Toward Massive Scale Air Quality Monitoring

Naser Hossein Motlagh, Eemil Lagerspetz, Petteri Nurmi, Xin Li, Samu Varjonen, Julien Mineraud, Matti Siekkinen, Andrew Rebeiro-Hargrave, Tareq Hussein, Tuukka Petäjä, Markku Kulmala, and Sasu Tarkoma

The authors present a research vision of real-time massive scale air quality sensing that integrates tens of thousands or even millions of air quality sensors to monitor air quality at fine spatial and temporal resolution. They highlight opportunities and challenges of their vision by discussing use cases, key requirements, and reference technologies.

Abstract

Dangers associated with poor air quality are driving deployments of air quality monitoring technology. These deployments rely on either professional-grade measurement stations or a small number of low-cost sensors integrated into urban infrastructure. In this article, we present a research vision of real-time massive scale air quality sensing that integrates tens of thousands or even millions of air quality sensors to monitor air quality at fine spatial and temporal resolution. We highlight opportunities and challenges of our vision by discussing use cases, key requirements, and reference technologies in order to establish a roadmap on how to realize this vision. We address the feasibility of our vision, introducing a testbed deployment in Helsinki, Finland, and carrying out controlled experiments that address collaborative and opportunistic sensor calibration, a key research challenge for our vision.

Introduction

Air pollution is one of the most significant health challenges of our time. According to the World Health Organization (WHO), in 2016 air pollution was linked to over 4.2 million deaths per year (11.6 percent of all deaths) with mortality in low and middle income countries particularly heavily affected by air pollution [1]. Air pollution is strongly associated with a broad spectrum of acute and chronic diseases [2]. Air pollution is also a significant economic burden worldwide with estimates suggesting 2–5 percent of GDP spent on treating related diseases. The significance of air pollution is exacerbated by increasing urbanization, with estimates suggesting that 96 percent of the world's population currently lives in areas where air pollution exceeds safe limits [3].

To counteract problems associated with poor air quality, it is fundamental to understand the characteristics of pollutants in the urban environment. Detailed information about air quality is essential for informing and assessing effectiveness of initiatives tackling poor air quality. This need to access air quality information is driving deployments of air quality monitoring technology. Current deployments are based on either professional-grade monitoring stations or low-cost air quality monitors integrated into public transportation or other parts of the urban infrastructure. The former provides highly accurate air quality information, but suffers from low spatial resolution and high deployment and maintenance costs. The latter suffers from poor accuracy unless sensors are periodically re-calibrated against professional-grade equipment. Currently, carrying out the calibration is time-consuming and laborious, limiting the scale at which these types of deployments can operate.

In this article, we envision real-time massive scale air quality sensing for metropolitan areas (Fig. 1) that integrates a large number (thousands or millions) of air quality sensors. Scaling up current deployments requires integrating air quality monitors with different characteristics, including granularity and quality of measurements. Monitoring devices ranging from expensive reference stations to low-cost sensors integrated into vehicles or carried by pedestrians are needed to achieve dense coverage. In this fashion, air pollution information can be gathered at fine spatial resolution — particularly from areas with higher population density where the health effects of air pollution are felt worst. To ensure the collected information is sufficiently accurate and robust for informing policy makers, pedestrians, and other actors, periodic re-calibration of all sensors operating within the urban environment is required [4]. However, truly massive scale air quality monitoring requires opportunistic calibration transfer as the scale of deployments renders manual calibration infeasible. This necessitates new solutions that take advantage of periods where the sensors are located in the vicinity of professional-grade measurement stations. For example, sensors deployed in taxis and garbage trucks can capture measurements as they pass by a reference station. By aggregating many measurements across several devices, a calibration model can be learned and propagated to all sensors located within proximity of the reference station or the vehicle performing the calibration.

Example Use Cases

Real-time massive scale air quality monitoring not only enables policy makers to receive accurate and fine-grained information about pollutants, but also enables new applications. We briefly discuss some next.

Localized Air Quality Forecast: High concentration of pollutants make outdoor activity hazardous and even counteract positive effects of an active lifestyle — particularly in megacities with high pollution levels. Our vision enables fine-grained and real-time air quality maps that iden-

Naser Hossein Motlagh, Eemil Lagerspetz, Petteri Nurmi, Xin Li, Samu Varjonen, Julien Mineraud, Matti Siekkinen, Andrew Rebeiro-Hargrave, and Sasu Tarkoma are with the Department of Computer Science, University of Helsinki; Tareq Hussein, Tuukka Petäjä, and Markku Kulmala are with the Institute for Atmospheric and Earth System Research, University of Helsinki.

Digital Object Identifier:
10.1109/MCOM.001.1900515

0163-6804/20/$25.00 © 2020 IEEE

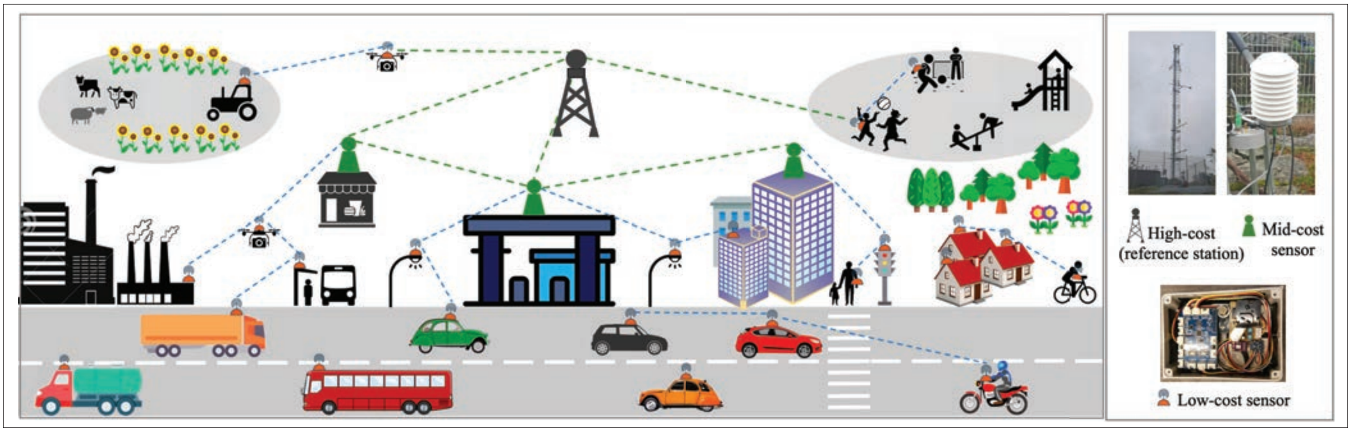

Figure 1. Vision of massive scale air quality monitoring with opportunistic sensor calibration. Green lines denote calibration links between reference and mid-cost sensors, and blue lines denote calibration links between low-cost sensors.

tify pollution hotspots and provide suggestions on optimal times to go outdoors and which routes to take for lower pollution levels. Current deployments do not have sufficient density or scale to capture spatial and temporal variations in pollutants with required accuracy.

Adaptive Air Filtering and Routing for Vehicles: Connected vehicles can use real-time air quality information to detect pollution hotspots and schedule optimal periods for managing air intake. Air vents can be closed in heavily polluted regions and re-opened in areas with cleaner air to reduce exposure to pollutants. Similarly, vehicles with recently calibrated data can be routed to pass by other vehicles to propagate air quality and calibration information across all vehicles moving in the city. These types of services are only feasible with air quality monitoring operating at a high spatial and temporal resolution.

Early Warning and Emergency Preparation: Air quality deterioration has been linked to sharp increase of respiratory problems and can lead to a sudden influx of respiratory patients. High resolution air quality information can be used for early warning of air quality issues, e.g., predicting haze [5] enable hospitals to prepare ahead of time. Knowledge of current air quality conditions can help doctors more precisely address patients' symptoms. This is not possible without real-time high-resolution air quality monitoring.

Requirements

Realizing the vision of massive scale air monitoring requires advances in devices, algorithms, and infrastructure to address limitations of current deployments and sensor technology. Next, we discuss key requirements for our vision.

High Spatial and Temporal Resolution: Achieving real-time measurements and capturing air quality variations at fine spatial resolution requires a dense deployment of measurement devices. Apte *et al.* [6] show that air quality can vary drastically within a distance of 150 ft. This resolution would require approximately 1000 sensors/mi^2, and covering an entire city district would require tens of thousands of sensors. With advances in sensing technology, deployments of this magnitude are becoming realistic. Affordable sensors costing less than $2500 are available for all pollutants included in current air quality indexes (SO_2, NO_2, PM_{10}, $PM_{2.5}$, O_3, CO), such as the Environmental Protection Agency (EPA) in the United States and the Air Quality Index of China. Dense networks of air quality sensors that combine sensors with different characteristics are needed to reach the required spatial and temporal resolution with sufficient measurement accuracy.

Regulatory Standards for Accuracy: High-cost reference stations have been designed to meet strict regulatory standards (e.g., EU Air Quality directive or U.S. EPA) that ensure the information provided is sufficiently accurate for actionable insights. Low- and mid-cost platforms rarely achieve required accuracy as they are sensitive to environmental conditions (e.g., temperature, humidity, and wind direction), and measurements drift over time [7]. These errors are caused by lower-quality receptors and saturation of sensitive material inside the sensor module, as well as local variations. For example, during traffic congestion, pollution levels may drastically increase and have an instant transient effect on a sensor reading, which may be mistaken as an anomaly. Ensuring high-resolution air quality information requires techniques that ensure regulatory accuracy for all types of sensors and mitigate effects of environmental and indirect causes of errors.

Low-Effort Deployment and Maintenance: Deploying massive amounts of air quality sensors requires considerable effort. For example, sensors installed on vehicles need to be packaged to have unobstructed air flows [8] and connected to a power source, whereas pedestrians carrying low-cost sensors may require training on how to ensure quality of the captured measurements. Besides deployment effort, operating the sensors induces costs (e.g., periodic maintenance, fault analysis, and electricity costs). To ensure long-term operation, best practices surrounding sensor design, easy deployment, and operation need to be drafted, and novel technical solutions are required. Similarly, solutions that facilitate management and operation of the devices need to be developed. We envision sensor deployment to be performed by cities, policy makers, and organizations operating within the city. Fixed sensors can be installed at bus stops, subway platforms, and buildings, where electricity is readily available. In our vision (Fig. 1), sensor deployment and maintenance are handled by city authorities and other sectors owning the property. Sensors that are mal-

Current communication technologies, such as LTE-4G and Narrow-Band IoT (NB-IoT) do not support a large amount of connections (4G) or suffer from insufficient bandwidth (NB-IoT). 5G networks are needed to enable deployments of this scale as they support up to millions of devices per square mile and provide necessary bandwidth to process the generated data.

functioning or need cleaning can be detected by a cross-calibration approach [9] and replaced or serviced by the responsible organizations.

Enabling Technologies

Initiatives for air quality monitoring are increasing in scale with deployments of hundreds or even thousands of sensors becoming a reality. Scaling up to deployments of hundreds of thousands of sensors, envisioned in this article, is currently difficult due to technological limitations. Emerging technological solutions are paving the way to overcome these limitations. We next briefly discuss some of these technologies.

Sensor Deployment

Massive scale air pollution monitoring requires combining high-cost reference stations with dense deployment of low- and mid-cost sensors. These should be deployed in pollution hotspots and densely populated or heavily industrialized urban areas. Finding optimal locations for sensor placement is a major consideration as air pollutants are affected by wind patterns, temperature, and humidity, and vary at different heights. In densely populated areas, such as near residential and office buildings, pollutants are a higher risk factor. To accurately cover a large area, sensors should not be placed directly in the vicinity of a point pollution source. Sensors should be protected from weather by a suitable casing and connected to a power source. Energy harvesting helps to reduce energy storage requirements of sensing units, and emerging cellular connectivity overcomes the need for power-consuming GPS modules, providing more flexibility regarding sensor deployments by reducing the form factor of sensing units.

Fine-Grained Positioning

Deployment and operation of thousands of mobile or fixed low-cost sensors requires time-consuming tracking. Additionally, precise calibration of air pollution sensors requires location information of each sensor to be accurate. Current deployments rely on GPS receivers integrated into every sensor, which increases power consumption. GPS is prone to errors such as signal reflection by skyscrapers. We envision overcoming these issues by taking advantage of 5G network positioning, which helps to reduce energy consumption and eases deployment of thousands of sensors in urban areas. 5G networks are being designed to enable 3D positioning at comparable (or even better) accuracy than GPS by integrating antenna arrays that increase the probability of line of sight between the sensors and 5G antennas, allowing accurate angle of arrival and position estimation.

Opportunistic Sensor Calibration

Periodic re-calibration of sensors against reference sensor data is required to ensure low-cost sensors provide air quality information that meets the regulatory standards. Current solutions require low-cost sensors to be located close to the reference station for a sufficient period of time (e.g., a week or fortnight) prior to calibration. Bringing hundreds of thousands of sensors regularly close to a reference sensor is not feasible. We envision opportunistic and collaborative re-calibration where less accurate sensors collect calibration information whenever an opportunity presents itself (e.g., when the sensor is located close to a reference station). Such calibration information can be exchanged [4] opportunistically between devices with re-calibration performed when a sufficient amount of (recent enough) information has been collected. To ensure re-calibration can accurately capture reference patterns from limited data, machine learning techniques that generalize across different contexts are needed. We envision calibration models to build on deep learning, as such techniques have been shown capable of generalizing across complex environmental and other contexts better than simpler models [10, 11]. Sensor calibration happens at the backend using data collected from similar sensors and the closest reference station (Fig. 1). Inference can be run at the edge, or even on the sensor devices requiring calibration, reducing the bandwidth load on the rest of the network.

Communication Technology for Massive Scale Deployment

Capturing local variations in air quality requires dense deployments that integrate upward of 1000 sensors/mi^2. Deployments at this scale would generate high data velocity, demanding bandwidth up to 1 GB/s/mi^2 from the network using current sensors. Data amounts are expected to grow with sensor capabilities and as more pollutants and their properties are measured. For example, hyperspectral cameras producing images of 30–300 MB in less than a second are used to estimate air quality over open areas. Supporting large-scale hyperspectral imaging is infeasible with current or even near-future communication technologies. Current communication technologies, such as LTE-4G and narrowband Internet of Things (NB-IoT) do not support a large amount of connections (4G) or suffer from insufficient bandwidth (NB-IoT). 5G networks are needed to enable deployments of this scale as they support up to millions of devices per square mile and provide necessary bandwidth to process the generated data.

Multi-Access Edge Computing

Traditionally, air quality parameters measured by sensors are transmitted to, and stored and processed at remote data centers. This approach is only suitable for small-to-moderate scale deployments where the volume of data collected from sensors is sufficiently small and does not exceed networking or processing bandwidth. Massive scale deployments integrating tens of thousands of sensors generate data at faster rates than can be transmitted and processed at remote infrastructure, requiring a faster way to operate. We envision massive scale air quality monitoring to take advantage of (mobile) edge computing, enabled by (5G) local connectivity to offer the required data storage and processing capability at the edge of the network [12]. Edge nodes are well suited to perform sensor calibration in the local area, or at least (re)train the necessary calibration models, and collecting and processing data in local areas.

Testbed Deployment

The scale of air quality deployments is increasing to a level where realizing our vision is becoming possible. Both the number of sensors and the density of sensor deployments are continually increasing. Cheng *et al*. [4] describe a deployment of 1000 low-cost sensors in Beijing, whereas the use of Google Street View cars for collecting air quality measurements has been explored as a way to increase spatial scale of measurements [6]. Our research is building toward multiple smaller-scale deployments carried out at locations with different characteristics.

Figure 2 shows a map of our current deployments, consisting of professional-grade measurement stations, industrial-grade sensors mounted at fixed locations, and mobile sensors carried by volunteer participants (see the right side of Fig. 1 for examples of the sensors used in our deployments). We have deployed sensors at three sites in Helsinki, Finland, chosen as representative examples of urban environments with different characteristics, enabling dense capture of air quality measurements. Our current deployment locations are (1) a shipping district with congested traffic; (2) a residential area located away from industry or congestion; and (3) a mixed residential and university area located close to congested roads. In total, our deployments currently include 100 mobile sensors, 12 NB-IoT sensors in a fixed location, and several test sensors co-located with reference stations at the first and third deployment sites. The mobile sensors capture environmental variables, including relative humidity and temperature, particulate matter concentrations in the form of $PM_{2.5}$ and PM_{10}, and gaseous pollutants, including NO_2, and CO. The NB-IoT sensors additionally measure SO_2 (traffic pollutants). To test generality across locations and to ensure sufficient diversity in the measurements, we carried out small-scale deployments in Beijing. The deployment locations in Beijing are shown at the top right corner of Fig. 2.

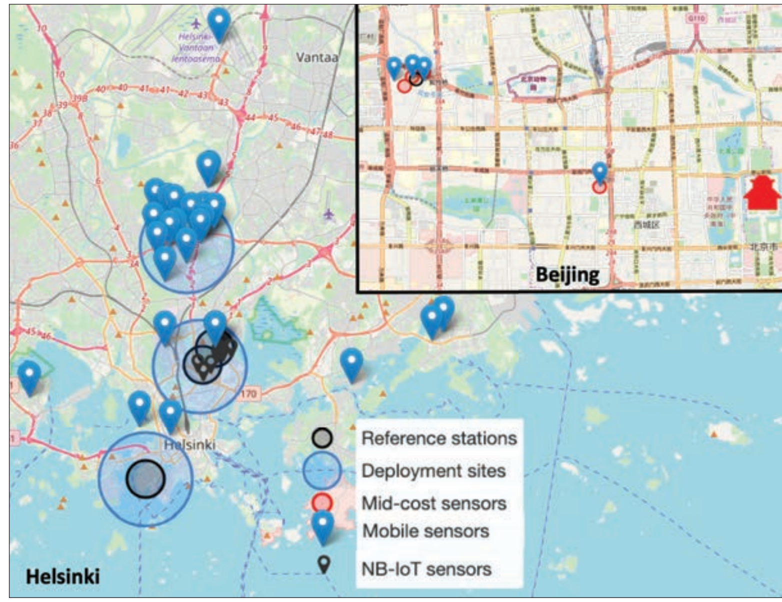

Figure 2. Maps of the Helsinki and Beijing deployments.

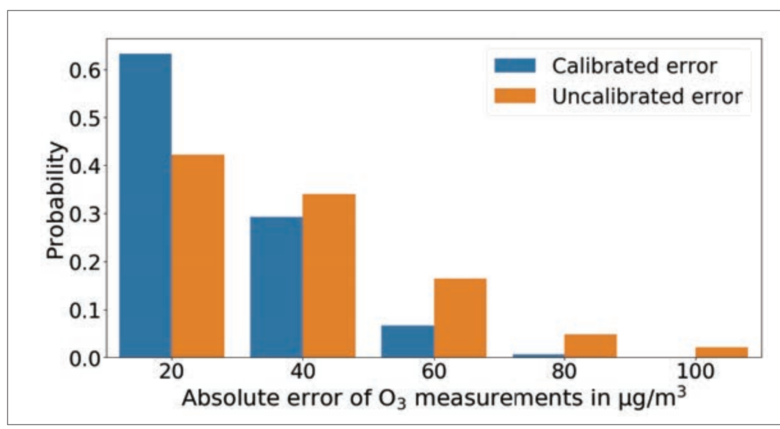

Figure 3. Calibration dramatically decreases the error between the mid-cost sensor and the reference station.

Feasibility Evaluation

Parallel to increasing size of deployments, advances in air quality sensing have led to development of methods for improving sensor accuracy through machine-learning-based calibration [7], [13] and transferring calibration models across large-scale deployments [4]. Existing works assume that at least one of the low-cost sensors is constantly located close to the reference station. In practice, this is difficult to satisfy, and ensuring good performance across all sensors is far from trivial. In this section, we demonstrate how small amounts of data from heterogeneous sensors can be used to support opportunistically learned calibration models. To ensure accurate ground truth, we conduct our experiments through measurements collected from a controlled benchmark that is part of our overall testbed, which integrates a high-cost reference station, a mid-cost sensor, and a low-cost air quality sensor in close proximity of each other.

Calibration Performance: We first assess overall feasibility of calibration by assessing performance of the mid-cost sensor after calibrating it against a reference station. For calibration, we use a deep learning model consisting of convolutional layers, fully connected neural network layers, and long short-term memory (LSTM) layers that model temporal dependencies. The structure of our model is similar to the one used for calibrating thermal camera measurements [10]. By fusing individual atmospheric variables together to form representations in the convolutional layer, we form feature representations that are more effective than conventional hand-crafted statistical features. The LSTM layers help in capturing long-term dependencies and variable-length effects within observations to take into account temporal correlations in air quality measurements. Figure 3 shows the error of low-cost sensor O_3 measurements before and after calibration. With two-and-a-half days of training data, the model is able to reduce the error by 56 percent. With more diverse, longer datasets, the error can be reduced even further.

Impact of Available Training Data: Our proposed vision relies on the ability to quickly calibrate low-cost sensors with data obtained from short visits in the vicinity of reference stations, such as high-quality measurement towers. To show the feasibility of calibration with small amounts of data, we demonstrate that the performance of our calibration model is robust and performs well with reduced training data. We carry out experiments on two subsets, referred to as high and low, containing data from periods where pollutant concentrations are high or low, respectively. The full dataset consists of meteorological variables (temperature, relative humidity, wind

Requiring all calibrated sensors to pass in the vicinity of a reference station is unrealistic, and in practice our vision requires creating calibration models out of data taken by different sensors. These sensors can rely on different sensing technologies, which differ in cost and accuracy.

	PM2.5		PM10	
Train: Test:	High Low	Low High	High Low	Low High
10%	1.38	9.09	2.88	17.84
20%	1.38	8.71w	3.23	16.79
30%	1.40	8.67	3.14	15.41
40%	1.38	8.29	3.40	15.04
50%	1.39	8.26	4.79	15.58
60%	1.79	8.48	3.69	15.89
70%	1.79	8.30	4.42	14.84
80%	1.98	7.95	3.86	15.15
90%	3.00	7.64	3.52	15.32
100%	1.38	7.96	4.63	15.88
Mixed	2.33	7.59	6.91	14.42
Orig. error	5.43	10.84	21.34	30.04

Table 1. Results of feasibility evaluation.

speed, and wind direction), and two pollutants, $PM_{2.5}$ and PM_{10}, from 50 days of measurements. The subsets were selected based on the sum of observed concentrations (higher or lower w.r.t. each other) within five-day windows of the dataset. The sub-division of data was done to ensure short-term correlations in pollution patterns are preserved, while effects from seasonal patterns can be broken.

To evaluate model performance with different amounts of training data, we progressively drop data in increments of 10 percent until only 10 percent of measurements were left (i.e., 10–100 percent). For each subset size, we randomly pick the corresponding subset and repeat training data selection and testing 10 times. Errors are then calculated as averages over the 10 repetitions (Table 1). The test dataset is the same for all the models, and the original errors (mid-cost sensor vs reference measurement before calibration) are shown in the last row. We train and test the models separately with low and high pollutant concentrations, as shown on the two top rows of Table 1. When trained with high concentration data, performance is over four times better than when the model has only seen low concentrations in the training phase (train: low). In particular, the differences between training data amounts (rows) are much smaller than the differences between low and high pollution concentration partitions of the same size (adjacent cells on the same row). This shows that the quality of training data is much more important than using massive quantities of data for training. Calibration can be performed with even a week of training data, but re-calibration should be carried out whenever pollutant concentrations change significantly.

Training with Mixed Sensor Data: Requiring all calibrated sensors to pass in the vicinity of a reference station is unrealistic, and in practice our vision requires creating calibration models out of data taken by different sensors. These sensors can rely on different sensing technologies that differ in cost and accuracy. For example, cars parked close to reference stations could be used to supplement short-term measurements taken by people passing by reference stations. To test the feasibility of calibration with heterogeneous data, we next assess calibration performance when a mix of measurements from a low-cost and a mid-cost sensor from different time periods are used to train the calibration model. The meteorological variables are obtained from a reliable weather station, while the pollutants are measured by the low-cost sensor in 50 percent of the data and by the mid-cost sensor in the other 50 percent. The data consisted of five randomly selected non-overlapping time windows from each sensor's data. The process was repeated 10 times. The averaged results are shown in the second-last row (labeled Mixed) in Table 1. From the results, we observe better performance with mixed data than with the reduced data. The accuracy of a mid-cost sensor (row Orig. error) can be improved with calibration data obtained from low-cost sensors, demonstrating that mixing calibration data even across different types of sensing technology is feasible.

Related Work: Urban Air Quality Monitoring

Conventionally, urban-scale air quality measurements have relied on professional-grade measurements stations that incorporate high-precision sensing instruments mounted on sensing towers (see Fig. 1 for an example). These stations are capable of measuring hundreds of parameters (per 1–10 s measurement window) with high accuracy. A state-of-the-art example is the Station for Measuring Ecosystem-Atmosphere Relations [14]. While accurate, these stations suffer from being bulky, and costly to deploy and maintain. The deployment cost of a single monitoring station can reach hundreds of thousands of dollars, with further costs resulting from maintenance and operation of the station. Due to the high cost, professional-grade equipment can only be deployed sparsely, with one station typically responsible for several city regions. Industrial scenarios rely on cheaper mid-cost sensors characterized by a smaller size and the ability to measure tens of parameters per second with high accuracy. One such device is the Vaisala AQT 420 air pollution sensor. These sensors tend to cost in the range of a few thousand dollars, similarly limiting density of deployments [15].

To provide information at finer spatial granularity, there are initiatives to equip pedestrians, public transportation, and vehicles with low-cost air quality sensors [6], [13]. These deployments rely on cheaper portable air monitoring stations, such as the sensor shown in Fig. 1. These sensors typically cost under $2500 and are able to estimate air pollution concentrations in continuous fashion. However, low-cost sensors are prone to errors resulting from cross-sensitivity between pollutants, variations in weather conditions, and sensor drift over time [7]. The accuracy of these sensors can be improved by periodically calibrating them against a reference sensor. Unfortunately, current calibration solutions require extensive amounts of manual labor, including taking sensors close to a reference station and re-deploying them, which limits the scale and granularity of air quality monitoring. However, their small size, the ability to measure dozens of air quality parameters, low power requirements, and the ability to communicate using cellular systems (e.g., LTE-4G and 5G networks) makes them likely to become the best candidates to be deployed massively in our envisioned massive air pollution sensing system.

Summary and Conclusion

We develop a vision of massive scale air quality monitoring that delivers accurate air quality information at high spatial and temporal resolution. By contrasting our vision against current deployments, we identify key research challenges and reference technologies required to deliver the vision and scale up current deployments. These challenges include solutions and best practices for supporting design, deployment, and maintenance of air quality sensors; opportunistic and collaborative sensor calibration techniques that help reaching accuracy levels close to regulatory standards; and positioning and networking solutions that facilitate analysis of data generated by massive amounts of air quality sensors. Besides highlighting the key challenges and technologies, we introduce new types of applications and use cases that our vision enables. To demonstrate the feasibility of our vision, we present results of small-scale experiments carried out using two testbed deployments that seek to address the research challenges of our vision.

We develop a vision of massive scale air quality monitoring that delivers accurate air quality information at high spatial and temporal resolution. By contrasting our vision against current deployments, we identify key research challenges and reference technologies required to deliver the vision, and scale up current deployments.

Acknowledgments

This work is supported by the MegaSense program, Helsinki Center for Data Science program, Academy of Finland grant 297741, and European Union Urban Innovative Action Healthy Outdoor Premises for Everyone (UIA03-240).

References

[1] WHO, "Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease," Tech. Rep., 2016.

[2] M. Chen *et al.*, "Urban Healthcare Big Data System Based on Crowdsourced and Cloud-Based Air Quality Indicators," *IEEE Commun. Mag.*, vol. 56, no. 11, Nov. 2018, pp. 14–20.

[3] A. Lewis and P. Edwards, "Validate Personal Air-Pollution Sensors," *Nature*, vol. 535, 2016, pp. 29–31.

[4] Y. Cheng *et al.*, "ICT: In-Field Calibration Transfer for Air Quality Sensor Deployments," *Proc. ACM Interact. Mob. Wearable Ubiquitous Tech.*, vol. 3, no. 1, Mar. 2019, pp. 6:1–6:19.

[5] Y. Zheng *et al.*, "Forecasting Fine-Grained Air Quality Based on Big Data," *Proc. 21th ACM SIGKDD Int'l. Conf. Knowledge Discovery and Data Mining*, ser. KDD '15, 2015, pp. 2267–76.

[6] J. S. Apte *et al.*, "High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data," *Environmental Science & Technology*, vol. 51, no. 12, 2017, pp. 6999–7008.

[7] M. I. Mead *et al.*, "The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks," *Atmospheric Environment*, vol. 70, 2013, pp. 186–203.

[8] Y. Gao *et al.*, "Mosaic: A Low-Cost Mobile Sensing System for Urban Air Quality Monitoring," *Proc. IEEE INFOCOM 2016*, Apr. 2016, pp. 1–9.

[9] L. Johansson *et al.*, "Fusion of Meteorological and Air Quality Data Extracted from the Web for Persononalized Environmental Information Services," *Environmental Modelling & Software*, vol. 64, 2015, pp. 143–155.

[10] T. Malmivirta *et al.*, "Hot or Not? Robust and Accurate Continuous Thermal Imaging on FLIR cameras," *Proc. IEEE International Conf. Pervasive Computing and Commun.*, Kyoto, Japan, Mar. 11–15, 2019.

[11] Y. Lin, W. Dong, and Y. Chen, "Calibrating Low-Cost Sensors by a Two-Phase Learning Approach for Urban Air Quality Measurement," *Proc. ACM Interact. Mob. Wearable Ubiquitous Tech.*, vol. 2, no. 1, Mar. 2018, pp. 18:1–18:18.

[12] E. Lagerspetz *et al.*, "Pervasive Data Science on the Edge," *IEEE Pervasive Computing*, vol. 18, no. 3, July–Sept. 2019, pp.40–49.

[13] D. Hasenfratz *et al.*, "Deriving High-Resolution Urban Air Pollution Maps Using Mobile Sensor Nodes," *Pervasive and Mobile Computing*, vol. 16, 2015, pp. 268–85.

[14] M. Kulmala, "Build a Global Earth Observatory," *Nature*, vol. 553, 2018, pp. 21–23.

[15] E. Lagerspetz *et al.*, "Megasense: Feasibility of Low-Cost Sensors for Pollution Hot-Spot Detection," *2019 IEEE 17th Int'l. Conf. Industrial Informatics*, vol. 1, July 2019, pp. 1083–90.

Biographies

Naser Hossein Motlagh (naser.motlagh@helsinki.fi) is a postdoctoral researcher in the Department of Computer Science, University of Helsinki, Finland. He completed his D.Sc. degree in networking technology at Aalto University, Finland, in 2018. His research interests include IoT, WSN and UAVs.

Eemil Lagerspetz (eemil.lagerspetz@cs.helsinki.fi) is a postdoctoral researcher in the Department of Computer Science, University of Helsinki. He completed his Ph.D. in computer science at the University of Helsinki in 2014. His research interests include large-scale data analysis (big data), ubiquitous computing, and energy efficiency.

Petteri Nurmi (petteri.nurmi@cs.helsinki.fi) is an associate professor at the University of Helsinki. He completed his Ph.D. in computer science at the University of Helsinki in 2009. His research interests include distributed systems, pervasive data science, and sensing systems.

Xin Li (xin.li@cs.helsinki.fi) is a research assistant in the Department of Computer Science, University of Helsinki. His research interests include sensing data analysis and sensor calibration with machine learning approaches.

Samu Varjonen (samu.varjonen@cs.helsinki.fi) is a postdoctoral researcher at the University of Helsinki. He completed his Ph.D. in computer science at the University of Helsinki in 2012. His research interests include IoT, overlay networks, and sensors.

Julien Mineraud (julien.mineraud@cs.helsinki.fi) is a postdoctoral researcher in the Department of Computer Science, University of Helsinki. He completed his Ph.D. at the TSSG and the Waterford Institute of Technology, Ireland, in 2013. His research interests include IoT and big data analytics.

Matti Siekkinen (matti.siekkinen@cs.helsinki.fi) is a senior researcher at the University of Helsinki. He completed his Ph.D. at EURECOM/University of Nice Sophia-Antipolis in 2006. His research interests include IoT, mobile computing, and multimedia systems.

Andrew Rebeiro-Hargrave (andrew.rebeiro-hargrave@helsinki.fi) is a senior researcher at the University of Helsinki. He completed his Ph.D. in geography at King's College London in 2001. His research interests include IoT, cognitive computing, and multimedia systems.

Tareq Hussein (tareq.hussein@helsinki.fi) is a visiting professor at the University of Helsinki. He completed his Ph.D. in physics at the Division of Atmospheric Sciences, University of Helsinki in 2005. His research interests include urban and indoor air quality and exposure.

Tuukka Petäjä (tuukka.petaja@helsinki.fi) is a full professor at the Institute for Atmospheric and Earth System Research, University of Helsinki. He completed his Ph.D. in physics at the University of Helsinki in 2006. His research interests include atmospheric aerosol particles and their role in climate change and air quality.

Markku Kulmala (markku.kulmala@helsinki.fi) is an academy professor and head of the Institute for Atmospheric and Earth System Research, University of Helsinki. He completed his Ph.D. in theoretical physics at the University of Helsinki in 1988. His research interests include nanoparticle formation in the atmosphere, and biosphere-atmosphere interactions and feedback.

Sasu Tarkoma (sasu.tarkoma@cs.helsinki.fi) is a full professor and head of the Department of Computer Science, University of Helsinki. He completed his Ph.D. in computer science at the University of Helsinki in 2006. His research interests include mobile computing, Internet technologies, and AI.