

Air quality in Africa research at Columbia University

Prof. Dan Westervelt, Celeste McFarlane

Congo-Kinshasa: Paulson Kasereka Isevulambire, Dr. Raymond Sinsi
Lumbuenamo

Congo-Brazzaville: Arnold Murphy Elouma Ndinga

Kampala: AirQo Team

Lomé: Prof. Sabi Kokou

Accra: Prof Alison Hughes, Dr. Emmanuel Appoh

Nairobi: Prof. Michael Gatari, Prof. Godwin Opinde, Prof. Paul Njogu

USA: R Subramanian, Mike Giordano, V. Faye McNeill, Albert Presto,
Randall Martin

Many others...

Lamont-Doherty Earth Observatory

COLUMBIA UNIVERSITY | EARTH INSTITUTE

Westervelt group

Celeste McFarlane (CU undergrad)



Garima Raheja (PhD student)



Funding:

US Dept of State
National Science Foundation
Columbia University Center for Climate and Life
USAID

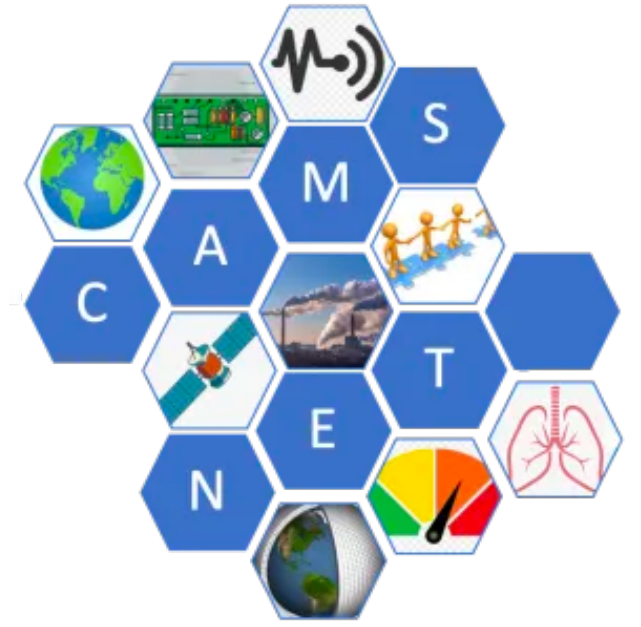


https://www.nsf.gov/awardsearch/showAward?AWD_ID=2020677&HistoricalAwards=false

Accelerating Research through International
Network-to-Network Collaborations (AccelNet)

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



Clean Air Monitoring and Solutions Network (CAMS-Net)

- Website: www.camsnet.org
- To get involved, reach out to danielmw@ideo.columbia.edu
- Example outputs from project: Tutorial on Multiple Linear Regression for air quality colocation study in R

The screenshot shows a web browser window with the URL `127.0.0.1:47078/?view=markdown`. The page title is "PurpleAir & Reference Grade Monitor Regressor" and it includes a "Source Code" link. The interface is divided into several sections:

- PurpleAir Data:** Includes a "Choose CSV File(s)" section with a "Browse..." button and "No file selected" text. Below are input fields for "Data Time Zone (3 Letter Code)" (set to "UTC"), "PurpleAir 'Sensor A' Column Name" (set to "pm2_5_atm"), "PurpleAir 'Sensor B' Column Name" (set to "pm2_5_atm_b"), "Date Column Name" (set to "UTCDateTime"), and "Relative Humidity Column Name" (set to "current_humidity").
- Reference Data:** Includes a "Choose CSV File(s)" section with a "Browse..." button and "No file selected" text. Below are input fields for "Data Time Zone (3 Letter Code)" (set to "UTC"), "Reference Value Column Name" (set to "Raw.Conc."), and "Date Column Name" (set to "date_UTC").
- Explanatory Variables:** Includes a section for "Explanatory Variables" with a checked checkbox for "PurpleAir Concentration".
- Corrected Data:** This section is currently empty.
- Model Summary:** This section is currently empty.

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`, `purrr`, `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

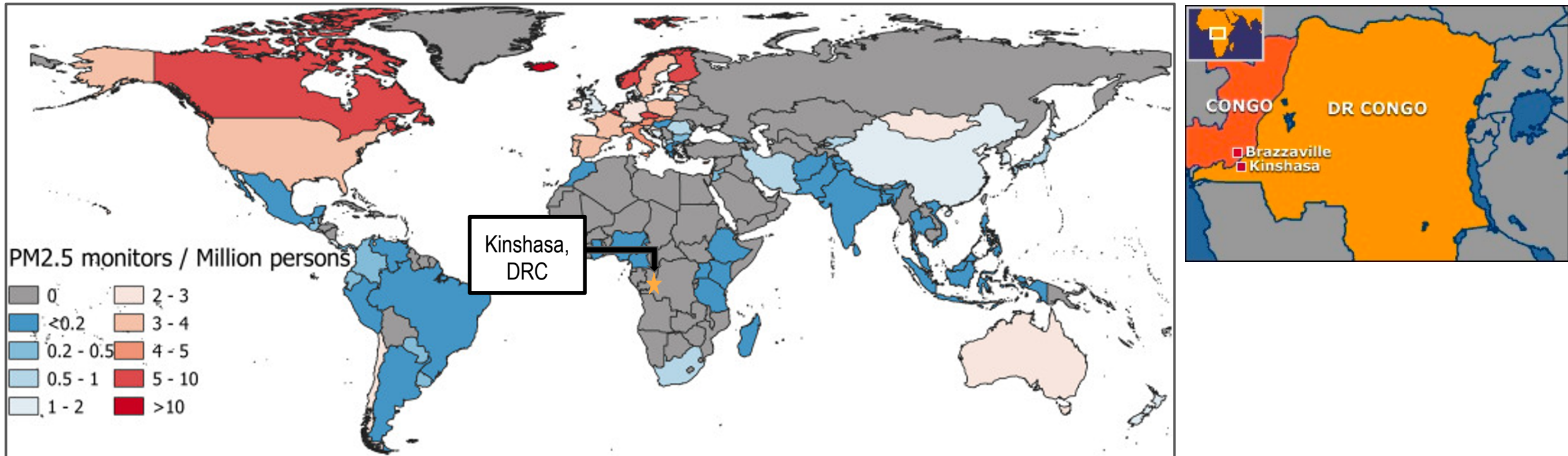
International Networking, Knowledge Sharing, and Capacity Building for Improved Air Quality in East Africa

- Scope of project is East Africa
 - Kenya, Uganda, Rwanda, Ethiopia
- Co-develop and co-implement an air quality management certificate program in at least one local university in each country
 - University of Nairobi, Kenyatta University, JKUAT, CMU-Africa, Addis Ababa University, Makerere University, others
 - Online in July 2021, in person soon after
 - Register here: <https://camsnet.org/air-quality-in-east-africa-certificate-program/>
- Co-develop and co-implement an air quality management plan
 - City governments and environmental protection agencies are involved in the project
- A locally owned air monitoring network for Mombasa, Kenya
 - LCS deployments in July 2021
 - BAM-1022 in Fall 2021
- With Prof. Solomon Bililign at NCAT



ECOTECH
NOx
analyzer

Air pollution data is sparse globally

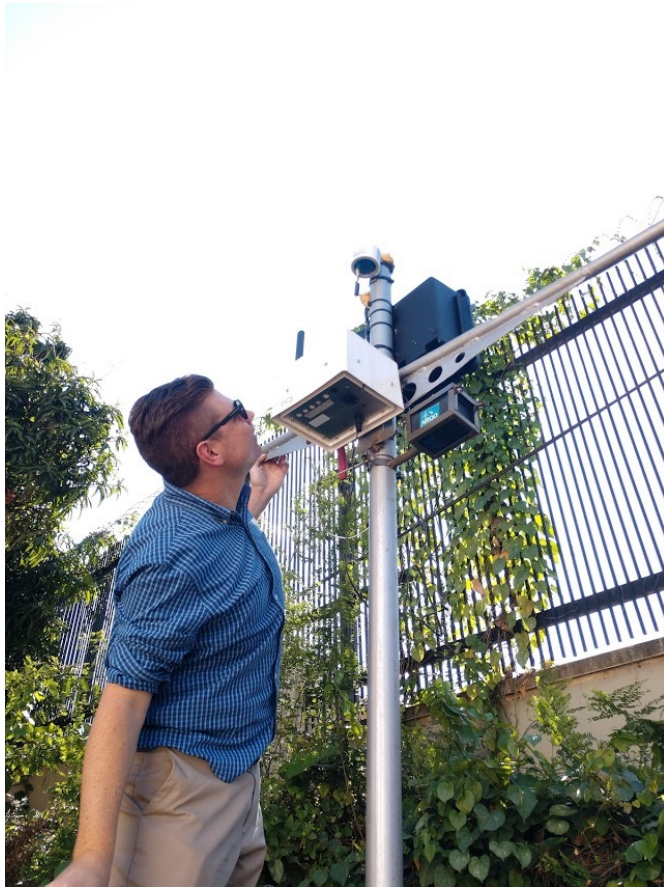


Martin et al. (2019)

- Kinshasa: population 14.3 million; Brazzaville: population 2.4 million
- No ambient air quality monitoring to date
- We deployed and calibrated a small network of 5 low cost air quality sensors starting in 2018, covering both cities

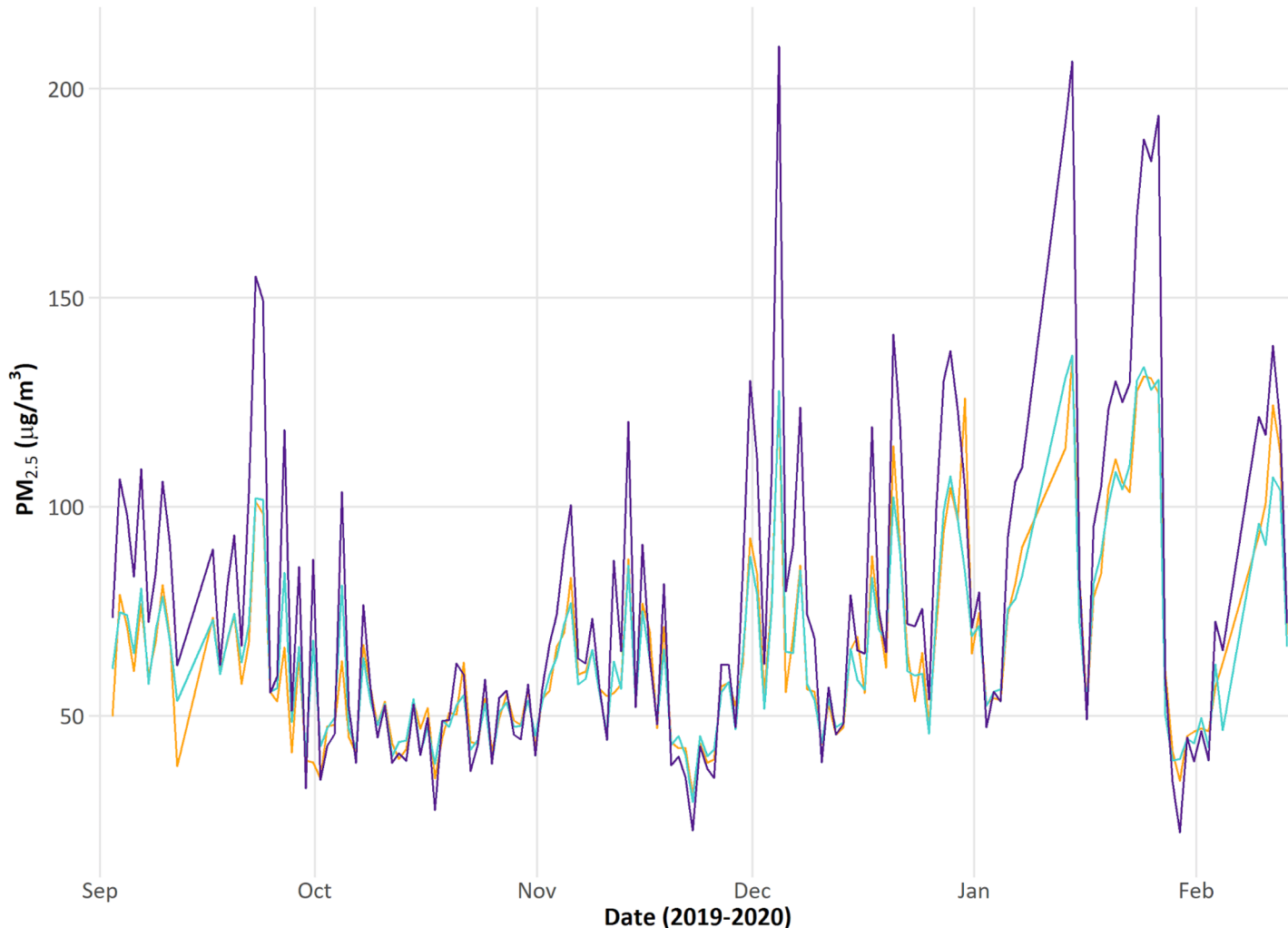
Kampala, Uganda

- Co-location of a PurpleAir at US Embassy starting in September 2019
- Several other sensor groups also co-located



Co-located Sensors Produce High Accuracy Model

— PM_{2.5} (Corrected Purple Air Data) — PM_{2.5} (Embassy Data) — PM_{2.5} (Purple Air Data)



- Multiple Linear Regression following the method of Malings et al. (2019)
- 75/25 training/test data split

$$PM_{2.5} = \beta_0 + \beta_1 \times \text{purpleair}PM_{2.5} + \beta_2 \times T(^{\circ}C) + \beta_3 \times RH(\%)$$

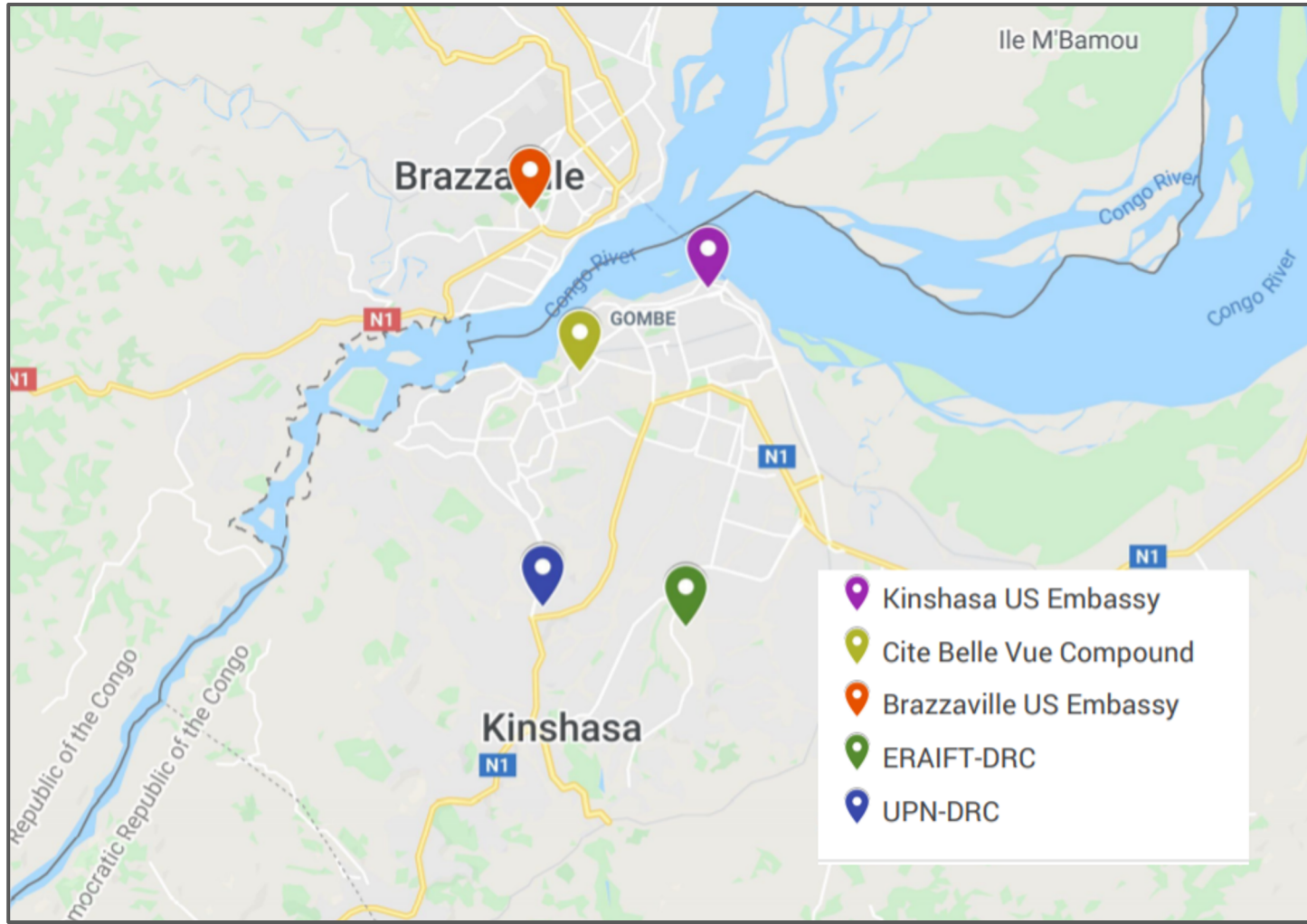
$$PM_{2.5} = 64.7 + 0.52 * [\text{PurpleAir}PM_{2.5}] - 0.234 * [T] - 0.59 * [RH]$$

Kampala colocation (Sep 2019 – Mar 2020)

$$PM_{2.5}^* = 64.7 + 0.52 * \text{purpleair}PM_{2.5} - 0.234 * T(^{\circ}C) - 0.59 * RH(\%)$$

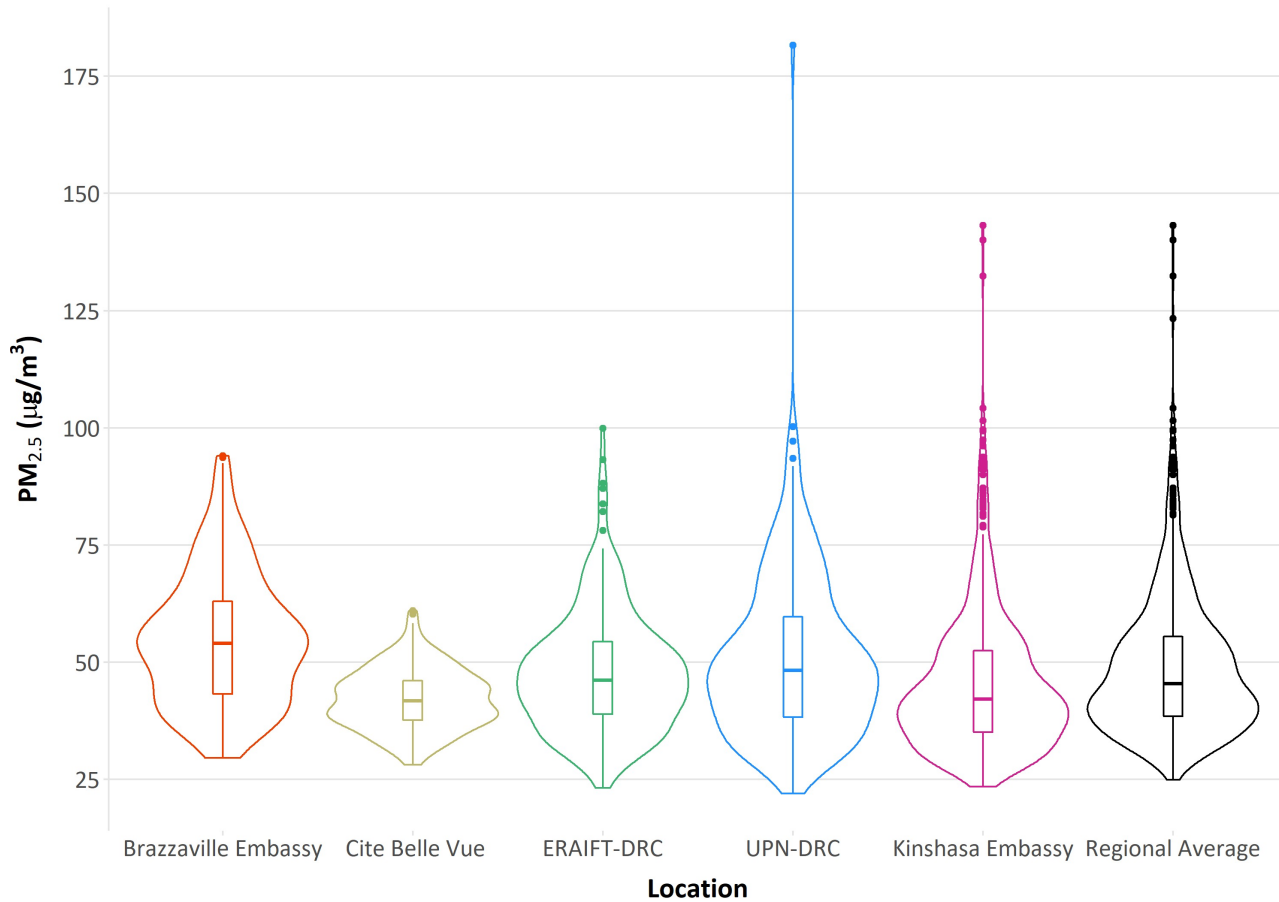
Model	Averaging time period	R ²	MAE (μg/m ³)
Raw PurpleAir Data	Daily	0.88	14.8
	Hourly	0.88	20.3
Multiple Linear Regression*	Daily	0.96	3.4
	Hourly	0.90	7.3
Random Forest	Daily	0.86	5.8
	Hourly	0.91	7.2

Five Low-Cost Sensors Deployed in Kinshasa, DRC and Brazzaville ROC



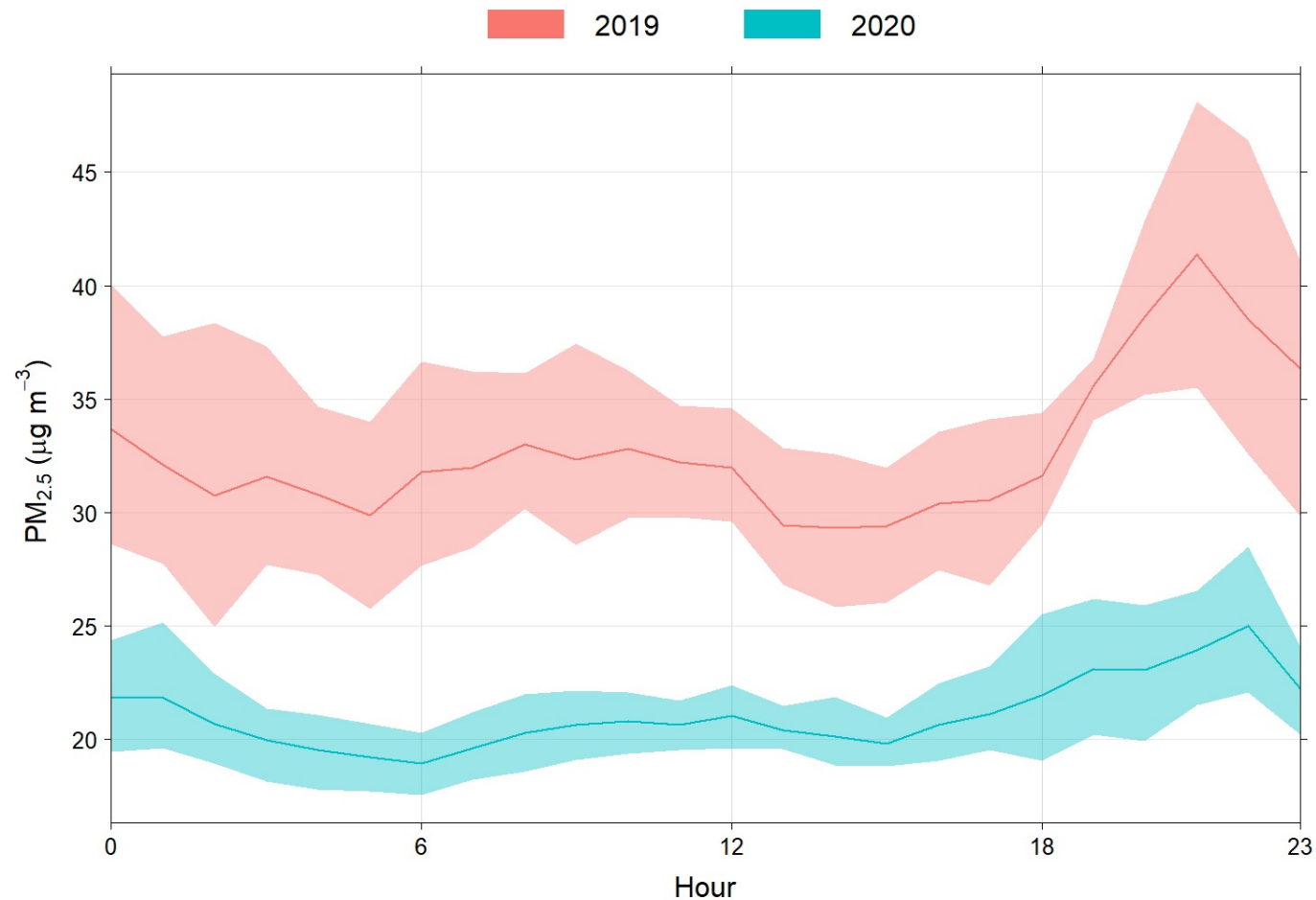
1. Kinshasa US Embassy, March 2018
2. Cite Belle Vue Compound, November 2019
3. Brazzaville US Embassy, February 2020
4. L'Ecole Régionale Postuniversitaire d'Aménagement et de Gestion intégrés des Forêts et Territoires tropicaux (ERAIFT), November 2019
5. L'université Pédagogique Nationale, November 2019

Median Daily PM_{2.5} Values are Quadruple the WHO Annual Mean Guideline



- 5-site average median daily PM_{2.5} value is 42 µg m⁻³
- Highest median and quartile ranges located at UPN-DRC and Brazzaville Embassy
- Excluding CBV, “long tails” with nonzero density near 100 µg m⁻³, indicate a high frequency of poor air quality
- Discrepancies can be described by variation in location type (residential vs educational vs diplomatic)

COVID19 Lockdown Lowers Mean Hourly PM_{2.5} By 15 μg m⁻³



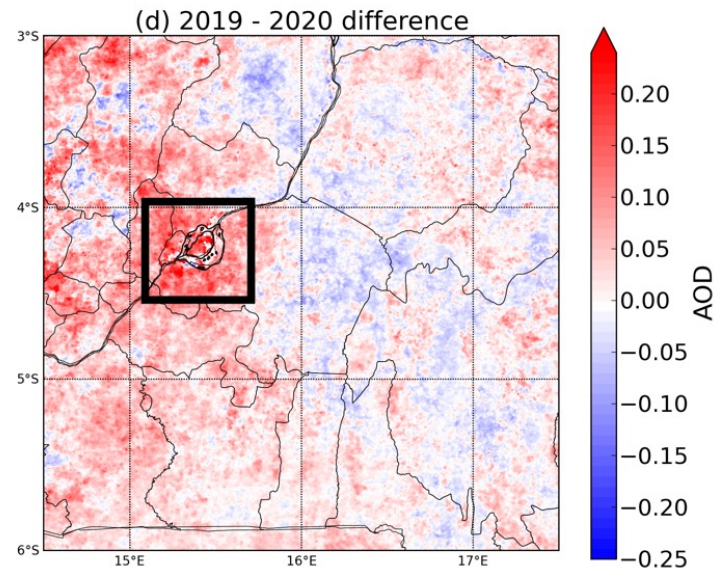
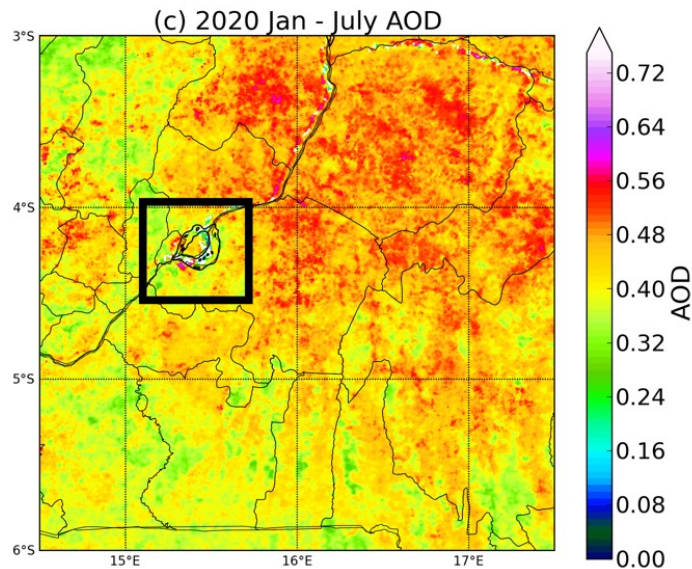
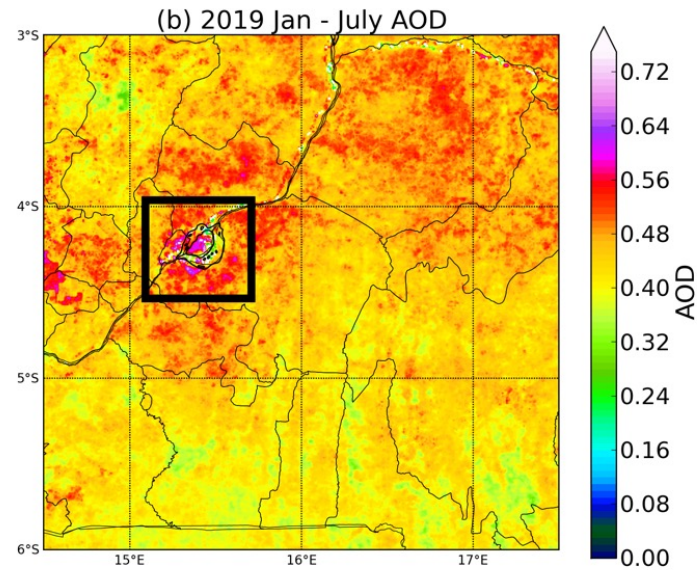
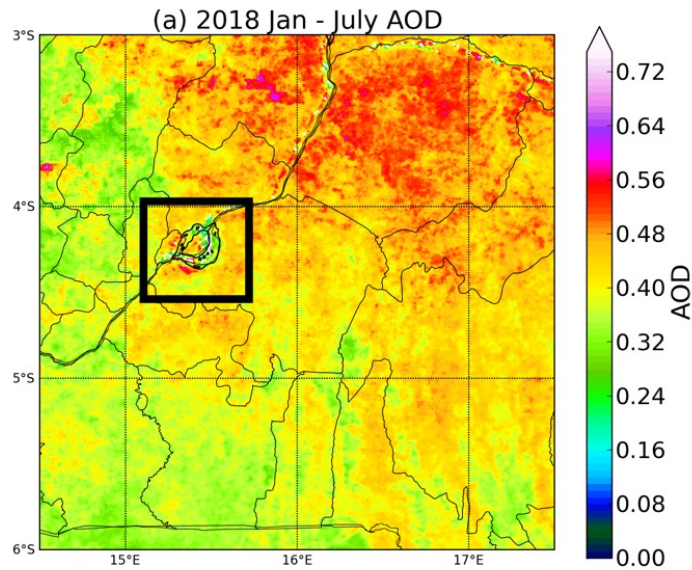
- Evening PM_{2.5} peak from 2019 flattened in 2020
- Average and maximum hourly PM_{2.5} increased between April 2018 & 2019
 - 32.18 vs 36 μg m⁻³ & 98.42 vs 103.5 μg m⁻³
- Average and maximum hourly PM_{2.5} decreased between April 2019 & 2020
 - 36 vs 21.28 μg m⁻³ & 98.42 vs 40.02 μg m⁻³
- Similar decrease observed in 1-km aerosol optical depth data over the region

Summary

- Air pollution data is sparse in Africa, though air quality seems to be poor. Several new projects involving capacity building and new research on LCS will aim to improve data quality and spur innovative solutions
- Deployed a 5 node LCS network in Kinshasa and Brazzaville, the first ever ambient air quality data in these cities
- Corrected towards FEM using surface-based calibrations from a collocation in Kampala using MLR
 - MAE reduced from $12 \mu\text{g m}^{-3}$ to close to zero
- PM_{2.5} in the Kinshasa area exceeds WHO guidelines by a factor of 4
 - Network wide mean of $45 \mu\text{g m}^{-3}$
- COVID-19 restrictions lowered PM_{2.5} by about $15 \mu\text{g m}^{-3}$

Extra slides

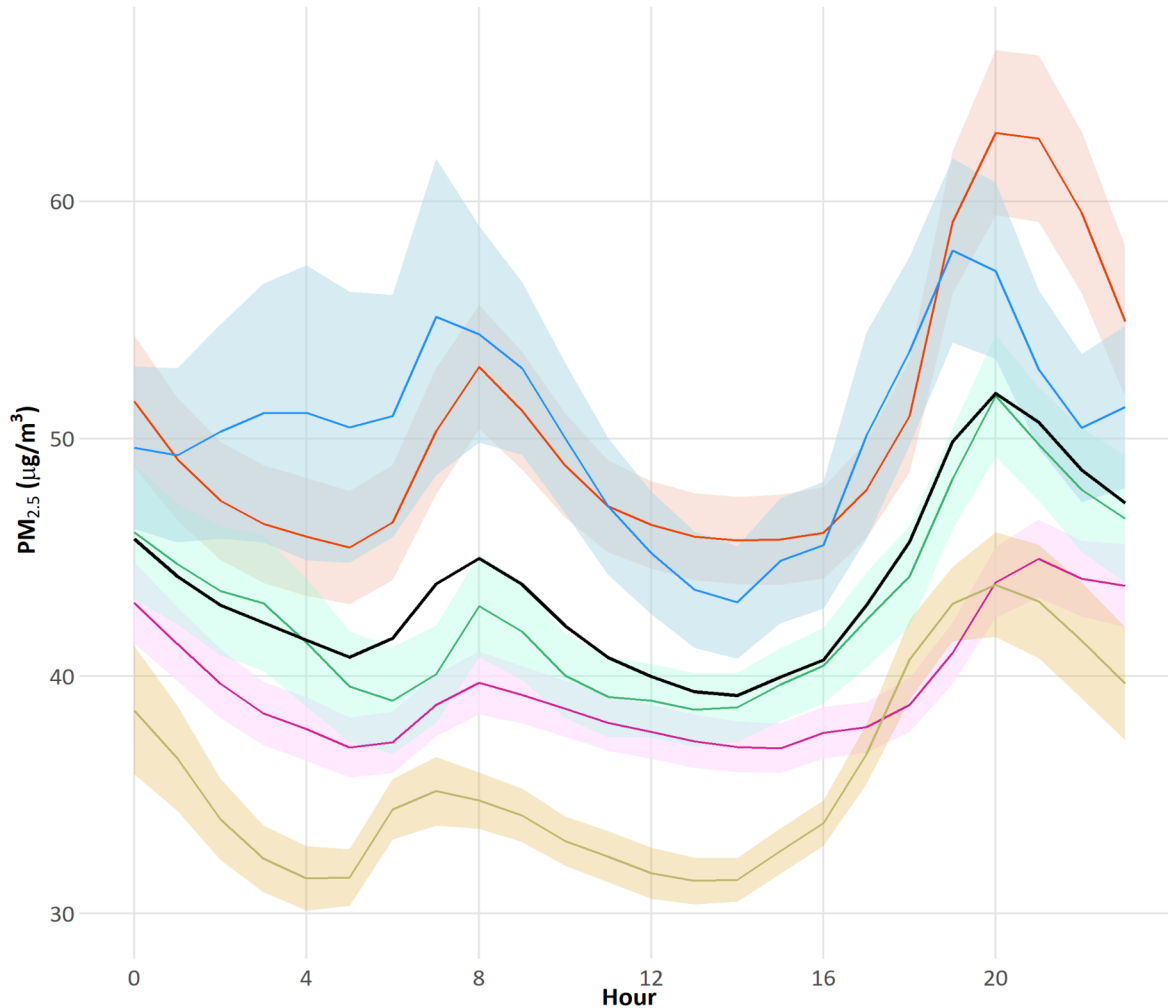
1-km Level 2 MAIAC AOD retrievals over Kinshasa



Diurnal Peaks Coincide With Evening Vehicle Traffic and Cooking

Location

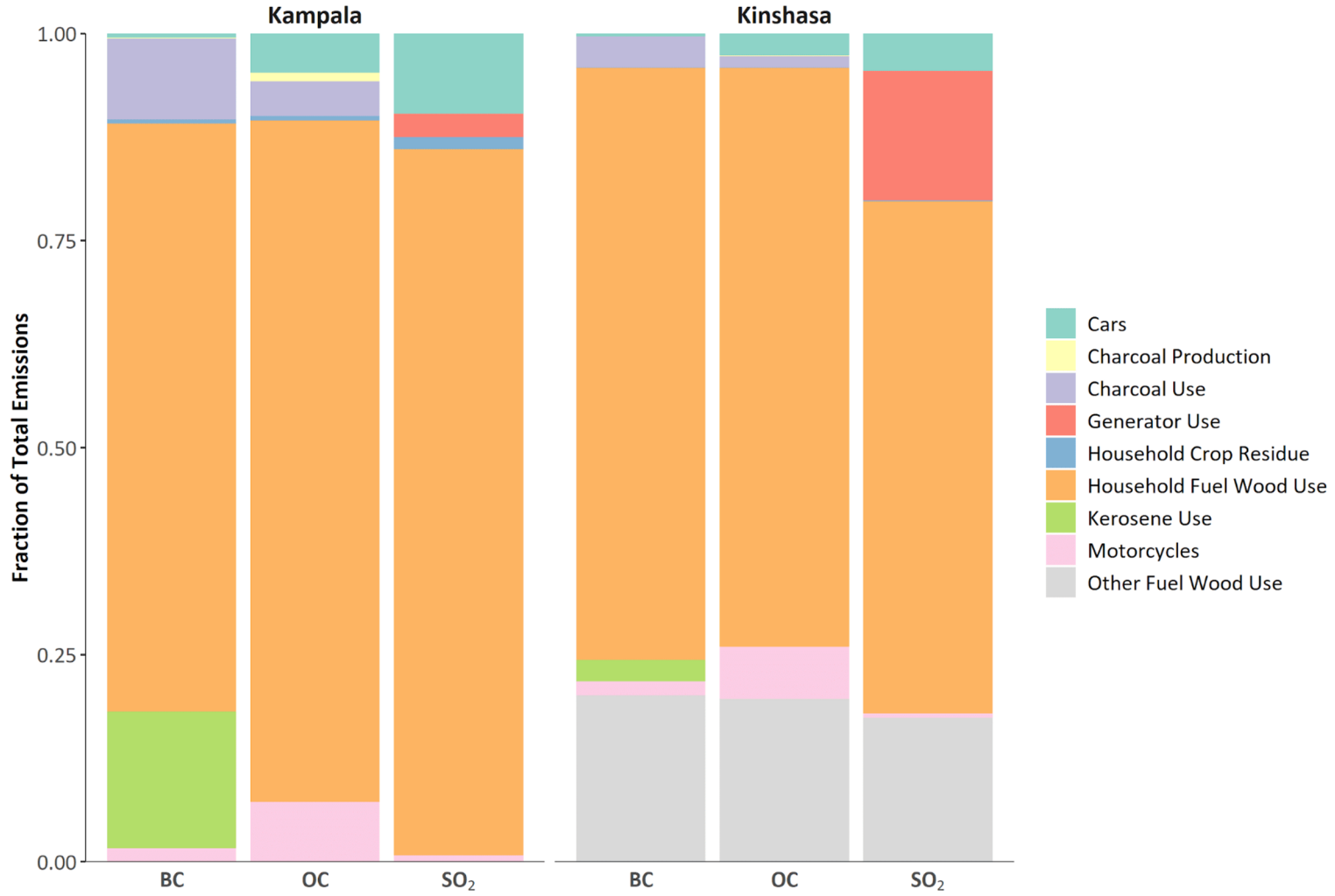
- Brazzaville Embassy
- ERAIFT-DRC
- Regional Average
- Cite Belle Vue
- Kinshasa Embassy
- UPN-DRC



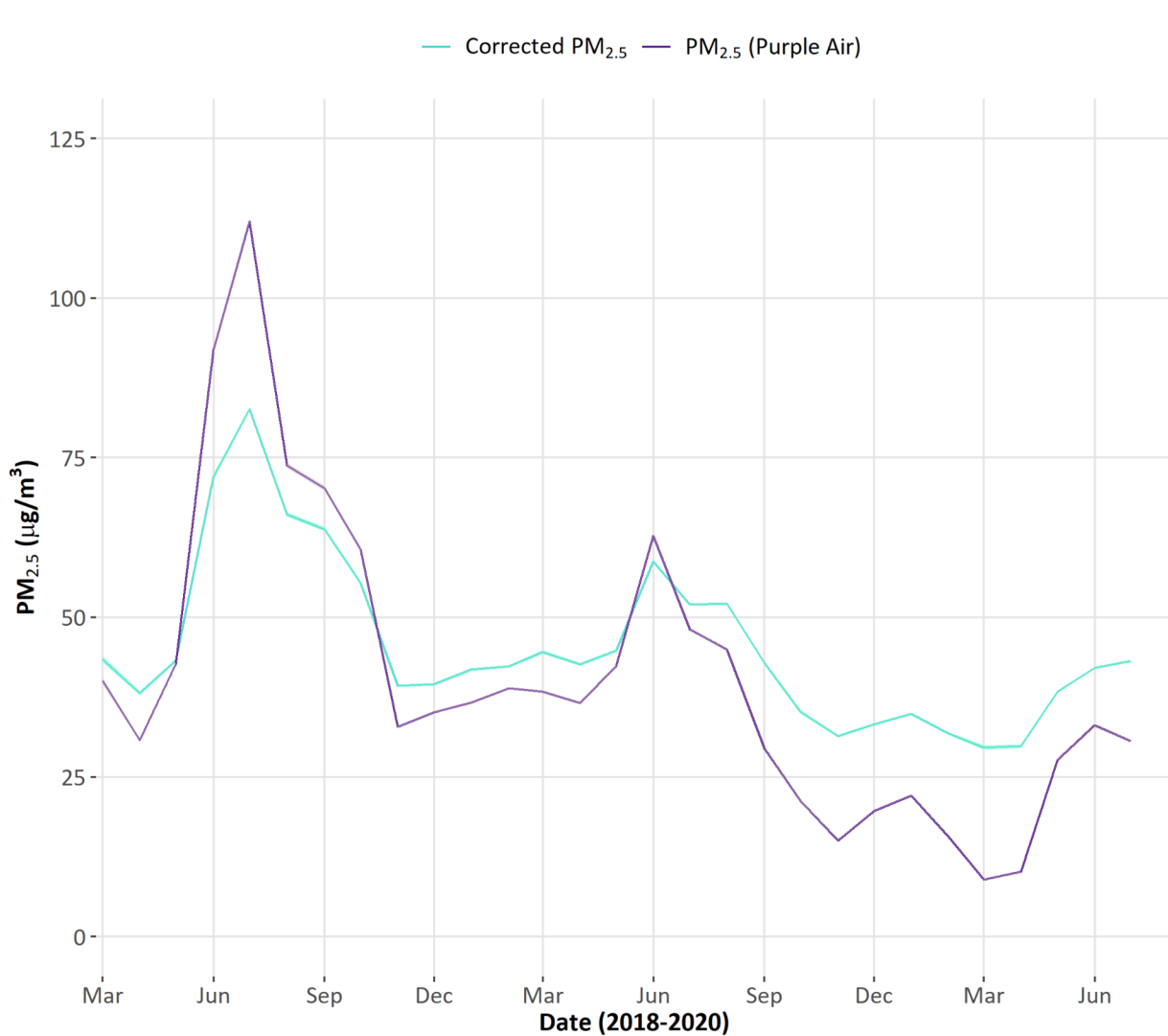
- Sites have similar diurnal PM_{2.5} variability with varying PM_{2.5} magnitude
- Morning and Evening peaks likely correlated with increase in traffic and also increased fire activity for cooking, waste burning, etc
- Highest PM_{2.5} values occur at Brazzaville Embassy and UPN-DRC ($\cong 60 \mu\text{g m}^{-3}$ peak in the evening)



Kampala and Kinshasa Have Similar Combustion Emission Profiles and are also similar climatologically



Kinshasa Dry Season PM_{2.5} Peaks are Decreasing Annually



- $\cong 10 \mu\text{g m}^{-3}$ difference in calibrated and raw data for PM_{2.5} values between $25 \mu\text{g m}^{-3}$ and $100 \mu\text{g m}^{-3}$
- Annual means of daily PM_{2.5} decrease from $54.4 \mu\text{g m}^{-3}$ (2018) to $43.5 \mu\text{g m}^{-3}$ (2019) to $35.7 \mu\text{g m}^{-3}$ (through July 2020)
 - Meteorological differences do not explain decreases in PM_{2.5}
 - No known PM_{2.5} emissions control measures in the last 3 years
 - Decreasing PM_{2.5} values from 2019 to 2020 can be partially attributed to COVID19