MegaSense : Feasibility of Low-Cost Sensors for Pollution Hot-spot Detection

Lagerspetz, Eemil

Lagerspetz , E , Hossein Motlagh , N , Zaidan , M , Fung , P L , Mineraud , J , Varjonen , S , Siekkinen , M , Nurmi , P , Matsumi , Y , Tarkoma , S & Hussein , T 2019 , MegaSense : Feasibility of Low-Cost Sensors for Pollution Hot-spot Detection . in Proceedings : 2019 IEEE 17th International Conference on Industrial Informatics (INDIN) . vol. 1 , IEEE International Conference on Industrial Informatics (INDIN). , vol. 1 , IEEE , pp. 1083-1090 , International Conference on Industrial Informatics , Helsinki- Espoo , Finland , 22/07/2019 . https://doi.org/10.1109/IN

MegaSense: Feasibility of Low-Cost Sensors for Pollution Hot-spot Detection

Eemil Lagerspetz†, Naser Hossein Motlagh†, Martha Arbayani Zaidan‡, Pak L. Fung‡, Julien Mineraud†,
Samu Varjonen†, Matti Siekkinen†, Petteri Nurmi†, Yutaka Matsumi††, Sasu Tarkoma†, and Tareq Hussein‡

‡ Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, Finland

† Department of Computer Science, University of Helsinki, Finland

†† Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Japan

Emails: †, ‡ firstname.lastname@helsinki.fi, †† matsumi@nagoya-u.jp

Abstract—Air pollution is a major problem in urban areas, where high population density is accompanied with excess anthropomorphic emissions impacting the environment and increasing health effects. Highly accurate air quality monitoring stations have been used to monitor the severity of the problem and warn citizens. However, air quality can vary sharply even within the same city block, and pollution exposure can vary even 30% between individuals living in the same residence. Therefore, a dense deployment of air quality sensors is needed to detect these variations, and protect citizens from overexposure. Low-cost air quality sensors make it possible to densely instrument a city and detect hot spots as they happen. However, thus far limited information exists on their accuracy and practicability. In this paper, we conduct a 44 day measurement campaign to assess performance of low-cost air quality monitors under different environmental conditions. As practical use case we consider pollution hot spot detection. Our results show that the mean error of low-cost sensors is small, but the variation in error is significantly larger than with reference sensors. We also show that the accuracy is sufficient for applications relying on variations in air quality index values, such as hotspot detection.

Index Terms—Air Quality, Low-Cost Sensors, Reference Station, Particulate Matter, Gas and Aerosol Sensor, Accuracy.

(a) $PM_{2.5}$ portable sensor equipped with a GPS, WiFi module and mobile phone connectivity for data logging and visualization [10].

(b) $PM_{2.5}$, gas, and ambient conditions sensor package equipped with microSD memory for data logging and LCD display visualization.

Fig. 1. Examples of portable aerosol & gas pollution sensors.

I. 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% of all deaths) with mortality in low and middle income countries particularly heavily affected by air pollution [1]. Air pollution is particularly problematic in urban areas with cities exhibiting the highest levels of pollution and suffering worst from adverse health effects. Indeed, over three thousand anthropogenic air pollutants have been identified in urban areas [2] with most relating to combustion sources, such as traffic activities and biomass burning. The complex nature of urban air pollution has prompted attempts to quantify the levels of air pollution in cities across the globe [3].

Air quality has been conventionally measured using professional-grade measurements stations that incorporate high precision sensing instruments mounted on tall sensing towers. A state-of-the-art example of such a station is the Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) [4], illustrated in Figure 2. While accurate, these stations are expensive to install and operate (one station typi-

cally costs in excess of $100,000). This limits resolution of air quality monitoring as a single station is typically responsible for an entire city region. Such resolution is not sufficient as air pollutant levels can vary drastically even within a city block [5]. Furthermore, pollution levels are highly dependent on the density of pollution sources, weather and altitude [6].

Low-cost air quality sensors costing a few thousand dollars or less have recently emerged as alternative to professional-grade measurement stations. Thanks to their affordability, low-cost sensors can be deployed densely to ensure a high monitoring resolution. Indeed, urban deployments of low-cost sensors are increasingly common with examples including Chicago Array of Things, which uses low-cost air pollution sensors attached to street lights [7], and Helsinki metropolitan Air Quality testbed (HAQT) [8]. Besides increased deployment, the commercial market for low-cost sensors is growing with estimates suggesting the market value of gaseous pollutant sensors will exceed $3 billion USD by 2027 [9].

While popularity of low-cost sensors is increasing, little information exists on the quality of the measurements they provide, particularly in real-time monitoring (see Table IV in Sec. V). Moreover, the information that exists has predominantly focused on establishing correspondence between low-cost sensors and professional-grade stations used as reference,

without quantifying the effects of weather or monitoring resolution. The present paper contributes by *quantifying performance of low-cost sensors for real-time pollution hot-spot detection*. We perform our analysis considering measurements from a 44-day measurement campaign conducted in Helsinki, Finland. We collect data using two representative examples of low-cost sensors (Figure 1) directly next to an air quality measurement reference station, the Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) [4]. Contrary to existing campaigns, we consider (near) real-time reporting accuracy $(1 - 5$ minutes) instead of aggregated measurements. Our focus is on low cost sensors (less than $2500) and on understanding how their performance is affected by environmental conditions. Our results show the measurements of the low-cost sensor to follow the same general trend as the reference station and the overall accuracy of the low-cost sensors to be high. However, we also find variation in sensor values to be significant which results in low correspondence when directly compared to reference station values. These results suggest that low-cost sensors can indeed be used for real-time applications, such as pollution hot-spot detection, as long as the error margin of the low-cost sensors remains sufficiently small.

The contributions of this paper are summarized as follows.

1) **Measurements** A 44-day measurement campaign from 7th May - 19th July 2018 (weekdays only) using both low-cost air pollution sensors and a reference monitoring station. Both aerosols and gaseous pollutants were measured.
2) **Analysis** Performance analysis of low-cost sensors including accuracy estimation and correlation analysis against reference station.
3) **New Insights** Our low-cost sensor results have high accuracy but high variation. Its MAE is 3.4 for $PM_{2.5}$, or 1.36% to 34.0% in terms of Air Quality Status Ranges (See III), and 7.24 for O_3 (1.43%-23.35%). This motivates the use of dense low-cost sensor deployments for local real-time air quality status information.

II. MEASUREMENT CAMPAIGN

To determine the potential for continuous air quality sensing of low-cost sensors, we estimate the accuracy of two low-cost sensors in the field: a GPS-enabled portable device (see Figure 1(a)), and a self-assembled sensor device (Figure 1(b)). Both of the sensors measure particulate matter (the former $PM_{2.5}$, and the latter includes a dust particle counter), and the latter also measures CO_2, O_3, CO, relative humidity RH, and temperature T. We focus on the accuracy of $PM_{2.5}$, the main particulate matter component of current air quality indexes, and O_3, one of the gaseous pollutants. $PM_{2.5}$ is linked to respiratory problems [11] and visible as smog. O_3 exposure can cause lung damage [12].

Experimental Devices: The two low-cost sensor units that we used are shown in Figure 1. Both are custom-built devices with commodity components:

Fig. 2. Example of a reference station: SMEAR III in Helsinki.

- A portable sensor unit designed to measure $PM_{2.5}$, equipped with a GPS, WiFi module and mobile phone connectivity for data logging and visualization. This kind of portable device has been used for real-time and spatial $PM_{2.5}$ monitoring [10]. The $PM_{2.5}$ sensor module was a Panasonic model GA1 that has a thermal resistor to induce an internal upward air flow to facilitate continuous sampling.
- A movable multiple output sensor unit, capable of measuring dust, O_3, CO, relative humidity RH, and temperature T. It was equipped with microSD memory for data logging and LCD display visualization. In the unit, all sensor modules were installed inside a transparent box equipped with a fan and a hole inlet for air sampling. The sensor module used to measure dust (correlated with $PM_{2.5}$) was a Sharp GP2Y1010AU0F attached to the hole inlet. The temperature and humidity sensor was an $AM2315$ capacitive sensor. The gas sensors were $FC-22-1$ series modules with MQ type for O_3 and CO.

Experiment Setup: During the measurement campaign, we operated the above described low-cost sensors and compared their monitoring results to those obtained with the SMEAR III as an example for a reference monitoring station (See Figure 2). In total, we performed 44 monitoring sessions, lasting approximately 2 hours each, on different days during 8th May - 19th July 2018 (excluding weekends and holidays). The monitoring was conducted at a distance of two meters from the SMEAR III station. The sensors were protected from the effects of sun and wind. The results of SMEAR III were downloaded from the open-access data portal AVAA[1]. Data from the two sensors was aggregated to one-minute averages, matching the time granularity of SMEAR III data.

Environmental Conditions: Figure 3 shows temperature and

---

[1]https://avaa.tdata.fi

TABLE I
COMPARISON OF DIFFERENT AIR QUALITY SENSOR UNITS.

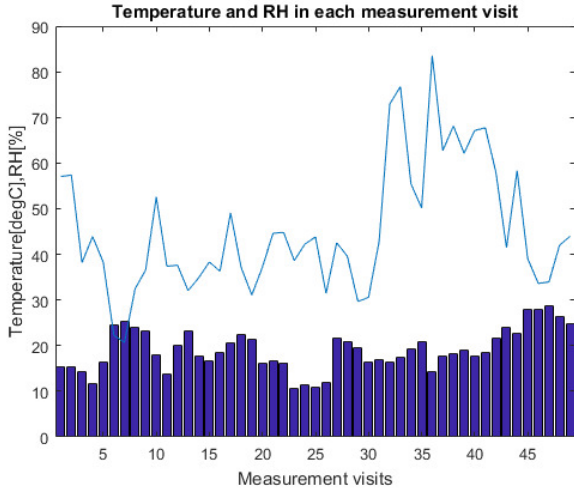| Sensor Type	Parameters	Accuracy	Deployment	Mobility	Connectivity	Power Drain	Cost (USD)	Management
Reference	300 to 800	Very High	1 to 2 per city	No	Fixed Line/Cellular	High	$\geqslant 1.5 \cdot 10^6$	Professional
Mid-cost	10 to 50	High	$\approx < 1000$	Yes	Cellular	Little	$(5 \; to \; 10) \cdot 10^3$	Semi-Professional
Low-cost	≈ 10	Low	$\approx > 10^5$	Yes	Cellular	Low	$\leqslant 2500$	Amateur

Fig. 3. Temperature and Relative Humidity during the measurement period (May 8 - July 19, 2018 (excluding weekends and holidays).

relative humidity during the measurement period. Temperature ranged from 10°C (50°F) to over 25°C (77°F) while relative humidity ranged from 20% to 85%. These values are typical to the location and season, allowing realistic evaluation of the performance of the low-cost sensors.

Data Cleaning: Our low-cost sensors were not able to measure extremely low particle concentrations below 1 $\mu g/m^3$, and therefore we only considered data with concentrations higher than this. In addition, as the sensor modules used heat to establish airflow, data values recorded directly after turning on the low-cost sensors were discarded until the sensor output was stable. Standard outlier removal was done to eliminate extremely high values.

III. LOW-COST SENSOR ACCURACY

We next assess performance of the two low-cost sensors used in the measurement campaign. We separately consider overall performance, and effects of environmental conditions. We end the section with a discussion on practicability using pollution hotspot detection as an example application that could operate using low-cost sensors.

A. Overall Result

We first consider overall performance of the low-cost sensors. A natural measure of accuracy is the average error of measurements from the reference values, or mean absolute

error (MAE). Unlike RMSE it is an unambiguous measure of average error magnitude [13]. To compare low-cost accuracy with the reference station, we use an x-y scatterplot similar to previous works on sensor calibration [14]. This plot can visually show the correspondence of the sensor and reference. In addition, we use color to show the effect of wind speed, temperature, and relative humidity, as they commonly affect sensor measurements [15].

The mean absolute error of our low-cost sensor measurements is shown in Table II. The MAE was relatively low, 3.4 for PM$_{2.5}$ (MAE of the median 2.80), and 7.24 for O$_3$ (MAE for median 6.45). This is quantified in terms of air quality status ranges in Table III in Section III-D.

TABLE II
MEAN ABSOLUTE ERROR OF MEASUREMENTS.

Measurement	MAE
PM$_{2.5}$	3.40
PM$_{2.5}$ (Median)	2.80
O$_3$	7.24
O$_3$ (Median)	6.45

B. Linearity and Wind

Besides temperature and relative humidity, wind may affect PM$_{2.5}$ and O$_3$ concentrations. For PM$_{2.5}$, we show the correspondence of our low-cost sensor data with SMEAR III data in regression plots in Figure 4. Figure 4(a) displays the scatter plot of 5 minutes measurements, indicating these two measurements were nearly linearly correlated (Pearson's $R \approx 0.31$). Furthermore, an enhanced linear correlation was obtained by considering the median of each experimental day as illustrated in Figure 4(b) ($R \approx 0.49$). As expected, the PM$_{2.5}$ was negatively correlated with wind speed (indicated by color) ($R \approx -0.45$) as shown in Figure 4. The results indicate that our low-cost sensors respond to PM$_{2.5}$ and wind speed similarly as the reference station, but have a high variance.

Figure 5 displays regression plots for O$_3$. Unlike PM$_{2.5}$, for O$_3$ 5 minute data and median per session show good linear correlations at $R \approx 0.65$ and $R \approx 0.67$, respectively. We also find wind speed to influence the correlation between the gas sensor and the SMEAR III data. For example, the variation of O$_3$ concentration was larger when the O$_3$ concentration was high and the wind speed was also high, as shown in Figures

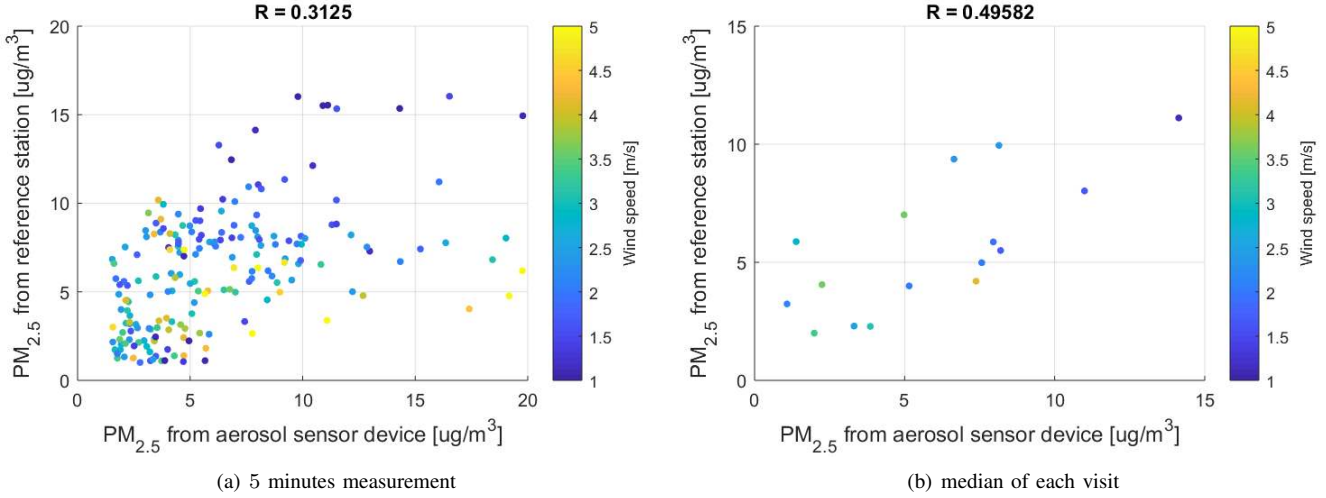(a) 5 minutes measurement   (b) median of each visit

Fig. 4. The scatter plot of $PM_{2.5}$ measured by low cost sensor (horizontal axis) and reference station (vertical axis)

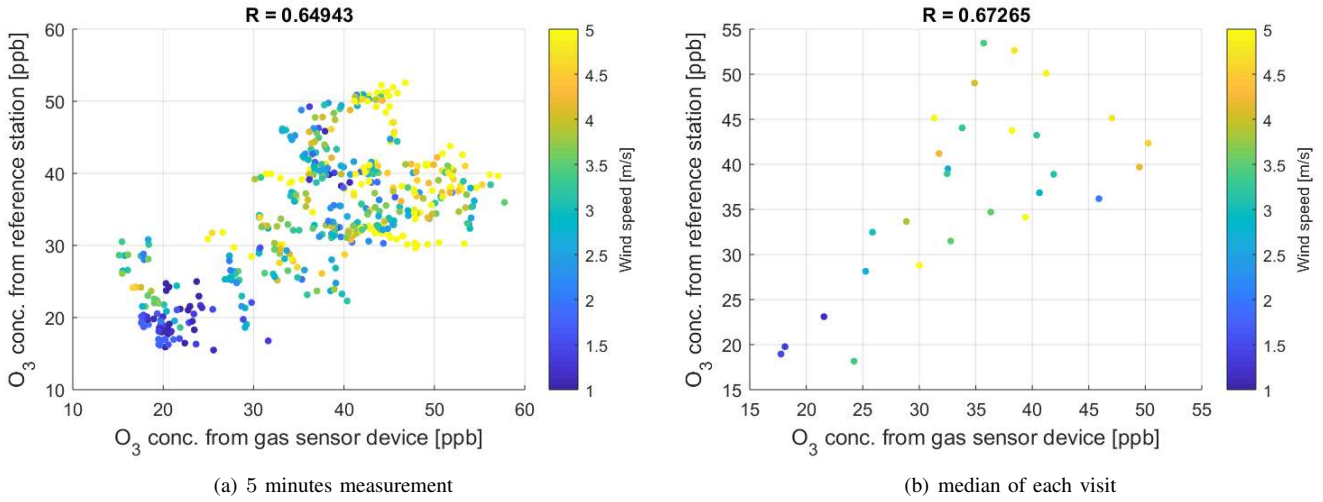(a) 5 minutes measurement   (b) median of each visit

Fig. 5. The scatter plot of O_3 measured by low cost sensor (horizontal axis) and reference station (vertical axis)

5(a) and 5(b). Accordingly, the O_3 concentration measured by the (b) sensor deviated from the linear trend when wind speed is high. As with $PM_{2.5}$ measurements, this may be explained by the variance in the data.

C. Impact of Temperature and Relative Humidity

Figure 6 shows the $PM_{2.5}$ values from sensor (a) on the x-axis, and the reference sensor data on the y-axis. The relative humidity is shown in color, with warmer colors representing higher humidity. When relative humidity rises, the relationship between the low-cost sensor's values and reference values seems to improve. Figure 7 shows the $PM_{2.5}$ values from sensor (a) on the x-axis, and the reference sensor data on the y-axis. The ambient temperature is shown as a color, with warmer colors representing higher temperatures. Temperature does not seem to have an impact on the linearity of the measurements, but higher temperatures seem to lead to higher $PM_{2.5}$ values, as expected.

D. Practicability

In the measurement location, typical values for $PM_{2.5}$ range from 0 to 40, and for O_3 from 0 to 60. To put our results in perspective, Table III shows the air quality status and corresponding pollutant value range for all of the statuses of the air quality indexes of Finland, USA, and China. For each of the statuses, we consider the range to its right, and compute the relative size of our MAE with respect to the range in the columns under $\frac{MAE}{AQ\ Status\ Range}$. For example, the range of the *Satisfactory* level of the Finnish AQI for $PM_{2.5}$ is $25-10 = 15$. Our MAE of 3.4 then corresponds to 22.67% of the range. This means that if the true $PM_{2.5}$ value is within the *Satisfactory* level, the measurement value of our low-cost sensors would on average fall to a higher or lower AQ Status (but not both) 22.67% of the time. The error decreases as we move down the table to worse AQ statuses. We can see from the table that our low-cost sensor is capable of estimating the air quality

TABLE III
RELATIONSHIP OF LOW-COST MAE TO THE LEVELS OF FINNISH, USA, AND CHINESE AIR QUALITY INDEXES. VALUES FOR USA AND CHINA OBTAINED FROM THE COMPARISON BY GAO [16].

| | Air Quality Status | | | | | | $\frac{MAE}{AQ\ Status\ Range}$ | | | |
Pollutant	Finland (FI)		USA (US)		China (CN)		FI	US	CN	Error Range
$PM_{2.5}$	Good	0-10	Good	0-15	Excellent	0-35	34.00%	22.67%	9.71%	9.71%–34.00%
$PM_{2.5}$	Satisfactory	10-25	Moderate	15-35	Good	35-75	22.67%	17.00%	8.50%	8.50%–22.67%
$PM_{2.5}$	Fair	25-50	USG*	35-65	Lightly Polluted	75-115	13.60%	11.33%	8.50%	8.50%–13.60%
$PM_{2.5}$	Poor	50-75	Moderately Polluted	65-150	Unhealthy	115-150	13.60%	4.00%	9.71%	4.00%–13.60%
$PM_{2.5}$	Very Poor	> 75	Heavily Polluted	150-250	Very Unhealthy	150-250	0.00%	3.40%	3.40%	0.00%–3.40%
$PM_{2.5}$	–	–	Hazardous	250-500	Severely Polluted	250-500	–	1.36%	1.35%	0.00%–1.36%
O_3	Good	0-60	Good	0-116	Excellent	0-100	12.07%	6.24%	7.24%	6.24% − 12.07%
O_3	Satisfactory	60-100	Moderate	116-147	Good	100-160	18.10%	23.35%	12.07%	12.07% − 23.35%
O_3	Fair	100-140	USG*	147-186	Lightly Polluted	160-215	18.10%	18.56%	13.16%	13.16% − 18.56%
O_3	Poor	140-180	Moderately Polluted	186-225	Unhealthy	215-265	18.10%	18.56%	14.48%	14.48% − 18.56%
O_3	Very Poor	> 180	Heavily Polluted	255-733	Very Unhealthy	265-800	0.00%	1.43%	1.35%	0.0% − 1.43%
O_3	–	–	Hazardous	–	Severely Polluted	–	–	–	–	–

*Unhealthy for Sensitive Groups
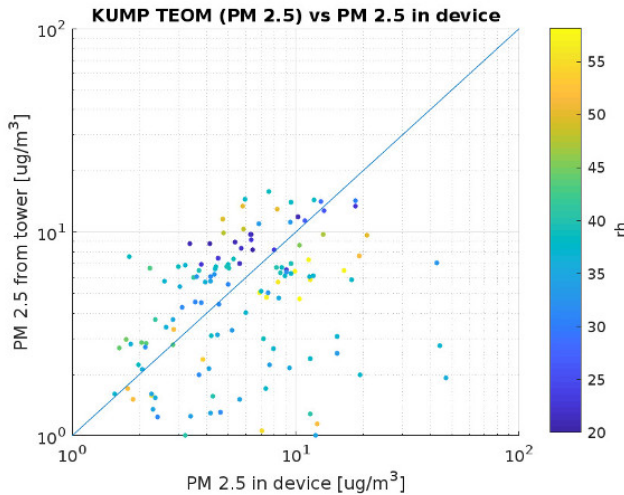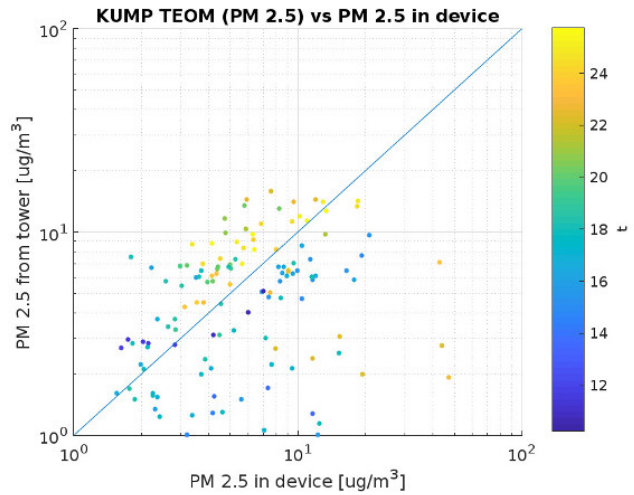

Fig. 6. Impact of Relative Humidity on $PM_{2.5}$.

Fig. 7. Impact of Temperature on $PM_{2.5}$.

status at least 80.6% of the time for all levels of the Chinese air quality index (largest error ±9.71%). For the stricter USA air quality index, the status would be correct for a randomly chosen true value 66% of the time, and for the Finnish air quality index 54.66% of the time. Note that the MAE for the best air quality status needs to be counted only once, as the error is never negative in that case.

For O_3, the relative error within air quality status ranges is more stable across air quality indexes and statuses. This is due to the smaller differences between the sizes of the ranges. Only the *Moderate* US AQI status makes our MAE of 7.24 larger than ±18.56%.

E. Application: Pollution Hotspot Detection

Our results have shown that low-cost sensors can be used to indicate the AQ status for $PM_{2.5}$ and O_3 as described by common air quality indexes with an error of ±1.43% − −34.00%, depending on the pollutant and the strictness of the air quality index. We observed that the relative error is smaller at higher pollutant concentrations, such as those that occur during a pollution hotspot. For example, if the ambient pollution in the city is the best status of Table III, and in a $PM_{2.5}$ hotspot drops to the next status by only 3.5 $\mu g/m^3$, on average our low-cost sensor will report the correct status and identify the hotspot. The higher the pollution level, the wider the value range of air quality indices. MAE of 3.5 is significantly smaller than any air quality index range, suggesting that the devices are capable of reliably identifying the corresponding pollution category of an area. In practice, pollution hotspots have $PM_{2.5}$ differences exceeding the MAE of our sensor. For example, Targino et al. (See Targino et al., Figure 4) [17] demonstrated $6 − 10\ \mu g/m^3$ differences for most of their $PM_{2.5}$ hotspots. This motivates the deployment of low-cost sensors into locations of interest in the city for $PM_{2.5}$ pollution hotspot detection.

IV. DISCUSSION

Our results show that low-cost devices can be used to monitor the air pollution in the urban areas, and greatly improve

resolution of air quality monitoring. However, employing low-cost air pollution sensor devices at outdoor environments also introduces new challenges. In the following we discuss some of these issues.

Large-Scale Data Collection: For short measurement campaigns, data can be stored directly on the device, but in large scale deployment, one must equip the devices with positioning (GPS and/or network based) and long range radio and take into consideration the possibly limited energy budget. The reliability of communications must also be considered in order to maintain an up-to-date view of the air quality. Currently available 4G cellular IoT technologies, such as NB-IoT and LTE-M, and future 5G mMTC technology [18] can provide suitable solutions also for air pollution sensing data communication.

Placement: Optimizing placement of low-cost sensor devices to measure the air pollutants at outdoor environments remains challenging. When outdoor temperature reaches $25\ °C$, the temperature inside the sensor device box could reach up to $50\ °C$. While internal temperature monitoring can help to discount these types of errors, it nevertheless limits the conditions that the devices can be operated in. This affects the functioning of battery as well as the stability of the sensors and can cause anomalies in sensor readings. Similar observations have been made with other types of portable air quality sensors [19].

Calibration: Accuracy of low-cost sensors can be improved through calibration, both in the laboratory before deployment, and also at regular intervals while in the field. Laboratory calibration has shown that low-cost air quality sensors have high potential [20], [21], but the process is costly and time-consuming. Low-cost sensors also suffer from cross-sensitivity to other pollutants in-field, and changing environmental conditions such as strong winds, high temperature, and humidity. As a consequence, more complex calibration methods must be developed to address these challenges [22]–[24]. While in-field calibration techniques have shown positive performances, in sub-part per billion for some gaseous components, more research is needed to provide data that meets regulatory requirements [25]. Nevertheless, while low-cost sensors cannot meet air quality regulation requirements, as we have shown, they can be used to give an indication of air quality status.

Power: Power consumption of sensor devices is another important concern for longer deployments. Suitable locations with available grid power are required for pollution hotspot detection. In our experiments, the battery of the sensor unit (Fig. 1(b)) was drained by the 5 sensors in less than 4 hours. However, the battery life can be easily improved by reducing the number of sensors. For example, when monitoring only O_3 concentrations with a similar low-cost sensor, one can have a battery life of 50 hours [26].

V. RELATED RESEARCH

Long-term air monitoring is currently conducted using reference stations – stationary and professional measurement sites – such as the SMEAR stations and the U.S. Environmental Protection Agency (EPA) monitoring sites. These stations are costly to build and require professional engineers to operate. Because of their size and price point, they cannot be used for dense air pollution monitoring deployments. Table I presents these reference stations along with lower-cost options.

To address the need for wider deployments, semi-professional measurement stations (Mid-cost in the table) have been deployed in the last decade [32], but these sophisticated multi-parameter, fully autonomous sensor systems can easily cost more than $10,000$ USD. Nowadays, low-cost sensor units are available, at prices below $2,500$ USD [25]. In this paper, we consider even lower-cost sensor units, termed *low-cost*, typically priced below 250 USD and capable of measuring a handful of variables. Because of their small size, high availability, and low cost, these sensors can be easily used for dense coverage measurement of air quality in urban areas [33]. The Table I demonstrates the classification and comparison of these senors. In this table, the terms *low-cost*, and *mid-cost* refer to the initial purchase cost of a single sensor unit when compared against the purchase cost of a single reference station that measures the same or similar air pollutants. The reasons for the price differences stem from, for instance, the power source requirement, communication hardware, processing unit of the sensor, the casing, sensor quality and the limits of pollutant detection [34]. In addition, air quality monitoring is not limited to a specific type of sensor module. A variety of sensor devices with different sensing capabilities can be employed to measure the air quality.

Regarding the measurement of the air quality using the low-cost air quality sensors, previous work [35] has evaluated the performance of low-cost sensors in high-pollution urban environments. In the work, low-cost sensors of the same type (PUWP) were used to monitor $PM_{2.5}$ in seven different areas in Xi'an, China. The authors conclude that the employed PUWP sensor is a suitable choice for monitoring $PM_{2\cdot5}$ in urban areas, especially in the areas with high concentration, e.g. industrial areas.

There are many types of low-cost sensor modules available to measure $PM_{2.5}$, e.g. the sensors used in [36], and in our experiment we chose two representative devices, built independently by research teams in Helsinki and Japan. In addition to $PM_{2.5}$, we also consider O_3 in our measurement campaign. Furthermore, we test the accuracy of our sensors against the SMEAR III station located in Helsinki.

Besides fixed monitoring approaches, mobile monitoring is a good measurement approach to quickly cover a city area and reveal urban air pollution patterns [5], [37]. The work of Apte et al. [5] employed two Google street view mapping vehicles with onboard professional pollution measurement instruments to monitor Black Carbon (BC), NO, and NO_2 in an area of $30\ km^2$ in Oakland, CA, USA. The authors of Yu et al. [37]

TABLE IV
COMPARISON OF DIFFERENT MEASUREMENT CAMPAIGNS.

Study/Campaign/Duration	Sensor Module	Measured Variable	Data variations Reasons	Remarks
[27] Oakland, California 15–23.4.2013	Shinyei: PPD42NS	$PM_{2.5}$	NA	Low-cost, R2 = 0.01-0.59
[28] Houston, Texas 4–27.09.2013 Denver, Colorado 14.7–12.08.2014	CairClip: O_3, NO_2	O_3, NO_2	NA	Low-cost, O_3:R2 = 0.83-0.85, NO_2: R2 = 0.13-0.28
[29] Multiple Cities, USA 20.6.2013–31.3.2014	$PM_{2.5}$ monitor (PDR-1500), O_3 monitor (OEM-106), Wind (09101) RM Young, Temp RH (HMP60) Vaisala	$PM_{2.5}$, O_3, Wind, Temp, RH	Lower operational time in wintertime, RH and Temp in multiple regression equation, slightly improves the correlation coefficients.	Mid-cost O_3:R2 = 0.76, $PM_{2.5}$: R2 = 0.79
[30] Atlanta, USA 01.10.2013–40.10.2013 21.11.2013–16.12.2013 Hyderabad, India 30.01.2014–10.02.2014	Shinyei: PPD42NS, PPD20V, and PPD60PV Sensirion: SHT 15 (Temp and RH)	$PM_{2.5}$, Temp, RH	Lower correlation is found in lower concentration range (40ug/m3)	Low-cost (high conc.): R2 $\geqslant$ 0.81 (low conc): R2 = 0.21, 0.3
[31] Helsinki, Finland 2004 - Present	TEOM 1405-D, TEI 49, TEI 42S, Horiba APMA 370, Horiba APSA 360, Metek USA-1, Vaisala DPA500, HMP243, Hauke-type DMA & TSI Model	$PM_{2.5}$, PM_{10}, O_3, CO, CO_2, Wind, Temp, RH, etc.	Wind direction, weekdays/weekends, seasons, etc.	Reference station: SMEAR III, Not a campaign

employed another mobile monitoring approach using a variety of sensors to measure BC, PNC, CO, and $PM_{2.5}$. The sensors were carried by two persons along parallel pathways in specific streets in the city of Newark, NJ, USA. The experiment was conducted to asses spatial gradients in traffic-related air pollutants, while controlling temporal variations.

The mobile measurement approach used in previous works [5], [37] has been able to capture pollution variations within the city, and discover pollution hotspots. They have highlighted the problem of highly localized, dramatic pollution variations even within the same city block. Effective air pollution monitoring to keep citizens informed of this would require continuous measurements. In urban areas, pollution hotspots can occur quickly, and therefore, real-time local information of each city block is required for effective monitoring to help citizens. However, continuous mobile measurement is expensive and laborious, but continuous measurement can be achieved cost-effectively with stationary local measurement using low-cost sensors.

Finally, while the error observed by our low-cost sensors was relatively low, we could not measure extremely low concentrations that occur regularly in Helsinki. In a megacity such as Beijing, these low concentrations happen extremely rarely, and the error of low-cost sensors decreases significantly.

VI. CONCLUSION

We contributed by quantifying performance of low-cost through a 44-day in-field measurement campaign. We measured the concentration of $PM_{2.5}$ and O_3 using two low-cost sensor devices, and estimated their accuracy against a reference measurement station (SMEAR III). The results achieved from the experiments indicate that low-cost sensor

measurements follow the expected pattern of reference data, but have a noticeable error. However, the accuracy of the two devices was sufficient to indicate air quality status, such as *Good*, *Fair*, or *Unhealthy*. This level of performance is suitable for detecting pollution hotspots in urban areas. In particular, the mean absolute error is significantly smaller than the difference between a hotspot and outside of it as reported in previous work. In addition, as ambient air quality worsens, the error is smaller relative to pollutant concentrations, further improving accuracy. In light of our results, we consider low-cost sensors a viable alternative for outdoor air pollution monitoring when there is a need to increase the spatio-temporal measurement resolution, e.g., placing sensors in each city block in an urban area to detect pollution hotspots.

As future work, we plan to use deep learning-based approaches for on-line calibration of low-cost sensors, in order to deliver more accurate local real-time air quality information within an urban area. Accurate real-time air quality information is required for regulation-grade pollution measurement as well as e.g. pollution-aware navigation. Beyond the linear regression analysis of the measurement data, a non-linear correlation method, such as [38], could yield further patterns in the results.

REFERENCES

[1] World Health Organization, "Ambient air pollution: A global assessment of exposure and burden of disease," World Health Organization, Tech. Rep., 2016.

[2] J. Fenger, "Urban air quality," *Atmospheric Environment*, vol. 33, no. 29, pp. 4877 – 4900, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1352231099002903

[3] Z. J. Andersen, L. C. Kristiansen, K. K. Andersen, T. S. Olsen, M. Hvidberg, S. S. Jensen, M. Ketzel, S. Loft, M. Sørensen, A. Tjønneland, K. Overvad, and O. Raaschou-Nielsen, "Stroke and Long-Term Exposure to Outdoor Air Pollution From Nitrogen Dioxide," *Stroke*, vol. 43, no. 2, pp. 320–325, 2012. [Online]. Available: http://stroke.ahajournals.org/content/43/2/320

[4] M. Kulmala, "Build a global earth observatory," *Nature*, vol. 553, pp. 21 – 23, 2018. [Online]. Available: https://www.nature.com/articles/d41586-017-08967-y

[5] J. S. Apte, K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lunden, J. D. Marshall, C. J. Portier, R. C. Vermeulen, and S. P. Hamburg, "High-resolution air pollution mapping with google street view cars: exploiting big data," *Environmental science & technology*, vol. 51, no. 12, pp. 6999–7008, 2017.

[6] M. Zuurbier, G. Hoek, M. Oldenwening, K. Meliefste, P. van den Hazel, and B. Brunekreef, "Respiratory effects of commuters' exposure to air pollution in traffic," *Epidemiology*, vol. 22, no. 2, pp. 219–227, 2011. [Online]. Available: http://www.jstor.org/stable/23047537

[7] Array of Things, "Array of Things Expands with Partner Projects Around U.S." https://medium.com/array-of-things/array-of-things-expands-with-partner-projects-around-u-s-c0ef9171f314, 2018, [Online; accessed 15-Jan-2019].

[8] HAQT, "Helsinki metropolitan Air Quality Testbed (HAQT)," http://fmispace.fmi.fi/index.php?id=haqt/, 2017, [Online; accessed 15-Jan-2019].

[9] D. Pugh, "The growing demand for low cost environmental sensors in smart cities," https://www.idtechex.com/research/articles/the-growing-demand-for-low-cost-environmental-sensors-in-smart-cities-00010361.asp, 2016, [Online; accessed 12-March-2019].

[10] T. Nakayama, Y. Matsumi, K. Kawahito, and Y. Watabe, "Development and evaluation of a palm-sized optical PM2.5 sensor," *Aerosol Science and Technology*, vol. 52, no. 1, pp. 2–12, 2018. [Online]. Available: https://doi.org/10.1080/02786826.2017.1375078

[11] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian, "The impact of PM2.5 on the human respiratory system," *Journal of thoracic disease*, vol. 8, no. 1, p. E69, 2016.

[12] F. Broeckaert, K. Arsalane, C. Hermans, E. Bergamaschi, A. Brustolin, A. Mutti, and A. Bernard, "Lung epithelial damage at low concentrations of ambient ozone," *The Lancet*, vol. 353, no. 9156, pp. 900–901, 1999.

[13] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.

[14] X. Zhai, A. G. Russell, P. Sampath, J. A. Mulholland, B.-U. Kim, Y. Kim, and D. D'Onofrio, "Calibrating r-line model results with observational data to develop annual mobile source air pollutant fields at fine spatial resolution: Application in atlanta," *Atmospheric Environment*, vol. 147, pp. 446 – 457, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1352231016308111

[15] T. Hussein, A. Karppinen, J. Kukkonen, J. Härkönen, P. P. Aalto, K. Hämeri, V.-M. Kerminen, and M. Kulmala, "Meteorological dependence of size-fractionated number concentrations of urban aerosol particles," *Atmospheric Environment*, vol. 40, no. 8, pp. 1427–1440, 2006.

[16] F. Gao, "Evaluation of the Chinese new air quality index (GB3095-2012): based on comparison with the US AQI system and the WHO AQGs," 2013.

[17] A. C. Targino, M. D. Gibson, P. Krecl, M. V. C. Rodrigues, M. M. dos Santos, and M. de Paula Corrêa, "Hotspots of black carbon and PM2.5 in an urban area and relationships to traffic characteristics," *Environmental Pollution*, vol. 218, pp. 475 – 486, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0269749116305978

[18] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: current status and future perspectives toward 5G systems," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 10–17, Sep. 2015.

[19] J. Mineraud, F. Lancerin, S. Balasubramaniam, M. Conti, and S. Tarkoma, "You are AIRing too Much: Assessing the Privacy of Users in Crowdsourcing Environmental Data," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1, Aug 2015, pp. 523–530.

[20] M. Gerboles, L. Spinelle, A. Kotsev, and M. Signorini, "Airsenseur: An open-designed multi-sensor platform for air quality monitoring," Tech. Rep., 06 2015.

[21] D. Oletic and V. Bilas, "Design of sensor node for air quality crowd-sensing," in *IEEE Sensors Applications Symposium (SAS)*, April 2015, pp. 1–5.

[22] B. Maag, Z. Zhou, and L. Thiele, "W-air: Enabling personal air pollution monitoring on wearables," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 24:1–24:25, Mar. 2018. [Online]. Available: http://doi.acm.org/10.1145/3191756

[23] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: No, co and co2," *Sensors and Actuators B: Chemical*, vol. 238, pp. 706–715, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092540051631070X

[24] N. Zimmerman, A. A. Presto, S. P. N. Kumar, J. Gu, A. Hauryliuk, E. S. Robinson, A. L. Robinson, and R. Subramanian, "A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring," *Atmospheric Measurement Techniques*, vol. 11, no. 1, pp. 291–313, 2018. [Online]. Available: https://www.atmos-meas-tech.net/11/291/2018/

[25] A. C. Lewis, E. von Schneidemesser, and R. E. Peltier, "Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications," World Meteorological Organization, Geneva, Switzerland, Technical Report 1215, May 2018.

[26] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory Air Pollution Monitoring Using Smartphones," *2nd International Workshop on Mobile Sensing*, pp. 1–5, 2012.

[27] D. M. Holstius, A. Pillarisetti, K. R. Smith, and E. Seto, "Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California," *Atmospheric Measurement Techniques*, vol. 7, no. 4, pp. 1121–1131, 2014. [Online]. Available: https://www.atmos-meas-tech.net/7/1121/2014/

[28] R.M. Duvall, R.W. Long, M.R. Beaver, K.G. Kronmiller, M.L. Wheeler and J.J. Szykman, "Performance Evaluation and Community Application of Low-Cost Sensors for Ozone and Nitrogen Dioxide," *Sensors*, vol. 16, no. 10, 2016.

[29] W. Jiao, G. S. W. Hagler, R. W. Williams, R. N. Sharpe, L. Weinstock, and J. Rice, "Field Assessment of the Village Green Project: An Autonomous Community Air Quality Monitoring System," *Environmental Science & Technology*, vol. 49, no. 10, pp. 6085–6092, 2015, pMID: 25905923. [Online]. Available: https://doi.org/10.1021/acs.est.5b01245

[30] K. K. Johnson, M. H. Bergin, A. G. Russell, and G. S. W. Hagler, "Field Test of Several Low-Cost Particulate Matter Sensors in High and Low Concentration Urban Environments," 2018.

[31] L. Järvi, H. Hannuniemi, T. Hussein, H. Junninen, P.P. Aalto, R. Hillamo, T. Mäkelä, P. Keronen, E. Siivola, T. Vesala and M. Kulmala, "The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface–atmosphere interactions in Helsinki, Finland."

[32] Tareq Hussein and Brandon E. Boor and Vanessa N. dos Santos and Juha Kangasluoma and Tuukka Petäjä and Heikki Lihavainen, "Mobile aerosol measurement in the eastern mediterranean - a utilization of portable instruments," *Aerosol and Air Quality Research*, vol. 17, no. 7, pp. 1875–1886, 2017. [Online]. Available: https://doi.org/10.4209/aaqr.2016.11.0479

[33] Y. Wang, J. Li, H. Jing, Q. Zhang, J. Jiang, and P. Biswas, "Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement," *Aerosol Science and Technology*, vol. 49, no. 11, pp. 1063–1077, 2015. [Online]. Available: https://doi.org/10.1080/02786826.2015.1100710

[34] C. Lung, R. Jones, C. Zellweger, A. Karppinen, M. Penza, T. Dye, and D. Hagan, *Low-cost sensors for the measurement of atmospheric composition: Overview of topic and future applications*, World Meteorological Organization, Geneva, Switzerland, 2018.

[35] M. Gao, J. Cao, and E. Seto, "A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China," *Environmental Pollution*, vol. 199, pp. 56

– 65, 2015. [Online]. Available: http://www.sciencedirect.com/science/ article/pii/S0269749115000160

[36] N. Hossein Motlagh, M. Arbayani Zaidan, E. Lagerspetz, S. Varjonen, J. Toivonen, J. Mineraud, A. Rebeiro-Hargrave, M. Siekkinen, T. Hussein, P. Nurmi, and S. Tarkoma, "Indoor air quality monitoring using infrastructure-based motion detectors," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, Helsinki, Finland, July 2019.

[37] C. H. Yu, Z. Fan, P. J. Lioy, A. Baptista, M. Greenberg, and R. J. Laumbach, "A novel mobile monitoring approach to characterize spatial and temporal variation in traffic-related air pollutants in an urban community," *Atmospheric Environment*, vol. 141, pp. 161 – 173, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/ pii/S1352231016304770

[38] M. A. Zaidan, V. Haapasilta, R. Relan, P. Paasonen, V.-M. Kerminen, H. Junninen, M. Kulmala, and A. S. Foster, "Exploring non-linear associations between atmospheric new-particle formation and ambient variables: a mutual information approach," *Atmospheric Chemistry and Physics*, vol. 18, no. 17, pp. 12 699–12 714, 2018.