SENSOR PLACEMENT OPTIMIZATION FOR AIR QUALITY MONITORING IN UGANDA

Anonymous authors
Paper under double-blind review

ABSTRACT

Outdoor air pollution is a major environmental challenge and a significant health risk with 91% of air pollution-related deaths occurring in low- and middle-income countries. In a bid to determine the parishes in which to place air quality monitoring sensors, we propose to use four methods i.e. random sampling, conditioned Latin hypercube sampling, K-Means clustering and Gaussian process modelling. Our aim will be to maximize the diversity of sample data collected. Subsequently, the performance of these techniques will be evaluated by calculating and plotting the correlation matrices of the resultant locations suggested by the algorithms as well as using the cluster silhouette method for the K-Means algorithm. Geocensus data from the Uganda Bureau of Statistics (UBOS) and time-series data from the sensors that have already been deployed will be used as input.

1 INTRODUCTION

Air pollution is one of the world's biggest environmental problems (Ritchie & Roser, 2020) and it is estimated that it accounts for over 8 million deaths annually worldwide (Lelieveld et al., 2019). According to a 2018 World Health Organization report, Kampala was the 15th most polluted city in the world between 2008 and 2017. One of the major causes of air pollution in Uganda is vehicle emissions (Nsereko, 2019). According to UNEP, the average age of vehicles being imported into the country is 16.5 years (Mutenyo, 2018) which may then be driven for another 20 years. Additionally, Uganda has a very chaotic public transportation system which is mainly made up of mini-buses and boda-bodas which are very rampant in the urban areas seeing as they are the second highest source of employment after agriculture(Kigambo, 2017). Furthermore, rubbish burning is the commonest means of domestic garbage disposal and biomass is still the most used source of fuel at an estimated 94% with charcoal and firewood very predominant(Lukiiza, 2019). Also, while the country's road network is over 129,469 km, only about 5,224 km is paved which results in fugitive road dust (FRD) particles that contribute to particulate matter levels (Chen et al., 2019).

Given the fact that Uganda is a developing country with very limited resources, has a varying topography and a mostly tropical climate, and very little research exists concerning a tropical low-income country's air quality, there was need to build a low-cost network of Internet of Things (IoT) devices to monitor fine particulate matter ($PM_{2.5}$) levels across the country which can withstand these conditions. Also, the use of many low-cost sensors instead of one reference monitor would help to pinpoint the exact sources of air pollution and the magnitude of their contribution which could in turn be used to inform decision making and policy.

2 THE PROBLEM

Uganda's admistrative units include districts, counties, sub-counties, parishes and villages in that order. The aim of this research is to, given a district, find suitable parishes in which to place sensors to get a good picture of the state of the air quality in that district. The number of the sensors will be determined by a cost constraint. Calibration of the low-cost sensors will be done by two reference Beta Attenuation Monitors (BAM).

3 LITERATURE REVIEW

Design-based Sampling

Wu et al. (2012) used soil moisture features together with the Gaussian assumption to optimally place sensors and for field estimation for soil moisture sensing. The locations were classified into clusters based on the ordering of the mean, with the number of sensors placed in each cluster determined by the ordering of the variances.

In their paper, Sarrate et al. (2014) proposed a reduction in the number of candidate sensors by grouping the initial sensors into clusters and applying Evidential C-Means (ECM) algorithm. The number of groups were determined by means of a study of the evolution of the validity index provided by the ECM algorithm for different number of groups. This research was specific to water distribution.

Sun et al. (2019) proposed a citizen-centric approach to optimal sensor placement. The city was divided into discrete, equal-sized square grids, in which sensors would be placed with at most one sensor in each grid based on the percentage of the population per grid. This technique targeted locations with a high population which is not a necessity for our research as our aim is to place sensors in locations that are diverse in terms of various features including population.

Model-based Sampling

Considering a mobile sensor network formed by sensing devices carried by individuals to monitor their exposure to air pollution, Xiang et al. (2012) considered the optimal placement of stationary sensors to assist the calibration of the mobile ones. The Mixed Integer Linear Programming (MILP) solution proposed was based on a stochastic human mobility model. In contrast, our work focuses on a stationary sensor network.

4 METHODOLOGY

Data Collection: Geo-census parish-level data obtained from UBOS for Kampala district and that includes several features relevant to air quality such as fuel choices, population density and km of road will be used to build the model which will then be transferrable to other districts/cities in Africa with a somewhat similar administrative structure. Data from sensors that are already installed in Kampala will be used as well.

Proposed Methods: As a baseline, the parishes in which sensors will be placed will be chosen at random. Given a district X with N parishes and in which n air quality monitors are to be installed, n parishes will be chosen at random where $n \leq N$.

Conditional Latin hypercube sampling (Minasny & McBratney, 2006), which is a sampling method for real-world multivariate data and was designed based on the Latin Hypercube Sampling method, will be used to choose sample parishes from the data.

Thirdly, K-Means algorithm will be used to group the parishes into clusters whereby the number of clusters will be equal to the number of sensors to be placed. A parish will then be chosen from each cluster for sensor deployment. The input data will be normalized by removing the mean and scaling to unit variance.

Lastly, using Gaussian process regression (Williams & Rasmussen, 2006), we will be able to model the data from the sensors already deployed in a simple spatial-only manner.

A correlation matrix will be plotted to compare the performance of each method. Since the aim is to maximize the diversity of the locations in which the sensors are installed, the less the correlation between the features of the locations chosen, the better the method will be assumed to perform. Moreover, for the k-means clustering, the cluster silhouette method will be used as an additional performance indicator. Also, using the air quality monitors already deployed, we shall compute the correlation between their measurements to get a better picture on performance using existing data.

REFERENCES

Siyu Chen, Xiaorui Zhang, Jintai Lin, Jianping Huang, Dan Zhao, Tiangang Yuan, Kangning Huang, Yuan Luo, Zhuo Jia, Zhou Zang, et al. Fugitive road dust pm2. 5 emissions and their potential health impacts. *Environmental science & technology*, 53(14):8455–8465, 2019.

Gaaki Kigambo. Kampala: moving on boda boda power. URL: `https://www.theeastafrican.co.ke/magazine/Kampala-moving-on-boda-boda-power/434746-4059576-94g5v3/index.html`, 2017. Accessed: 2020-02-11.

Jos Lelieveld, Klaus Klingmüller, Andrea Pozzer, Ulrich Pöschl, Mohammed Fnais, Andreas Daiber, and Thomas Münzel. Cardiovascular disease burden from ambient air pollution in europe reassessed using novel hazard ratio functions. *European heart journal*, 40(20):1590–1596, 2019.

David Lukiiza. Uganda's first renewable energy demonstration site opened. *The New Vision*, 2019. URL `https://www.newvision.co.ug/new_vision/news/1498366/ugandas-renewable-energy-demonstration-site`.

Budiman Minasny and Alex B McBratney. A conditioned latin hypercube method for sampling in the presence of ancillary information. *Computers & geosciences*, 32(9):1378–1388, 2006.

John Mutenyo. UN clean mobility week: A presentation on feebate system in uganda, and policy recommendations. URL: `https://wedocs.unep.org/bitstream/handle/20.500.11822/25187/FuelEconomyPolicyImpactTool_Uganda.pdf?sequence=2&isAllowed=y`, 2018. Accessed: 2020-02-11.

Patience Nsereko. Air pollution, a global concern of local significance. *NEMA NEWS*, pp. 4–5, 2019. URL `https://nema.go.ug/sites/all/themes/nema/docs/June%202019_NEMA%20Newsletter%20(1).pdf`.

Hannah Ritchie and Max Roser. Air pollution. *Our World in Data*, 2020. `https://ourworldindata.org/air-pollution`.

Ramon Sarrate, Joaquim Blesa, and Fatiha Nejjari. Clustering techniques applied to sensor placement for leak detection and location in water distribution networks. In *22nd Mediterranean Conference on Control and Automation*, pp. 109–114. IEEE, 2014.

Chenxi Sun, Victor OK Li, Jacqueline CK Lam, and Ian Leslie. Optimal citizen-centric sensor placement for air quality monitoring: A case study of city of cambridge, the united kingdom. *IEEE Access*, 7:47390–47400, 2019.

Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

Xiaopei Wu, Mingyan Liu, and Yue Wu. In-situ soil moisture sensing: Optimal sensor placement and field estimation. *ACM Transactions on Sensor Networks (TOSN)*, 8(4):1–30, 2012.

Yun Xiang, Lan S Bai, Ricardo Pledrahita, Robert P Dick, Qin Lv, Michael Hannigan, and Li Shang. Collaborative calibration and sensor placement for mobile sensor networks. In *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 73–83. IEEE, 2012.