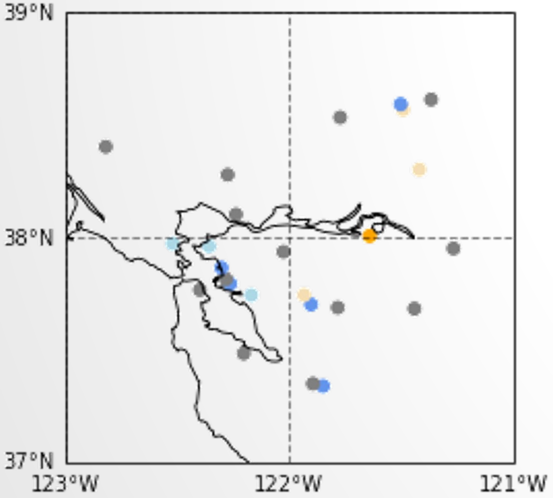


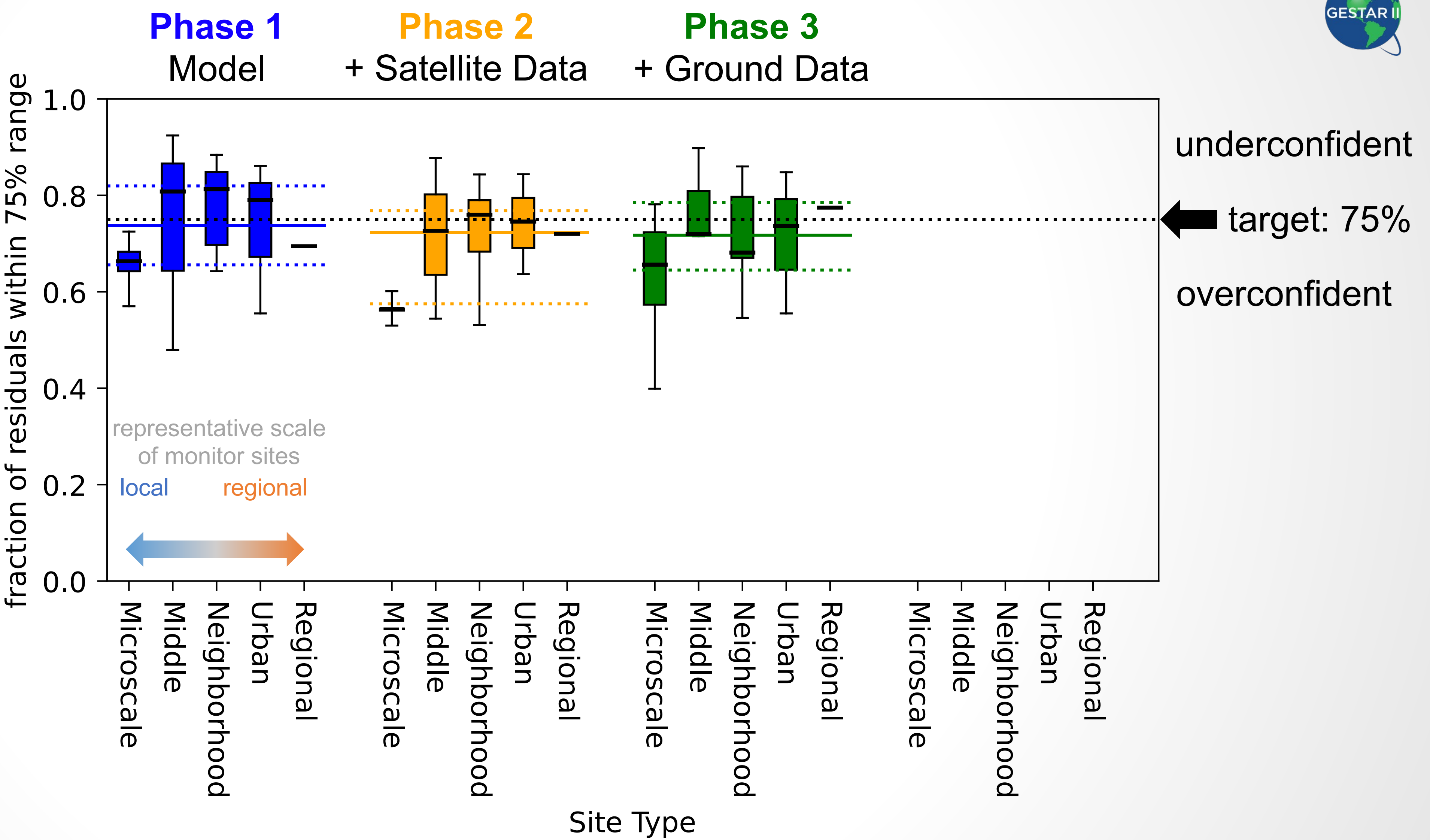


Case Study Results

Case Study Details
San Francisco
September 2019
Surface NO_2
Lognormal distribution
Cross-validation test
25 ground monitors

Ground Sites

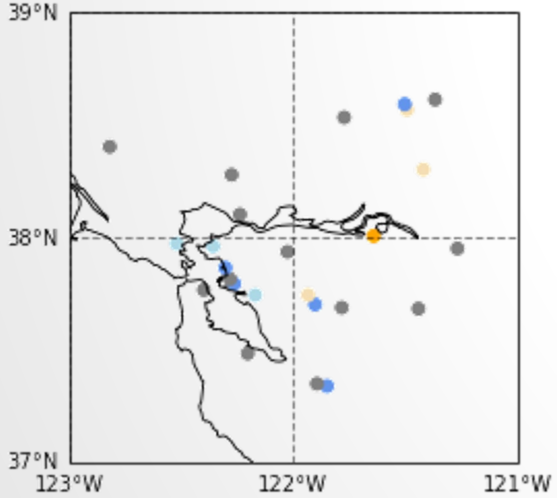


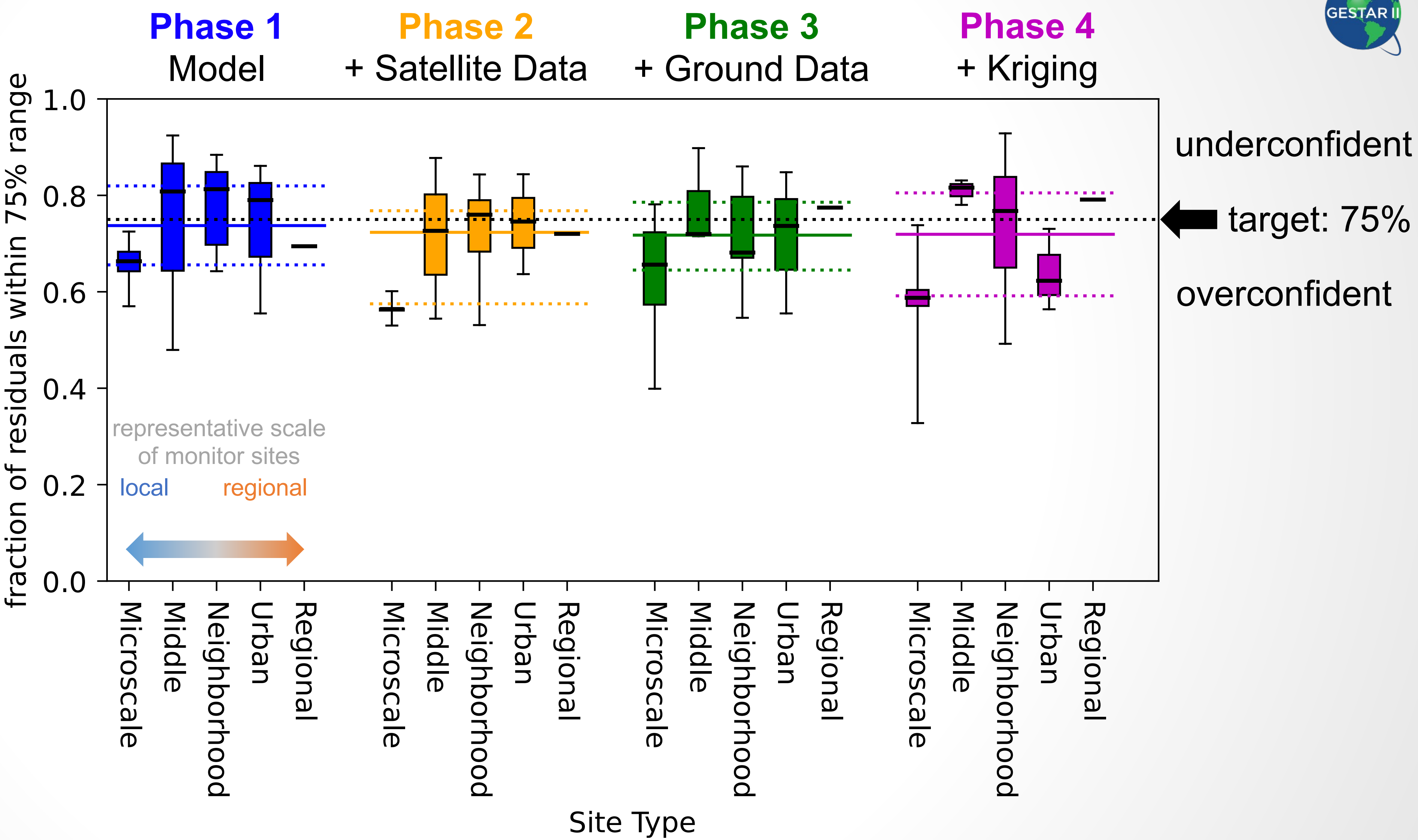


Case Study Results

Case Study Details
San Francisco
September 2019
Surface NO_2
Lognormal distribution
Cross-validation test
25 ground monitors

Ground Sites





NASA Applied Remote Sensing Training (ARSET)

https://appliedsciences.nasa.gov/arset

ARSET provides accessible, relevant, and cost-free training on remote sensing satellites, sensors, methods, and tools.

APPLIED REMOTE SENSING TRAINING PROGRAM

TRAIN - EMPOWER - ADVANCE

Our trainings are:

- Online and in-person
- Open to everyone
- Live, instructor-led, or self-guided
- Provided at no cost, with materials and recordings available from our website
- Often multi-lingual
- Tailored to those with a range of experience in remote sensing, from **introductory** to **advanced**

ARSET offers trainings for:

- Disasters
- Health & Air Quality
- Land Management
- Water Resources
- Climate

EARTH**DATA** Offers

The Air Quality Data Pathfinder

for Your Research & Applications

Air pollution is one of the largest global environmental and health threats. NASA provides data resources to better understand the movement of pollutants and the impact of events leading to poor air quality. This Pathfinder helps you access, and leverage data acquired from NASA's satellite, airborne, and ground-based missions and campaigns.

Available Data Types:

- Aerosols
- Trace Gases (e.g., Nitrogen Dioxide, Sulfur Dioxide, Carbon Monoxide, etc.)
- Weather (e.g., Air Temperature, Clouds, Precipitation, etc.)
- Land Surface (e.g., Soil Moisture, Surface Reflectance, Topography, etc.)
- Human Dimensions

Data are from satellites, airborne and ground-based platforms, and models, including:

- AIRS
- AMSR2
- GPM
- MODIS
- OLI/TIRS
- OMI
- OMPS
- SMAP
- TROPOMI
- VIIRS
- GEOS
- MERRA-2

Visit the EARTH**DATA**
Air Quality Data Pathfinder
for more information:

- Commonly Used Datasets for Air Quality Research and Applications
- Tools for Using Data
- Resources for Applying and Connecting NASA Data
- GIS Resources
- Tips for Getting Help and Connecting with NASA experts
- Tutorials and more!

Health and Air Quality Applied Science Team (HAQAST)

https://haqast.org/

"Our goal is to use NASA's data and satellites to pursue cutting edge applied research in order to keep you healthy and safe."

- Use NASA satellite & other data to help solve real-world public health and air quality problems.
- Work around the world on diverse issues related to health and air quality.
- Collaborate with public stakeholders to help guide long-term research.
- "Tiger Teams" pursue short-term, high-impact projects in small groups.

Getting started with NASA satellite data for health and air quality: https://haqast.org/getting-started/

Thank You!

presenter email: carl.a.malings@nasa.gov