

# A Personalized Air Quality Sensing System – A preliminary study on assessing the air quality of London Underground Stations

Ruizhe Zhang, Daniele Ravi, Guang-Zhong Yang and Benny Lo

**Abstract**— Recent studies have shown that air pollution has a negative impact on people’s health, especially for patients with respiratory and cardiac diseases (e.g. COPD, asthma, ischemic heart disease). Although there are already many air quality monitoring stations in major cities, such as London, these stations are sparsely located, and the periodic collection of information is insufficient to provide the granularity needed to assess the environmental risk for an individual (e.g. to avoid exacerbation). Wearable devices, on the other hand, are more suitable in this context, providing a better estimation of the air quality in the proximity of the person. Therefore, relevant warnings and information on health risks can be provided in real-time. As a proof of concept, we have developed a wearable sensor for continuous monitoring of air quality around the user, and a preliminary study was conducted to validate the sensor and assess the air quality in London underground stations. Based on the PM<sub>2.5</sub> (particulate matter with a diameter of 2.5  $\mu\text{m}$ ), temperature and location information, a model is generated for predicting the air quality of each station at different times. Our preliminary results have shown that there are significant differences in air quality among stations and metro lines. It also demonstrates that wearable sensors can provide necessary information for users to make travel arrangements that minimize their exposure to polluted air.

## I. INTRODUCTION

Studies have shown that air pollution has a direct impact on people’s health. The World Health Organization (WHO) estimated that 23% of all deaths globally in 2002 were caused by environmental factors [1]. WHO has also indicated that exposure to air pollution is one of the leading risk factors for mortality, which is linked to around 7 million deaths in 2012 [2]. The effects of air pollution on health range from subtle symptoms to more critical illness, such as shortness of breath, coughing, and aggravation of asthma. Exposure to air pollutants is also a major risk factor for chronic cardiac and respiratory diseases, such as stroke, COPD (Chronic Obstructive Pulmonary Disease) and bronchitis. The effect of air pollution is even more significant in emerging economies, such as China and India. Moreover, the lack of a proper air quality monitoring system can put the population’s health at risk, worsening the effect of the air pollution since most people are not aware that they are breathing hazardous substances in a polluted environment. Although there are already public air pollution monitoring stations distributed across major cities, these data are insufficient to evaluate the environmental risk for an

individual in real-time. Especially for patients who are suffering from chronic respiratory diseases, this information can be vital for them to avoid polluted areas/routes and prevent exacerbation or coughing. It has been demonstrated, in previous research, that body worn sensors [3] can sense physiological activity information to provide personalized health information in real-time. With personalized environmental sensing capability, an integrated wearable sensor system can help users to better understand their surrounding environment and the potential environmental impacts on their health, thus improving the quality of life in the long-term. In this paper, we present a personalized sensor system that can monitor the air quality of the surrounding environment (e.g. measuring the PM<sub>2.5</sub>), extract and analyze the sensory data and provide information for people to assess the potential health risks.

Particulate matter (PM) are solid or liquid particles found in the air such as dust, smoke, dirt, or soot. These particles come in many sizes and shapes and they can be emitted from construction, road paving, smokestacks, fire, etc. Studies have shown that high concentration of PM could affect climate and human health.

Our system focuses on monitoring the air pollutants based on these particulate matters and the sensor data are transmitted to a mobile phone app for analysis. The obtained data are pre-processed to reduce noises and remove irrelevant samples, the health risk is then quantified by combining air quality and physiological data using an artificial Neural Network (NN), and the result information can be shown to the user via the mobile app.

## II. RELATED

Wearable sensors have been proposed for sensing air quality with the aim of improving health and wellbeing. In [4] a smartphone has been used to receive data from a set of wireless sensors, which continuously monitor the air quality around the user. More recent studies have shown that the condition of personal inhalation exposure to air pollutants also depends on the daily activity patterns. For example, systems such as ExposureSense [5] and CalFit [6] have used activity recognition and air quality sensing, to estimate the user’s air pollution exposure. Activity recognition is performed using inertial sensors whilst air quality is measured using air pollutant sensors (e.g. PM<sub>10</sub>). With the available GPS system integrated within smartphones, a spatial-temporal map can be generated to find the correlation between the activity patterns and air pollution exposure. For example, Wu *et al.* [7] have introduced an incremental diagnosis method (IDM) which uses wireless sensors to monitor a patient’s activity in real-time with a Naïve

Bayesian network. In the system, the wearable sensor was configured as the monitor device, collecting the data from the ambient environment to enable early symptoms detection. Instead, in Tsow *et al.* [8] and Li *et al.* [9], sensors are developed to sense the targeted air pollutants and provide a comprehensive view of the air quality condition.

Several commercial products are also available for air quality sensing, such as Air Quality Egg [10] and TZOA [11], which were designed explicitly to improve the quality of life of the users.

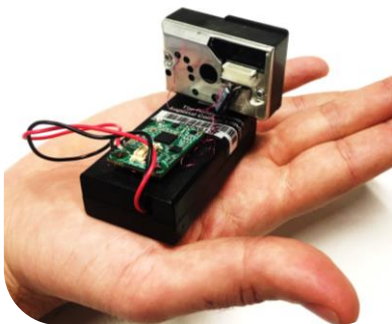


Figure 1. The BSN personalized air quality sensor

A more complete platform has been designed by Rizea *et al.* [12], which can measure a wide domain of parameters such as CO<sub>2</sub>, NO, PM, humidity, temperature, etc. The platform has also been enhanced with a real-time data processing system, which can process, store and visualize the air quality information. Kim *et al.* [13] have investigated the quality of air in the metro system of Seoul, South Korea. The data was collected by spatially distributed sensors, detecting air pollutants such as PM<sub>2.5</sub>. Due to the large volume of the data acquired, principal component analysis (PCA) was used to reduce the dimensionality of the multi-variate feature set. However, this study mainly focused on the data manipulation and lacked further analysis of the air pollution impacts on an individual. Hosseini *et al.* [14] have embedded dust and spirometer sensors into their mHealth system to determine the environmental impacts on asthma patients. The system uses a smartwatch to collect data from the embedded physiological sensors, and the data are processed by a cloud server. A random forest classifier is used to classify data samples into high, medium or minimal risk categories. The algorithm has the advantage of reducing overfitting and improving predicting accuracy.

### III. METHODS

The main component of our device is an optical dust sensor Sharp GP2Y1010AU0F, which is used to monitoring fine particle concentration (i.e. the PM<sub>2.5</sub>) in the air. The sensor is integrated into Hamlyn Centre’s BSN node, as shown in Fig. 1. The sensor mainly consists of an infrared emitting diode (IRED) and a phototransistor that measures the dust concentration by detecting the intensity of the reflected light. The data are sampled at a fixed rate (~25Hz) and transmitted to a smartphone via the Bluetooth Low Energy (BLE) link provided by the BSN node. Contextual information, such as time, GPS information and the name of the metro line are

also collected and annotated using the associated mobile app. This information is used to study the spatial-temporal distribution of PM<sub>2.5</sub> concentration and enable accurate prediction of air quality. An extensive study was conducted to assess the air quality of the London underground stations. Seven metro lines were considered in our study and each of them was labeled with a line index. Identifying which metro line the user takes is important since several characteristics, such as the volume of passengers, the depth level or the ventilation designs, could affect the air quality measurements.

By using the proposed device, we have collected data from 10 metro stations and 16 different platforms in the period from 20<sup>th</sup> to 22<sup>nd</sup> June in 2016.

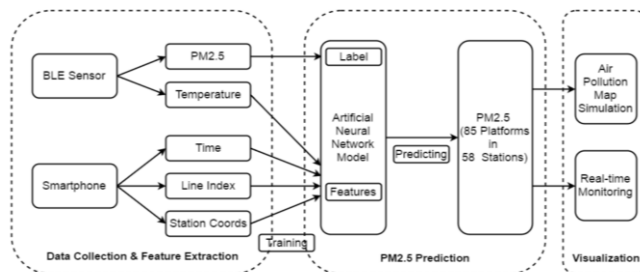


Figure 2. The proposed framework for air quality analysis

Fig. 2 shows the proposed framework of the wearable sensor system. It consists of three main modules. At the first stage, air quality data and other user information are collected via the smartphone app. The air quality data are pre-processed with a low-pass filter to reduce noise effect. In the second block, the data are analyzed with an artificial NN. Specifically, a multi-layered perceptron network is trained to predict PM<sub>2.5</sub> level on all the 85 different platforms of the 58 stations of London underground. To train the NN, we have divided the dataset in a training set (80%) and a test set (20%), and a 10-fold cross-validation was applied to validate our results. In the last block, the data are summarized and presented to the user as visual results.

#### Health Impact Index

To quantify the health impact from the air quality measurement, we defined a physiological parameter Health Impact Index (HII). It describes the impact on the user’s health based on the air quality measurement, age, gender, body height, energy expenditure, and body mass information of the user.

First, we calculate the daily Basal Metabolic Rate (BMR) using the Mifflin St Jeor’s equation [15]:

$$BMR = 10.0m + 6.25h - 5.0a + s \quad (1)$$

where BMR is measured in kcal/day,  $m$  represents the body mass (kg),  $h$  represents the body height (cm),  $a$  represents the age (years) and  $s$  is 5 for males and  $-161$  for females. Second, the energy expenditure (EE) is estimated by:

$$EE = MET * BMR \quad (2)$$

where MET is a factor determined by the metabolic

equivalent of a task and measures the energy cost of physical activities. Since the BMR expresses the minimum metabolic rate in resting states, we need to multiply it with the MET factor to estimate the energy cost in different physical activities. In our study, the MET factor is determined by our previous engine implemented in the ActiveMiles app [16].

Finally, the HII is calculated using the following equation:

$$HII = \frac{EE * PMC}{k} \quad (3)$$

where PMC is the average PM2.5 concentration measured in minute by the system, and the constant  $k = 60 \times 24$  converts the daily energy expenditure into EE per minute.

The main advantage of using the proposed HII instead of using only PM2.5 is that the HII considers both the air quality and the actual physical activities of the user providing meaningful information in evaluating the environmental effect on people's health. For example, Fig. 3 shows that performing physical activity in a polluted area can have a more drastic effect on health.

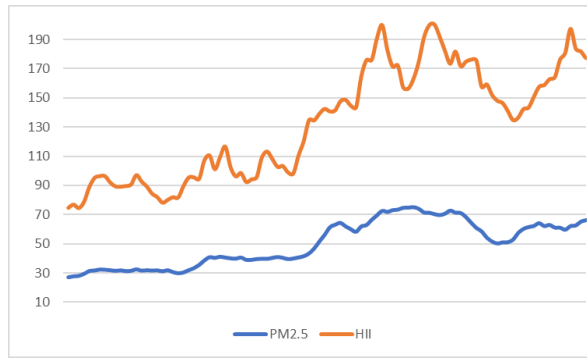


Figure 3. PM2.5 vs HII on Central Line

#### IV. RESULTS

The first experiment, conducted in our work, aims to estimate the PM2.5 concentration for all the platforms at three different time points (9:00, 13:00 and 18:00) based on the collected data. Fig. 4 shows the results grouped in metro lines. The dashed, dotted and solid lines represent PM2.5 level at different times. Most of the lines have highest pollution level at 18:00, which is especially obvious on the Piccadilly line. For Jubilee line and Circle line, pollution level does not change much along the day. This could be due to the fact that Jubilee line is relatively new, with platform screen doors which block away much of the dust generated by the train in the tunnel. In addition, many stations of circle lines are above ground stations; hence, the air quality is relatively stable throughout the day. Moreover, stations in the city center are always much more polluted compared to remote stations.

Second, we compare the air pollution level of a metro line with each other. Fig. 5 shows the predicted PM2.5 levels grouped in lines at time 18:00. Central, Piccadilly, Victoria and Jubilee lines have a relatively high average pollution level compared to other lines, due to the high volume of passengers, while Circle line has the best air quality as most of its stations are above ground. Though the average PM2.5

in Northern line is lower at around 38, its maximum pollution level is close to the Piccadilly line as it passes through most of the busiest stations, such as Leicester Square. Furthermore, the low variation of PM2.5 in Central line shows that most of its stations are highly polluted. This could be caused by the aged ventilation facility which leads to a relatively higher temperature in those platforms or stations.

The last experiments are conducted to compare the proposed classifier with other approaches. Specifically, the performance obtained by four new classifiers are evaluated using four error metrics. The considered classifiers are the M5p tree, the linear regression and two different versions of the Support Vector Machine (the epsilon-SVM and the nu-SVM).

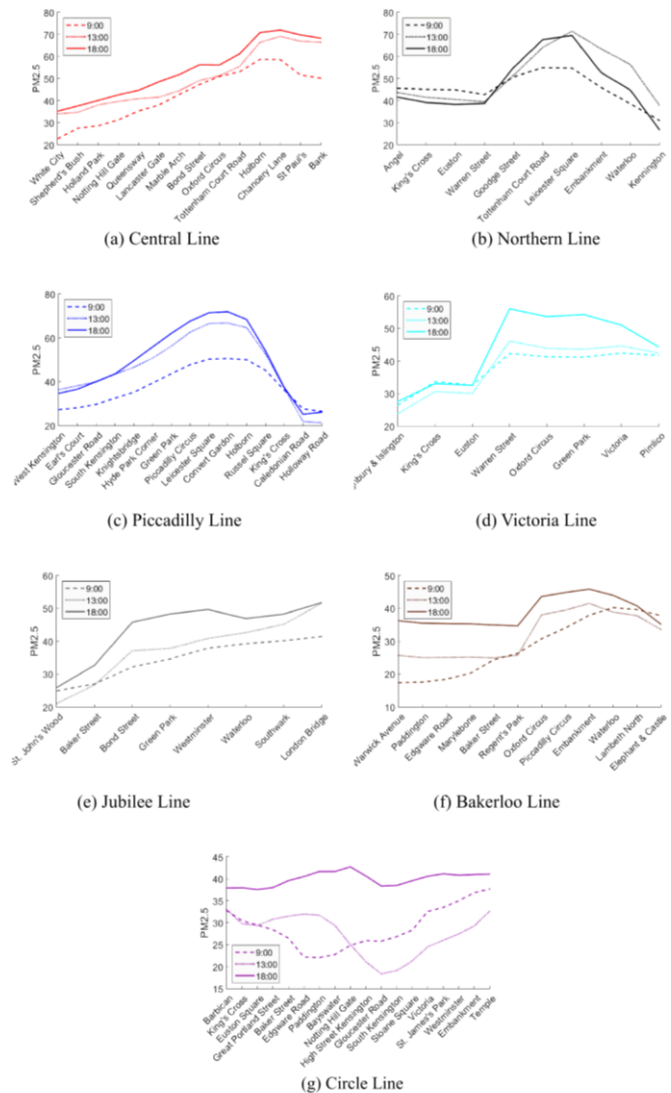


Figure 4. Simulated PM2.5 Concentration Results

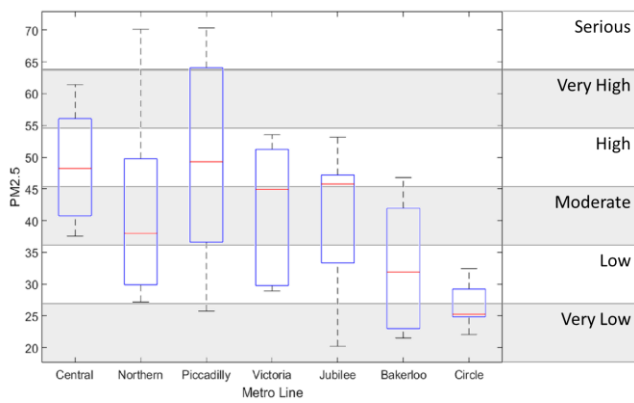


Figure 5. Mean and standard deviation of PM2.5 in different lines of London underground

As shown in Table 1, the NN has the lowest mean absolute error and the lowest root mean squared error among all five approaches, while nu-SVM has the lowest relative absolute error and lowest root relative squared error. Conversely, the linear regression approach got the worst results in all four types of errors. In general, we can conclude that the NN and both the SVMs can be considered valid solutions for our application.

Table 1: Comparative results of PM2.5 prediction with different classifiers

Approach	MAE	RMSE	Relative absolute error	Root relative squared error
NN	<b>3.20</b>	<b>4.07</b>	22.13%	24.87%
eps-SVM	3.28	4.08	22.23%	24.52%
Nu-SVM	3.22	4.06	<b>21.80%</b>	<b>24.45%</b>
M5P	3.66	4.53	24.77%	27.17%
Linear Regression	9.53	11.72	64.44%	70.30%

## V. CONCLUSION

In this paper, a wearable sensor system is developed to study the air pollution impacts on human health. A NN is used to estimate the air quality and the potential health impact in London underground stations. The preliminary results show that there is significant difference in air quality at different time among stations mainly due to the variation of the number of passengers in peak and off-peak hours. In addition, certain metro lines have relatively better air quality than others due to their infrastructural differences (i.e. platform screen doors or over ground stations), volume of passengers, and the efficiency of their ventilation systems. These results demonstrated the concept of personalized air quality monitoring and show the potential of such pervasive sensing system. By gathering such continuous information among users with a cloud server, accurate and real-time air quality information can be provided to the users and help them to better understand their surrounding environment, assess the potential health risk, and plan their routes to avoid heavily polluted stations, routes and areas. In addition, such

detailed information could also help the government to assess the environment in public transportation systems, and set plans to improve the air quality in the city.

## REFERENCES

- [1] World Health Organization (WHO). Preventing disease through healthy environments. [online] Available: [http://www.who.int/quantifying\\_ehimpacts/publications/preventindisease6.pdf](http://www.who.int/quantifying_ehimpacts/publications/preventindisease6.pdf).
- [2] World Health Organization (WHO). 7 million premature deaths annually linked to air pollution". 25 March 2014. [online] Available: <http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/>
- [3] G. Z. Yang, *Body Sensor Networks*, 2014, Springer
- [4] P. Dutta, P. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff. "Common Sense: participatory urban sensing using a network of handheld air quality monitors," *SenSys*, 2009, pp. 349–350.
- [5] B. Predić, Z. Yan, J. Eberle, D. Stojanovic and K. Aberer, "ExposureSense: Integrating daily activities with air quality using mobile participatory sensing," *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2013 IEEE International Conference on, San Diego, CA, 2013, pp. 303-305.
- [6] A. Nazelle, E. Seto, D. Donaire-Gonzalez, M. Mendez, J. Matamala, M. J. Nieuwenhuijsen, M. Jerrett, "Improving estimates of air pollution exposure through ubiquitous sensing technologies," *Environmental Pollution*, May 2013, Pages 92-99, Vol 176
- [7] W. H. Wu, A. A. T. Bui, M. A. Batalin, D. Liu and W. J. Kaiser, "Incremental Diagnosis Method for Intelligent Wearable Sensor Systems," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 5, pp. 553-562, Sept. 2007.
- [8] F. Tsow et al., "A Wearable and Wireless Sensor System for Real-Time Monitoring of Toxic Environmental Volatile Organic Compounds," in *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1734-1740, Dec. 2009.
- [9] H. Li et al., "Wearable autonomous microsystem with electrochemical gas sensor array for real-time health and safety monitoring," 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, 2012, pp. 503-506.
- [10] Air Quality Egg. [online] Available: <http://airqualityegg.com/>
- [11] TZOA. [online] Available: <http://www.tzoa.com/>
- [12] D. O. Rizea, A. C. Olteanu and D. S. Tudose, "Air quality data collection and processing platform," 2014 RoEduNet Conference 13th Edition: Networking in Education and Research Joint Event RENAM 8th Conference, Chisinau, 2014, pp. 1-4.
- [13] Y. S. Kim, In-Won Kim, Jo-Chun Kim and ChangKyoo Yoo, "Real-time multivariate monitoring and diagnosis of air pollutants in a subway station," *Control, Automation and Systems*, 2008. ICCAS 2008. International Conference on, Seoul, 2008, pp. 2610-2615.
- [14] A. Hosseini et al., "HIPAA Compliant Smartwatch Application for the Self-Management of Pediatric Asthma," in *IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Jun. 2016
- [15] Mifflin, Mark D., et al. "A new predictive equation for resting energy expenditure in healthy individuals." *The American journal of clinical nutrition* 51.2 (1990): 241-247.
- [16] Ravi, Daniele, et al. "A deep learning approach to on-node sensor data analytics for mobile or wearable devices." *IEEE Journal of Biomedical and Health Informatics* (2016).