Potential for well-calibrated air sensors to enhance utility of satellite retrievals for air quality management

Prof. Daniel M. Westervelt

Lamont-Doherty Earth Observatory
Columbia University | Earth Institute

UM6P
University Mohammed VI Polytechnic

Westervelt Aerosol Group

Garima Raheja
(PhD student)

Tedy Mwendwa
(Research Engineer)

Joe Adabouk Amooli
(PhD student)

Yanda Zhang (Postdoc)

Sophia Roberts
(Summer intern, Barnard)

NSF

Victoria Owusu Tawiah
(visiting PhD Student)

Trevor Durning
(UG research assistant)

Paulson Kasereka
(PhD student, UM6P)

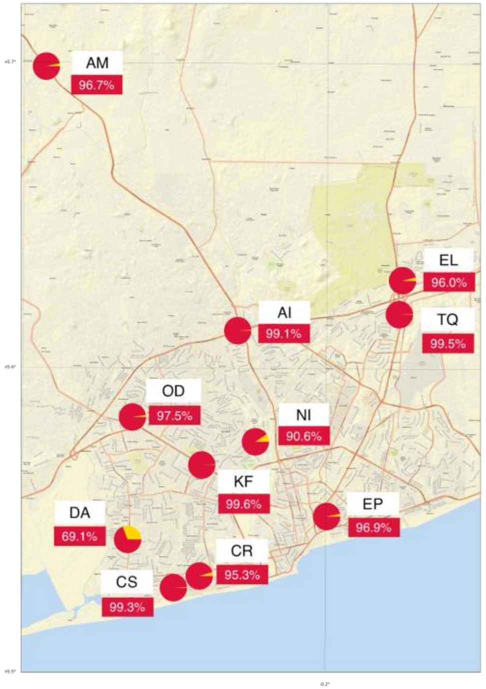
Andrea Belvis Aquino
(Summer intern, UPRM)

DEPARTMENT OF STATE · UNITED STATES OF AMERICA

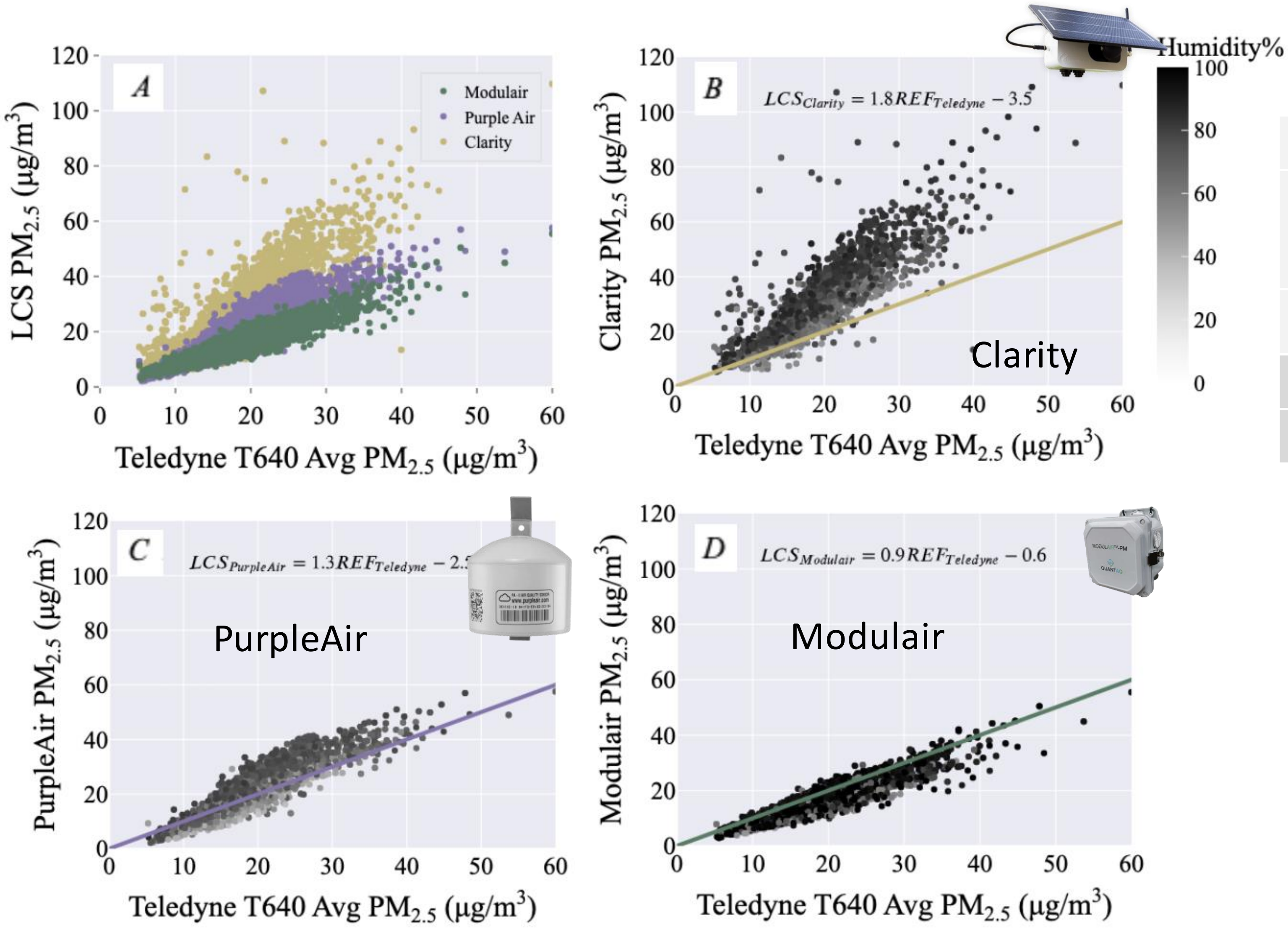
CLEAN AIR FUND

COLUMBIA UNIVERSITY
DATA SCIENCE INSTITUTE

Extracting useful, actionable data from air sensors to close air pollution data gaps around the world



Research projects – Sensor performance between different vendors against reference PM2.5



Model	Purple Air		Clarity		Modulair	
	R^2	MAE (µg/m³)	R^2	MAE (µg/m³)	R^2	MAE (µg/m³)
Manufacturer-Reported	0.82	4.54	0.69	13.68	0.84	3.04
MLR	0.87	1.96	0.74	2.49	0.86	2.15
GMR	0.89	1.76	0.79	2.27	0.87	1.89

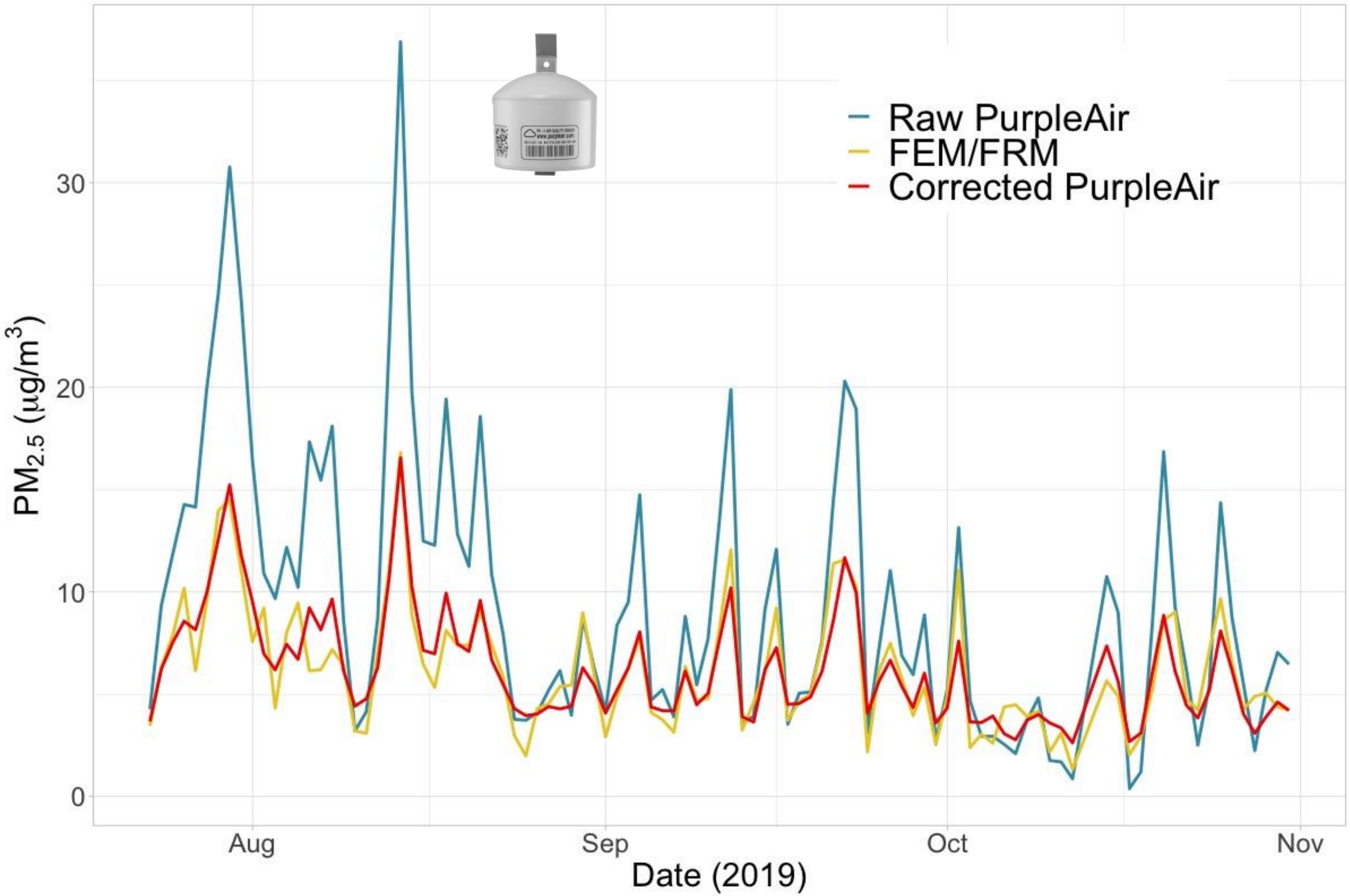
Set aside some fraction of the data to as "training data" to develop "correction factor"

AFRI-SET

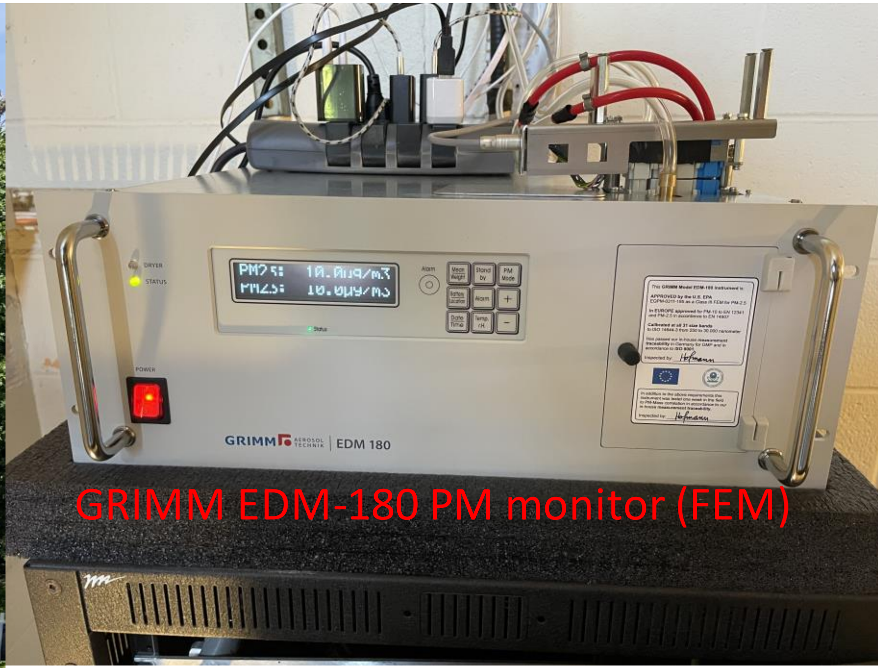
Raheja et al. (2023) *ES&T*

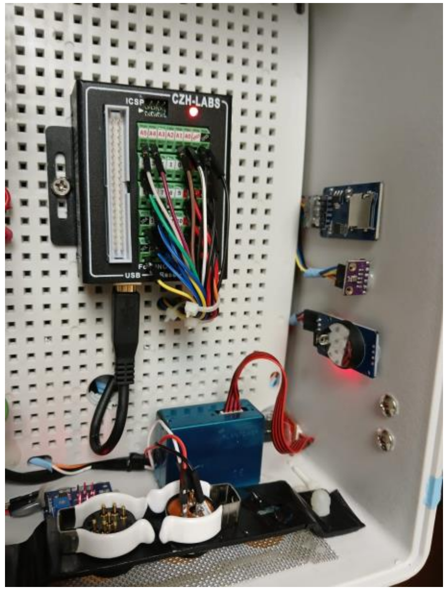
NYC calibrations at Near-Roadway site (Queens College)

	Spearman r	MAE ($\mu g\ m^{-3}$)
Raw-NR	0.801	4.02
Corr-NR	0.817	1.83
Raw-QB	0.757 (0.88)	5.01
Corr-QB	0.849 (0.87)	1.65



Columbia-LDEO Air Sensors lab





“homemade” air sensors: PM, tVOCs, NO2

Satellite data: what role for air sensors?

- **The problem**: satellite estimates provide excellent spatial coverage of pollutants at decent resolution (1 km^2 daily), but measure the entire atmospheric column, rather than "breathing level"

- **One solution**: geophysical model to convert AOD to surface PM2.5 (van Donkelaar, Martin, etc)

$$PM_{2.5} = \eta\,(x, y, t) \times AOD_{BE}$$

GEOS-Chem

- **Another solution**: use statistical models to develop an empirical relationship between surface PM2.5 and column AOD

$$PM2.5 = f(AOD, T, Pr, NO2, CO)$$

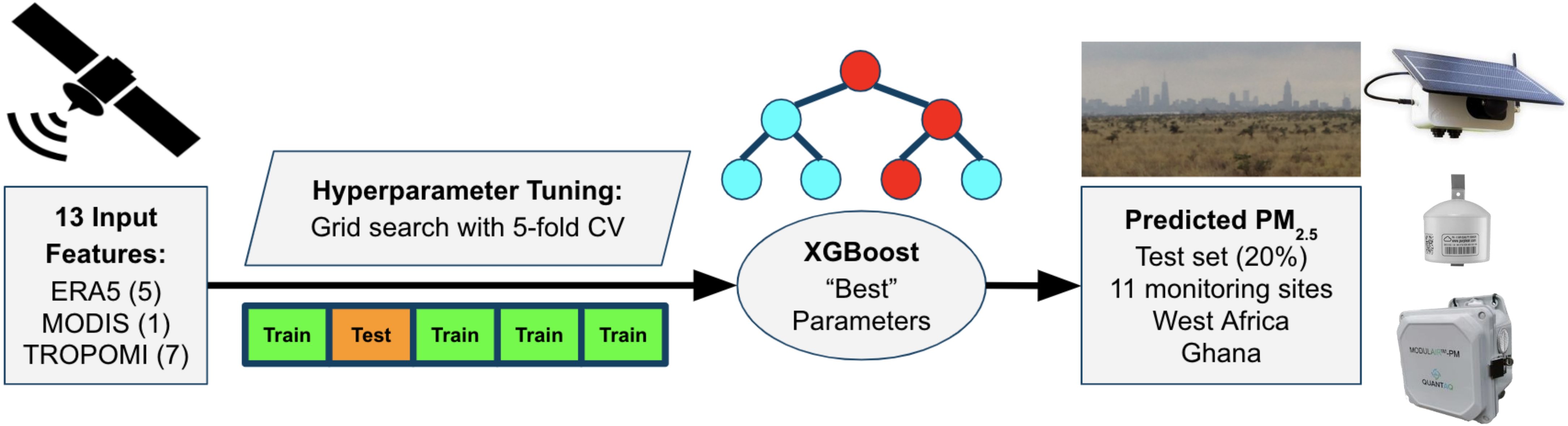
High Spatiotemporal Resolution Modeling of $PM_{2.5}$ in West Africa Using Satellite Data and Machine Learning

Goal: Estimate $PM_{2.5}$ at high spatiotemporal resolution (1 km^2, daily) in West Africa over the past 2 decades

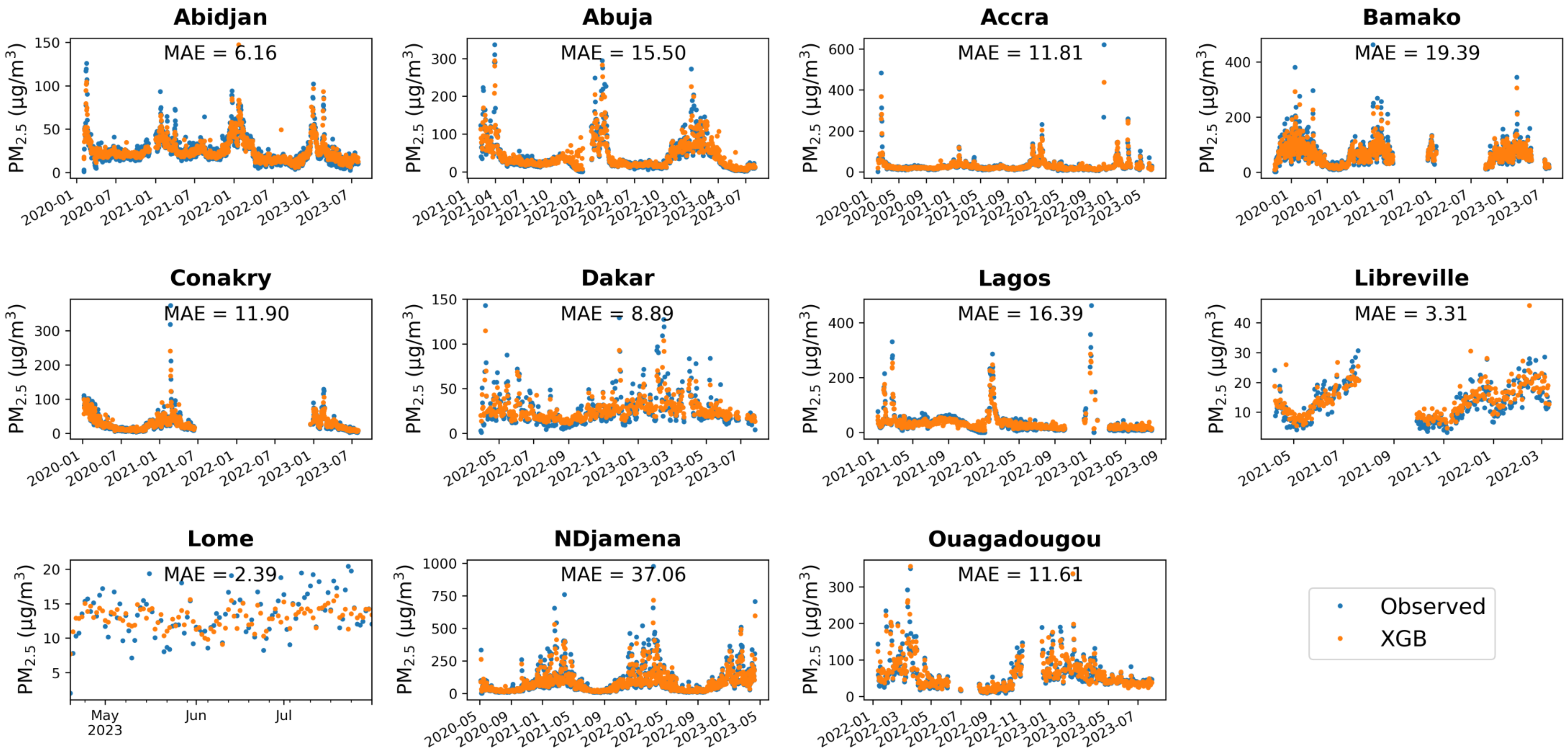
Data & Methods: Trained, tested, and fine-tuned a machine learning (XGBoost) model:

- ✓ $PM_{2.5}$ from 11 U.S. Embassies & Consulates monitoring sites
- ✓ AOD from MODIS MAIAC satellite retrievals
- ✓ 5 meteorological features from ERA5
- ✓ 7 trace gas or aerosol features from TROPOMI satellite retrievals

Applications: Epidemiological, policy, and environmental justice studies (Ghana)

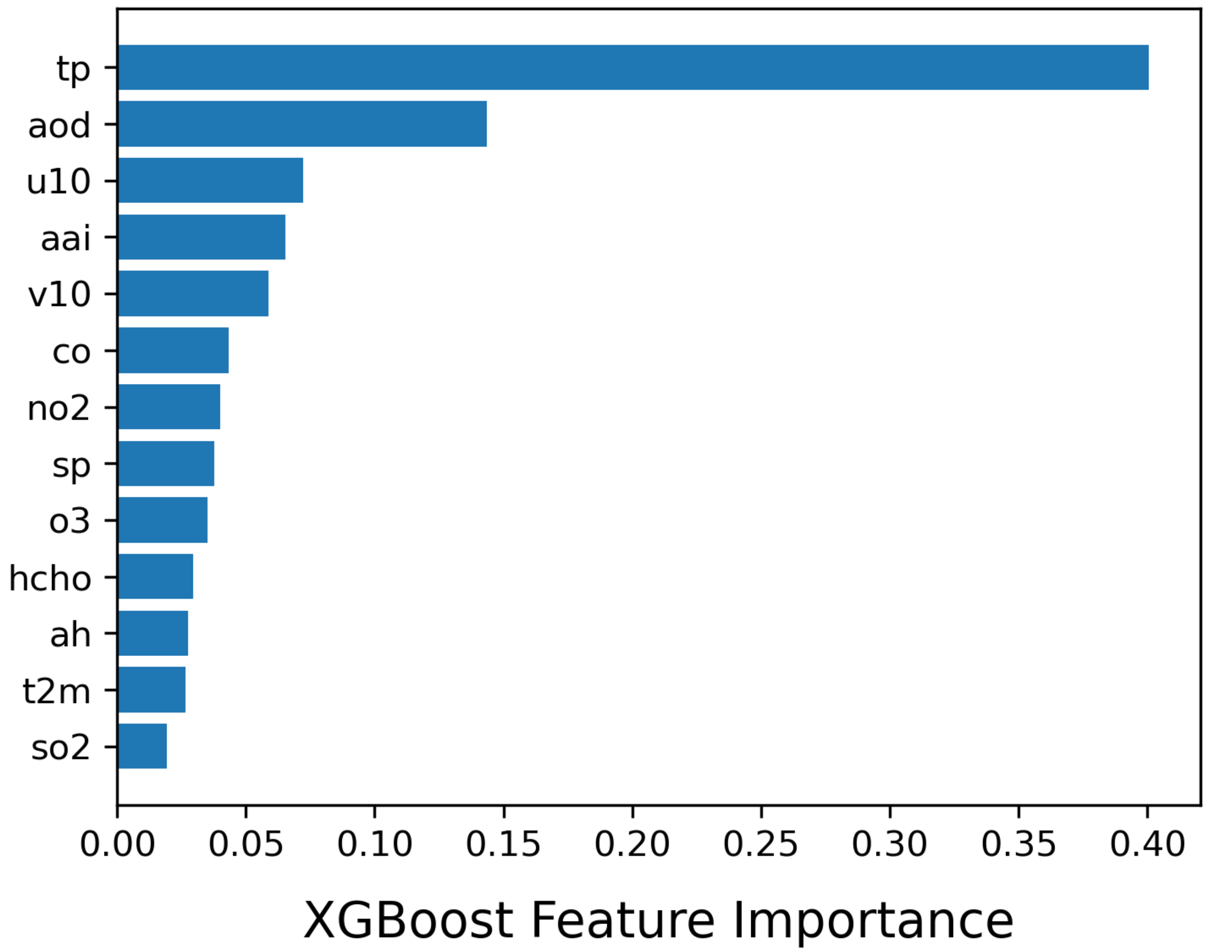


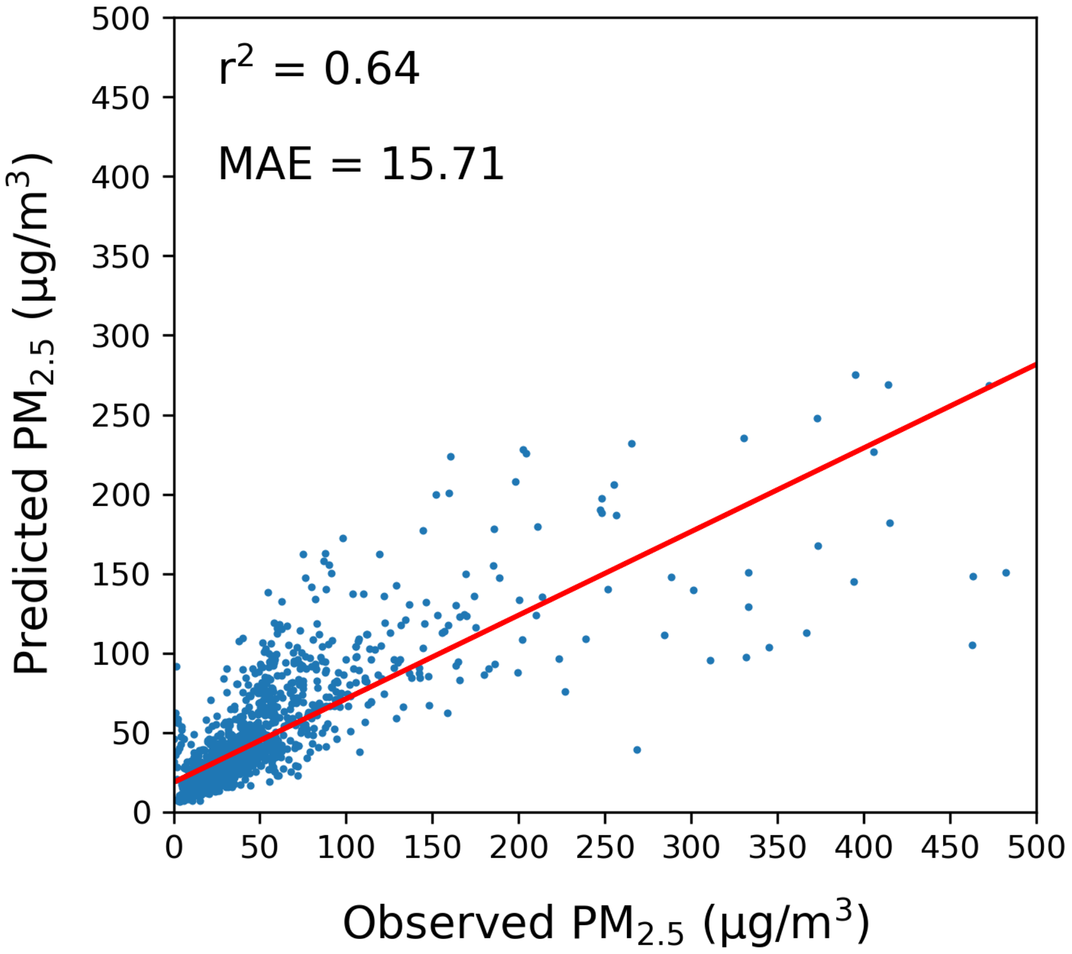
Predicted ML-satellite PM2.5 versus observations



- R^2 ~ 0.7
- Most important features: AOD, precipitation, UVAI, windspeed

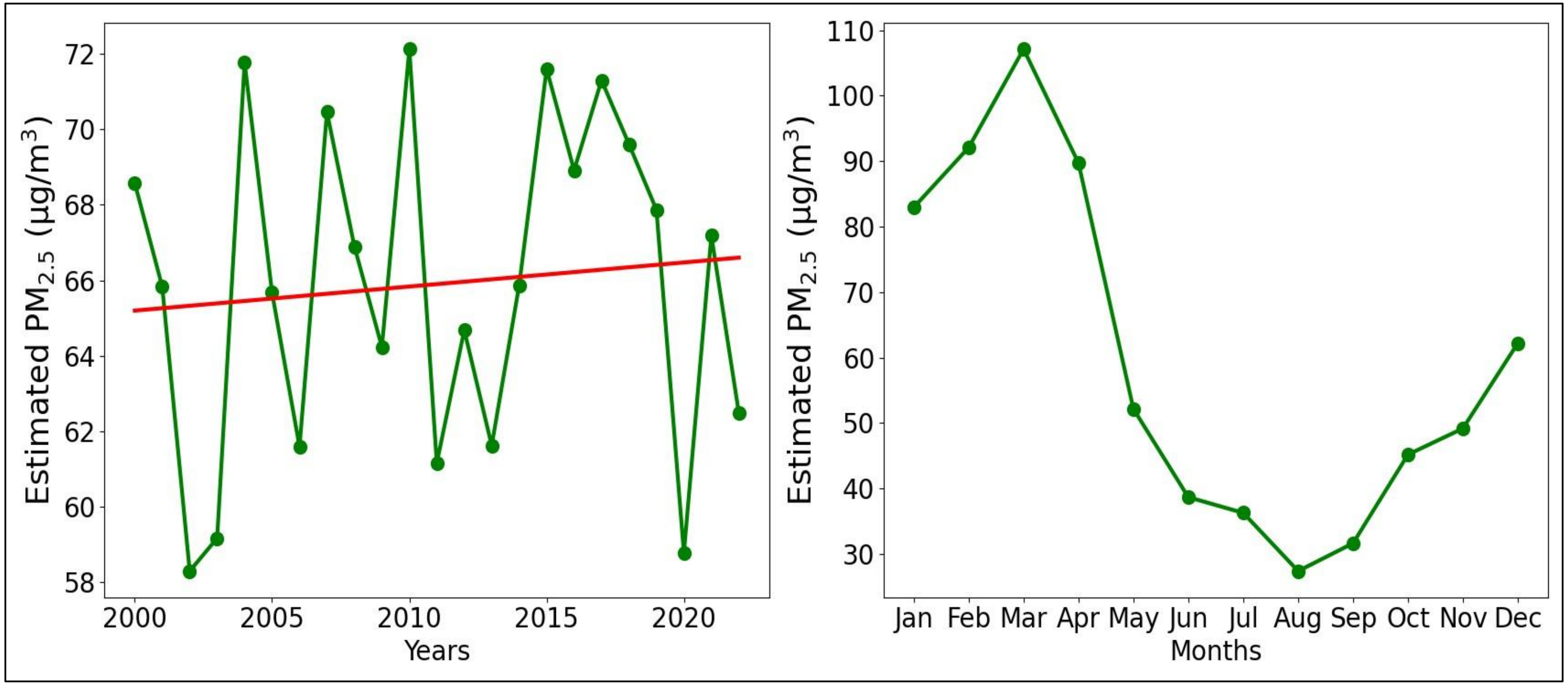
Predicted ML-satellite PM2.5 versus observations (daily)





Westervelt et al., 2024, in prep

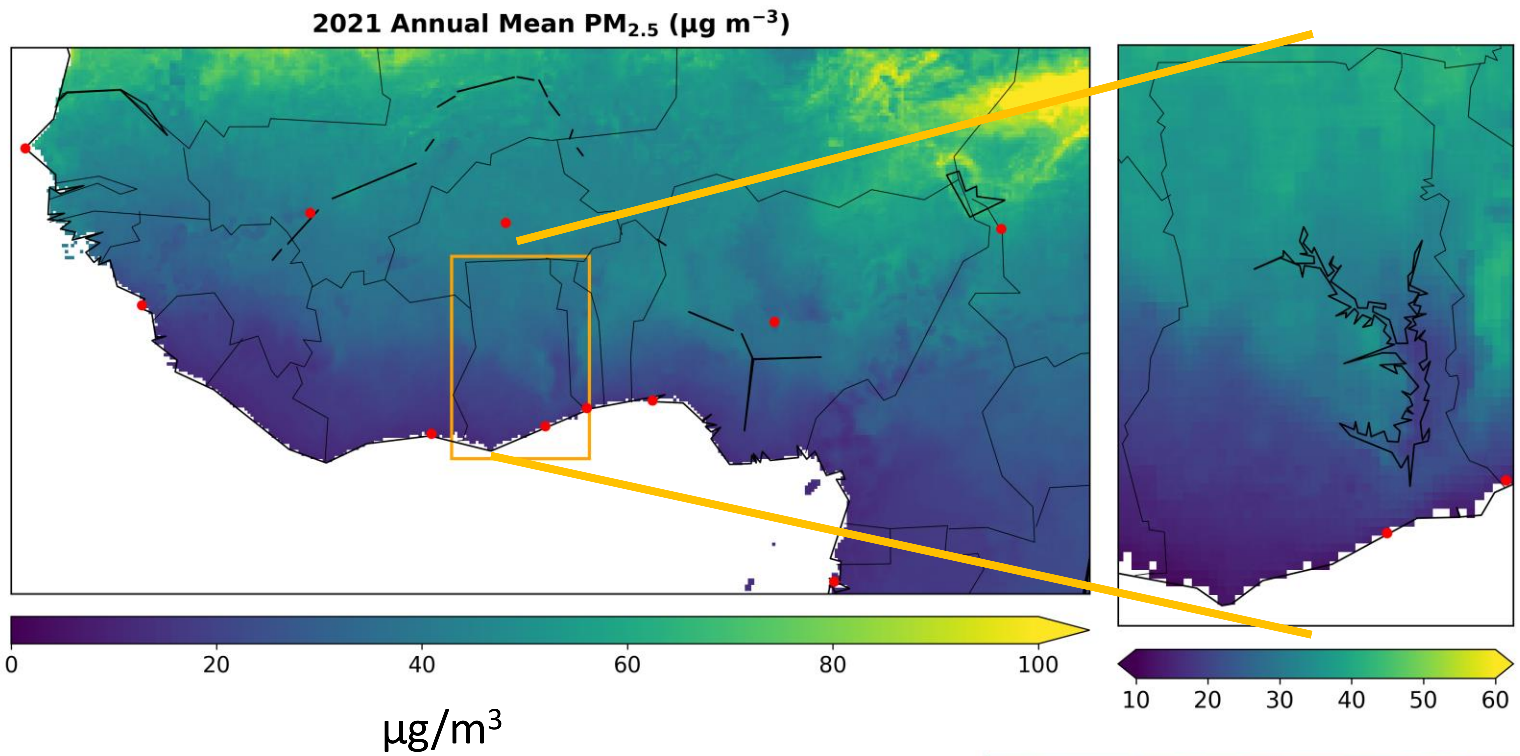
Results from Burkina Faso

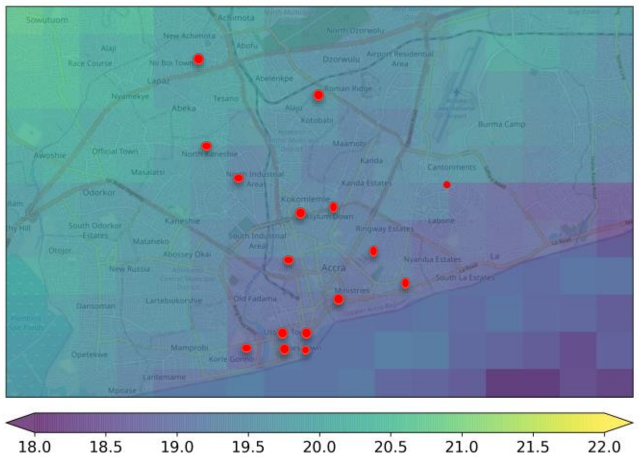


- XGBoost model has r^2 = 0.87 and nRMSE ~ 20%
- Evidence of slight increases in PM2.5 since 2000 (beginning of satellite record), with substantial interannual variability

Amooli et al., 2024, submitted to ES: Atmospheres

1 km^2 spatial resolution





- Can apply statistical relationship back through the satellite record to get a long term (2000-present), daily, 1km^2 dataset

Applications: Epidemiological, policy, and environmental justice studies (Ghana)

Icahn School of Medicine at Mount Sinai

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Summary

- Growing hybrid observational networks of surface PM2.5 over Africa and other traditionally data-sparse areas can be used to convert columnar AOD to “breathing level” PM2.5
  - Recommend that this is applied and benchmarked more in the Global South as novel datasets are coming online

- We demonstrate satellite-ML predictions of PM2.5, trained against air sensor data, with high r^2 (>0.7) and low error (20%)
  - Suitable for use in long-term trends, policy evaluation, epidemiology