

# Low-Cost Sensor Performance Intercomparison, Correction Factor Development, and 2+ Years of Ambient PM<sub>2.5</sub> Monitoring in Accra, Ghana

Garima Raheja,\* James Nimo, Emmanuel K.-E. Appoh, Benjamin Essien, Maxwell Sunu, John Nyante, Mawuli Amegah, Reginald Quansah, Raphael E. Arku, Stefani L. Penn, Michael R. Giordano, Zhonghua Zheng, Darby Jack, Steven Chillrud, Kofi Amegah, R. Subramanian, Robert Pinder, Ebenezer Appah-Sampong, Esi Nerquaye Tetteh, Mathias A. Borketey, Allison Felix Hughes, and Daniel M. Westervelt\*



Cite This: *Environ. Sci. Technol.* 2023, 57, 10708–10720



Read Online

ACCESS |



Metrics & More

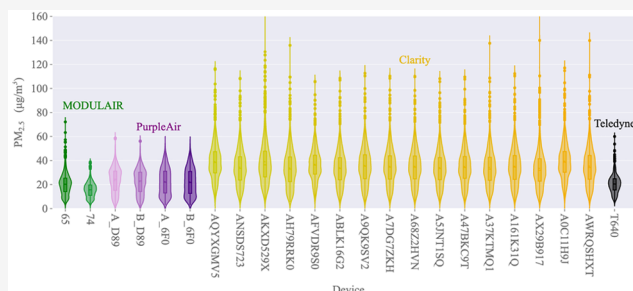


Article Recommendations



Supporting Information

**ABSTRACT:** Particulate matter air pollution is a leading cause of global mortality, particularly in Asia and Africa. Addressing the high and wide-ranging air pollution levels requires ambient monitoring, but many low- and middle-income countries (LMICs) remain scarcely monitored. To address these data gaps, recent studies have utilized low-cost sensors. These sensors have varied performance, and little literature exists about sensor intercomparison in Africa. By colocating 2 QuantAQ Modulair-PM, 2 PurpleAir PA-II SD, and 16 Clarity Node-S Generation II monitors with a reference-grade Teledyne monitor in Accra, Ghana, we present the first intercomparisons of different brands of low-cost sensors in Africa, demonstrating that each type of low-cost sensor PM<sub>2.5</sub> is strongly correlated with reference PM<sub>2.5</sub>, but biased high for ambient mixture of sources found in Accra. When compared to a reference monitor, the QuantAQ Modulair-PM has the lowest mean absolute error at 3.04 μg/m<sup>3</sup>, followed by PurpleAir PA-II (4.54 μg/m<sup>3</sup>) and Clarity Node-S (13.68 μg/m<sup>3</sup>). We also compare the usage of 4 statistical or machine learning models (Multiple Linear Regression, Random Forest, Gaussian Mixture Regression, and XGBoost) to correct low-cost sensors data, and find that XGBoost performs the best in testing ( $R^2$ : 0.97, 0.94, 0.96; mean absolute error: 0.56, 0.80, and 0.68 μg/m<sup>3</sup> for PurpleAir PA-II, Clarity Node-S, and Modulair-PM, respectively), but tree-based models do not perform well when correcting data outside the range of the collocation training. Therefore, we used Gaussian Mixture Regression to correct data from the network of 17 Clarity Node-S monitors deployed around Accra, Ghana, from 2018 to 2021. We find that the network daily average PM<sub>2.5</sub> concentration in Accra is 23.4 μg/m<sup>3</sup>, which is 1.6 times the World Health Organization Daily PM<sub>2.5</sub> guideline of 15 μg/m<sup>3</sup>. While this level is lower than those seen in some larger African cities (such as Kinshasa, Democratic Republic of the Congo), mitigation strategies should be developed soon to prevent further impairment to air quality as Accra, and Ghana as a whole, rapidly grow.



**KEYWORDS:** air quality, low-cost sensors, PurpleAir, clarity, Modulair-PM, statistical methods, machine learning, urban air, sensor network, Ghana

## 1. INTRODUCTION

In 2019, over 1 million deaths in Africa were attributable to exposure to ambient particulate matter with an aerodynamic diameter less than 2.5 μm (PM<sub>2.5</sub>).<sup>1</sup> Ambient PM<sub>2.5</sub> pollution, from sources including agricultural and waste burning, transportation, and residential cooking, is linked to adverse health impacts such as stroke, cardiovascular and respiratory diseases such as lung cancer, obstructive pulmonary disease, and myocardial infarction.<sup>2–6</sup> High levels of PM<sub>2.5</sub> exposure during pregnancy can lead to preterm birth, low birth weight, and reduced cognitive function in infants.<sup>7–9</sup> While cities in the United States, Europe, and parts of China have observed much

improvement in terms of PM<sub>2.5</sub> levels, many cities in Africa have seen PM<sub>2.5</sub> levels worsen over time (though lack of data makes this challenging to ascertain).<sup>10–12</sup> The Ghana Environmental Protection Agency estimates that about 2800 deaths annually in

Received: December 7, 2022

Revised: June 1, 2023

Accepted: June 2, 2023

Published: July 12, 2023



**Table 1. Summary of Low-Cost Monitoring Device Attributes**

Company	Monitor Name (version)	Time Resolution (min)	Purchase Cost (USD)	Cost per year (USD)
Purpleair, Inc.	PurpleAir (PA-II SD)	2	\$269 (note: product discontinued)	\$0
Clarity Movement Co.	Clarity (Node-S Generation II)	20	\$1000	\$1400
Quant-AQ, Inc.	Modulair-PM (Modulair-PM)	1	\$1500	\$300

the Greater Accra Metropolitan Area (GAMA) can be attributed to PM<sub>2.5</sub> pollution.<sup>13</sup> Anthropogenic activities such as open solid (household and agricultural) waste burning and emissions from vehicles have skyrocketed in recent decades, making low-income communities, children, pregnant women, and older adults increasingly at risk to higher outdoor PM<sub>2.5</sub> levels.<sup>14–21</sup> This is especially concerning during the dry dusty Harmattan season (December to February), which is marked by dust from the Sahara Desert blowing into Accra due to large-scale circulation patterns including the movement of the Intertropical Convergence Zone (ITCZ).<sup>22–24</sup> For example, Alli et al. observed that PM<sub>2.5</sub> concentrations during the Harmattan period raised PM<sub>2.5</sub> levels by 56 to 71  $\mu\text{g}/\text{m}^3$  during a city-wide measurement campaign in the GAMA.<sup>25</sup> Additionally, during the Harmattan, aerosol composition changes can also be expected in addition to mass concentration changes.<sup>26</sup>

Ambient PM<sub>2.5</sub> measurements enable governmental agencies to formulate policies and control strategies for PM<sub>2.5</sub>.<sup>27–30</sup> Current estimates in Africa, however, are limited due to a severe lack of both ground-based air quality observations and corresponding health outcome data. Globally, the population-weighted mean distance to the nearest air quality monitor is 220 km; monitors are especially sparse in Africa.<sup>3,4,31</sup> A recent study has shown that only two out of 15 countries under the economic community of West African States (ECOWAS) monitor PM<sub>2.5</sub> pollution.<sup>32</sup> The dearth of PM<sub>2.5</sub> monitoring sensors across African countries has been attributed to the high capital cost of the instruments, as well as the high installation and maintenance cost.<sup>33–35</sup> PM<sub>2.5</sub> monitors such as Teledyne T640 and MetOne BAM-1020 ( $\beta$  Attenuation Monitor), which have US EPA Federal Equivalent Method status, require high installation, operational and maintenance costs, and are thus not practical to install in the large numbers<sup>34,36</sup> that would be necessary to capture the strong spatial and temporal variation in cities. This poses challenges for air quality practitioners in Africa to monitor PM<sub>2.5</sub> levels and subsequently develop policies to control and regulate PM<sub>2.5</sub> pollution.

Therefore, more PM<sub>2.5</sub> ambient concentration measurement has become imperative in Africa to estimate environmental and health impacts to inform policies to protect vulnerable groups from the harmful effects of PM<sub>2.5</sub> pollution. Sub-Saharan Africa, in particular, is going through transitions in energy usage, urbanization, and population growth, causing economic activities such as infrastructure, industrialization, and motorization to surge.<sup>3,31,32,37</sup>

Within the last 10 years, there has been a dramatic rise in the use of real-time, off-the-shelf low-cost sensors (LCS), such as those manufactured by PurpleAir, Inc., Clarity Movement Co., and QuantAQ, Inc. for PM<sub>2.5</sub> monitoring. Many LCS are optical particle counters using Plantower PMS5003 nephelometers for laser light scattering-based estimates of ambient PM<sub>2.5</sub>.<sup>38</sup> With substantial validation and careful calibration to correct for inherent biases, these LCS have been demonstrated to be viable alternatives to the more expensive reference-grade monitors.<sup>2,39–43</sup> Further, the high spatial density of LCS networks can offer a more detailed view of a city than a single reference

monitor of the same cost, supporting continuous long-term monitoring of PM<sub>2.5</sub> exposures and potentially identifying hotspots. The performance and accuracy of LCS are greatly influenced by particle physicochemical properties including size distribution which vary by source, meteorological conditions such as temperature, and hygroscopic growth affected by humidity.<sup>44–47</sup> Hygroscopic growth occurs when particulates absorb water as a function of relative humidity (which increases as relative humidity increases), thus altering their size and structure affecting the amount of scattered light and therefore optically estimated PM<sub>2.5</sub> concentrations.

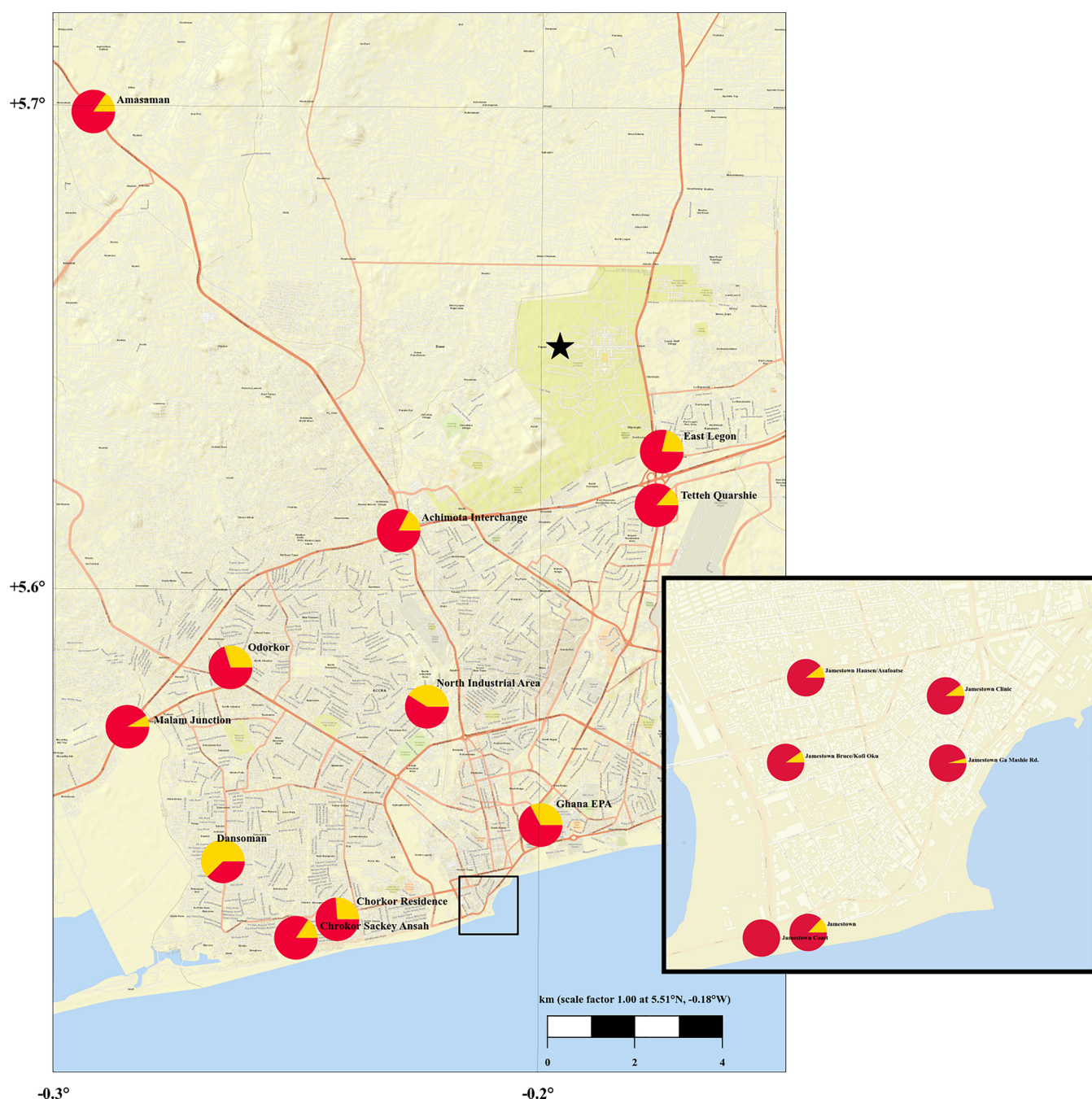
Comparing reference monitors and LCS can help develop calibration models to correct for LCS sensitivities like hygroscopic growth. While some literature exists on corrections derived from filter-based instruments, very little of the real-time sensor comparison or calibration work has been done in Africa, where environmental conditions and emissions sources are very different from the United States and Europe. Adong et al. colocated eight devices of a single LCS manufacturer (AirQo) and two MetOne BAM-1020 devices.<sup>48</sup> McFarlane et al. applied Gaussian Mixture Regression (GMR) to calibrate a PurpleAir colocated with a MetOne BAM-1020 at the US Embassy in Accra, Ghana, for one year.<sup>49</sup> The results revealed that the manufacturer-reported data from PurpleAir was biased slightly high compared to the BAM and only moderately correlated ( $r^2 = 0.53$ , MAE = 6.2  $\mu\text{g}/\text{m}^3$ ). However, using a GMR-based correction factor reduced the bias (MAE = 2.2  $\mu\text{g}/\text{m}^3$ ) and greatly improved the correlation ( $r^2 = 0.88$ ).

## 2. METHODS

Here, we seek to conduct, to our knowledge, the first large-scale LCS intercomparison study in Africa, by colocating 22 LCS from three different sensor brands (PurpleAir, Inc., Clarity Movement Co., and QuantAQ, Inc.) with a reference monitor (Teledyne T640). We then compare the performance of four types of statistical and machine learning models (Multiple Linear Regression, Gaussian Mixture Regression, XGBoost, and Random Forest) for calibrating and correcting data from each type of LCS. Finally, we apply the GMR correction factor to the network of 17 Clarity devices deployed around various neighborhoods and major roadways in Accra, Ghana, from 2018 to 2021, resulting in, to our knowledge, the longest and most spatially and temporally detailed survey of PM<sub>2.5</sub> in Accra to date.

**2.1. University of Ghana (UG) Collocation.** Low-cost sensors (LCS) manufactured by three different brands—Clarity Node-S Generation II, PurpleAir PA-II SD, and QuantAQ Modulair-PM—were colocated at the University of Ghana (UG), Department of Physics, Legon, Accra, Ghana. The monitors were mounted on metal poles and fastened with cable ties at a height of five meters. Data were inspected between individual sensor units for any possible defective units and none were found.

All of the low-cost air quality monitoring devices in this study have in-built Plantower PMS5003 nephelometers that estimate the concentration of fine particulate matter dispersed in ambient



**Figure 1.** Locations of Clarity nodes deployed in Accra, Ghana. The inset shows Jamestown. Pie charts show the percentage of measured days that exceeded WHO Daily  $PM_{2.5}$  Guidelines ( $15 \mu\text{g}/\text{m}^3$ ). The University of Ghana, where LCS were collocated as mentioned in Section S.I., is denoted with a black star.

air based on laser light scattering technology.<sup>50–52</sup> The low-cost monitoring devices also contain BOSCH BME280 sensors to measure internal meteorological parameters including temperature and relative humidity.<sup>50,51</sup> In addition to the Plantower nephelometer, Modulair-PMs include an Alphasense OPC-N3 (optical particle counter), which provides more realistic estimates of supermicron aerosol not captured by Plantower nephelometers, with a measurement range of  $0.35\text{--}40 \mu\text{m}$ .<sup>43,53–56</sup> Modulair-PM-reported  $PM_{2.5}$  includes a manufacturer-applied correction accounting for particle density, aspiration efficiency, and hygroscopic growth. Summary information about the different types of monitors is given in Table 1. Detailed information on the different types of low-cost

air quality monitors (Clarity Node-S, PurpleAir PA-II, and QuantAQ Modulair-PM) used in this research can be found at their respective websites, <https://www.clarity.io/>, <https://www2.purpleair.com/>, and <https://www.quant-aq.com/>. The  $PM_{2.5}$  ( $\mu\text{g}/\text{m}^3$ ) columns were extracted from each dataset; for the PurpleAir, the Sensor A  $PM_{2.5}$  ( $\text{cf} = \text{atm}$ ) and Sensor B  $PM_{2.5}$  ( $\text{cf} = \text{atm}$ ) columns were averaged.

The LCS were collocated with a Teledyne T640 (reference-grade monitor) placed on a two-story building rooftop located at  $5.65136^\circ \text{N}$ ,  $0.18566^\circ \text{W}$ , and elevation of 108 meters above sea level for a period of 4 months (11th May to 25th September, August 2021), recording data with a 1 min resolution. The Teledyne T640 contains an aerosol sample conditioner, a

sample flow controller, and a 5 lpm vacuum pump with a temporal resolution of 1 min. It uses broadband spectroscopy using 90° white-light scattering with a polychromatic light-emitting diode (LED), measures with a resolution of 256 sizes over 0.18–20 μm range, combined to 64 channels for mass calculation, and exceeds US EPA PM<sub>10</sub> FEM and Class III FEM PM<sub>2.5</sub> performance requirements for additive and multiplicative bias compared to FRM samplers.<sup>50,51</sup> Instrument installation, calibration, and training were held virtually on August 10–11, 2020 according to the US EPA Teledyne T640 Standard Operating Procedure.<sup>57</sup> Further information on the Teledyne model T640 can be found at <https://www.teledyne-api.com/>.<sup>58</sup>

The colocation site is an urban area with low-density housing, sparse trees, and low traffic flows. The distance to the nearest road is 500 meters. There are no known major burning or other emissions sources near the site.

**2.2. Comparison of Machine Learning Models for Correction Factors.** Hourly averaged low-cost sensor data is cleaned by only keeping measurements where PM<sub>2.5</sub> > 0 μg/m<sup>3</sup>, PM<sub>2.5</sub> < 1000 μg/m<sup>3</sup>, relative humidity > 0%, and in the case of PurpleAir, where |(Channel A-Channel B)/(Channel B)| < 20%. The uptime (percent of data where these criteria were met) during the University of Ghana colocation where these conditions are true is 98.7% for Clarity, 99.2% for PurpleAir, and 99.9% for Modulair-PM.

Four different models are tested for correcting low-cost sensor data: Multiple Linear Regression, Gaussian Mixture Regression, Random Forest, and XGBoost. Each model uses three measurements from the LCS as explanatory features: PM<sub>2.5</sub>, temperature (*T*, °C) and relative humidity (RH, %) (eq 1). In the case of PurpleAir, the Channel A and Channel B PM<sub>2.5</sub> readings are averaged. The reference value is the Teledyne T640 measurement. For all models, a 10-fold cross-validation is used for hyperparameter training, with an 80–20% training–testing data split via random subsampling without replacement.

$$\begin{aligned} \text{PM}_{2.5}[\text{LCS, calibrated}] \\ = f(\text{PM}_{2.5}[\text{LCS reported}], \text{RH}[\text{LCS reported, \%}], \\ T[\text{LCS reported, } ^\circ\text{C}]) \end{aligned} \quad (1)$$

The first is **multiple linear regression (MLR)**, which optimizes the best fit by minimizing the distance between the “true” *y*-values (in this case, the reference monitor PM<sub>2.5</sub>) and the “predicted” *y*-values. MLR is beneficial because of its ease of understanding, and because the model is easy to convey and reuse using an equation. Linear (and higher-order polynomial) fits are the least computationally intensive and simplest to code, and therefore the most common correction methods in low-cost sensor literature.<sup>33</sup>

The second is **Random Forest**, a supervised ensemble model that uses a combination of decision trees. Random forest (RF) is useful because each decision tree theoretically isolates errors. For this study, the RandomForestRegression from sklearn.ensemble method in Python<sup>59</sup> is used and is optimized using grid search (maximum features: 1,2,3 and maximum tree depth: 1,2,3,4,5). Optimal parameter results from grid search are in the [Supporting Information](#).

The third model is the **Gaussian Mixture Regression** model, which is a multivariate nonlinear regression method that models the probability density of the output data conditional to the input data. It is implemented here using the sklearn.mixture.GaussianMixture<sup>60</sup> method in Python and the gmr library

developed by AlexanderFabisch on Github.<sup>61</sup> GMR is beneficial because it can produce “components” which identify regimes under which regression is classified (see McFarlane et al.<sup>49</sup> for more information).

The final is **Extreme Gradient Boosting (XGBoost)**, which uses distributed gradient-boosted decision trees for regression. The models are optimized to maximize *R*<sup>2</sup>, which is the coefficient of determination, and minimize Mean Absolute Error (MAE), which is a measure of bias. In this study, XGBoost is implemented using the open-source xgboost Python library.<sup>62</sup>

**2.3. Accra Deployment.** Seventeen Clarity Node-S Generation I monitors were deployed across the city of Accra beginning in May 2018. Note that the set of monitors deployed for this set of the project is different from the set of monitors used for the colocation work in Section 5, Part I—University of Ghana (UG) Collocation. Four of them were deployed along major roadsides (Tetteh Quarshie, Amasaman, Malam Junction, and Achimota Interchange) on metal poles five meters above the ground with a metal basket to securely fasten the monitors. Another four of the monitors were mounted at already existing permanent monitoring sites (East Legon, Odorkor, Dansoman, and North Industrial Area) across the city and one at the Ghana EPA Head office. The rest were deployed on electric poles in Jamestown (Jamestown Bruce/Kofi Oku, Jamestown Clinic, Jamestown Coast, Ga Mashie Road, and Jamestown Hansen/Asafoatse) and Chorkor (Chorkor Residence and Chorkor Sackey Ansah) residential areas. The locations for the monitors can be seen in [Figure 1](#).

The time series of manufacture-reported and GMR-corrected PM<sub>2.5</sub> measurements from the Accra deployment are shown in [Figure S1](#). The GMR-corrected values are lower than the manufacturer-reporter values but maintain the same temporal trends.

[Figure S2](#) shows the timeline of all Clarity nodes deployed across Accra. A total of 11,001 valid days of data were retrieved for analysis from the network before the monitors were decommissioned.

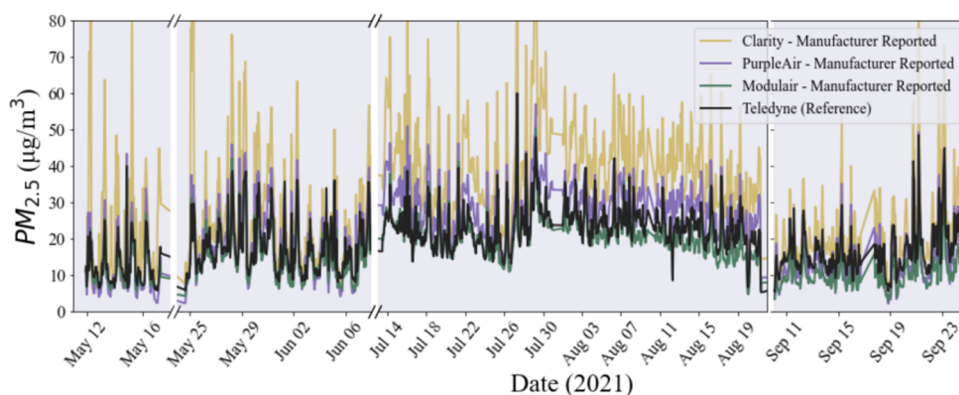
### 3. RESULTS

**3.1. Colocation at University of Ghana.** **3.1.1. Sensor Intercomparison.** [Table 2](#) summarizes the data collected during the University of Ghana colocation.

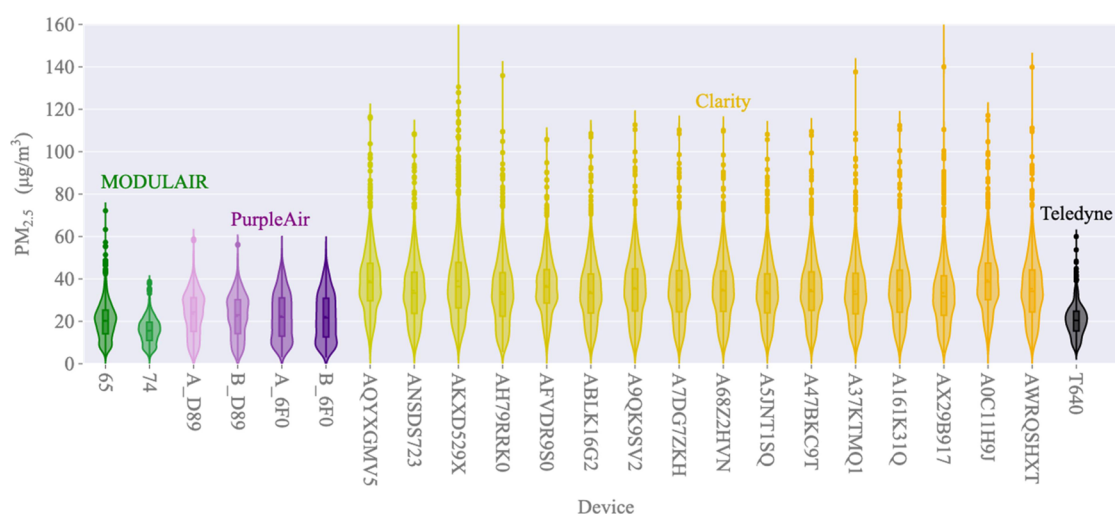
**Table 2. Summary of University of Ghana LCS Intercomparison (Averaged by LCS Brand)**

Low-cost Monitor	Reported PM <sub>2.5</sub> Range (μg/m <sup>3</sup> )	Reported Temperature range (°C)	Reported Relative Humidity range (%)
PurpleAir	2.1–57.6	23.1–41.0	29.5–87.6
Clarity	5.3–135.3	23.6–37.9	48.0–89.9
Modulair-PM	2.8–55.5	22.7–39.8	44.4–100.0
Teledyne	5.2–60.0	4.6–34.2	58.3–100.0

[Figure 2](#) shows the hourly average time series of PM<sub>2.5</sub> measured by the reference monitor, and the averages of manufacture-reported PM<sub>2.5</sub> (MR) measured by each LCS. The gaps in the data are due to monitor malfunction or scheduled maintenance. All intercomparison analysis is performed using only the hours (*n* = 3063) where at least one of all four types of monitors was functioning. Each low-cost monitor is temporally correlated with the reference monitor (black). Clarity devices (yellow line) tend to overestimate both



**Figure 2.** Time series of hourly averaged  $\text{PM}_{2.5}$  ( $\mu\text{g}/\text{m}^3$ ) LCS and reference measurements at the University of Ghana between May and September of 2021. Time gaps are due to sensor maintenance and/or failure. Reference monitor in black, PurpleAir in purple, Clarity in yellow, and Modulair-PM in green.



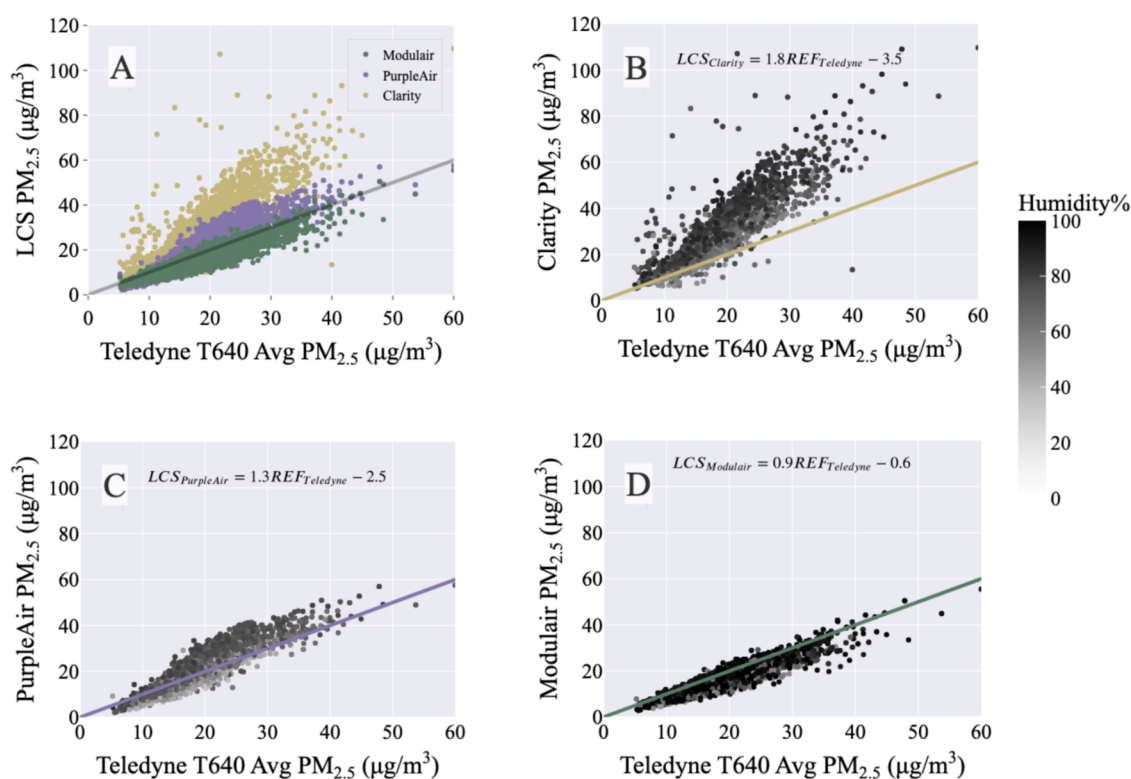
**Figure 3.** Violin plot of distributions of hourly averaged  $\text{PM}_{2.5}$  from each monitor colocated at University of Ghana, May to September 2021. Violins show data distribution with boxplots inside and outliers labeled with dots. Reference monitor in black, PurpleAir in purple, Clarity in yellow, and Modulair-PM in green.

the background  $\text{PM}_{2.5}$  levels and also the peaks, reporting maximum values of over  $100 \mu\text{g}/\text{m}^3$ . PurpleAir devices (purple line) are also biased high, but to a lesser extent, while Modulair-PMs (green line) show little bias. There are two distinct rainy seasons in Accra (May/June and September/October) which keep  $\text{PM}_{2.5}$  levels reasonably low compared to other seasons.<sup>41,63–65</sup> Note that the intercomparison time does not include measurements taken during the Harmattan (December–February), the effects of which are discussed later; further data collection will hope to address this gap.

Figure 3 shows the violin plots of hourly averaged MR measurements from each monitor. The mean of the Teledyne T640 measurements is  $20.4 \mu\text{g}/\text{m}^3$ . The Clarity monitors show a broader spread than the other monitors and have means ranging from  $33.5$  to  $39.0 \mu\text{g}/\text{m}^3$ . Clarity devices also have recorded hourly mean concentrations that exceed  $100 \mu\text{g}/\text{m}^3$ , which is not seen in any of the other devices. The PurpleAirs show similar measurement ranges between the two channels of each monitor and also between monitors, with a total mean of  $22.6 \mu\text{g}/\text{m}^3$ . The Modulair-PM monitors are within the same range as the PurpleAir and Teledyne but have different means due to the skewed distribution of outlier values above  $40 \mu\text{g}/\text{m}^3$ , with MOD\_65 having a mean of  $20.5 \mu\text{g}/\text{m}^3$  and MOD\_74 having a

mean of  $15.5 \mu\text{g}/\text{m}^3$ . Overall, the Modulair-PM distribution most closely matches that of Teledyne, followed by PurpleAir and Clarity.

Figure 4A shows a scatter plot of hourly LCS  $\text{PM}_{2.5}$  measurements compared to hourly reference monitor measurements, shaded by relative humidity. Figure 4B–D isolates each LCS from Figure 4A, and the equation for the line of best fit for each LCS is noted on each panel. (Figure S3 shows the scatter plots of hourly LCS  $\text{PM}_{2.5}$  measurements compared to hourly reference monitor measurements, shaded by temperature.) The drawn line indicates the line of equality between LCS and reference monitors. As seen in the scatter plots, the relationship between reference  $\text{PM}_{2.5}$  and manufacturer-reported  $\text{PM}_{2.5}$  is linear with moderate to strong correlation ( $R^2 = 0.82$  for PurpleAir,  $R^2 = 0.69$  for Clarity,  $R^2 = 0.84$  for Modulair-PM). The slopes in the relationship, however, vary between the three low-cost monitoring devices. Clarity has the highest slope (1.8), followed by PurpleAir (1.3) and Modulair-PM (0.9). Clarity data also show the most scatter, with MAE =  $13.68 \mu\text{g}/\text{m}^3$ , while PurpleAir MAE =  $4.54 \mu\text{g}/\text{m}^3$ , and Modulair-PM MAE =  $3.04 \mu\text{g}/\text{m}^3$ . Despite Clarity and PurpleAir using the same Plantower device to measure  $\text{PM}_{2.5}$ , Figure 4B,C shows that the reported values can be quite different. Clarity and PurpleAir report



**Figure 4.** LCS  $PM_{2.5}$  measurements compared to Teledyne T640 measurements. (A) Hourly average  $PM_{2.5}$  measurements from all three low-cost monitors, averaged by type, compared to the respective Teledyne hourly average  $PM_{2.5}$  measurements. The one-to-one line is shown for comparison. The coefficients of the line of best fit are denoted for each LCS type (B–D).

**Table 3.** Comparison of  $R^2$  and MAE of Test Sets of Machine Learning Models to Correct LCS with Reference Measurements

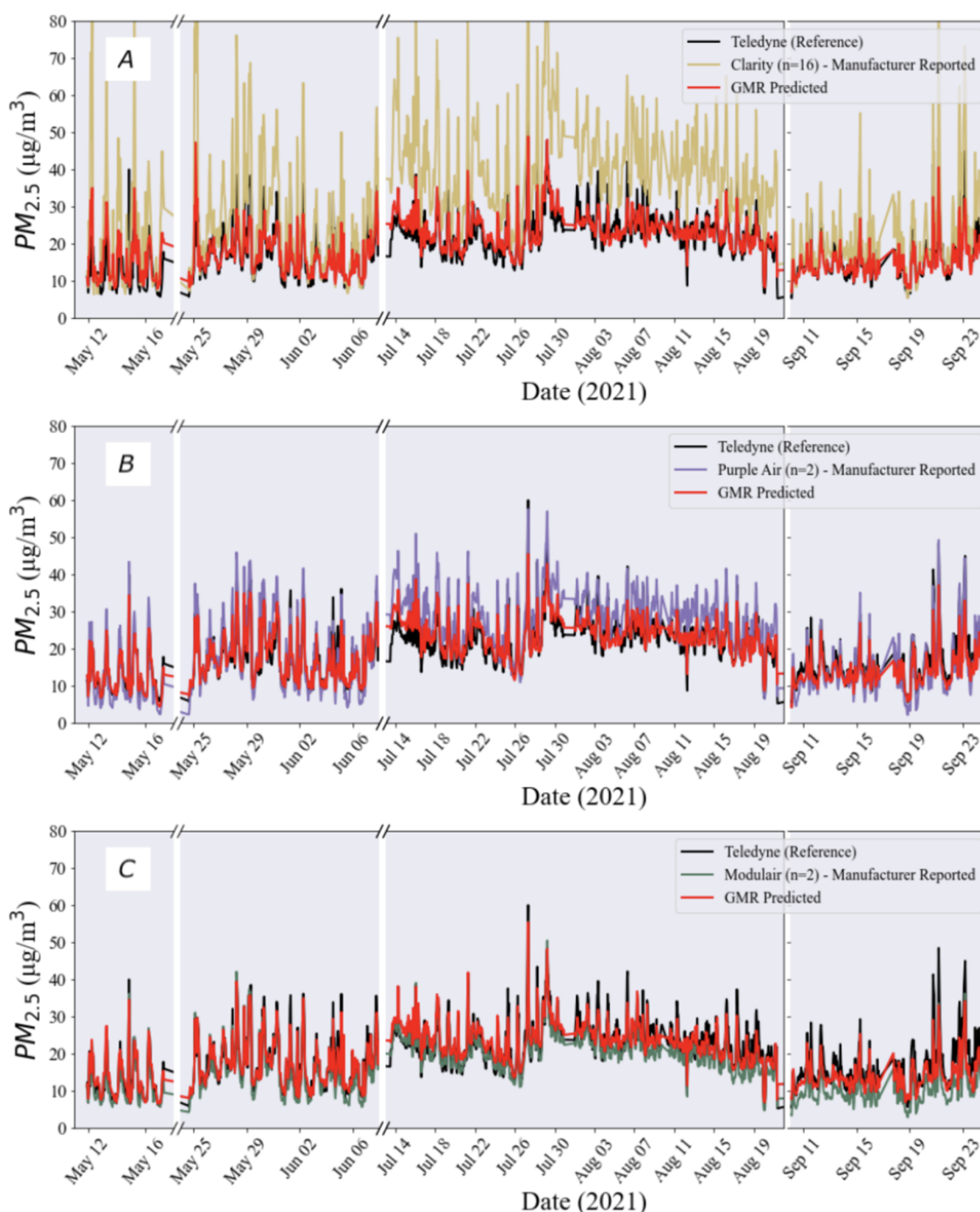
Model	Purple Air				Clarity				Modulair-PM			
	$R^2$	MAE ( $\mu\text{g}/\text{m}^3$ )	CV MAE	RMSE	$R^2$	MAE ( $\mu\text{g}/\text{m}^3$ )	CV MAE	RMSE	$R^2$	MAE ( $\mu\text{g}/\text{m}^3$ )	CV MAE	RMSE
Manufacturer-Reported	0.82	4.54	-	17.24	0.69	13.68	-	17.08	0.84	3.04	-	3.85
MLR	0.85	2.10	0.10	2.88	0.73	2.58	0.13	3.77	0.85	2.16	0.11	2.69
GMR	0.86	1.93	0.09	2.58	0.79	2.27	0.11	3.15	0.87	2.04	0.10	2.57
RF	0.89	1.93	0.10	2.54	0.83	2.17	0.11	2.98	0.86	2.06	0.10	2.66
XGBoost	0.97	0.56	0.03	1.24	0.94	0.80	0.05	1.81	0.96	0.68	0.03	1.42

Plantower “Beijing-calibrated” values as is,<sup>33</sup> while Modulair-PM combines Plantower and Alphasense OPC-N3 with assumptions about aerosol hygroscopicity to estimate  $PM_{2.5}$ . Each point is shaded by the relative humidity measurement from each LCS. PurpleAir, Clarity, and Modulair-PM mean measured relative humidity and standard deviation are  $73.30 \pm 10.49$ ,  $64.44 \pm 12.27$ , and  $84.42 \pm 15.32\%$ , respectively. The Modulair-PM relative humidity readings are significantly higher than the other LCS. This is a known issue of early Modulair-PM devices which has since been corrected by switching to a new relative humidity sensor manufactured by Sensirion.

**3.1.2. ML Model Correction Comparison.** Table 3 shows the results of applying the four different models to each LCS. Data from each monitor is hourly averaged. CvMAE (bias-corrected

mean normalized MAE) is calculated as defined by Equation 4 in Giordano, et al.<sup>33</sup> Selection of optimal model hyperparameters was conducted using parameter grid search (which is a process that exhaustively tests all combinations of hyperparameters)<sup>66</sup> and 10-fold cross-validation, scoring on minimum testing MAE. Modulair-PM has the best manufacturer-reported (MR) measurements, but the MAE is still improved by applying all four correction factor techniques. Clarity has the lowest  $R^2$  and MAE of raw data, but the model-improved  $R^2$  and MAE are comparable to those of other LCS. Root-mean-squared error (RMSE) is also shown, for comparison to US EPA methods.

Since MLR is the easiest to transfer to more use cases, the parameters for MLR correction for each monitor are given in



**Figure 5.** (A) Raw measurements (yellow line) and GMR-corrected average Clarity values (red), and Teledyne T640 reference measurements (black), from the UG collocation. (B) Raw measurements (purple line) and GMR-corrected average PurpleAir values (red), and Teledyne T640 reference measurements (black), from the UG collocation. (C) Raw measurements (green line) and GMR-corrected average Modulair-PM values (red), and Teledyne T640 reference measurements (black), from the UG collocation.

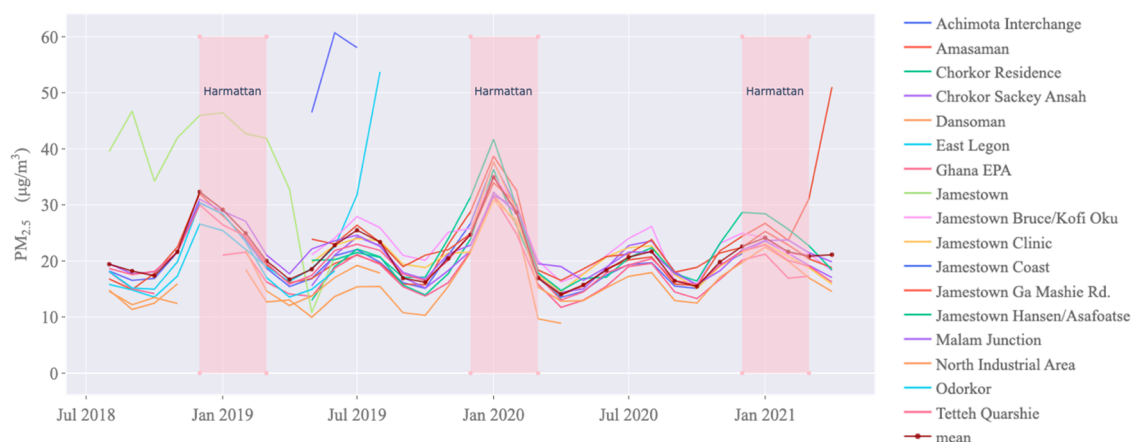
**Table S1.** To use the parameters to apply MLR, for PurpleAir, apply eq 2, and likewise for other LCS

$$\text{MLR Corrected PM}_{2.5} = 17.51 + 0.69\text{PA}_{\text{PM}_{2.5}} - 0.12 \text{PA}_{\text{temperature},^{\circ}\text{C}} - 0.15 \text{PA}_{\text{humidity},\%} \quad (2)$$

where  $\text{PA}_{\text{PM}_{2.5}}$  is PurpleAir manufacturer-reported  $\text{PM}_{2.5}$ ,  $\text{PA}_{\text{Temperature}^{\circ}\text{C}}$  is PurpleAir manufacturer-reported temperature in degree Celsius, and  $\text{PA}_{\text{Humidity}\%}$  is PurpleAir manufacturer-reported relative humidity.

Standard errors and test set predictions for MLR for each LCS type are shown in Figures S4–S6. Figures S7–S9 show the range and combinations of hyperparameters, and training and testing results, for XGBoost for each type of LCS. Figures S10–S12 show the Random Forest training and optimal hyperparameters. Figures S13–S15 show the Gaussian Mixture Regression training.

In Table 3, XGBoost shows the best performance, with highest  $R^2$  and lowest MAE, in all three LCS cases. While it would be intuitive to use the best-performing model, it is not



**Figure 6.** GMR-corrected monthly average time series at each site in the Clarity network deployed around Accra. The dotted red line shows the mean of all sites. Purple boxes denote Harmattan times (December to February).

suiting to our application. XGBoost and RF are both tree-based algorithms, which are limited in usability when applied to data that is outside the ranges of the testing set. Inside the University of Ghana colocation setting, the LCS  $PM_{2.5}$  measurements remain largely within the  $0\text{--}60\ \mu\text{g}/\text{m}^3$  range. However,  $PM_{2.5}$  measurements from LCS deployed around Accra show many raw values  $>60\ \mu\text{g}/\text{m}^3$  (Figure S16), and previous studies of Accra have demonstrated measured values greater than  $100\ \mu\text{g}/\text{m}^3$ .<sup>3,67,68</sup> Figure S17 shows that XGBoost performs well in the  $0\text{--}60\ \mu\text{g}/\text{m}^3$  but never extrapolates  $PM_{2.5}$  estimates outside of that range. Tree-based algorithms are very applicable to our study since they result in excellent increases in  $R^2$  and decreases in MAE and therefore create a high-performance correction model, and thus present an interesting advance in the field of LCS correction and could definitely be used in other studies where conditions are at least somewhat similar to this site. We note these caveats as a word of caution to those who might consider extrapolating this study using data from a different site.

However, the GMR and MLR corrections generally scale well outside the bounds of the training data. GMR is used to correct the Clarity measurements from the deployed Accra network discussed in Figures 1 and 6–8. Note that for minimal coding applications, MLR would be suitable to use since the  $R^2$  and MAE are in a comparable range to the GMR. However, since GMR is overall better, and can be implemented in just a few lines of code as well, it is used here.

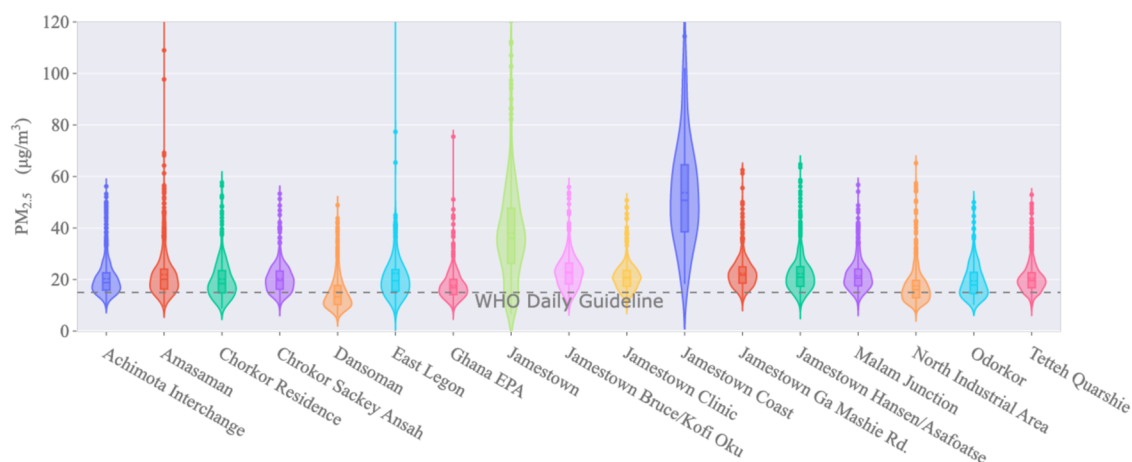
Figure 5A–C shows raw LCS measurements, reference monitor measurements, and GMR-corrected  $PM_{2.5}$  values for each LCS at the University of Ghana colocation. The GMR correction factor is trained on an 80% split of the colocation dataset, separately developed for each monitor, and is subsequently applied to all of the data. The correction factor generally decreases the raw concentrations, which is expected given the well-known high bias in optical sensors due to relative humidity impacts, but it maintains the trends observed by the LCS. An exception to this is the Modulair-PM which comes with an existing relative humidity correction applied by the manufacturer.

We note that the time period of the UG colocation does not include any measurements during the Harmattan; it is confirmed from other data sources<sup>25,41,69</sup> that particle concentrations during this season can be much higher in the Harmattan than during the wet season. Further data collection during the Harmattan is necessary to address this gap.

**3.2. Accra Deployment—17 Clarity Nodes.** Figure 1 shows the locations of Clarity nodes deployed across Accra, with the inset showing Jamestown, a neighborhood in Accra that was targeted for a higher network density due to interest in understanding the impacts of a local pollution reduction education campaign by EPA Ghana and Environment360 (a Ghanaian nonprofit) that began after the LCS were installed.<sup>70</sup> Data from these monitors is corrected as discussed in Section 4.a.ii. Each pie chart indicates the percentage of measured days exceeding the daily WHO  $PM_{2.5}$  guideline ( $15\ \mu\text{g}/\text{m}^3$ ), with sites city-wide reporting between 37.6 and 95.7% of days above WHO guidelines. In Jamestown, sites report 87.2 to 100% of days above WHO guidelines. The Dansoman site reports the least percentage of days above WHO guidelines.

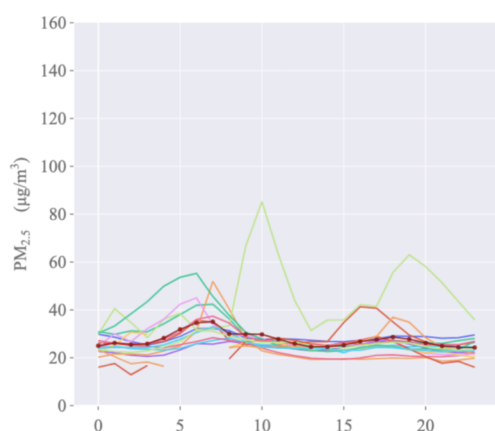
Figure 6 displays the time series of GMR-corrected monthly averages from each site in the Accra network, with the dotted red line showing the mean of all sites. The January 2019 and January 2020 Harmattan periods show a greater rise in ambient  $PM_{2.5}$  concentrations than the January 2021 Harmattan period. As shown in Figure 6, the December 2018 monthly average is  $1.9\times$  the April 2019 monthly average; the January 2020 monthly average is  $2.5\times$  the April 2020 monthly average. However, the January 2021 monthly average is only  $1.1\times$  the April 2021 monthly average. Other studies have confirmed the “weaker” effects of the January 2021 Harmattan. (The coronavirus lockdowns were not found to be associated with significant reductions in  $PM_{2.5}$  in this network; only 6 sites out of the 17 in the network had data and showed  $PM_{2.5}$  reductions in March–May 2019 compared to March–May 2020, with a monthly average reduction of 16.2%.) The small peaks during the summertime can be explained by precipitation seasonality in Accra, in which the center of the northward-advancing intertropical convergence zone (ITCZ) passes Accra in June and July and results in slightly reduced precipitation, resulting in less wet deposition and higher  $PM_{2.5}$  concentrations.<sup>71</sup> A similar pattern has been observed previously in Accra<sup>49</sup> and also in neighboring Lomé, Togo.<sup>39</sup>

All of the sites tend to follow similar regional trends, but some sites are consistently measuring higher  $PM_{2.5}$  (Malam Junction, Tetteh Quarshie) and some consistently lower  $PM_{2.5}$  (Dansoman, Ghana EPA). For the time they were online, the Jamestown and Jamestown Coast sites measured significantly higher  $PM_{2.5}$  than the rest of the network. The Jamestown Coast site is predominantly a commercial area characterized by thick

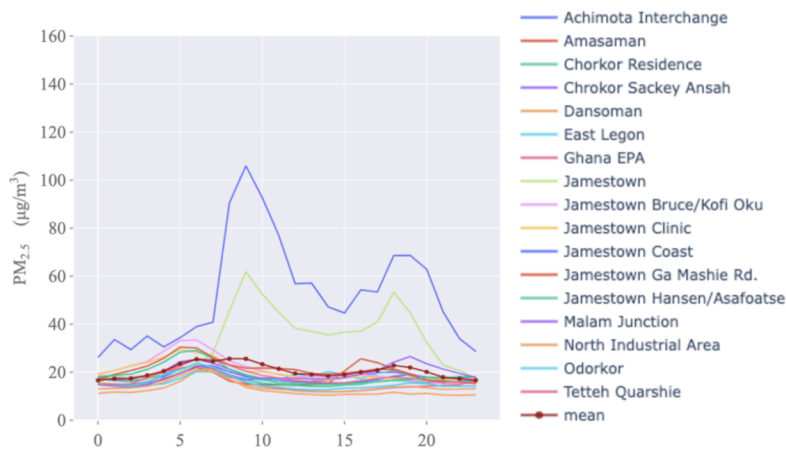


**Figure 7.** Violin plots of daily-averaged GMR-corrected  $PM_{2.5}$  across the Accra, Ghana Clarity network. The dashed gray line indicates the WHO Daily  $PM_{2.5}$  Guideline ( $15 \mu\text{g}/\text{m}^3$ ).

### (A) Harmattan



### (B) March to November



**Figure 8.** GMR-corrected diurnal averages, by hour of day, during the (A) Harmattan (December to February) and (B) outside (March to November). Note that the Jamestown Coast site has no data collected during the Harmattan.

smoke from singeing of slaughtered animals and fish smoking. The animal hides are usually burned using used car tires. Slaughterhouses, also known as abattoirs, have been linked to toxic air pollution emissions.<sup>72,73</sup>

Figure 7 shows the daily-averaged, GMR-corrected  $PM_{2.5}$  values from each Clarity monitor in the Accra network. Most locations show consistent daily averages surpassing the WHO Daily Guidelines, which is shown with a dashed gray line. For comparison, the distributions of manufacturer-reported  $PM_{2.5}$  from the University of Ghana colocation are shown in Figure S18. There is a drastic difference in the variability between collocated and deployed LCS, indicating that the variation in distributions of  $PM_{2.5}$  measured across the city are not purely factors of instrument noise.

The full network daily mean  $PM_{2.5}$  is  $23.4 \mu\text{g}/\text{m}^3$ , which is 1.6 times the WHO Daily  $PM_{2.5}$  guideline of  $15 \mu\text{g}/\text{m}^3$ . This is slightly lower but comparable to other studies which have found the mean daily  $PM_{2.5}$  in Accra to be  $26\text{--}37 \mu\text{g}/\text{m}^3$ .<sup>3,25,67,68</sup> The values in Figure 7 indicate unhealthy levels of ambient  $PM_{2.5}$ ; they are higher than mean daily concentrations in neighboring Lomé, Togo ( $23.5 \mu\text{g}/\text{m}^3$ )<sup>39</sup> but not as high as other cities in Africa such as Kinshasa (2019 average:  $43.5 \mu\text{g}/\text{m}^3$ ),<sup>41</sup> and other cities around the world such as Delhi (2007–2021 mean  $PM_{2.5}$ :

$125 \pm 86 \mu\text{g}/\text{m}^3$ )<sup>74</sup> or Isfahan, Iran (2014–2019 mean  $PM_{2.5}$ :  $29.9\text{--}50.9 \mu\text{g}/\text{m}^3$ ).<sup>75</sup> However, rapid development in the region, combined with evidence of Harmattan-linked decreasing visibility over the past 30 years,<sup>76</sup> creates a possibility that the ambient  $PM_{2.5}$  will rise in Accra in the coming years.

Monitor sites around Accra show some consistent trends. For example, at all sites outside of Jamestown, the upper quartile of daily  $PM_{2.5}$  measurements is below  $30 \mu\text{g}/\text{m}^3$ , but outliers are as high as  $334.5 \mu\text{g}/\text{m}^3$ . Some Jamestown sites show a larger spread than the rest of the network, despite having fewer measured days of valid data. The Jamestown Coast site is a clear outlier, and the relatively higher mean measurement could be due to the proximity of the site to a slaughterhouse. The Jamestown sites are within 1 km of each other but show quite different trends, further affirming the need for high-density LCS networks for assessing heterogeneity in ambient  $PM_{2.5}$  monitoring.

Figure S19 shows the annual averages across 4 years in the Accra network. Note that 2018 and 2021 are incomplete years, as shown in Figure S2. Across the 17 deployed Clarity nodes, 7 show decreasing annual  $PM_{2.5}$  averages, and 9 show increasing averages; 1 node only measured data in a single calendar year.

Figure 8 shows the difference in diurnal averages at each site during and outside the Harmattan. During the Harmattan, the

baseline PM<sub>2.5</sub> concentrations are elevated; the mean morning peak reaches 35.1  $\mu\text{g}/\text{m}^3$  at 7 AM and the mean evening peak is 28.6  $\mu\text{g}/\text{m}^3$  at 6 PM. Outside the Harmattan, the mean morning peak is 25.4  $\mu\text{g}/\text{m}^3$  at 6 AM, and the mean evening peak 22.8  $\mu\text{g}/\text{m}^3$  at 6 PM. The diurnal cycle showing peaks at hours of peak human activity indicates qualitatively that the main sources of pollution are anthropogenic. During the Harmattan season, there is a large background influence, possibly from dust aerosol, though local sources of pollution still drive the daily trends. Note that the Plantower PMS5003 sensors are limited in their ability to measure dust and other coarse particles larger than 1–2  $\mu\text{m}$ .<sup>77–80</sup>

#### 4. DISCUSSION

By comparing measurements from the collocation of 16 Clarity Node-S, 2 PurpleAir PA-II SD, and 2 QUANT-AQ Modulair-PM monitors over 4 months at the University of Ghana, we find that manufacturer-reported data from Modulair-PM monitors most closely correlates with data from a collocated Teledyne T640 reference monitor, followed by PurpleAir then Clarity nodes (Modulair-PM  $R^2 = 0.84$ , MAE = 3.04  $\mu\text{g}/\text{m}^3$ ; PurpleAir:  $R^2 = 0.82$ , MAE = 4.54  $\mu\text{g}/\text{m}^3$ ; Clarity:  $R^2 = 0.69$ , MAE = 13.68  $\mu\text{g}/\text{m}^3$ ). We then assess four machine learning models for correcting low-cost sensor data and find that while XGBoost and Random Forest raise the  $R^2$  and lower the mean absolute error the most, tree-based models like these fail at predicting PM<sub>2.5</sub> values outside the range of the training dataset, which in this case is the University of Ghana collocation. Though such models offer exciting possibilities for accurate correction of low-cost sensor data, they should only be used when the training dataset covers a broader range than the assumed range of the deployment measurements, since these models cannot extrapolate well beyond training ranges. Multiple linear regression can also be used for correcting low-cost sensor data; the trade-off for using this low-costing technique is slightly reduced performance. We recommend the usage of the Gaussian mixture regression technique, which is adept at extrapolating correction trends outside of the bounds of collocation training data and also provides large reductions in mean absolute error from the manufacturer-reported data. (Modulair-PM  $R^2 = 0.87$ , MAE = 2.04  $\mu\text{g}/\text{m}^3$ ; PurpleAir:  $R^2 = 0.86$ , MAE = 1.93  $\mu\text{g}/\text{m}^3$ ; Clarity:  $R^2 = 0.79$ , MAE = 2.27  $\mu\text{g}/\text{m}^3$ ). We use the Gaussian Mixture Regression technique to correct data from 17 Clarity monitors deployed around Accra, Ghana, to find that the mean daily average PM<sub>2.5</sub> in the city is 23.4  $\mu\text{g}/\text{m}^3$ , which is 1.6 times the WHO Daily PM<sub>2.5</sub> guideline of 15  $\mu\text{g}/\text{m}^3$ . We also find that during the Harmattan, mean morning peak concentrations can be elevated 1.2 times the non-Harmattan levels. Anthropogenic sources likely drive the PM<sub>2.5</sub> concentrations especially outside of the months of November through March. Stark heterogeneity between data from these monitors, some of which are within 1 km of each other but show vastly different measured distributions, demonstrates that, as Accra continues to grow, there will be a growing need for high-density ambient PM<sub>2.5</sub> monitoring, as well as further research about the performance (including degradation over time) of these monitors in this region. Potential caveats lie with the fact that the Harmattan season was not included in the collocation time period and that the entire 18-node Clarity network could not be collocated before deployment; however, these issues have been addressed in other work such as in McFarlane et al. (2021a).<sup>49</sup> The correction factors reported in this paper will be useful in future low-cost monitoring air quality studies in Accra, other parts of

West Africa, and other cities around the world with similar meteorologies, allowing researchers to deploy low-cost sensors and retrieve actionable data without performing their own unique collocations, which can be cumbersome to carry out. With the growing use of low-cost sensors on the African continent, it is vital that data from these be appropriately validated and calibrated.

#### ■ ASSOCIATED CONTENT

##### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.2c09264>.

Comparison of GMR-corrected PM<sub>2.5</sub> (Figure S1); timeline of clarity nodes deployed in Accra (Figure S2); scatter plots of LCS vs Reference PM<sub>2.5</sub> from UG collocation shaded by temperature (Figure S3); coefficients and standard errors for Multiple linear regression corrections of LCS (Table S1); multiple linear regression statistics for LCS (Figures S4–S6); XGBoost statistics for LCS (Figures S7–S9); random forest statistics for LCS (Figures S10–12); GMR statistics (Figures S13–S15); distribution of raw PM<sub>2.5</sub> during Accra deployment (Figure S16); comparison of machine learning-corrected LCS PM<sub>2.5</sub> (Figure S17); distributions of raw PM<sub>2.5</sub> during UG collocation (Figure S18); GMR-corrected annual PM<sub>2.5</sub> averages across Accra deployment (Figure S19); and Pearson correlations between Clarity nodes deployed around Accra (Figure S20) (PDF)

#### ■ AUTHOR INFORMATION

##### Corresponding Authors

**Garima Raheja** – Department of Earth and Environmental Sciences, Columbia University, New York, New York 10027, United States; Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, United States; [orcid.org/0000-0002-5037-7979](https://orcid.org/0000-0002-5037-7979); Email: [garima.raheja@columbia.edu](mailto:garima.raheja@columbia.edu)

**Daniel M. Westervelt** – Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, United States; NASA Goddard Institute for Space Science, New York, New York 10025, United States; [orcid.org/0000-0003-0806-9961](https://orcid.org/0000-0003-0806-9961); Email: [danielmw@ldeo.columbia.edu](mailto:danielmw@ldeo.columbia.edu)

##### Authors

**James Nimo** – Department of Physics, University of Ghana, Ghana, Ghana; African Institute of Mathematical Sciences, Kigali, Rwanda

**Emmanuel K.-E. Appoh** – Ghana Environmental Protection Agency, Accra, Ghana

**Benjamin Essien** – Ghana Environmental Protection Agency, Accra, Ghana

**Maxwell Sunu** – Ghana Environmental Protection Agency, Accra, Ghana

**John Nyante** – Ghana Environmental Protection Agency, Accra, Ghana

**Mawuli Amegah** – Ghana Environmental Protection Agency, Accra, Ghana

**Reginald Quansah** – School of Public Health, University of Ghana, Accra, Ghana

**Raphael E. Arku** – Department of Environmental Health Sciences, School of Public Health and Health Sciences,

University of Massachusetts, Amherst, Massachusetts 01003, United States

**Stefani L. Penn** – Industrial Economics, Inc, Cambridge, Massachusetts 02140, United States

**Michael R. Giordano** – Univ Paris Est Creteil, CNRS UMS 3563, Ecole Nationale des Ponts et Chaussées, Université de Paris, OSU-EFLUVE—Observatoire Sciences de L'Univers-Envelopes Fluides de La Ville à L'Exobiologie, F-94010 Créteil, France

**Zhonghua Zheng** – Department of Earth and Environmental Sciences, The University of Manchester, Manchester M13 9PL, U.K.; [orcid.org/0000-0002-0642-650X](https://orcid.org/0000-0002-0642-650X)

**Darby Jack** – Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States; [orcid.org/0000-0002-9932-0201](https://orcid.org/0000-0002-9932-0201)

**Steven Chillrud** – Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States

**Kofi Amegah** – University of Cape Coast, Cape Coast, Ghana

**R. Subramanian** – Univ Paris Est Creteil, CNRS UMS 3563, Ecole Nationale des Ponts et Chaussées, Université de Paris, OSU-EFLUVE—Observatoire Sciences de L'Univers-Envelopes Fluides de La Ville à L'Exobiologie, F-94010 Créteil, France; Kigali Collaborative Research Centre, Kigali, Rwanda; Present Address: Qatar Environment and Energy Research Institute, Hamad Bin Khalifa University, Doha, Qatar

**Robert Pinder** – Environmental Protection Agency, Raleigh, North Carolina 27709, United States

**Ebenezer Appah-Sampong** – Ghana Environmental Protection Agency, Accra, Ghana

**Esi Nerquaye Tetteh** – Ghana Environmental Protection Agency, Accra, Ghana

**Mathias A. Borketey** – Ghana Environmental Protection Agency, Accra, Ghana

**Allison Felix Hughes** – Department of Physics, University of Ghana, Ghana, Ghana

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.est.2c09264>

### Author Contributions

G.R.: data curation, formal analysis, methodology, software, validation, visualization, writing—original draft, writing—review and editing J.N.: writing—original draft, data curation, methodology E.K.-E.A.: conceptualization, data curation, funding acquisition, project administration, resources, supervision B.E.: data curation M.S.: data curation J.N.: data curation M.A.: data curation R.Q.: project administration, conceptualization R.E.A.: writing—review and editing, S.L.P.: conceptualization, project administration, funding acquisition, resources M.R.G.: writing—review and editing, data curation, methodology Z.Z.: methodology D.J.: conceptualization S.C.: conceptualization K.A.: data curation, resources, writing—review and editing R.S.: conceptualization, supervision, data curation R.P.: conceptualization, funding acquisition, resources, E.A.-S.: funding acquisition, supervision, E.N.T.: data curation M.A.B.: data curation A.F.H.: supervision, data curation, project administration, methodology D.M.W.: conceptualization, methodology, funding acquisition, project administration, supervision, resources, writing—review and editing.

### Funding

G.R. and D.M.W. acknowledge funding from the National Science Foundation Office of International Science and Engineering Grant Number 2020677. R.S. and M.R.G. acknowledge State funding managed by the Agence Nationale de la Recherche (ANR) under the “Programme d’Investissements d’Avenir” integrated into France 2030, under the reference ANR-18-MPGA-0011. The authors declare no conflict of interest. Corrected data from the Accra Clarity network can be found here.

### Notes

The authors declare no competing financial interest.

### REFERENCES

- (1) Fisher, S.; Bellinger, D. C.; Cropper, M. L.; Kumar, P.; Binagwaho, A.; Koudoukoupo, J. B.; Park, Y.; Taghian, G.; Landrigan, P. J. Air Pollution and Development in Africa: Impacts on Health, the Economy, and Human Capital. *Lancet Planet. Health* **2021**, *5*, e681–e688.
- (2) Abera, A.; Mattisson, K.; Eriksson, A.; Ahlberg, E.; Sahilu, G.; Mengistie, B.; Bayih, A. G.; Aseffaa, A.; Malmqvist, E.; Isaxon, C. Air Pollution Measurements and Land-Use Regression in Urban Sub-Saharan Africa Using Low-Cost Sensors—Possibilities and Pitfalls. *Atmosphere* **2020**, *11*, No. 1357.
- (3) Cohen, A. J.; Brauer, M.; Burnett, R.; Anderson, H. R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; Feigin, V.; Freedman, G.; Hubbell, B.; Jobling, A.; Kan, H.; Knibbs, L.; Liu, Y.; Martin, R.; Morawska, L.; Pope, C. A.; Shin, H.; Straif, K.; Shaddick, G.; Thomas, M.; van Dingenen, R.; van Donkelaar, A.; Vos, T.; Murray, C. J. L.; Forouzanfar, M. H. Estimates and 25-Year Trends of the Global Burden of Disease Attributable to Ambient Air Pollution: An Analysis of Data from the Global Burden of Diseases Study 2015. *Lancet* **2017**, *389*, 1907–1918.
- (4) Rao, S.; Pachauri, S.; Dentener, F.; Kinney, P.; Klimont, Z.; Riahi, K.; Schoepp, W. Better Air for Better Health: Forging Synergies in Policies for Energy Access, Climate Change and Air Pollution. *Glob. Environ. Change* **2013**, *23*, 1122–1130.
- (5) Balakrishnan, K.; Dey, S.; Gupta, T.; Dhaliwal, R. S.; Brauer, M.; Cohen, A. J.; Stanaway, J. D.; Beig, G.; Joshi, T. K.; Aggarwal, A. N.; Sabde, Y.; Sadhu, H.; Frostad, J.; Causey, K.; Godwin, W.; Shukla, D. K.; Kumar, G. A.; Varghese, C. M.; Muraleedharan, P.; Agrawal, A.; Anjana, R. M.; Bhansali, A.; Bhardwaj, D.; Burkart, K.; Cercy, K.; Chakma, J. K.; Chowdhury, S.; Christopher, D. J.; Dutta, E.; Furtado, M.; Ghosh, S.; Ghoshal, A. G.; Glenn, S. D.; Guleria, R.; Gupta, R.; Jeemon, P.; Kant, R.; Kant, S.; Kaur, T.; Koul, P. A.; Krish, V.; Krishna, B.; Larson, S. L.; Madhipatla, K.; Mahesh, P. A.; Mohan, V.; Mukhopadhyay, S.; Mutreja, P.; Naik, N.; Nair, S.; Nguyen, G.; Odell, C. M.; Pandian, J. D.; Prabhakaran, D.; Prabhakaran, P.; Roy, A.; Salvi, S.; Sambandam, S.; Saraf, D.; Sharma, M.; Shrivastava, A.; Singh, V.; Tandon, N.; Thomas, N. J.; Torre, A.; Xavier, D.; Yadav, G.; Singh, S.; Shekhar, C.; Vos, T.; Dandona, R.; Reddy, K. S.; Lim, S. S.; Murray, C. J. L.; Venkatesh, S.; Dandona, L. The Impact of Air Pollution on Deaths, Disease Burden, and Life Expectancy across the States of India: The Global Burden of Disease Study 2017. *Lancet Planet. Health* **2019**, *3*, e26–e39.
- (6) Cohen, A. J.; Anderson, H. R.; Ostro, B.; Pandey, K. D.; Krzyzanowski, M.; Künzli, N.; Gutschmidt, K.; Pope, A.; Romieu, I.; Samet, J. M.; Smith, K. The Global Burden of Disease Due to Outdoor Air Pollution. *J. Toxicol. Environ. Health A* **2005**, *68*, 1301–1307.
- (7) Johnson, M.; Shin, H. H.; Roberts, E.; Sun, L.; Fisher, M.; Hystad, P.; Van Donkelaar, A.; Martin, R. V.; Fraser, W. D.; Lavigne, E.; Clark, N.; Beaulac, V.; Arbuckle, T. E. Critical Time Windows for Air Pollution Exposure and Birth Weight in a Multicity Canadian Pregnancy Cohort. *Epidemiology* **2022**, *33*, 7–16.
- (8) Šrám, R. J.; Binkov, B.; Dejmeš, J.; Bobak, M. Ambient Air Pollution and Pregnancy Outcomes: A Review of the Literature. *Environ. Health Perspect.* **2005**, *113*, 375–382.

- (9) Woodruff, T. J.; Parker, J. D.; Kyle, A. D.; Schoendorf, K. C. Disparities in Exposure to Air Pollution during Pregnancy. *Environ. Health Perspect.* **2003**, *111*, 942–946.
- (10) Health Effects Institute. *State of Global Air 2020 Special Report*, 2020.
- (11) WHO. Report - WHO Air Quality Database (Update) 2022 <https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database>. (accessed March 22, 2023).
- (12) Singh, A.; Avis, W. R.; Pope, F. D. Visibility as a Proxy for Air Quality in East Africa. *Environ. Res. Lett.* **2020**, *15*, No. 084002.
- (13) Mudu, P. *Ambient Air Pollution and Health in Accra, Ghana*; World Health Organization: Geneva, 2021.
- (14) Marais, E. A.; Wiedinmyer, C. Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-Africa). *Environ. Sci. Technol.* **2016**, *50*, 10739–10745.
- (15) McDuffie, E. E.; Smith, S. J.; O'Rourke, P.; Tibrewal, K.; Venkataraman, C.; Marais, E. A.; Zheng, B.; Crippa, M.; Brauer, M.; Martin, R. V. A Global Anthropogenic Emission Inventory of Atmospheric Pollutants from Sector- and Fuel-Specific Sources (1970–2017): An Application of the Community Emissions Data System (CEDS). *Earth Syst. Sci. Data* **2020**, *12*, 3413–3442.
- (16) Robust relationship between air quality and infant mortality in Africa | Nature. <https://www.nature.com/articles/s41586-018-0263-3>. (accessed November 11, 2022).
- (17) Baayoun, A.; Itani, W.; El Helou, J.; Halabi, L.; Medlej, S.; El Malki, M.; Moukhadder, A.; Aboujaoude, L. K.; Kabakian, V.; Mounajed, H.; Mokalled, T.; Shihadeh, A.; Lakkis, I.; Saliba, N. A. Emission Inventory of Key Sources of Air Pollution in Lebanon. *Atmos. Environ.* **2019**, *215*, No. 116871.
- (18) Magnoudéwa, B. B. Rapport Sectoriel D' Inventaire Des Gaz A Effet De Serre Du Togo, 2021.
- (19) Anderson, C. M.; Kissel, K. A.; Field, C. B.; Mach, K. J. Climate Change Mitigation, Air Pollution, and Environmental Justice in California. *Environ. Sci. Technol.* **2018**, *52*, 10829–10838.
- (20) UNICEF. Silent Suffocation in Africa Air Pollution Is a Growing Menace, Affecting the Poorest Children the Most, 2019.
- (21) Air Pollution From Forest and Vegetation Fires in Southeast Asia Disproportionately Impacts the Poor - Reddington - 2021 - GeoHealth - Wiley Online Library. <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021GH000418>. (accessed November 11, 2022).
- (22) Schwanghart, W.; Schütt, B. Meteorological Causes of Harmattan Dust in West Africa. *Geomorphology* **2008**, *95*, 412–428.
- (23) He, C.; Breuning-Madsen, H.; Awadzi, T. W. Mineralogy of dust deposited during the Harmattan season in Ghana *Geogr. Tidsskr.-Dan. J. Geogr.* **107** 9 15 DOI: 10.1080/00167223.2007.10801371.
- (24) A long-term experimental study of the Saharan dust presence in West Africa - ScienceDirect 2008 <https://www.sciencedirect.com/science/article/pii/S0169809507001196>. (accessed November 11, 2022).
- (25) Alli, A. S.; Clark, S. N.; Hughes, A.; Nimo, J.; Bedford-Moses, J.; Baah, S.; Wang, J.; Vallarino, J.; Agyemang, E.; Barratt, B.; Beddows, A.; Kelly, F.; Owusu, G.; Baumgartner, J.; Brauer, M.; Ezzati, M.; Agyei-Mensah, S.; Arku, R. E. Spatial-Temporal Patterns of Ambient Fine Particulate Matter (PM<sub>2.5</sub>) and Black Carbon (BC) Pollution in Accra. *Environ. Res. Lett.* **2021**, *16*, No. 074013.
- (26) Weinstein, J. P.; Hedges, S. R.; Kimbrough, S. Characterization and Aerosol Mass Balance of PM<sub>2.5</sub> and PM<sub>10</sub> Collected in Conakry, Guinea during the 2004 Harmattan Period. *Chemosphere* **2010**, *78*, 980–988.
- (27) Israel, B. A.; Schulz, A. J.; Parker, E. A.; Becker, A. B. Community-Campus Partnerships for Health. Community-Based Participatory Research: Policy Recommendations for Promoting a Partnership Approach in Health Research. *Educ. Health* **2001**, *14*, 182–197.
- (28) Li, S.-T.; Shue, L.-Y. Data Mining to Aid Policy Making in Air Pollution Management. *Expert Syst. Appl.* **2004**, *27*, 331–340.
- (29) An integrated assessment of two decades of air pollution policy making in Spain: Impacts costs and improvements - ScienceDirect 2015 <https://www.sciencedirect.com/science/article/pii/S0048969715300735>. (accessed November 11, 2022).
- (30) Development of an integrated policy making tool for assessing air quality and human health benefits of air pollution control | SpringerLink 2015 <https://link.springer.com/article/10.1007/s11783-015-0796-8>. (accessed November 11, 2022).
- (31) Martin, R. V.; Brauer, M.; van Donkelaar, A.; Shaddick, G.; Narain, U.; Dey, S. No One Knows Which City Has the Highest Concentration of Fine Particulate Matter. *Atmos. Environ.: X* **2019**, *3*, No. 100040.
- (32) Mir Alvarez, C.; Hourcade, R.; Lefebvre, B.; Pilot, E. A Scoping Review on Air Quality Monitoring, Policy and Health in West African Cities. *Int. J. Environ. Res. Public Health* **2020**, *17*, No. 9151.
- (33) Giordano, M. R.; Malings, C.; Pandis, S. N.; Presto, A. A.; McNeill, V. F.; Westervelt, D. M.; Beekmann, M.; Subramanian, R. From Low-Cost Sensors to High-Quality Data: A Summary of Challenges and Best Practices for Effectively Calibrating Low-Cost Particulate Matter Mass Sensors. *J. Aerosol Sci.* **2021**, *158*, No. 105833.
- (34) Hagan, D. H.; Kroll, J. H. Assessing the Accuracy of Low-Cost Optical Particle Sensors Using a Physics-Based Approach. *Atmos. Meas. Tech.* **2020**, *13*, 6343–6355.
- (35) Morawska, L.; Thai, P. K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; Gao, J.; Hagler, G. S. W.; Jayaratne, R.; Kumar, P.; Lau, A. K. H.; Louie, P. K. K.; Mazaheri, M.; Ning, Z.; Motta, N.; Mullins, B.; Rahman, M. M.; Ristovski, Z.; Shafiei, M.; Tjondronegoro, D.; Westerdahl, D.; Williams, R. Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone? *Environ. Int.* **2018**, *116*, 286–299.
- (36) Raheja, G.; Harper, L.; Hoffman, A.; Gorby, Y.; Freese, L.; O'Leary, B.; Deron, N.; Smith, S.; Auch, T.; Goodwin, M.; Westervelt, D. M. Community-Based Participatory Research for Low-Cost Air Pollution Monitoring in the Wake of Unconventional Oil and Gas Development in the Ohio River Valley: Empowering Impacted Residents through Community Science. *Environ. Res. Lett.* **2022**, *17*, No. 065006.
- (37) Katoto, P. D. M. C.; Byamungu, L.; Brand, A. S.; Mokaya, J.; Strijdom, H.; Goswami, N.; De Boever, P.; Nawrot, T. S.; Nemery, B. Ambient Air Pollution and Health in Sub-Saharan Africa: Current Evidence, Perspectives and a Call to Action. *Environ. Res.* **2019**, *173*, 174–188.
- (38) Ouimette, J. R.; Malm, W. C.; Schichtel, B. A.; Sheridan, P. J.; Andrews, E.; Ogren, J. A.; Arnott, W. P. Evaluating the PurpleAir Monitor as an Aerosol Light Scattering Instrument. *Atmos. Meas. Tech. Discuss.* **2021**, 1–35.
- (39) Raheja, G.; Sabi, K.; Sonla, H.; Gbedjangni, E. K.; McFarlane, C. M.; Hodoli, C. G.; Westervelt, D. M. A Network of Field-Calibrated Low-Cost Sensor Measurements of PM<sub>2.5</sub> in Lomé, Togo, Over One to Two Years. *ACS Earth Space Chem.* **2022**, DOI: 10.1021/acsearthspacchem.1c00391.
- (40) Shusterman, A. A.; Teige, V. E.; Turner, A. J.; Newman, C.; Kim, J.; Cohen, R. C. The Berkeley Atmospheric CO<sub>2</sub> Observation Network: Initial Evaluation. *Atmos. Chem. Phys.* **2016**, *16*, 13449–13463.
- (41) McFarlane, C.; Isevlambire, P. K.; Lumbuenamo, R. S.; Ndinga, A. M. E.; Dhammapala, R.; Jin, X.; McNeill, V. F.; Malings, C.; Subramanian, R.; Westervelt, D. M. First Measurements of Ambient PM<sub>2.5</sub> in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-Calibrated Low-Cost Sensors. *Aerosol Air Qual. Res.* **2021**, *21*, No. 200619.
- (42) Baron, R.; Saffell, J. Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review. *ACS Sens.* **2017**, *2*, 1553–1566.
- (43) Crilley, L. R.; Shaw, M.; Pound, R.; Kramer, L. J.; Price, R.; Young, S.; Lewis, A. C.; Pope, F. D. Evaluation of a Low-Cost Optical Particle Counter (Alphasense OPC-N2) for Ambient Air Monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 709–720.
- (44) Kumar, V.; Sahu, M. Evaluation of Nine Machine Learning Regression Algorithms for Calibration of Low-Cost PM<sub>2.5</sub> Sensor. *J. Aerosol Sci.* **2021**, *157*, No. 105809.

- (45) Malings, C.; Tanzer, R.; Haurlyliuk, A.; Saha, P. K.; Robinson, A. L.; Presto, A. A.; Subramanian, R. Fine Particle Mass Monitoring with Low-Cost Sensors: Corrections and Long-Term Performance Evaluation. *Aerosol Sci. Technol.* **2020**, *54*, 160–174.
- (46) Nowack, P.; Konstantinovskiy, L.; Gardiner, H.; Cant, J. Towards Low-Cost and High-Performance Air Pollution Measurements using Machine Learning Calibration Techniques. *Atmos. Meas. Tech.* **2020**, *13*, 1–30.
- (47) Si, M.; Xiong, Y.; Du, S.; Du, K. Evaluation and Calibration of a Low-Cost Particle Sensor in Ambient Conditions Using Machine-Learning Methods. *Atmos. Meas. Tech.* **2020**, *13*, 1693–1707.
- (48) Adong, P.; Bainomugisha, E.; Okure, D.; Sserunjogi, R. Applying Machine Learning for Large Scale Field Calibration of Low-Cost PM<sub>2.5</sub> and PM<sub>10</sub> Air Pollution Sensors. *Appl. AI Lett.* **2022**, *3*, No. e76.
- (49) McFarlane, C.; Raheja, G.; Malings, C.; Appoh, E. K. E.; Hughes, A. F.; Westervelt, D. M. Application of Gaussian Mixture Regression for the Correction of Low Cost PM<sub>2.5</sub> Monitoring Data in Accra, Ghana. *ACS Earth Space Chem.* **2021**, *5*, 2268–2279.
- (50) PurpleAir, Inc. PurpleAir. Our Technology - PurpleAir Laser Particle Counters & Arduinos. 2021 <https://www2.purpleair.com/pages/technology>. (accessed October 18, 2021).
- (51) MODULAIR-PM. MODULAIR-PM 2022. <https://docs.quant-aq.com/modulair-pm>. (accessed November 11, 2022).
- (52) Air Quality Measurement Resources | Clarity Movement Co. <https://www.clarity.io/air-quality-monitoring-resources#documentation>. (accessed November 16, 2022).
- (53) Hagan, D. H.; Cross, E. S. Introduction to the MODULAIR-PM 2022 DOI: 10.5281/zenodo.7062168.
- (54) Bezantakos, S.; Costi, M.; Barmounis, K.; Antoniou, P.; Vouterakos, P.; Keleshis, C.; Sciare, J.; Biskos, G. Qualification of the Alphasense Optical Particle Counter for Inline Air Quality Monitoring. *Aerosol Sci. Technol.* **2021**, *55*, 361–370.
- (55) Sousan, S.; Koehler, K.; Hallett, L.; Peters, T. M. Evaluation of the Alphasense Optical Particle Counter (OPC-N2) and the Grimm Portable Aerosol Spectrometer (PAS-1.108). *Aerosol Sci. Technol.* **2016**, *50*, 1352–1365.
- (56) Particulate Matter Sensors | PM Sensor. Alphasense. <https://www.alphasense.com/products/optical-particle-counter/>. (accessed February 22, 2023).
- (57) Standard Operating Procedure Teledyne Model 640. 2019.
- (58) Model T640 with NumaView Software- T640 PM Mass Monitor 2022 <https://www.teledyne-api.com/products/particulate-instruments/t640>. (accessed November 11, 2022).
- (59) sklearnensembleRandomForestRegressor. scikit-learn. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>. (accessed November 11, 2022).
- (60) sklearnmixtureGaussianMixture. scikit-learn 2022 <https://scikit-learn/stable/modules/generated/sklearn.mixture.GaussianMixture.html>. (accessed November 11, 2022).
- (61) Fabisch, A. Gmr 2022 <https://github.com/AlexanderFabisch/gmr>. (accessed November 11, 2022).
- (62) XGBoost Documentation — xgboost 171 documentation 2022 <https://xgboost.readthedocs.io/en/stable/index.html>. (accessed November 11, 2022).
- (63) Kwak, H.-Y.; Ko, J.; Lee, S.; Joh, C.-H. Identifying the Correlation between Rainfall, Traffic Flow Performance and Air Pollution Concentration in Seoul Using a Path Analysis. *Transp. Res. Procedia* **2017**, *25*, 3552–3563.
- (64) Tian, X.; Cui, K.; Sheu, H.-L.; Hsieh, Y.-K.; Yu, F. Effects of Rain and Snow on the Air Quality Index, PM<sub>2.5</sub> Levels, and Dry Deposition Flux of PCDD/Fs. *Aerosol Air Qual. Res.* **2021**, *21*, No. 210158.
- (65) Yoo, J.-M.; Lee, Y.-R.; Kim, D.; Jeong, M.-J.; Stockwell, W. R.; Kundu, P. K.; Oh, S.-M.; Shin, D.-B.; Lee, S.-J. New Indices for Wet Scavenging of Air Pollutants (O<sub>3</sub>, CO, NO<sub>2</sub>, SO<sub>2</sub>, and PM<sub>10</sub>) by Summertime Rain. *Atmos. Environ.* **2014**, *82*, 226–237.
- (66) 32Tuning the hyper-parameters of an estimator. scikit-learn. [https://scikit-learn/stable/modules/grid\\_search.html](https://scikit-learn/stable/modules/grid_search.html). (accessed February 22, 2023).
- (67) Arku, R. E.; Vallarino, J.; Dionisio, K. L.; Willis, R.; Choi, H.; Wilson, J. G.; Hemphill, C.; Agyei-Mensah, S.; Spengler, J. D.; Ezzati, M. Characterizing Air Pollution in Two Low-Income Neighborhoods in Accra, Ghana. *Sci. Total Environ.* **2008**, *402*, 217–231.
- (68) Amegah, A. K.; Dakuu, G.; Mudu, P.; Jaakkola, J. J. K. Particulate Matter Pollution at Traffic Hotspots of Accra, Ghana: Levels, Exposure Experiences of Street Traders, and Associated Respiratory and Cardiovascular Symptoms. *J. Exposure. Sci. Environ. Epidemiol.* **2022**, *32*, 333–342.
- (69) Aboh, I. J. K.; Henriksson, D.; Laursen, J.; Lundin, M.; Ofofu, F. G.; Pind, N.; Selin Lindgren, E.; Wahnström, T. Identification of Aerosol Particle Sources in Semi-Rural Area of Kwabenya, near Accra, Ghana, by EDXRF Techniques. *X-Ray Spectrom.* **2009**, *38*, 348–353.
- (70) Stefani, P. Jamestown Open Burning Neighborhood Pilot Data Analysis Draft Report 2021.
- (71) Nicholson, S. E. The ITCZ and the Seasonal Cycle over Equatorial Africa. *Bull. Am. Meteorol. Soc.* **2018**, *99*, 337–348.
- (72) Singh, A. L.; Jamal, S.; Baba, S. A.; Islam, M. M. Environmental and Health Impacts from Slaughter Houses Located on the City Outskirts: A Case Study. *J. Environ. Prot.* **2014**, *05*, 566–575.
- (73) Jonah, A. E. Determination of Some Air Pollutants And Meteorological Parameters in Abattoir, NTAK INYANG in UYO L.G.A of AKWA IBOM State in Nigeria. *Int. J. Sci. Manag. Stud.* **2020**, *3*, 1–8.
- (74) Chetna; Dhaka, S. K.; Longiany, G.; Panwar, V.; Kumar, V.; Malik, S.; Rao, A. S.; Singh, N.; Dimri, A. P.; Matsumi, Y.; Nakayama, T.; Hayashida, S. Trends and Variability of PM<sub>2.5</sub> at Different Time Scales over Delhi: Long-Term Analysis 2007–2021. *Aerosol Air Qual. Res.* **2022**, *22*, No. 220191.
- (75) Hajizadeh, Y.; Jafari, N.; Mohammadi, A.; Momtaz, S. M.; Fanaei, F.; Abdollahnejad, A. Concentrations and Mortality Due to Short- and Long-Term Exposure to PM<sub>2.5</sub> in a Megacity of Iran (2014–2019). *Environ. Sci. Pollut. Res.* **2020**, *27*, 38004–38014.
- (76) Balarabe, M. A. The Thirty Years Trend Analysis of Harmattan Season Visibility and Temperature in Sahel Zone of Nigeria. *Phys. Mem. - J. Theor. Appl. Phys.* **2019**, *1*, 15–21.
- (77) Stavroulas, I.; Grivas, G.; Michalopoulos, P.; Liakakou, E.; Bougiatioti, A.; Kalkavouras, P.; Fameli, K. M.; Hatzianastassiou, N.; Mihalopoulos, N.; Gerasopoulos, E. Field Evaluation of Low-Cost PM Sensors (Purple Air PA-II) Under Variable Urban Air Quality Conditions, in Greece. *Atmosphere* **2020**, *11*, No. 926.
- (78) Robinson, D. L. Accurate, Low Cost PM<sub>2.5</sub> Measurements Demonstrate the Large Spatial Variation in Wood Smoke Pollution in Regional Australia and Improve Modeling and Estimates of Health Costs. *Atmosphere*. 2023 <https://www.mdpi.com/2073-4433/11/8/856>. (accessed February 22, 2023).
- (79) Kosmopoulos, G.; Salamalikis, V.; Pandis, S. N.; Yannopoulos, P.; Bloutsos, A. A.; Kazantzidis, A. Low-Cost Sensors for Measuring Airborne Particulate Matter: Field Evaluation and Calibration at a South-Eastern European Site. *Sci. Total Environ.* **2020**, *748*, No. 141396.
- (80) Kuula, J.; Friman, M.; Helin, A.; Niemi, J. V.; Aurela, M.; Timonen, H.; Saarikoski, S. Utilization of Scattering and Absorption-Based Particulate Matter Sensors in the Environment Impacted by Residential Wood Combustion. *J. Aerosol Sci.* **2020**, *150*, No. 105671.