 doi:10.20944/preprints202006.0158.v1

Using a Network of Locally Developed Low Cost Particulate Matter Sensors for Land Use Regression Modeling of PM2.5 in Urban Uganda

Eric S. Coker[1]*, Joel Ssematimba[2], Engineer Bainomugisha[2]

[1]Department of Environmental and Global Health, College of Public Health and Health, Professions, University of Florida, 1255 Center Dr., Gainesville, FL.; eric.coker@phhp.ufl.edu

[2]AirQo, Department of Computer Science, College of Computing and Information Sciences, Makerere University, Plot 56 Pool Road, Kampala, Uganda.

*Corresponding Author

Acknowledgements

We would like to acknowledge the tireless efforts of the AirQo staff who assisted with data collection and data management.

ABSTRACT

Background

There are major air pollution monitoring gaps in sub-Saharan Africa. Developing capacity in the region to conduct air monitoring in the region can help estimate exposure to air pollution for epidemiology research. The purpose of our study is to develop a land use regression (LUR) model using low-cost air quality sensors developed by a research group in Uganda (AirQo).

Methods

Using these low-cost sensors, we collected continuous measurements of fine particulate matter (PM2.5) between May 1, 2019 and February 29, 2020 at 22 monitoring sites across urban municipalities of Uganda. We compared average monthly PM2.5 concentrations from the AirQo sensors with measurements from a BAM-1020 reference monitor operated at the US Embassy in Kampala. Monthly PM2.5 concentrations were used for LUR modeling. We used eight Machine Learning (ML) algorithms and ensemble modeling; using 10-fold cross validation and root mean squared error (RMSE) to evaluate model performance.

Results

Monthly PM2.5 concentration was 60.2 μg/m^3 (IQR: 45.4-73.0 μg/m^3; median= 57.5 μg/m^3). For the ML LUR models, RMSE values ranged between 5.43 μg/m^3 - 15.43 μg/m^3 and explained between 28% and 92% of monthly PM2.5 variability. Generalized additive models explained the largest amount of PM2.5 variability (R^2=0.92) and produced the lowest RMSE (5.43 μg/m^3) in the held-out test set. The most important predictors of monthly PM2.5 concentrations included monthly precipitation, major roadway density, population density, latitude, greenness, and percentage of households using solid fuels.

Conclusion

To our knowledge, ours is the first study to model the spatial distribution of urban air pollution in sub-Saharan Africa using air monitors developed from the region itself. Non-parametric ML for LUR modeling performed with high accuracy for prediction of monthly PM2.5 levels. Our analysis suggests that locally produced low-cost air quality sensors can help build capacity to conduct air pollution epidemiology research in the region.

KEYWORDS

land use regression, low-cost sensors, machine learning, particulate matter, Africa

1. Introduction

Data gaps in lower and middle-income countries (LMICs) related to environmental pollution is limiting environmental policy development and governance as well as our understanding of health impacts from pollution in LMICs. Low-cost sensors (LCS) hold great promise for being able to bridge these environmental pollution data gaps in LMICs. (Amegah, 2018) The widespread use of LCSs in LMIC settings, however, is yet to be realized. This underutilization of LCS in LMICs is due to both technical and non-technical reasons, including: (1) limitations of data quality collected by LCSs; (2) a lack of downstream data analytics applications for LCSs; and (3) a lack of consideration for sustainable operating mechanisms and physical and socioeconomic contexts in LMICs.(Amegah, 2018; Mao et al., 2019) Despite their current limitations, low-cost air quality sensors (LCAQS) have made substantial progress in terms of acceptance for their use in certain air pollution measurement and research applications. (Amegah, 2018; Clements et al., 2017; Malings et al., 2020; Masiol et al., 2019, 2018; McKercher and Vanos, 2018; Weissert et al., 2020, 2019)

Emergent LCAQS applications include the capability to enhance air quality regulatory monitoring by improving spatial and temporal resolution of current air monitoring programs, (Malings et al., 2020; McKercher and Vanos, 2018) and identifying particulate matter (PM) sources in complex urban environments. (Hagan et al., 2019) Recent studies conducted in the U.S. suggest that air pollution data collected using LCAQS can also help with generating spatio-temporal models that can reliably predict fine spatial-scale urban air pollution concentrations. (Masiol et al., 2019, 2018; Weissert et al., 2020, 2019) The present study builds off of these recent advances in air pollution-modeling by using LCAQS data for a spatial air pollution-prediction model. Where our study differs, however, is we implement the study in the LMIC context of urban Uganda.

What makes our study particularly unique is that we are using a spatially dense network of LCAQS that have been designed and fabricated locally in Uganda. These LCAQS developed by

AirQo are the first, to our knowledge, to originate from a sub-Saharan Africa (SSA) country. Such locally designed and produced LCAQS can plausibly address several limitations of other LCAQS, including offering a more sustainable operating mechanism as well as creating an LCAQS designed to operate in the challenging urban SSA infrastructural, socioeconomic and environmental context. For instance, the LCAQS used in our study, named AirQo, are designed and optimized to work in places characterized by sporadic internet connectivity, irregular power supply, high temperatures and dusty environments. The devices include a custom designed filtration system to minimize clogging, dust deposition, and reduce insect infestation common in the SSA region. Therefore, this study is motivated by a proof-of-concept in terms of using locally-sourced LCAQS for developing a LUR model to be employed in future ambient air pollution epidemiology research in the SSA region.

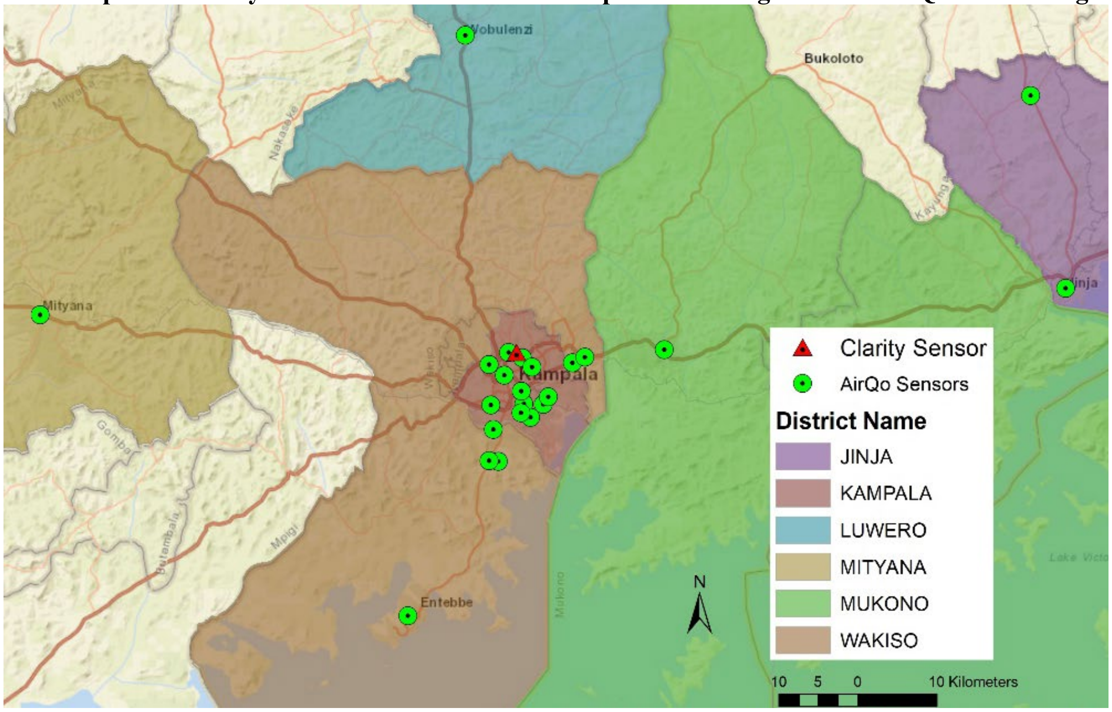
Moreover, conventional LUR air pollution modeling is implemented using multivariable linear regression and often applies K-fold cross-validation to validate the prediction model. (Brokamp et al., 2017; Eeftens et al., 2012; Mao et al., 2012; Sahsuvaroglu et al., 2006) Recent advances in LUR air pollution modeling suggests that Machine Learning (ML) algorithms, such as Random Forests (RF), helps deal with overfitting and relaxing assumptions of linearity. (Araki et al., 2018; Beckerman et al., 2013; Brokamp et al., 2017; Di et al., 2019; Rahman et al., 2020; Weissert et al., 2020, 2019) Hence, our study takes a ML approach to LUR modeling, using the data generated from the LCAQS network described in this study.

2. Materials and Methods

 doi:10.20944/preprints202006.0158.v1

The country of Uganda straddles the equator and is located in the East Africa region of SSA. The study area (Figure 1) encompasses six urban sites of Uganda's central and eastern region, including Jinja, Kampala, Luwero, Mityana, Mukono, and Wakiso districts. Uganda's capital city, Kampala, where nearly two-thirds (n=14) of the LCAQS were placed in our study, is located along the northern shores of Lake Victoria at an altitude of approximately 1,140 meters above sea level. (Fuhrimann et al., 2015) The districts included in the study have a wide range of population sizes; ranging from ~2.0 million, 1.5 million, 0.6 million, 0.47 million, 0.46 million, and 0.33 million, for Wakiso, Kampala, Mukono, Jinja, Luwero, and Mityana, respectively. (UBOS, 2014)

Figure 1. Map of the Study Area's Six Districts and the Spatial Coverage of the LCAQS Monitoring Sites.



 doi:10.20944/preprints202006.0158.v1

2.2. PM2.5 Measurements

2.2.1. Sensor Network

The AirQo devices measure particulate matter (PM) PM2.5 and PM10 using a nephelometer (light-scattering) technology. The devices also measure location (latitude, longitude) and meteorology parameters including internal and external temperature, atmospheric pressure, and humidity. The AirQo devices transmit data over a local Global System for Mobile Communications (GSM) network every 90 seconds and can run off solar or mains. Currently, we have deployed devices at static locations and mobile monitors (e.g., motorcycle taxis) thereby forming a network of both fixed and dynamic nodes. Currently, the sensor network includes 65 nodes with 40 in Kampala area and 25 in other urban areas of Uganda. In this study we use the data from the fixed monitoring locations only and have restricted the data to monitors that have been in operation for at least 75% of the study period (n=22 AirQo sensors). We installed devices between 2.5 and 4 meters high. Sensor placement is determined on a number of spatial features including population density, land use, road network, pollution sources and receptors, economic activities, and practical limitations, among others. The fixed installation locations include private property, schools, buildings, and lighting poles. Depending on the installation location, we fabricated custom mountings to support and secure the air quality monitor. To ensure data quality, at least one AirQo devices is co-located near (~10 meters) a Beta Attenuation Mass Monitor (BAM)1020 reference monitor currently installed and operated at the U.S. Embassy in Kampala. Additionally, for internal data quality assurance, each device includes two PM sensors. This dual sensor approach enables us to rapidly compare a given sensor against its twin sensor in order to detect any problems for the sensor. The collected data are transmitted in near real-time to a cloud-based platform. In addition to the AirQo sensors, we also used another LCAQS known as Clarity Node (n=1). The Clarity sensor uses a similar nephelometer technology that the AirQo device uses to detect PM2.5. Additionally, the Clarity sensor transmits PM monitoring data over a local GSM network in near real-time to a cloud. (Pantelic et al., 2019) The majority of the LCAQS used in this study are AirQo sensors (n=22) while only one Clarity Node-S sensor was used.

2.3. $PM_{2.5}$ Estimation

2.3.1. Predictor Variables

We assembled 18 predictor variables for LUR modeling. We define these variables using four broad categories, including spatial variables, meteorological variables, land use variables, and demographic variables. Table 1 summarizes the relevant information for each of the predictor variables in terms of their range of buffer sizes, spatial resolution, data format, and references.

Table 1. Predictor Variables used for LUR Modeling.

Variable	**Buffer Size/Resolution/Spatial Unit**	**Data Format**	**Reference**
Meteorological and Spatial Predictors			
Precipitation	Inches (monthly averages; 2005-2015)	Tabular/Vector	(https://www.timeanddate.com/weather/uganda/entebbe/climate, n.d.)
Latitude		Tabular/Vector	
Longitude		Tabular/Vector	
Elevation	100m	Raster values transformed into Vector for analysis	(USGS, n.d.)
Land Use Predictors			
Major Roadways	250m buffer	Raster values transformed into Vector for analysis	(OpenStreetMap, n.d.)
Major Roadways	500m buffer	Raster values transformed into	(OpenStreetMap, n.d.)

		Vector for analysis	
Major Roadways	750m buffer	Raster values transformed into Vector for analysis	(OpenStreetMap, n.d.)
Greenness (NDVI)	250m buffer/250m	Raster values transformed into Vector for analysis	(FEWSNET, n.d.)
Greenness (NDVI)	500m buffer/250m	Raster values transformed into Vector for analysis	(FEWSNET, n.d.)
Greenness (NDVI)	750m buffer/250m	Raster values transformed into Vector for analysis	(FEWSNET, n.d.)
Demographic and Household Predictors			
Number of People	Parish-level	Tabular/Vector	(UBOS, n.d.)
Number of Households	Parish-level	Tabular/Vector	(UBOS, n.d.)
Household Density (number of households/Parish area)	Parish-level	Raster values transformed into Vector for analysis	(UBOS, n.d.)
Percent Households Solid Fuel Use	Parish-level	Tabular/Vector	(UBOS, n.d.)
Population Density	~2km	Tabular/Vector	(HDX, n.d.)
Population Density	250m buffer/~2km	Tabular/Vector	(HDX, n.d.)
Population Density	500m buffer/~2km	Tabular/Vector	(HDX, n.d.)
Population Density	750m buffer/~2km	Tabular/Vector	(HDX, n.d.)

2.3.2. Statistical Analysis and Land Use Regression Modeling

We used PM2.5 concentration data from 23 LCAQS in total, including 22 sensors from the AirQo network and one Clarity sensor. We used sensor data collected between May 1, 2019 and February 29, 2020. Monthly PM2.5 air concentration averages were computed (n=218 observations) and combined with covariates for the LUR modeling. We calculated summary statistics for the monthly averages overall for the study area and stratified by month and district.

 doi:10.20944/preprints202006.0158.v1

We then calculated Pearson correlation coefficients between monthly PM2.5 averages and land use variables. Since the distribution of monthly averaged PM2.5 measurements were highly right-skewed, we log-transformed PM2.5 concentrations for the machine learning LUR (ML-LUR) modeling described in turn. We used monthly PM2.5 averages for modeling since the intended purpose of these exposure estimates is for predicting trimester-specific and entire pregnancy PM2.5 exposure averages for a future birth cohort study, as has been done in previous studies. (Coker et al., 2015)

For the ML-LUR algorithms, the combined PM2.5 and LUR dataset was first split into a training set (90%) and validation test set (10%). Next, we performed 10-fold cross-validation on the training set (n=198 observations) only, using root mean squared error (RMSE) to guide each base learner model. Eight different ML algorithms were fit in order to compare each learner's performance. These models include linear regression model (LM), Support Vector Machines with Radial Basis Function Kernel (SVM), Random Forest (RF), Quantile Random Forest (QRF), eXtreme Gradient Boosting (xgbTree), Generalized Additive Model (GAM), Lasso and Elastic-Net Regularized Generalized Linear Models (GLMNET), and Least Angle Regression (LARS). All 18 covariates described in Table 1, which included land use variables (e.g., major roadway density, greenspace), population demographic variables, and historical precipitation data, were included in the analysis. We implemented the base ML algorithms using the *caret* package in R, with the *'caretList'* command used to fit all ML algorithms in parallel. In addition to the individual base learner models already mentioned, we performed ensemble modeling using the caret package in order to assess whether improved ML-LUR model performance is achieved through ensemble modeling as seen in (Di et al., 2019 and Lim et al., 2019). (Di et al., 2019; Lim et al., 2019) We implemented ensemble modeling with the *caretEnsemble* package in R, using approaches offered by the '*caretEnsemble'* (CE) and *'caretStack'* (CS) commands. For the CE approach, we applied GLM to create a linear combination of all base learner models. Whereas in the CS approach we applied a stacked caret approach that combined the results from multiple component caret models. Since there were strong correlations between results from the base learner models, we used GLMNET when applying the stacked approach. Our final assessment of RMSE and R^2 applied to the 10% held-out test set only since this should better represent the ability of the ML-LUR to predict monthly PM2.5 concentrations at unmeasured locations for our study area.

3. RESULTS

3.1. PM2.5 Monitoring Results

Average monthly PM2.5 concentrations for the entire study area was 60.2 μg/m^3 (IQR: 45.4-73.0 μg/m^3; median= 57.5 ug/m^3). According to Figure 2a, monitoring sites in Luwero and Mukono Districts exhibited the highest PM2.5 levels. As expected, according to Figure 2b, elevated PM2.5 concentrations were observed to be lowest during the wet season and highest during the dry season.

3.1.1. Comparison of AirQo with a reference monitor

For comparison, we co-located an AirQo sensor with a BAM1020 reference monitor located at the US Embassy in Kampala. The mean monthly PM2.5 concentrations were 63.1 μg/m^3 and 60.2 μg/m^3 for the BAM1020 and AirQo monitors, respectively. Figure 3 plots the monthly PM2.5 averages of the BAM1020 embassy monitor versus the AirQo sensor. With an RMSE of 5.58 μg/m^3, normalized RMSE of 8.8%, and an R^2 of 0.87, the AirQo sensor compare well with the BAM1020 in terms of monthly averages.

Figure 2. PM2.5 Concentrations (May 2019-February 2020) by (a) District and (b) Month.

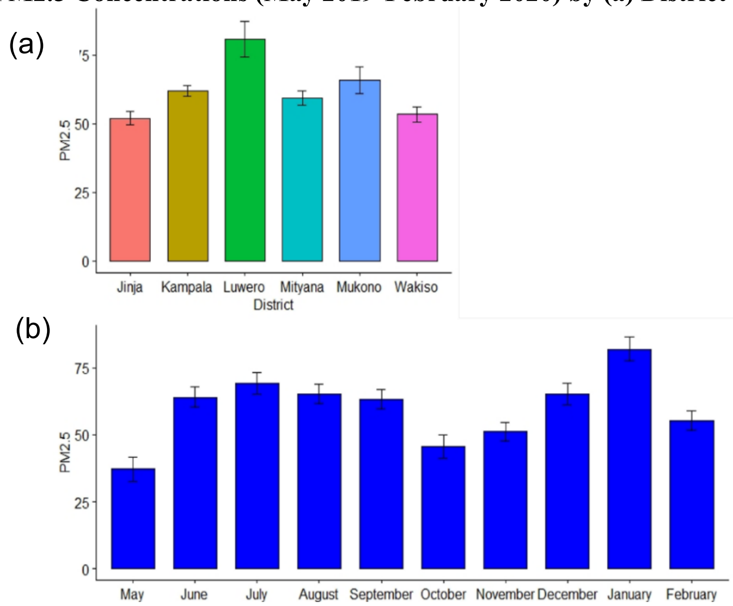
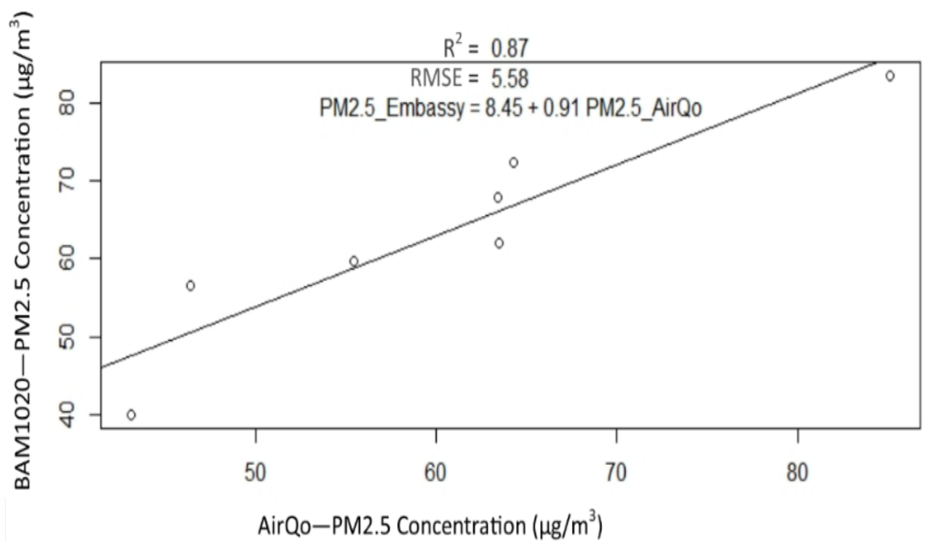


Figure 3. Average Monthly PM2.5 Concentrations from AirQo and BAM1020 Monitors.



3.2. ML-LUR Results

We summarize the RMSE and R^2 values for the base learner models and ensemble models in Table 2 (for log-transformed and exponentiated values). These values were computed using the held-out test set (N=20) only. The GAM resulted in the lowest RMSE as well as highest R^2 values (R^2=0.94) for the log-transformed values. Even the ensemble models performed quite

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2020 doi:10.20944/preprints202006.0158.v1

well, as shown in Table 2, both the GAM and xgbTree models outperformed the ensemble models. The exponentiated predictions exhibit a similar pattern as the log-transformed values, indicating the GAM with the lowest RMSE (5.43 µg/m^3) and highest R^2 (0.92) values.

Table 2. Performance (RMSE and R^2) for Models

Model	RMSE [a]	$R^{2\,a}$	RMSE [b]	$R^{2\,b}$
GAM	0.083	0.94	5.43	0.92
xgbTree	0.111	0.87	6.75	0.85
Stacked Ensemble (glmnet)	0.117	0.85	7.10	0.84
Ensemble (lm)	0.121	0.83	7.35	0.83
RF	0.148	0.81	7.60	0.82
QRF	0.183	0.70	9.62	0.69
SVM	0.195	0.54	12.6	0.45
LM	0.198	0.54	13.2	0.42
LARS	0.210	0.50	13.9	0.39
GLMNET	0.242	0.37	15.4	0.28

[a]Log-transformed (not exponentiated)
[b]Exponentiated

3.2.1. Variable Importance

As shown in Figure A1 in the Appendix, several of the LUR variables are moderately to highly correlated with one another. After extracting the variable importance values of study variables, as calculated from the top performing model (GAM), we are able to rank the ML-LUR variables in terms of predicting monthly PM2.5 concentrations. According to Figure A2, precipitation, greenness (NDVI), roadway density, latitude, and solid fuel usage are the top-ranking variables. When restricting our analysis to the top-5 predictor categories (precipitation, NDVI, roadway density, latitude, and solid fuel use) only, the GAM model explained 88% of the monthly PM2.5 concentration variability using the entire data set (data not shown).

Discussion

In our study, we leveraged air quality data from a network of mostly locally designed and produced LCAQS that were then used to predict estimates of PM2.5 in urban districts of Uganda. Moreover, we applied ML to optimize the LUR model. Importantly, we find that the AirQo sensors compare well against a BAM1020 reference monitor co-located at the U.S. Embassy. In general, when using predictors typically used in LUR modeling, the non-parametric ML algorithms performed the best in terms of being able to accurately predict monthly PM2.5 concentrations when compared to parametric modeling (e.g., linear model).

Of the land use predictors considered in our study, several stood out as strong predictors. The strongest predictors include precipitation, greenness, density of major roadways, latitude, solid fuel usage, and population density. To our knowledge, our study is the first to use population census data on solid fuel usage (at the Parish-level) in a LUR model. Specifically, we find that higher solid fuel usage is positively correlated with monthly PM2.5 concentrations. This finding suggests that area-level solid fuel use data can help inform LUR prediction models for PM2.5 in lower income SSA urban areas, and potentially other regions with high levels of solid fuel usage. Previous LUR prediction models for PM2.5 in SSA have been shown to have relatively poorer performance(Saucy et al., 2018; Tularam, 2019) compared to gaseous pollutant models for SSA or PM2.5 models developed in higher income regions. Given our results, as well as other air pollution research conducted in urban SSA that also show strong correlations between neighborhood-level solid fuel use and outdoor PM concentrations(Zhou et al., 2011), we suggest future modeling efforts in this region should incorporate solid fuel use data to improve PM2.5 modeling predictions.

To our knowledge, this is the first study to use LCAQS for LUR modeling in SSA. As suggested by previous authors(Amegah, 2018), we demonstrate that LCAQS hold strong potential for providing highly spatially resolved PM2.5 measurement data that can be harnessed for exposure estimation in air pollution epidemiology research. While data can be integrated to improve model performance, such as aerosol optical depth (AOD) remote sending data, we are encouraged by our findings. Future analyses will focus on optimizing calibration approaches for the AirQo PM sensor data. Since accurate measurement of PM2.5 with light scattering sensors can be limited by accuracy errors caused by environmental parameters such as relative humidity and temperature and may be subject to drift(US EPA, n.d.), we will use a co-located reference method (e.g., BAM1020) and model the influence of relative humidity (RH) and temperature on measurement accuracy; which can then be used in turn for regression-based calibration purposes in future epidemiology research. (Wang et al., 2019)

Conclusion

Deploying LCAQS can help address the urgent and growing need for expanding and improving air quality monitoring in resource-limited settings of SSA. With reasonably accurate predictions of PM2.5 using ML-LUR with 10-fold cross-validation, data from the locally developed AirQo sensors used in the present study provided evidence suggesting that they can be used for modeling exposures for a birth cohort study.

References

Amegah, A.K., 2018. Proliferation of low-cost sensors. What prospects for air pollution epidemiologic research in Sub-Saharan Africa? Environ. Pollut. 241, 1132–1137. https://doi.org/10.1016/j.envpol.2018.06.044

Araki, S., Shima, M., Yamamoto, K., 2018. Spatiotemporal land use random forest model for estimating metropolitan NO2 exposure in Japan. Sci. Total Environ. 634, 1269–1277. https://doi.org/10.1016/j.scitotenv.2018.03.324

Beckerman, B.S., Jerrett, M., Martin, R.V., van Donkelaar, A., Ross, Z., Burnett, R.T., 2013. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. Atmos. Environ. 77, 172–177. https://doi.org/10.1016/j.atmosenv.2013.04.024

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G., Ryan, P., 2017. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. Atmos. Environ. 151, 1–11. https://doi.org/10.1016/j.atmosenv.2016.11.066

Clements, A.L., Griswold, W.G., Rs, A., Johnston, J.E., Herting, M.M., Thorson, J., Collier-Oxandale, A., Hannigan, M., 2017. Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). Sensors 17, 2478. https://doi.org/10.3390/s17112478

Coker, E., Ghosh, J., Jerrett, M., Gomez-Rubio, V., Beckerman, B., Cockburn, M., Liverani, S., Su, J., Li, A., Kile, M.L., Ritz, B., Molitor, J., 2015. Modeling spatial effects of PM2.5 on term low birth weight in Los Angeles County. Environ. Res. 142, 354–364. https://doi.org/10.1016/j.envres.2015.06.044

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz, J., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909. https://doi.org/10.1016/j.envint.2019.104909

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dėdelė, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Mölter, A., Nádor, G., Nieuwenhuijsen, M., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E., Sugiri, D., Tsai, M.-Y., Yli-Tuomi, T., Varró, M.J., Vienneau, D., Klot, S. von, Wolf, K., Brunekreef, B., Hoek, G., 2012. Development of Land Use Regression Models for $PM_{2.5}$, $PM_{2.5}$ Absorbance, PM_{10} and PM_{coarse} in 20 European Study Areas;

Results of the ESCAPE Project. Environ. Sci. Technol. 46, 11195–11205. https://doi.org/10.1021/es301948k
FEWSNET, n.d. Products | Early Warning and Environmental Monitoring Program [WWW Document]. URL https://earlywarning.usgs.gov/fews/product/448 (accessed 2.24.20).
Fuhrimann, S., Stalder, M., Winkler, M.S., Niwagaba, C.B., Babu, M., Masaba, G., Kabatereine, N.B., Halage, A.A., Schneeberger, P.H.H., Utzinger, J., Cissé, G., 2015. Microbial and chemical contamination of water, sediment and soil in the Nakivubo wetland area in Kampala, Uganda. Environ. Monit. Assess. 187, 475. https://doi.org/10.1007/s10661-015-4689-x
Hagan, D.H., Gani, S., Bhandari, S., Patel, K., Habib, G., Apte, J.S., Hildebrandt Ruiz, L., Kroll, J.H., 2019. Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India. Environ. Sci. Technol. Lett. 6, 467–472. https://doi.org/10.1021/acs.estlett.9b00393
HDX, n.d. High Resolution Population Density Maps - Humanitarian Data Exchange [WWW Document]. URL https://data.humdata.org/dataset/highresolutionpopulationdensitymaps (accessed 2.24.20).
https://www.timeanddate.com/weather/uganda/entebbe/climate, n.d. Climate & Weather Averages in Entebbe, Uganda [WWW Document]. URL

https://www.timeanddate.com/weather/uganda/entebbe/climate (accessed 2.24.20).

Lim, C.C., Kim, H., Vilcassim, M.J.R., Thurston, G.D., Gordon, T., Chen, L.-C., Lee, K., Heimbinder, M., Kim, S.-Y., 2019. Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. Environ. Int. 131, 105022. https://doi.org/10.1016/j.envint.2019.105022

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P.K., Robinson, A.L., Presto, A.A., Subramanian, R., 2020. Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation. Aerosol Sci. Technol. 54, 160–174. https://doi.org/10.1080/02786826.2019.1623863

Mao, F., Khamis, K., Krause, S., Clark, J., Hannah, D.M., 2019. Low-Cost Environmental Sensor Networks: Recent Advances and Future Directions. Front. Earth Sci. 7, 221. https://doi.org/10.3389/feart.2019.00221

Mao, L., Qiu, Y., Kusano, C., Xu, X., 2012. Predicting regional space–time variation of PM2.5 with land-use regression model and MODIS data. Environ. Sci. Pollut. Res. 19, 128–138. https://doi.org/10.1007/s11356-011-0546-9

Masiol, M., Squizzato, S., Chalupa, D., Rich, D.Q., Hopke, P.K., 2019. Spatial-temporal variations of summertime ozone concentrations across a metropolitan area using a network of low-cost monitors to develop 24 hourly land-use regression models. Sci. Total Environ. 654, 1167–1178. https://doi.org/10.1016/j.scitotenv.2018.11.111

Masiol, M., Zíková, N., Chalupa, D.C., Rich, D.Q., Ferro, A.R., Hopke, P.K., 2018. Hourly land-use regression models based on low-cost PM monitor data. Environ. Res. 167, 7–14. https://doi.org/10.1016/j.envres.2018.06.052

McKercher, G.R., Vanos, J.K., 2018. Low-cost mobile air pollution monitoring in urban environments: a pilot study in Lubbock, Texas. Environ. Technol. 39, 1505–1514. https://doi.org/10.1080/09593330.2017.1332106

OpenStreetMap, n.d. OpenStreetMap [WWW Document]. OpenStreetMap. URL https://www.openstreetmap.org/ (accessed 2.24.20).

Pantelic, J., Dawe, M., Licina, D., 2019. Use of IoT sensing and occupant surveys for determining the resilience of buildings to forest fire generated PM2.5. PLOS ONE 14, e0223136. https://doi.org/10.1371/journal.pone.0223136

Rahman, M.M., Karunasinghe, J., Clifford, S., Knibbs, L.D., Morawska, L., 2020. New insights into the spatial distribution of particle number concentrations by applying non-parametric land use regression modelling. Sci. Total Environ. 702, 134708. https://doi.org/10.1016/j.scitotenv.2019.134708

Sahsuvaroglu, T., Arain, A., Kanaroglou, P., Finkelstein, N., Newbold, B., Jerrett, M., Beckerman, B., Brook, J., Finkelstein, M., Gilbert, N.L., 2006. A Land

Use Regression Model for Predicting Ambient Concentrations of Nitrogen Dioxide in Hamilton, Ontario, Canada. J. Air Waste Manag. Assoc. 56, 1059–1069. https://doi.org/10.1080/10473289.2006.10464542

Saucy, A., Röösli, M., Künzli, N., Tsai, M.-Y., Sieber, C., Olaniyan, T., Baatjies, R., Jeebhay, M., Davey, M., Flückiger, B., Naidoo, R., Dalvie, M., Badpa, M., de Hoogh, K., 2018. Land Use Regression Modelling of Outdoor NO2 and PM2.5 Concentrations in Three Low Income Areas in the Western Cape Province, South Africa. Int. J. Environ. Res. Public. Health 15, 1452. https://doi.org/10.3390/ijerph15071452

Tularam, H., 2019. Land use Regression Model for Exposure Assessment in the Mace Birth Cohort Study in Ethekwini: Environ. Epidemiol. 3, 401. https://doi.org/10.1097/01.EE9.0000610480.85058.4a

UBOS, 2014. National Population and Housing Census. Kampala, UG.

UBOS, n.d. Visualizations – Uganda Bureau of Statistics. URL https://www.ubos.org/data-portals-2/visualizations/ (accessed 2.24.20).

US EPA, n.d. collocation_instruction_guide.pdf [WWW Document]. URL https://www.epa.gov/sites/production/files/2018-01/documents/collocation_instruction_guide.pdf (accessed 5.12.20).

USGS, n.d. EarthExplorer - Home [WWW Document]. URL https://earthexplorer.usgs.gov/ (accessed 2.24.20).

Wang, Y., Du, Y., Wang, J., Li, T., 2019. Calibration of a low-cost PM2.5 monitor using a random forest model. Environ. Int. 133, 105161. https://doi.org/10.1016/j.envint.2019.105161

Weissert, L., Alberti, K., Miles, E., Miskell, G., Feenstra, B., Henshaw, G.S., Papapostolou, V., Patel, H., Polidori, A., Salmond, J.A., Williams, D.E., 2020. Low-cost sensor networks and land-use regression: Interpolating nitrogen dioxide concentration at high temporal and spatial resolution in Southern California. Atmos. Environ. 223, 117287. https://doi.org/10.1016/j.atmosenv.2020.117287

Weissert, L.F., Alberti, K., Miskell, G., Pattinson, W., Salmond, J.A., Henshaw, G., Williams, D.E., 2019. Low-cost sensors and microscale land use regression: Data fusion to resolve air quality variations with high spatial and temporal resolution. Atmos. Environ. 213, 285–295. https://doi.org/10.1016/j.atmosenv.2019.06.019

Zhou, Z., Dionisio, K.L., Arku, R.E., Quaye, A., Hughes, A.F., Vallarino, J., Spengler, J.D., Hill, A., Agyei-Mensah, S., Ezzati, M., 2011. Household and community poverty, biomass use, and air pollution in Accra, Ghana. Proc. Natl. Acad. Sci. 108, 11028–11033. https://doi.org/10.1073/pnas.1019183108