Poster: Low-Cost Sensor Correlation based on Urban Form

Alex Cabral
Georgia Tech
Atlanta, Georgia, USA
acabral30@gatech.edu

Jim Waldo
Harvard University
Boston, Massachusetts, USA

Amy V. Mueller
Northeastern University
Boston, Massachusetts, USA

ABSTRACT

There is an increasing need to use low-cost sensors to monitor environmental harms that urban residents are exposed to around the globe. One of the hurdles in achieving this goal is optimizing where to place a limited number of sensor nodes to gather useful data. This poster describes an exploration of low-cost sensor reading correlation with respect to the heterogeneity of urban environments. Using data from Boston and Chicago, USA, the findings show that sensor nodes on a medium flushing road show expected correlation levels within about 750 meters from each other. These findings lay the groundwork for future urban environmental sensor network design considerations and techniques.

CCS CONCEPTS

• **Computer systems organization** → **Sensor networks**; • **Applied computing** → *Environmental sciences.*

KEYWORDS

sensor networks, urban sensing, environmental sensing, smart cities

1 INTRODUCTION

Billions of people around the globe live in cities where they are exposed to environmental hazards such as extreme heat and air pollution [8, 13]. These harms are currently most typically measured by geographically sparse regulatory monitors, making it difficult to determine the areas and residents most affected given the fine-grained heterogeneity of these harms. Dense networks of low-cost environmental sensors can help identify urban areas that are most negatively impacted by environmental hazards [3], paving the way for targeted policies and mitigation strategies [4]. However, city governments generally have a limited budget and determining where to place a finite set of nodes to gather a representative set of city data remains an open question. Under the assumption that environmental hazards are generally static within a certain radius, nodes are often optimized for placement based purely on distance from other nodes [12]. However, the presence and variation of urban form, hetereogeneity of emissions types, and changing wind patterns caused by street canyons and varying road types have all been shown to affect the levels of environmental hazards [7] and can all vary greatly within 1 square kilometer.

This work aims to evaluate the correlation of fine particulate matter, $PM_{2.5}$, readings as a function of distance to explore 1) how the street flushing type affects the correlation of readings between sensors and 2) what density of sensors is required in different areas to achieve a desirable level of data representativeness. Using data from the Eclipse sensor network in Chicago, Illinois [3] and a set of field experiments in Boston, Massachusetts, the correlations of near-simultaneous data from pairs of $PM_{2.5}$ sensors were calculated. The distance between and surrounding urban area of the sensors were then considered to determine which physical urban features might affect the flow of pollutants and thus the design of an urban environmental sensor network. The expectation was that sensor reading correlations would decrease as distance between the sensors increased, and that the rate of decrease would differ based on surrounding urban features. The findings indicate that sensor pairs within 750 meters of each other on a single medium flushing street exhibit a small linear decrease in reading correlation with increased distance. Sensor pairs on high and low flushing streets reveal greater variance in reading correlations, offering insight into the density of sensor nodes required in different areas based on urban form.

2 RELATED WORKS

Capturing ground-truth data to measure the heterogeneity of environmental hazards in a city is virtually impossible given the fine-grained heterogeneity of these hazards [6]. Thus, researchers have begun to examine methods to estimate how representative low-cost sensor network data are based on predicted values from computational models utilizing data fusion strategies [5, 14]. However, these approaches result in residual uncertainty due to simplifications and estimates in proxies used. For example, these techniques often do not account for urban form and features such as buildings, instead viewing the monitoring area as a two-dimensional plane with traffic patterns [5]. This is at odds with the inherently three-dimensional nature of cities and with prior research indicating that urban form can complicate the accuracy of urban environmental models [7].

In particular, the *street flushing type*, which is calculated based on the road width to building height ratio, has been classified into three categories based on the fundamental physics of how the ratio affects the *flushing* of air pollutants in urban settings [14]. In *low flushing streets* air pollution can be much higher than the average found in other locations in a city [13] due to the trapping of eddies that limits ventilation of air in this area to the upper atmosphere [9, 13]. Conversely, *high flushing streets* are wide with low, or perhaps no, buildings on both sides, allowing for a high level of communication with the overlying atmosphere and thus dilution of pollution into this larger air mass [9, 11]. Medium flushing streets are characterized by a road width to building height ratio that falls in the middle of low and high flushing streets, and thus experiences moderate levels of pollutant dilution into a larger air mass.

3 METHODOLOGY

Two distinct datasets were analyzed. The datasets were intended to simultaneously capture air quality and information about the surrounding urban area and street flushing types. The first dataset consists of $PM_{2.5}$ data collected from July 1, 2021 through June 30, 2022 from 106 nodes of the Eclipse sensor network deployed in Chicago, Illinois [3]. The $PM_{2.5}$ data were combined with estimated road width data based on Google Street View images and building height information from OSM Buildings [10]. The second dataset contains $PM_{2.5}$ readings from two Dylos DC1700 monitors that were held at different nearby locations around Boston, Massachusetts across several weeks of field experiments. The $PM_{2.5}$ data were combined with road width data from the Massachusetts Department of Transportation [1] and building height information from the City of Boston [2]. Street flushing types were then calculated based on the road width to building height ratio, based on mapping values suggested in prior work [14].

For each pair of sensors that were within 1 km of each other, a set of $PM_{2.5}$ readings were correlated with each other and the mean correlation values were then correlated with features such as the distance between sensors and the street flushing types. Correlations were calculated using both Pearson's correlation and adjusted R-squared values. The mean correlation for each set of readings was then correlated with the distance between the sensors and examined with consideration for street flushing type and the orientation of the sensor nodes, being either on the same street or around the corner from each other.

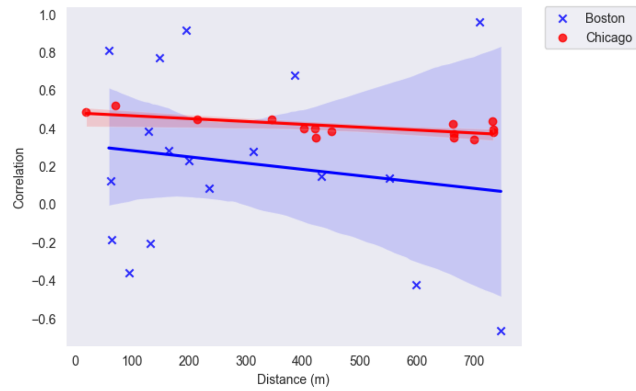

Figure 1: This scatterplot show the mean Pearson's correlation for a set of near-simultaneous $PM_{2.5}$ readings versus the distance between sensors on medium flushing streets for the Eclipse dataset [3] as red dots and the Boston field experiments as blue X marks. Only sensor pairs within 750 meters of each other are included in this plot.

4 RESULTS AND DISCUSSION

Across both datasets, the correlation for sensor pairs on the same medium flushing street was moderately negative, with correlation values of −0.45 for the Boston dataset and −0.39 for the Eclipse dataset. These values match the expected relationship of decreasing correlation between sensors with an increase in distance. Conversely, the results of correlation between sensors on low and high flushing streets were extremely variable, suggesting that a pattern may not exist for these types of streets because of the unique hyper-local trapping of pollutants that can occur on low flushing streets or the undisturbed wind flow patterns on high flushing streets.

Additionally, the expected relationship was observed only for sensor pairs within 750 meters of each other on medium flushing streets, as show in Fig. 1. For sensor pairs further than 750 meters from each other, the correlation between sensors exhibited more variable results. This indicates that there if there is a distance threshold at which sensor correlation drops off significantly, it may fall around 750 meters on medium flushing roads. Collecting additional data in field experiments where the sensors are between 750 meters and 1 km away from each other would help to identify where that threshold is, especially because prior work has suggested using 1 km as a radius for data "representativeness" [12]. Furthermore, given the variation in results between medium flushing and low and high flushing streets, there is a strong possibility that the distance threshold will differ based on the street flushing type, driving the need for additional data collection to determine the distance threshold in different urban settings.

5 CONCLUSIONS

This poster presents an in-depth analysis of two distinct datasets to identify which physical urban features affect the correlation of $PM_{2.5}$ readings between two low-cost sensors. By examining urban features that generalize to all cities, this work explores network design techniques that move beyond pure distance-based placement. Furthermore, the work highlights methods by which to identify the density of nodes needed in different neighborhoods using easily collected urban information.

REFERENCES

[1] Analyze Boston. 2022. City of Boston managed streets - datasets - analyze Boston. https://data.boston.gov/dataset/city-of-boston-managed-streets

[2] Analyze Boston. 2024. Parcels 2022 - dataset - analyze Boston. https://data.boston.gov/dataset/parcels-2022

[3] Madeleine IG Daepp, Alex Cabral, et al. 2022. Eclipse: an end-to-end platform for low-cost, hyperlocal environmental sensing in cities. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 28–40.

[4] Madeleine IG Daepp, Alex Cabral, Tiffany M Werner, Raed Mansour, et al. 2023. The "Three-Legged Stool": Designing for Equitable City, Community, and Research Partnerships in Urban Environmental Sensing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing System (CHI)*. ACM.

[5] Shaoming He et al. 2020. Distributed estimation over a low-cost sensor network: A Review of state-of-the-art. *Information Fusion* 54 (Feb. 2020), 21–43. https://doi.org/10.1016/j.inffus.2019.06.026

[6] Denise Hertwig, Sue Grimmond, Simone Kotthaus, Christina Vanderwel, Hannah Gough, Martial Haeffelin, and Alan Robins. 2021. Variability of physical meteorology in urban areas at different scales: implications for air quality. *Faraday Discussions* 226 (2021), 149–172.

[7] Piotr Holnicki and Zbigniew Nahorski. 2015. Emission data uncertainty in urban air quality modeling—case study. *Environmental Modeling & Assessment* 20 (2015), 583–597.

[8] Tord Kjellstrom et al. 2007. Urban environmental health hazards and health equity. *Journal of urban health* 84 (2007), 86–97.

[9] Xian-Xiang Li, Chun-Ho Liu, Dennis YC Leung, and Kit Ming Lam. 2006. Recent progress in CFD modelling of wind field and pollutant transport in street canyons. *Atmospheric Environment* 40, 29 (2006), 5640–5658.

[10] OSMBuildings. [n. d.]. OSMBuildings: 3D building geometry viewer based on OpenStreetMap Data. https://github.com/OSMBuildings/OSMBuildings

[11] Zhi Shen, Bobin Wang, Guixiang Cui, and Zhaoshun Zhang. 2015. Flow pattern and pollutant dispersion over three dimensional building arrays. *Atmospheric Environment* 116 (2015), 202–215.

[12] Alexis A. Shusterman et al. 2018. Observing local CO2 sources using low-cost, near-surface urban monitors. *Atmospheric Chemistry and Physics* 18, 18 (Sept. 2018), 13773–13785. https://doi.org/10.5194/acp-18-13773-2018

[13] Sotiris Vardoulakis, Keith Dear, and Paul Wilkinson. 2016. Challenges and opportunities for urban environmental health and sustainability: the HEALTHY-POLIS initiative. *Environmental health* 15, 1 (2016), 1–4.

[14] Eun-Hye Yoo, Andrew Zammit-Mangion, and Michael G Chipeta. 2020. Adaptive spatial sampling design for environmental field prediction using low-cost sensing technologies. *Atmospheric Environment* 221 (2020), 117091.