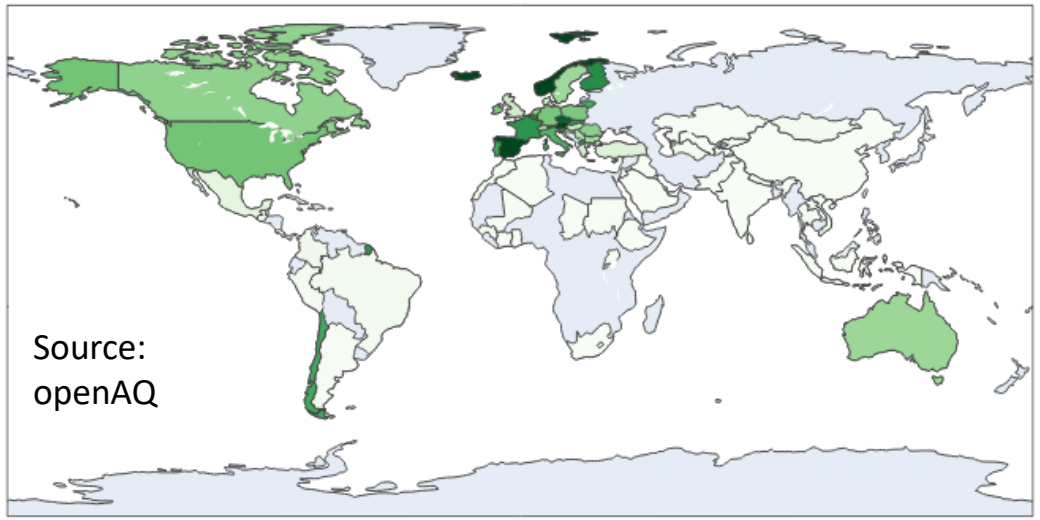
Getting useful, actionable data out of low-cost sensors: Examples from CAMS-Net and beyond

Prof. Dan Westervelt

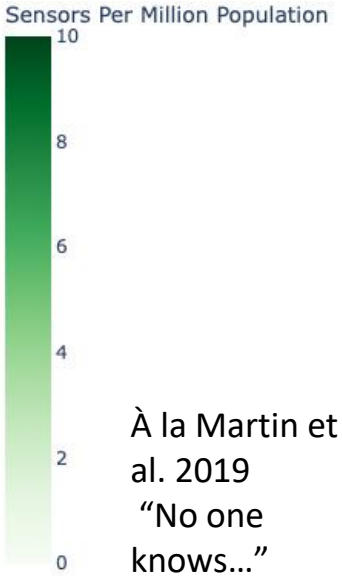
Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE

Why low cost sensors? Air quality data is sparce globally

Density of PM2.5 Reference monitors in 2023



Density of PM2.5 Low cost sensors in 2023



À la Martin et al. 2019 "No one knows…"

"Reference monitors" (FEM)

"Low cost sensors"

A global network for getting useful, actionable data out of low cost sensors

- Clean Air Monitoring and Solutions network (CAMS-Net)
- Create an international network of networks that provides a forum for exchange of knowledge, ideas, and data among scientists, decision-makers, citizen groups, the private sector, and other stakeholders towards the goal of improved usage and application of low-cost sensor (LCS) data for air quality
- Getting useful, actionable data out of LCS and exploring uses of this data for air quality modeling, satellite observations, policy recommendations, and health studies

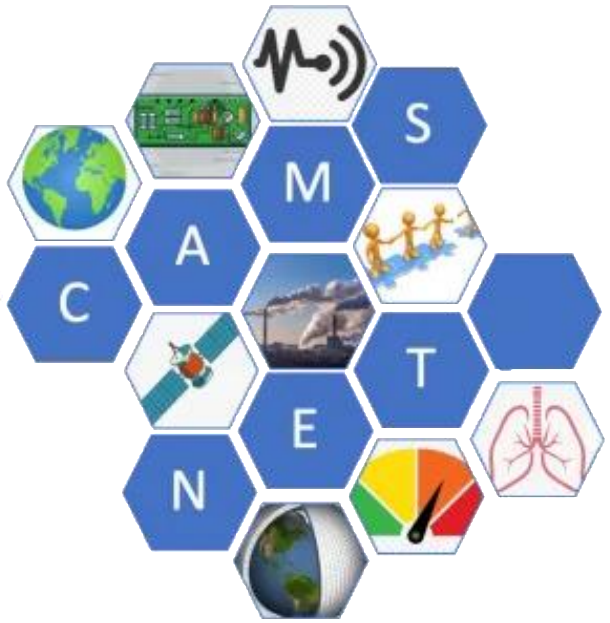
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

NSF

Carnegie Mellon University

WASHINGTON UNIVERSITY IN ST. LOUIS 1853

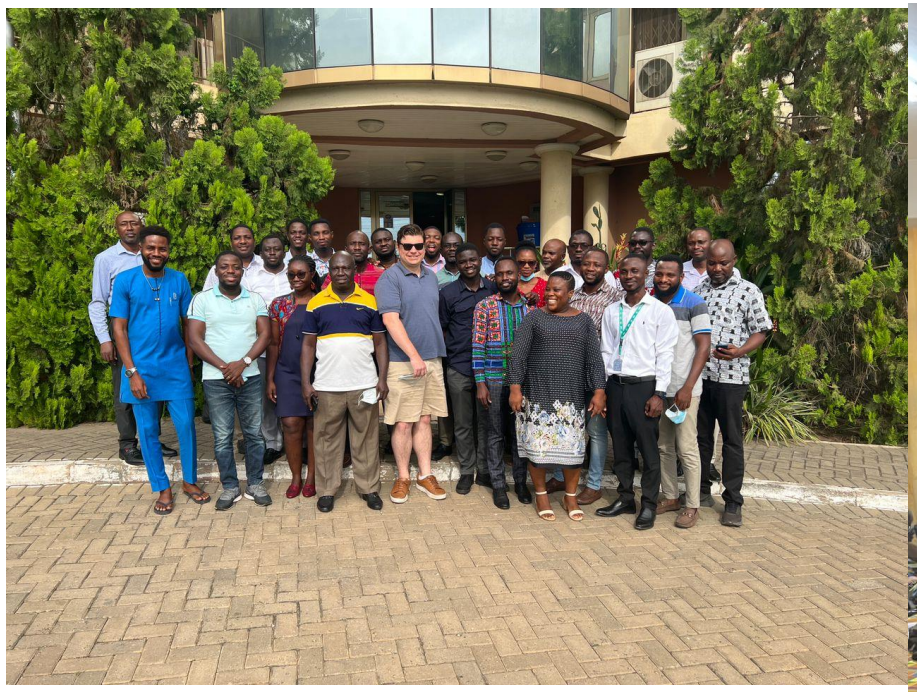
SPARTAN

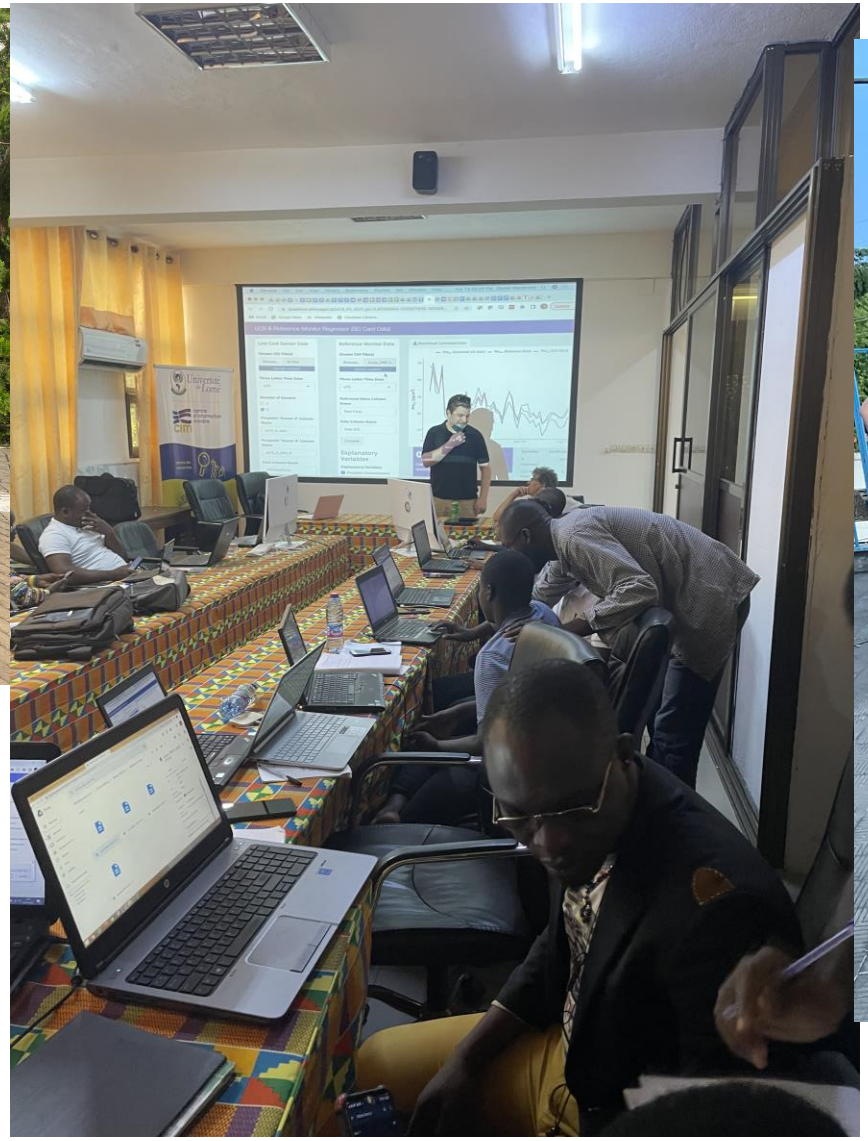


Clean Air Monitoring and Solutions Network (CAMS-Net)

- Website: www.camsnet.org
- Over 50 partner networks including universities, government agencies, non-profits, media, citizen science groups, private companies
- Global scope with emphasis on Global South, including South Asia and Africa
- >150 members

Capacity building and field work -- installation of FEM PM2.5 monitors and low cost sensors





International meetings

- May 10-13, 2022 as part of Air Sensors International Conference (ASIC)
- March 7-10, 2023, CAMS-Net + AfriqAir general meeting in Kigali, RW

Air Sensors International Conference

Pushing ahead: application and communication in science

NEW Parallel Symposium!

We are excited to announce a partnership with the Clean Air Monitoring and Solutions Network that will bring more air monitoring discussions based on their international network. This parallel symposium will take place as a daily session at the Pasadena Convention Center with all current ASIC sessions. Read the description below and submit an abstract through the portal.

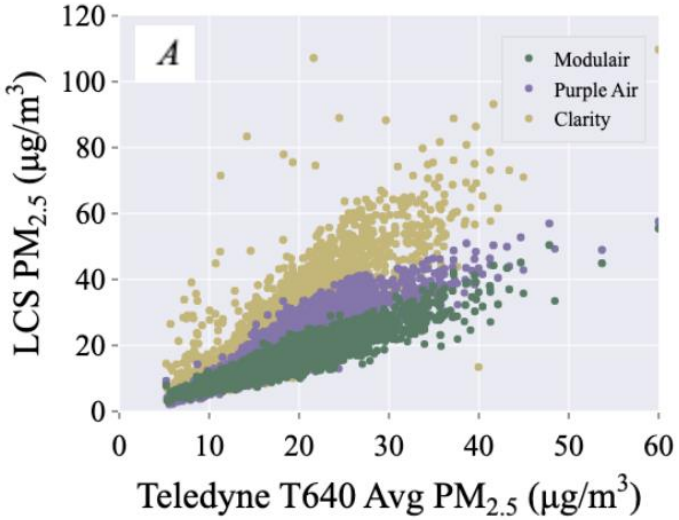
Clean Air Monitoring and Solutions Network: getting useful, actionable data out of low cost sensors for air quality action

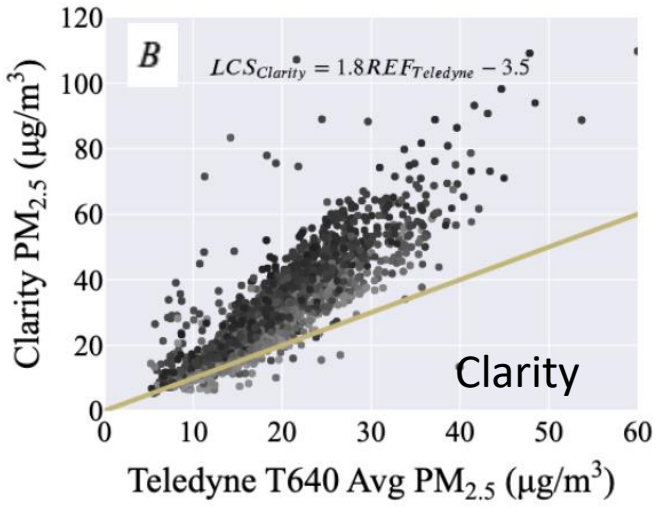
Locally led pilot projects, funded by CAMS-Net

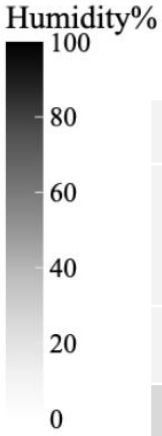
- Initiated 4 pilot projects for member networks to carry out air quality sensing project. Emphasis on forming new collaborations

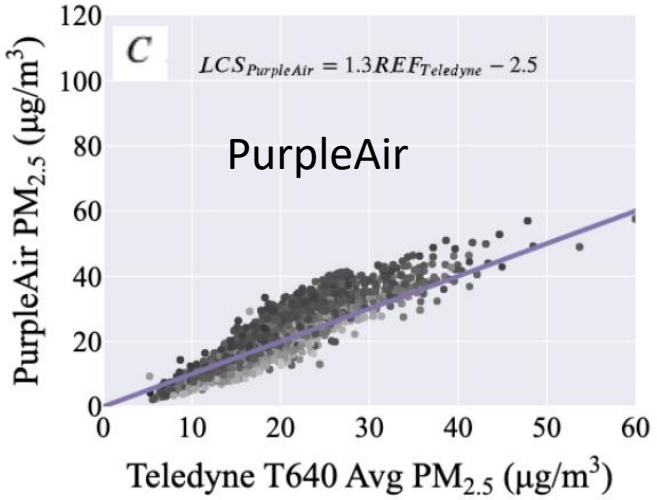
Member networks	**Project Title**
AfriqAir, Alioune Diop University (Senegal), AirQo, University of Douala (Cameroon)	DESIGN AND TESTING OF NETWORKS OF LCS FOR AIR POLLUTANTS MONITORING IN WEST AND CENTRAL AFRICAN CITIES
Clean Air One Atmosphere, Univeristy of Ghana, Clarity, AfriqAir	Exploring the robustness of LCS for understanding the impacts of location specific agricultural practices on local air quality.
AfriqAir, AirQo, ISGlobal, Manhica Center for Health Research	Expanding air quality monitoring, capacity, and health research in Mozambique
GMET, Ghana EPA, UGhana, Clean Air one Atmosphere	Schools Air Quality Outreach

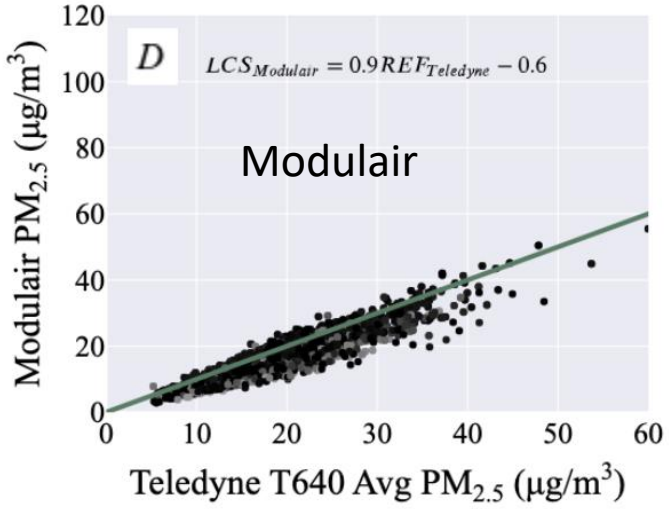
Performance between different low cost sensor vendors against FEM PM2.5







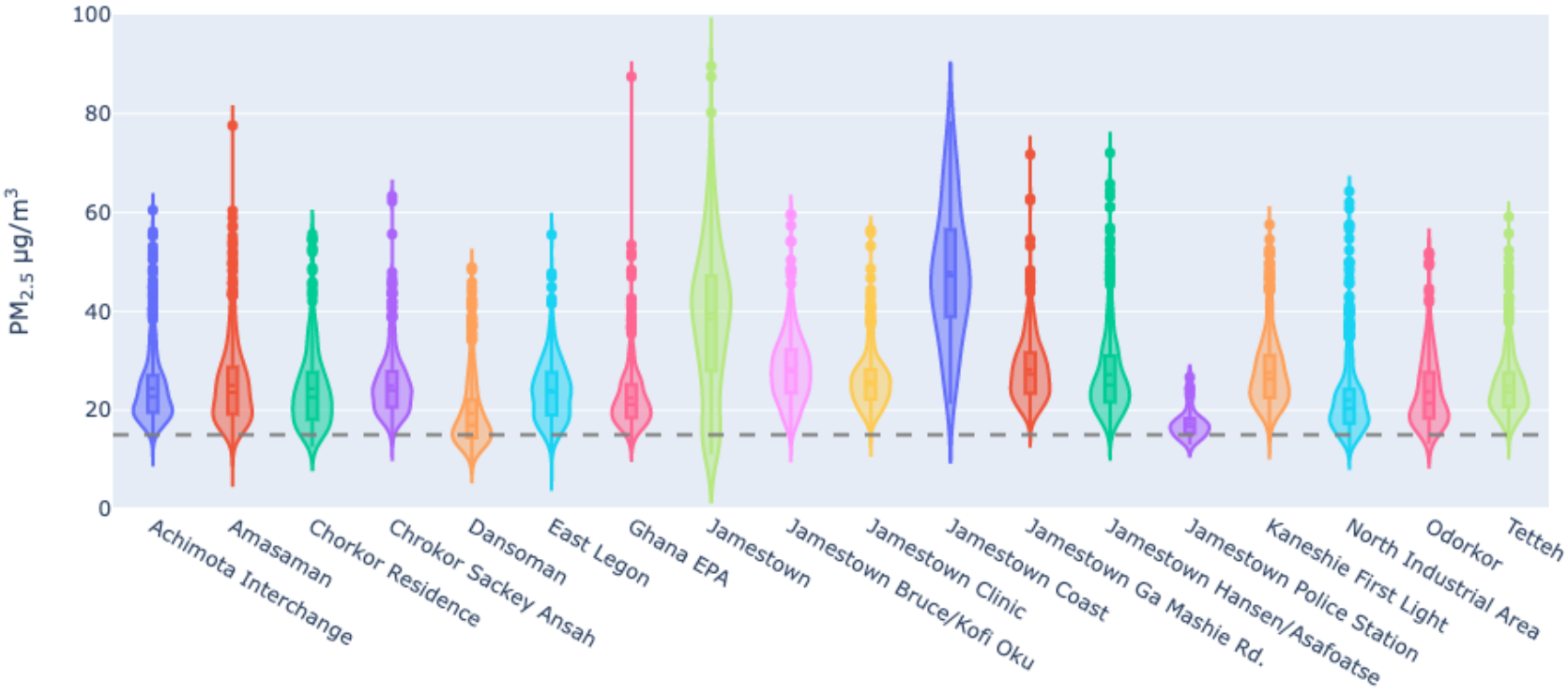


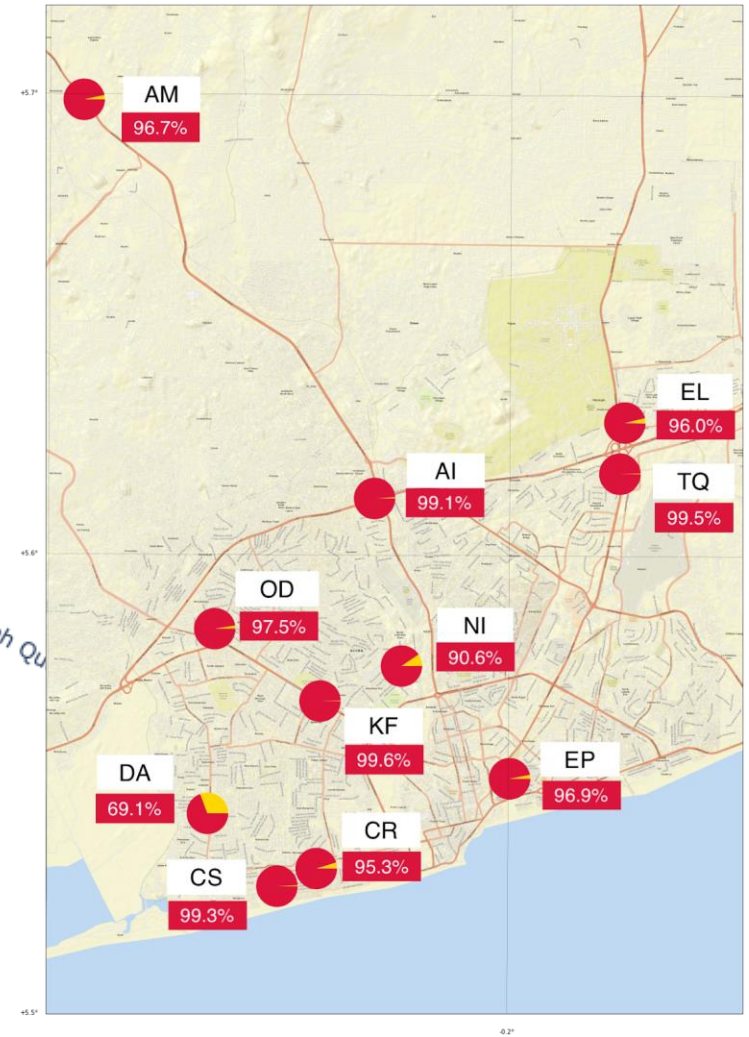


Model	Purple Air		Clarity		Modulair	
	R^2	MAE (µg/m^3)	R^2	MAE (µg/m^3)	R^2	MAE (µg/m^3)
Manufacturer-Reported	0.82	4.54	0.69	13.68	0.84	3.04
MLR	0.87	1.96	0.74	2.49	0.86	2.15
GMR	0.89	1.76	0.79	2.27	0.87	1.89

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

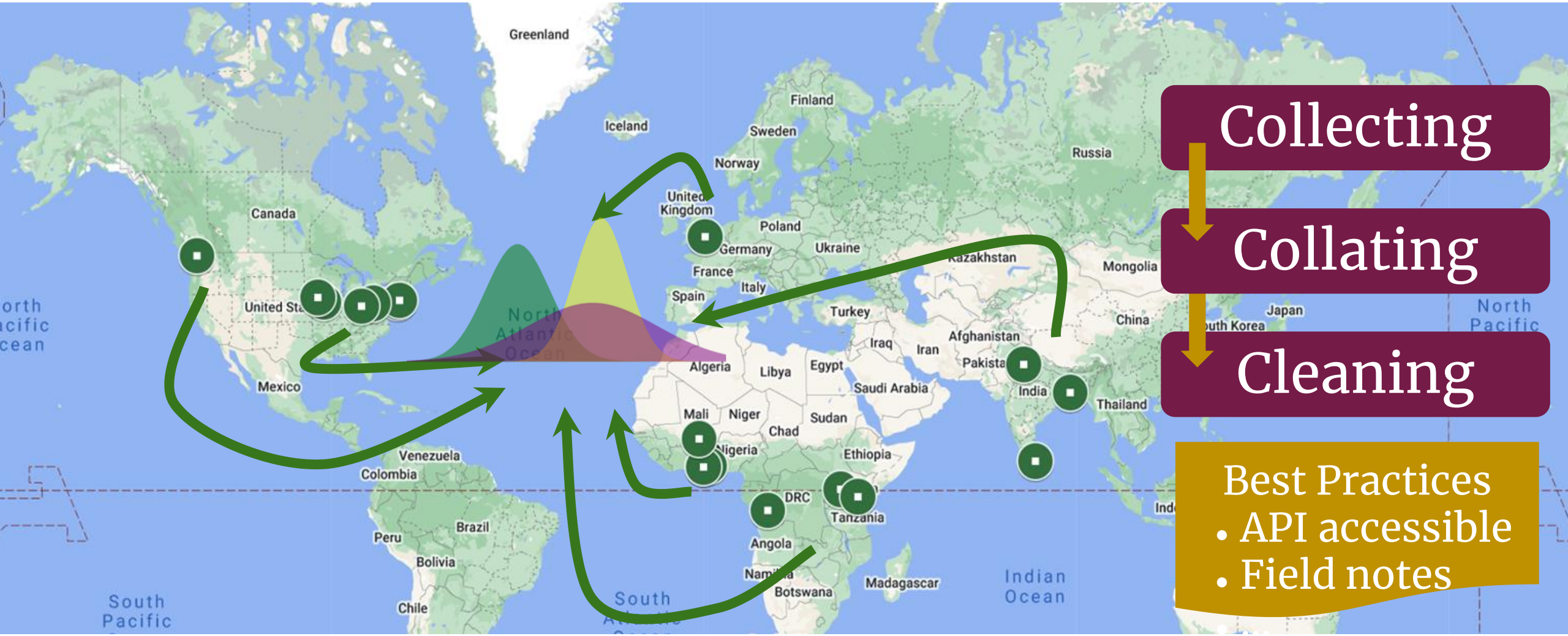
Application of the local correction factor to PM2.5 LCS networks in Accra, Ghana



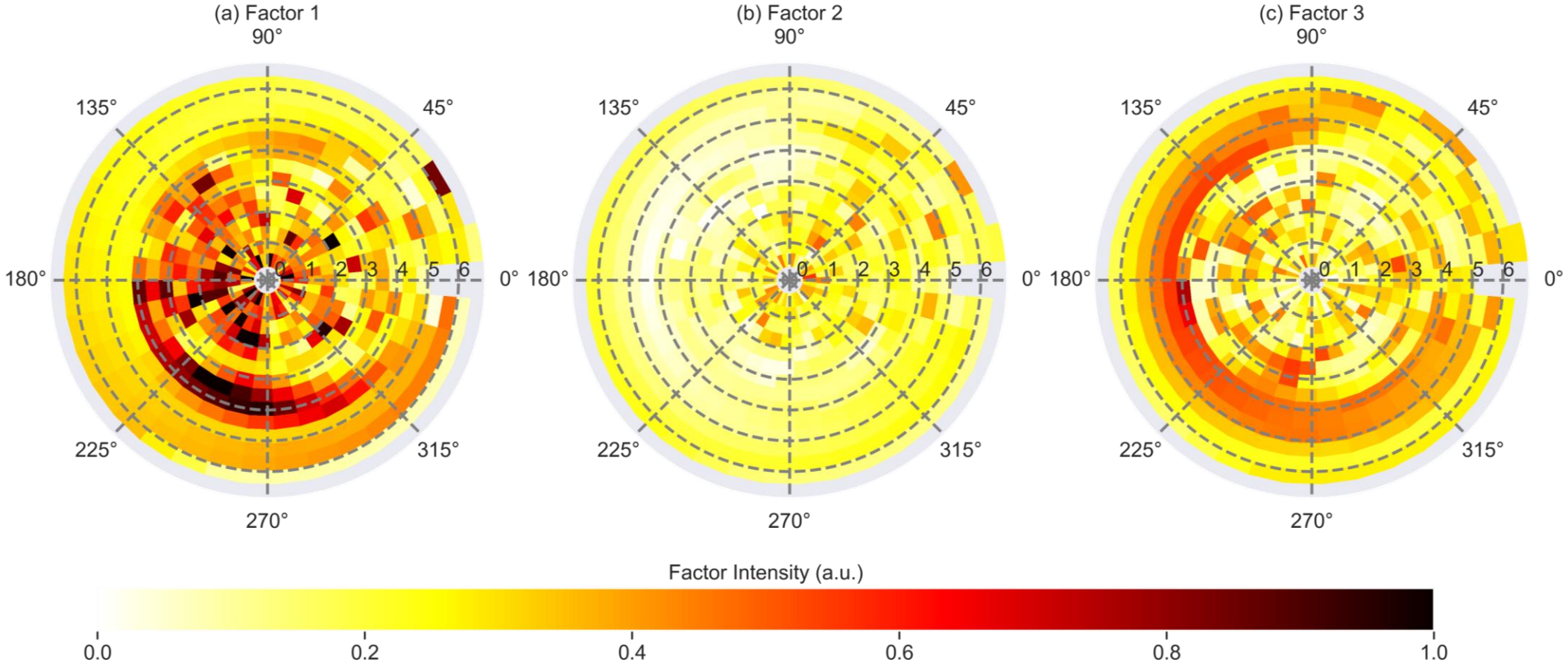


Raheja et al. (2023) *submitted to ES&T*

Building a more widely applicable (regional, global?) LCS correction factor model



Multipollutant sensor packages may be useful for source attribution (example from Kinshasa)



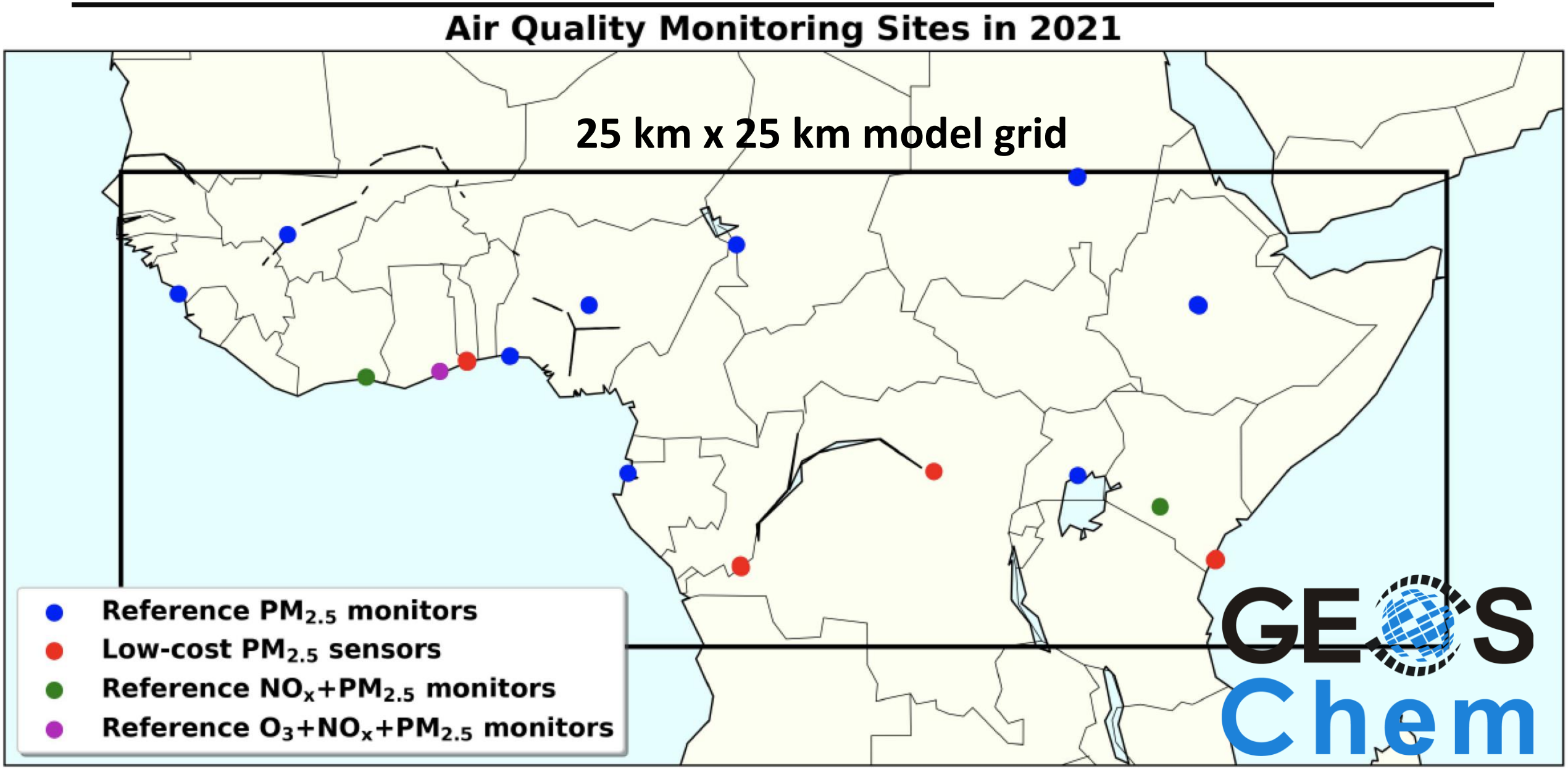
QUANTAQ

Non-negative matrix factorization of PM size distribution and EC gas sensor data reveals physically interpretable factors:

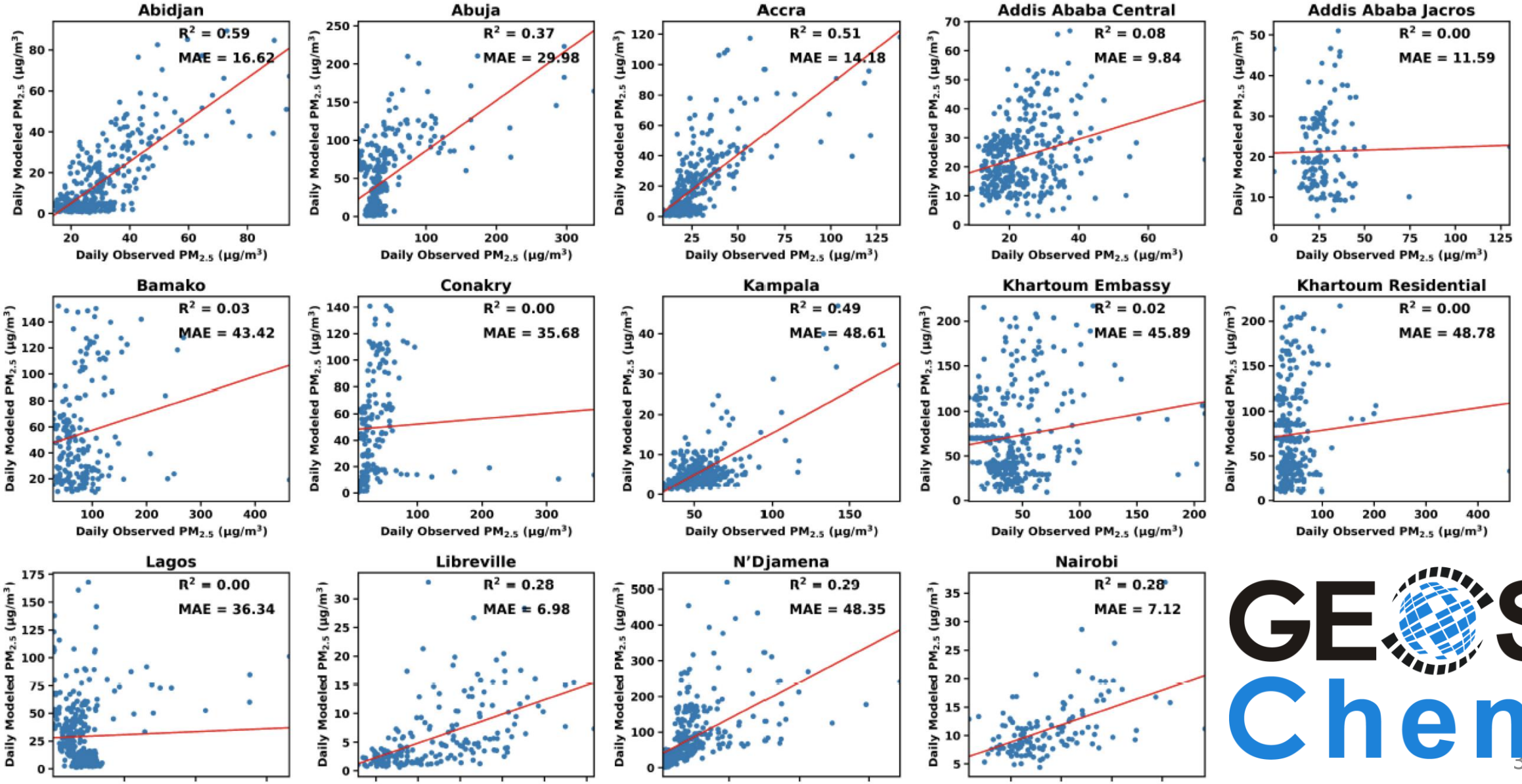
Factor 1: CO-dominated
Factor 2: Particle-dominated
Factor 3: NO2-dominated

Westervelt et al (2023) *in prep*

Well-calibrated low-cost sensors as model evaluation data

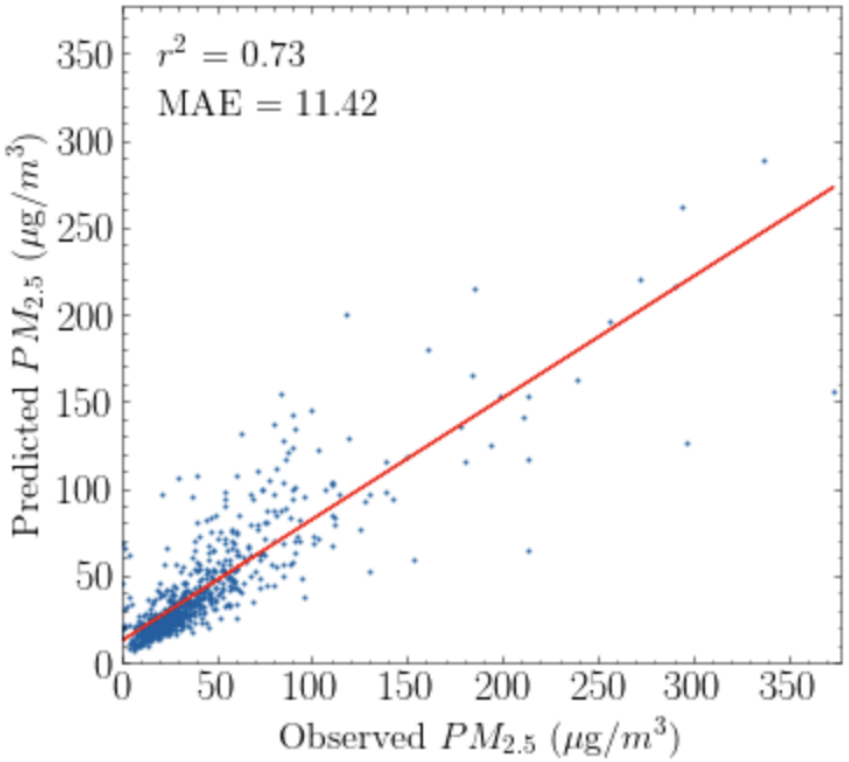


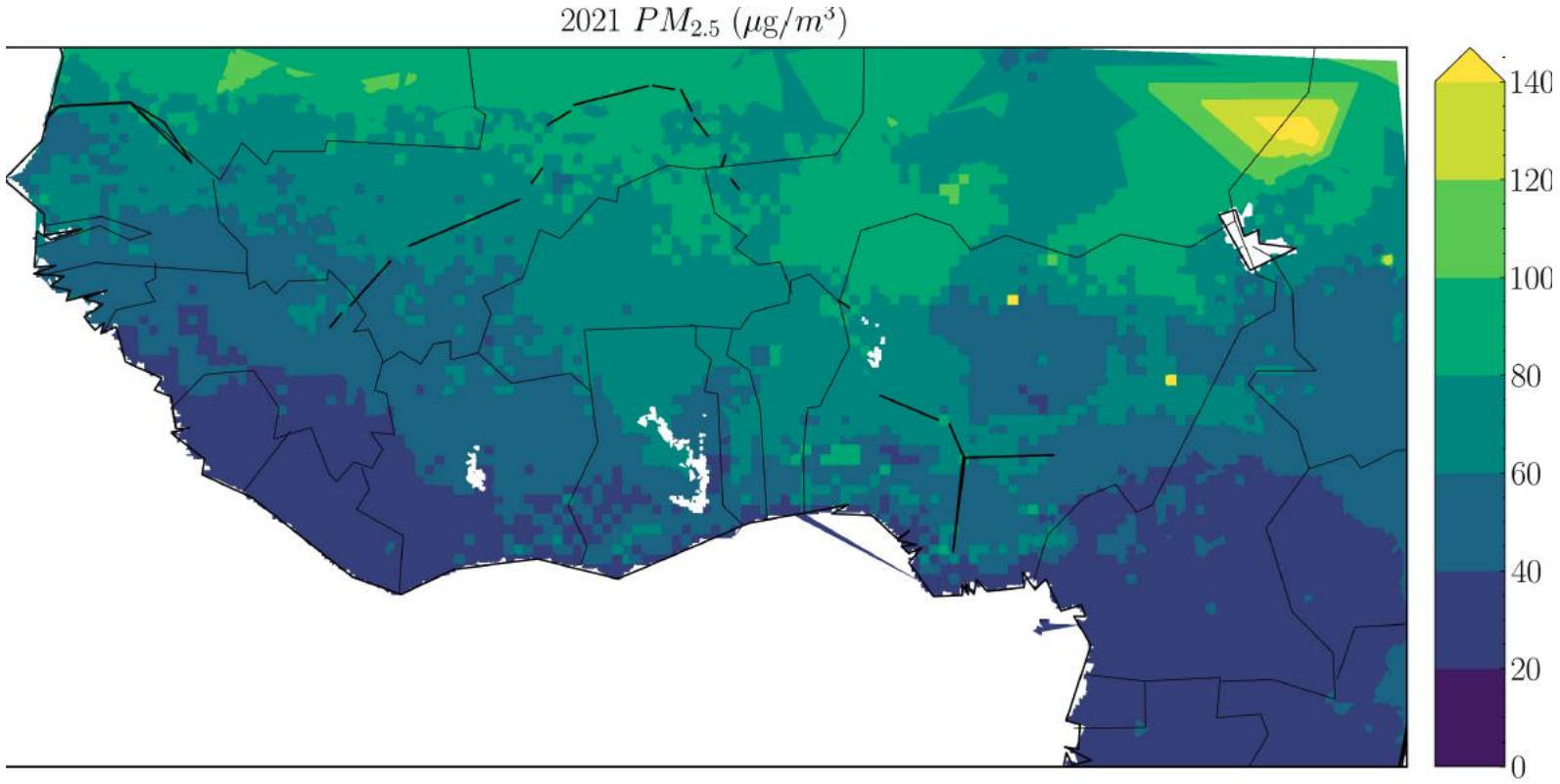
GC mostly underpredicts observed surface $PM_{2.5}$



Low-cost sensors as training data for AOD-to-$PM_{2.5}$ conversion

- XGBoost ML model
- Input: MAIAC AOD, TROPOMI NO2, CO, HCHO, SO2, precip, temp
- Output: PM2.5
- Training + testing data: Reference and calibrated low-cost sensor data





Summary

- CAMS-Net is an international network of networks that seeks to accelerate novel research into use and application of low cost sensors
  - SPARTAN is a partner network
- Low cost sensors have useful applications in surface monitoring networks, air quality modeling, training data for satellite-derived ML surface PM2.5 models, others
  - Data should be well calibrated, often locally
- Synergies with SPARTAN?
  - Co-location of sensors at SPARTAN sites

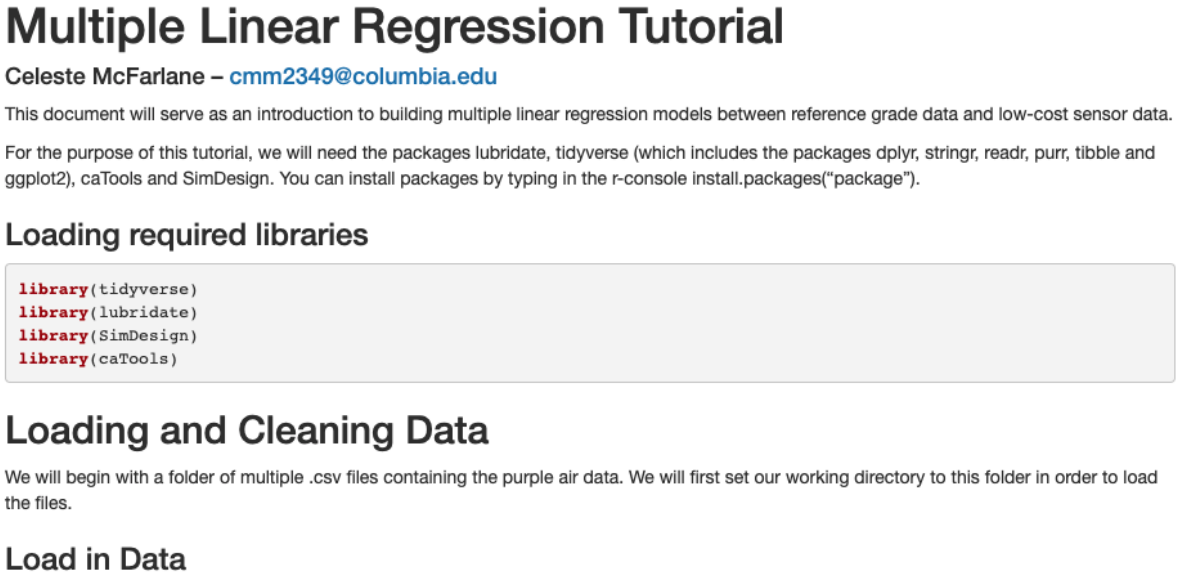
Extra slides

First African School on Atmospheric Sciences

- African School on Atmospheric Science
  - November 2022 in person in Morroco at UM6P
  - CAMS-Net will provide training materials and lecturers
  - https://asas2022.sciencesconf.org/

Current activities: capacity building

- Capacity building for academics and decision-makers
- Example: Calibration tool and tutorial for low cost sensors



Multiple Linear Regression Tutorial

Celeste McFarlane – cmm2349@columbia.edu

This document will serve as an introduction to building multiple linear regression models between reference grade data and low-cost sensor data.

For the purpose of this tutorial, we will need the packages lubridate, tidyverse (which includes the packages dplyr, stringr, readr, purr, tibble and ggplot2), caTools and SimDesign. You can install packages by typing in the r-console install.packages("package").

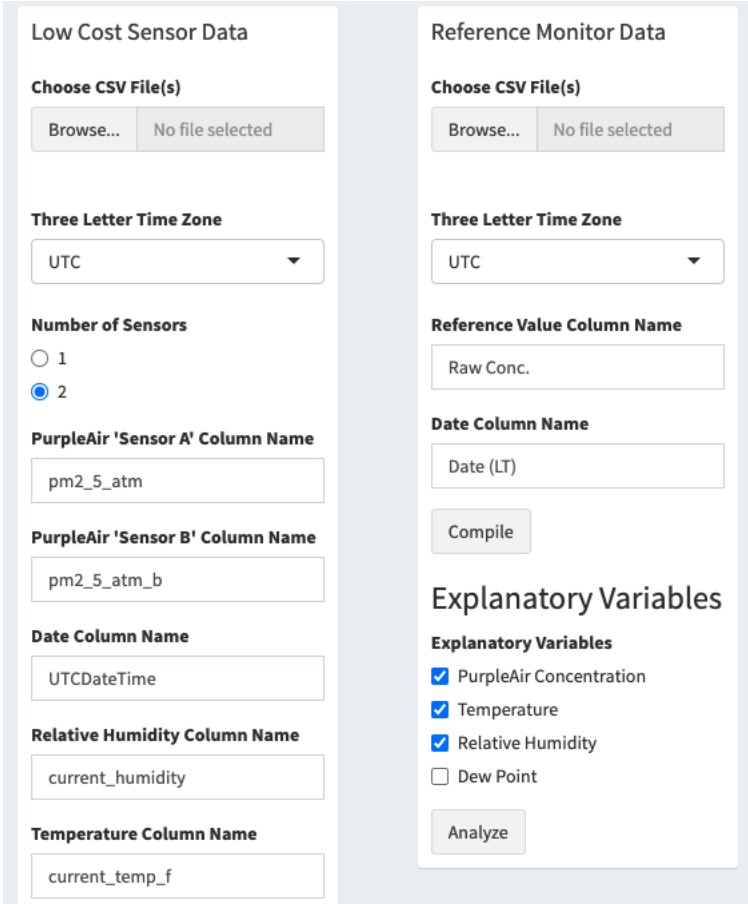
Loading required libraries

```
library(tidyverse)
library(lubridate)
library(SimDesign)
library(caTools)
```

Loading and Cleaning Data

We will begin with a folder of multiple .csv files containing the purple air data. We will first set our working directory to this folder in order to load the files.

Load in Data



Low Cost Sensor Data

Choose CSV File(s)

Browse... No file selected

Three Letter Time Zone

UTC

Number of Sensors

1

2

PurpleAir 'Sensor A' Column Name

pm2_5_atm

PurpleAir 'Sensor B' Column Name

pm2_5_atm_b

Date Column Name

UTCDateTime

Relative Humidity Column Name

current_humidity

Temperature Column Name

current_temp_f

Reference Monitor Data

Choose CSV File(s)

Browse... No file selected

Three Letter Time Zone

UTC

Reference Value Column Name

Raw Conc.

Date Column Name

Date (LT)

Compile

Explanatory Variables

Explanatory Variables

- [x] PurpleAir Concentration
- [x] Temperature
- [x] Relative Humidity
- [ ] Dew Point

Analyze

Both freely available as ShinyApps (written in R)

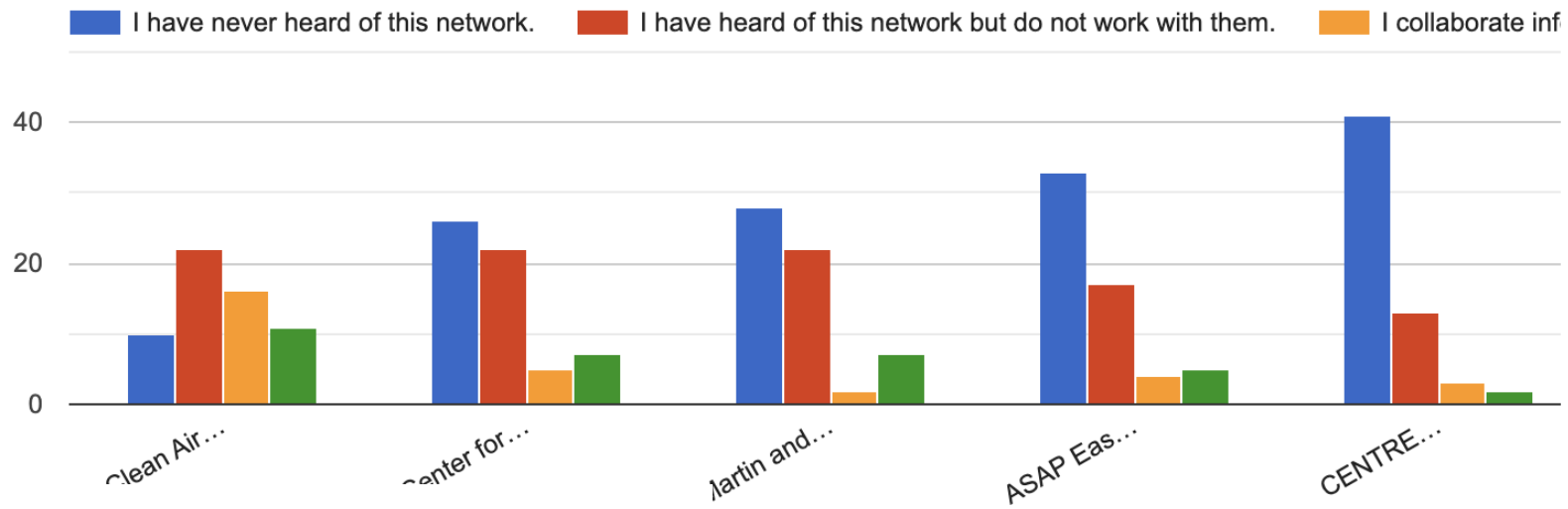
Current activities: networking

- ”Meet the Networks” series: monthly (or biweekly) events introducing networks to one another

Surveys

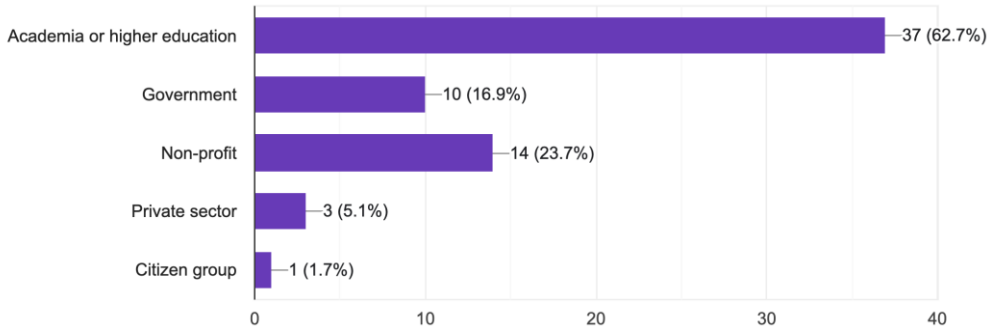
Please select which of the following statements best describes your relationship with each partner network within CAMS-Net.

Copy



Which sector do you work in?

59 responses



Are you an early career scientist? (student or within 7 years of terminal degree)

59 responses