PhD School: Community-Driven Low-Cost Environmental Sensor Networks

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

ABSTRACT

Low-cost environmental sensor networks can provide essential data to monitor processes, hazards, and changes in various physical environments. However, the sensing technology is often designed in isolation and deployed for short time periods, resulting in technology that is inaccessible for the communities the sensors are meant to serve. I address this shortcoming by working with underrepresented communities to identify and address the unique challenges associated with real-world sensor network deployments. This work will build toward a framework to enable widespread creation and deployment of effective low-cost environmental sensor networks.

CCS CONCEPTS

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

KEYWORDS

environmental sensing, smart cities, low-cost sensing

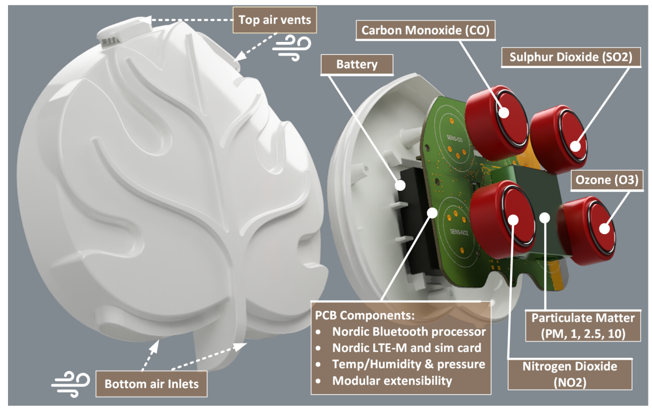

Figure 1: An overview of the embedded sensing system used in my prior research to measure air quality in Chicago, Illinois from July 2021 through April 2023 [9].

1 INTRODUCTION

Billions of people around the globe face environmental challenges, ranging from heat exposure and rising sea levels to unhealthy levels of pollution or noise. To monitor environmental hazards, researchers, regulators, and policymakers rely on data from highly accurate regulatory equipment managed by government agencies and research institutes. However, these regulatory monitors are expensive, large, and require special expertise for maintenance [20, 22]. As a result, environmental regulatory networks are geographically sparse and many neighborhoods, cities, and larger geographic regions lack monitors. The decreasing cost of sensing technologies has enabled environmental scientists, activists, and concerned communities to use dense, low-cost embedded wireless sensor networks for environmental monitoring [27]. Recent environmental sensing initiatives that have utilized participatory and community-driven techniques have resulted in novel low-cost sensing devices [8], a diverse array of network deployments [26], and more equitably distributed sensor networks [10]. Despite the promise of community-driven environmental sensor networks, however, there are few examples of real-world networks that have been deployed for long periods. In fact, environmental sensor networks are rarely seen in the real-world "except when engineered by professionals at significant cost" [26]. Budgetary constraints naturally make it difficult to deploy these networks. However, another main hindrance is a lack of guidance on integrating embedded wireless sensing systems into existing environmental monitoring workflows. Thus, several questions arise including where to deploy sensor nodes, how many nodes are needed to achieve environmental monitoring goals, and how to create data visualizations that are engaging, intuitive, and actionable by incorporating local knowledge and open data.

My work aims to provide meaningful contributions for two main research gaps that arise in community-driven environmental sensor networks: 1) Identifying and addressing the issues and challenges that surface when deploying low-cost embedded wireless sensor networks at scale in real-world settings, and 2) Answering calls to create a generalizable framework for the processes involved in designing and deploying low-cost environmental sensor networks with non-technical community partners [26]. I approach this work via two main research projects: 1) my PhD work focused on designing and deploying a large-scale, urban air quality sensor network with city and community partners [9, 10], and 2) my postdoctoral project co-designing low-cost sensors with Native American communities to monitor environmental factors affecting the growth of wild rice in rural areas [13]. I address these research gaps and open questions by exploring generalizable techniques and methods to ensure reliability, data accuracy, and accessible visualizations for low-cost environmental sensor networks.

2 SENSOR NETWORK RELIABILITY

Prior research indicates that to be successful an environmental sensor network must be reliable [16, 21], easy to maintain [9, 11], and low-cost [9, 16, 21]. Reliability and maintenance are especially important because environmental sensors may be sited in difficult to reach locations and community partners often do not have the technical expertise to troubleshoot sensors or conduct repairs. Additionally, environmental regulators have stated that all nodes in a low-cost environmental sensor network should continue to operate and transmit data for at least 75% of the total network deployment time to meet performance standards [28].

Two key features of a sensor network design that help achieve these goals are connectivity and power. *Connectivity* is essential for data transmission, real-time node monitoring, and software updates, while *power* provides for reliable operation. Cellular networks are the appropriate connectivity choice for most environmental sensor networks given the widespread global availability and the lower cost and ease of setup and scaling compared to low-power wide-area networks (LPWAN), such as LoRaWAN [14, 15, 17]. Similarly, solar power is the most ubiquitous form of renewable energy for sensor networks and will remain prevalent in the coming years because of its relative low-cost and ease of scalability and maintainability.

In my dissertation research, I found that the placement of nodes in a city can also affect a network's reliability with relation to connectivity and solar power [3]. In particular, urban form such as buildings can block the path between sensor nodes and charging or connectivity sources, such as the sun and cellular towers. Additionally, rural areas, such as those in which my current research is focused, can also face issues with cellular connectivity [25]. Thus, nodes must be strategically placed to account for structures and impediments that may affect reliability by incorporating open data of the physical elements in a study area, cellular tower locations, etc. I have previously developed machine learning models that utilize open building data to predict urban locations that will have solar charging or cellular connectivity issues with 77% and 75% accuracy respectively [5]. However, open data cannot always help predict reliability issues, as 1) these data are often inaccurate, and 2) other forms such as trees, which are not captured in open data, may also cause reliability issues. Hence I will work to utilize various open data sources, such as Google Street Map and Google Earth captures, to explore the possibility of developing digital twins that can provide useful information on reliable node locations for future environmental sensor network deployments.

3 DATA ACCURACY AND UTILITY

A primary goal of low-cost environmental sensor networks is the collection of data to enable scientific inquiry, policy changes, and potential legal challenges for environmental harms [10]. However, achieving this aims remains difficult 1) because of accuracy issues with low-cost sensor readings [18, 23] and 2) because previously deployed networks have been small-scale and thus lack the spatial and/or temporal comparisons needed to identify problem areas and therein identify potential mitigation strategies [9, 12].

To address the issue of sensor accuracy, researchers often focus on calibrating low-cost sensor data to reference monitors via sensor collocation and various machine learning techniques [19]. In my prior work focused on low-cost urban air pollution sensing [9], our gradient boosting calibration models achieved an R^2 value of over 0.7 when comparing readings from three low-cost sensors each co-located at three EPA reference monitors [9]. However, because the reference monitors are placed on rooftops and towards the city outskirts, it is unclear how well these efforts align with the ground truth in the urban areas most residents live and work in. Furthermore, there are several environments that lack reference monitors with which to collocate, creating the need for alternative means to verify the accuracy of low-cost sensor data. In my current research project, I am exploring the potential of developing a blind calibration strategy based on the principles of distributed estimation [7] and consensus agreement [1], relying on near-collocation and agreement with multiple low-cost sensors to compute and update the uncertainty level for each sensor over time.

Ensuring that networks can collect useful data across spatial and temporal variables requires careful consideration of where and how sensor nodes are deployed [29]. Thus, my research focuses on developing sensor network deployment strategies that consider the unique physical environment of the study area and characteristics of the environmental harms being monitored. In my PhD dissertation, I explored developing a metric for data accuracy that estimates uncertainty based on the effect of urban form on low-cost sensors and environmental hazards, such as air pollution and noise [4]. Based on the road-width to building-height ratio (W/H)—classified as low, medium, or high [24]—and the number and distance of nearby sensors, I calculated the correlation of sensor readings at each block and used these values to determine which factors may affect the accuracy of sensor network data in different environmental conditions.

4 ACCESSIBILITY OF SENSOR NETWORK DATA

One of the primary goals of environmental sensor networks is the production of data for use and interpretation by diverse stakeholders [10]. To account for different levels of data literacy, it is essential to design interfaces that are familiar and accessible to community members [26], a task that can be challenging given the diverse experiences of community members and the difficulty in visualizing changes across both space and time. In my PhD, I helped to create a novel situated visualization [10] in which sensor network data was shown in real-time near locations with sensor nodes, as shown in Fig. 2. This unique visualization resulted in engaging and educationally-fulfilling experiences for community members and researchers [10], laying the foundation for future work in sensor network data interfaces that are accessible for diverse partners and community members.

With diverse stakeholders and data literacy also comes a diverse set of use cases for data. Because it is impossible to predict what users may want to use the data for [10], my future research agenda includes work to utilize LLMs to generate spatiotemporal data plots based on user search terms. Additionally, I recognize the importance of communicating the uncertainty or confidence levels of low-cost sensor data, and thus plan to incorporate prior research in uncertainty visualization [2] as I develop that tool.

Finally, when considering data accessibility, one must determine who can access environmental sensor network data. Data governance is often a concern when large amounts of data are collected and shared [6], but it becomes even more essential when working with groups such as Native American tribes, who have sovereignty over their lands and may seek to protect the privacy of locations, people, and phenomena on those lands [26]. As part of my postdoctoral research, I am thus investigating ways to incorporate data sovereignty into large, heterogeneous datasets that include sensor

Figure 2: People interacting with a situated visualization showing real-time urban sensor network data in Chicago [10].

network data from both tribal and non-tribal lands. This work will enable the successful adoption of low-cost environmental sensor networks by underrepresented communities who may lack trust in researchers due to a long history of exploitation and other communities who strive to maintain privacy.

5 OPEN CHALLENGES

I face several technical and non-technical challenges in achieving success in my research agenda. One of the primary technical challenges is accommodating for the lack of open data, primarily for cellular connectivity, in determining where to place sensor nodes. The success of future and widespread environmental sensor network deployments is dependent on reliable connectivity, revealing a need for more novel data collection techniques and tools to identify cellular signal strength. An additional technical challenge is identifying the right balance in creating sustainable technology but also utilizing machine learning techniques that may improve research outcomes but negatively impact sustainability efforts. The key non-technical challenge I face in my research is tailoring my real-world, applied computing work and translating the contributions for successful publication in academic and scientific journals.

REFERENCES

[1] Philip J Boland. 1989. Majority systems and the Condorcet jury theorem. *Journal of the Royal Statistical Society: Series D (The Statistician)* 38, 3 (1989), 181–189.

[2] Ken Brodlie, Rodolfo Allendes Osorio, and Adriano Lopes. 2012. A review of uncertainty in data visualization. *Expanding the frontiers of visual analytics and visualization* (2012), 81–109.

[3] Alex Cabral, Vaishnavi Ranganathan, and Jim Waldo. 2023. Cellular LTE and Solar Energy Harvesting for Long-Term, Reliable Urban Sensor Networks: Challenges and Opportunities. *arXiv e-prints* (2023), arXiv–2307.

[4] Alex Cabral and Jim Waldo. 2023. Designing Large-Scale Wireless Sensor Networks for Urban Environmental Sensing. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 253–257. https://doi.org/10.1145/3594739.3610768

[5] Alex Cabral, Jim Waldo, and Amy V Mueller. 2024. Towards a Predictive Model for Improved Placement of Solar-Powered Urban Sensing Nodes. In *2024 IEEE/ACM Ninth International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 215–216.

[6] Stephanie Carroll, Ibrahim Garba, et al. 2020. The CARE principles for indigenous data governance. *Data science journal* 19 (2020).

[7] Chee-Yee Chong. 2017. Forty years of distributed estimation: A review of noteworthy developments. In *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 1–10.

[8] Kim M Cobb, Russell Clark, et al. 2022. Research to action frameworks for equitable coastal resilience: a case study from savannah, Georgia. In *102nd American Meteorological Society Annual Meeting*. AMS.

[9] 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.

[10] 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.

[11] Ahmad H Dehwah et al. 2015. Lessons learned on solar powered wireless sensor network deployments in urban, desert environments. *Ad Hoc Networks* 28 (2015), 52–67.

[12] Graham Dove, Shiyu Chen, et al. 2022. From Environmental Monitoring to Mitigation Action: Considerations, Challenges, and Opportunities for HCI. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–31. https://doi.org/10.1145/3555093

[13] Eric Greenlee, Blaine Rothrock, Hyeonwook Kim, Ellen Zegura, and Josiah Hester. 2024. " The Devil You Know": Barriers and Opportunities for Co-Designing Microclimate Sensors. *ACM Journal on Computing and Sustainable Societies* (2024).

[14] Jetmir Haxhibeqiri, Eli De Poorter, Ingrid Moerman, and Jeroen Hoebeke. 2018. A survey of LoRaWAN for IoT: From technology to application. *Sensors* 18, 11 (2018), 3995.

[15] Mohammad Istiak Hossain and Jan I Markendahl. 2021. Comparison of LPWAN Technologies: Cost Structure and Scalability. *Wireless Personal Communications* 121, 1 (2021), 887–903.

[16] Mohamed Amine Kafi, Yacine Challal, et al. 2013. A Study of Wireless Sensor Networks for Urban Traffic Monitoring: Applications and Architectures. *Procedia Computer Science* 19 (2013), 617–626. https://doi.org/10.1016/j.procs.2013.06.082

[17] Ala' Khalifeh et al. 2019. A Survey of 5G Emerging Wireless Technologies Featuring LoRaWAN, Sigfox, NB-IoT and LTE-M. In *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. 561–566. https://doi.org/10.1109/WiSPNET45539.2019.9032817

[18] Alastair Lewis and Peter Edwards. 2016. Validate personal air-pollution sensors. *Nature* 535, 7610 (2016), 29–31.

[19] Balz Maag, Zimu Zhou, and Lothar Thiele. 2018. A Survey on Sensor Calibration in Air Pollution Monitoring Deployments. *IEEE Internet of Things Journal* 5, 6 (Dec. 2018), 4857–4870. https://doi.org/10.1109/JIOT.2018.2853660

[20] M.I. Mead, O.A.M. Popoola, et al. 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment* 70 (May 2013), 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060

[21] Rohan Narayana Murty, Geoffrey Mainland, et al. 2008. Citysense: An urban-scale wireless sensor network and testbed. In *2008 IEEE conference on technologies for homeland security*. IEEE, 583–588.

[22] National Oceanic and Atmospheric Administration. 2024. Instruments. https://www.ncei.noaa.gov/access/crn/instruments.html

[23] Gwen Ottinger. 2010. Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology, & Human Values* 35, 2 (2010), 244–270.

[24] Daniel T O'Brien and Amy V Mueller. 2023. In Pursuit of Local Solutions for Climate Resilience: Sensing Microspatial Inequities in Heat and Air Pollution within Urban Neighborhoods in Boston, MA. *Sustainability* 15, 4 (2023), 2984.

[25] Beatriz Palacios Abad et al. 2022. Note: Towards Community-Empowered Network Data Action. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*. 585–588.

[26] Blaine Rothrock, Eric Greenlee, and Josiah Hester. 2023. Community-Driven Mobile and Ubiquitous Computing. *GetMobile: Mobile Computing and Communications* 27, 3 (2023), 5–9.

[27] Emily G Snyder, Timothy H Watkins, et al. 2013. The changing paradigm of air pollution monitoring. *Environmental science & technology* 47, 20 (2013), 11369–11377.

[28] United States Environmental Protection Agency. 2024. Air Sensor Performance Targets and Testing Protocols. https://www.epa.gov/air-sensor-toolbox/air-sensor-performance-targets-and-testing-protocols

[29] 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.