Air pollution monitoring network using low-cost sensors, a case study in Hanoi, Vietnam

T N T Nguyen*, D V Ha, T N N Do, V H Nguyen, X T Ngo, V H Phan, N D Nguyen and Q H Bui

Center of Multi-disciplinary Integrated Technologies for Field Monitoring, VNU University of Engineering and Technology, Hanoi, Vietnam

E-mail: thanhntn@fimo.edu.vn

Abstract. Air pollution is a serious problem in Vietnam, especially in urban areas with high pressures of population, traffic, construction, and industrial development. Besides high accurate measurements from automatic and continuous monitoring ground stations and high-cost sensor devices, low-cost sensors have recently utilized to extend air pollution monitoring networks although their data quality is still argumentative. In this paper, we present a low-cost device, named FAirKit, which measured 6 basic air pollutants including PM$_{2.5}$, PM$_{10}$, CO, O$_3$, NO$_2$, and SO$_2$, and temperature and relative humidity. The sensors are calibrated with standard devices to improve their data quality. FAirKits are installed and transferred data in real-time to servers where an information system based on Sensor Web Enablement (SWE) standard of Open Geospatial Consortium (OGC) has been developed to store, process, and visualize real-time air pollution information. Currently, the low-cost sensors network has been deploying in Hanoi, Vietnam to enhance public awareness and alert local people to air pollution.

1. Introduction

In the last years, rapid economic growth has negative impacts on the global environment. Air pollution is considered as a major factor contributing to climate change, global warming, ozone depletion and acid rain. In Vietnam, air pollution is rapidly increasing in recent years. Specific to Vietnam, a recent report from the Environmental Performance Index (EPI) suggests that the quality of the environment in Vietnam has steadily dropped compared to other nations (EPI 2018). As the report, Vietnam EPI is ranked at 132 out of 180. Meanwhile, air quality in Vietnam is lagging with a rank of 161 out of 180 (EPI 2018).

At present, many major cities are facing high levels of air pollution. Monitoring data in recent years have shown that air pollution levels in urban areas are generally high and exceed national standards many days in a year, especially at urban areas in the north [1]. For the purpose of improving the quality of air for Hanoi City, Hanoi People's Committee, Hanoi Department of Natural Resources and Environment (Hanoi DONRE) and Hanoi Environmental Protection Agency (Hanoi EPA) have gradually deployed the network of air monitoring stations in Hanoi city. Currently, Hanoi has 10 air monitoring stations located in Thanh Cong, Tay Mo, Tan Mai, Trung Yen, Pham Van Dong, Nhon, My Dinh, Kim Lien, Hoan Kiem, Hang Dau (8 sensor and 2 fixed stations) which require a high investment and operation costs. However, the number of existing stations is not sufficient to warn people of local detailed air quality status (e.g. district- or even street-level). Therefore, the number of monitoring nodes in the city needs to be increased for warning and active prevention from air pollution. Low-cost sensors are considered to
be a good approach beside standard measurement instruments in Hanoi, Vietnam. It is a recent research and application trend in the world [2]. The development of low-cost devices for air pollution monitoring is being matured and commercialized such as AirVisual (Swiss) [3], Alphasense (United Kingdom - UK) [4], or Airbox (Taiwan) [5], etc. The monitoring network based on low-cost sensors have been deployed and operated in United State (Air Quality Egg) [6], UK (SNAQ) [7], and European countries (Capitor, Everyaware, CityOS) [8]–[10].

FAirNet, a sensor network for air pollution monitoring, has been developed by FIMO center, University of Engineering and Technology, Vietnam National University Hanoi (VNU UET). FAirNet includes four components, which are (i) FAirKit device measures up to six basic air quality parameters (PM$_{2.5}$, PM$_{10}$, CO, NO$_2$, SO$_2$, and O$_3$) and relative humidity and temperature using low-cost sensors. For enhancing accuracy, the FAirKit device is equipped with a calibration algorithm; (ii) FairServer is computer server for FAirKit data storage and processing services; (iii) FAirWeb and FAirApp are website and mobile application for displaying information of air pollution measured by FAirKits in real time.

This study aims to introduce the FAirNet measurement and its application for air pollution monitoring in some selected areas of Hanoi city. FAirNet has highlighted the usage of low-cost air pollution sensors with quality guarantees using sufficient calibration methods to provide online monitoring information to raise public awareness and alert local people to air pollution levels in a complex and dense urban area.

2. The study area

FAirKit devices are planning to install at Hoan Kiem district, Hanoi, Vietnam. Hoan Kiem district is the administrative, political, economic and cultural central of Hanoi city. Besides, Hoan Kiem is the historic inner of the city where many important railway, waterway and road traffic hubs are located to link Hoan Kiem with other districts and provinces. Hoan Kiem district is divided into 4 areas including: the old town, the Sword lake and its surrounding area, the old quarter and the outside of Red river dike (Figure 1a).

As of October 2016, the population of Hoan Kiem is 155,900 people with an area of 5.29 km$^2$. The average population density is about 29,500 people km$^{-2}$ (Figure 1b). Over the past years, Hoan Kiem economy has developed with a high and sustainable growth rate. Its economic structure has shifted towards services, trade and tourism.

Hoan Kiem has different characteristics from other Hanoi districts such as very high population density, degradation of old housing areas, the diversity of roads (large roads, narrow roads, ...) (Figure 1c), services, tourism and restaurant activities (e.g. walking streets on weekends, ...). Therefore, air quality in Hoan Kiem will bring its own characteristics. The installation of air quality monitoring network is necessary to assess the current status and report the air pollution level to people. It also provides the basis for proposing appropriate air pollution control policy of the governmental office.
3. The air pollution monitoring network

3.1. FAirKit devices

3.1.1. FAirKit Configuration. FAirKit supports measurements of PM$_{2.5}$, PM$_{10}$, CO, NO$_2$, SO$_2$, O$_3$, relative humidity, and temperature using low-cost sensors. The architecture of a FAirKit device is shown in Figure 2. Raspberry Pi Zero W, a device that supplies power to other components, will collect data from sensors for local storage and send it to the central server system (i.e. FAirServer). MCP3008 is an IC to convert analog signals from sensors into digital signals which the Raspberry Pi Zero W can read. DHT22 is a temperature and humidity sensor which consists of a humidity sensing component, an NTC temperature sensor (or thermistor), and an IC on back. Particulate matter concentrations (PM$_{2.5}$ and PM$_{10}$) are measured by PMS7003 dust meter that is based on laser scattering principle. MICS-4514 sensor for NO$_2$, MQ-7 sensor for CO, MQ-136 sensor for O$_3$, and MQ-136 sensor for SO$_2$ are metal oxide gas sensors which follow the same measurement principle. The resistance of the detecting layer in sensor changes if there is presence of the target gases. The reduction of gases removes insulative oxygen species at the grain boundaries, thus causes the overall resistance going down. Otherwise, oxidising gases add to insulative oxygen species and cause resistance increasing.
3.1.2. FAirKit Calibration. The main challenge of air quality monitoring devices using low-cost sensors is quality of data. There are multiple error sources for low-cost sensors which can be divided into 2 groups: internal errors and external errors [11]. Internal errors are related to sensor working principle, poor sensitivity in low concentration environments, systematic measuring error, nonlinear correlation with standard measurement, and sensor sensitive drift after a certain time of operation. External errors are caused by effect of environmental factors such as temperature and humidity, the diversity and complexity of substances in the air leading to the "confusion" of sensor.

Therefore, the low-cost sensors are required to be calibrated with standard devices to improve their accuracy. Many different calibration methods were applied to low-cost sensors and their networks in two phases: pre-calibration and post-calibration. The pre-calibration is to identify all the internal and external error sources of the sensor and control them before putting the sensor into operation. The general principle of data calibration is based on building a model for estimating the relationship between low-cost sensor’s dataset, ancillary data, and reference sensor’s dataset using regression method (e.g. ordinary least squares [12], [13]), 2nd order curve fitting regression [14], multiple least squares [15] [16], k nearest neighbors (KNN) [17], non-linear curve fitting [18], neural networks [19]). The post calibration is applied after deployment of the sensor devices. However, it is difficult in this stage because lack of reference devices (sensors) to calibrate for each sensor node in the network. Some calibration methods were proposed for the whole sensor network, including blind calibration [15][20], collaborative calibration [21], and transfer calibration [22][23][24][25].

FAirKit devices will be subjected to a two-stage process of data calibration to ensure quality of measurement data. The first level adjustment is carried out before installation. FAirKit devices and reference equipment will be co-located for a sufficiently long time. Then, the data from these two devices will be used to calibrate the FAirKit device using regression methods. The periodic calibration will be implemented when the device is active. The procedure of FAirKit calibration is presented in Figure 3.
Figure 3. Calibration procedure of FAirKit

Statistical parameters are used to assess the quality of FAirKit data and calibration model, which are coefficient of determination ($R^2$), Root Mean Square error (RMSE) and Relative error (RE)

$$R^2 = \frac{\sum_{t=1}^{n} (STA_t - \overline{STA})(FK_t - \overline{FK})^2}{\sum_{t=1}^{n} (STA_t - \overline{STA})^2 \sum_{t=1}^{n} (FK_t - \overline{FK})^2}$$  \hspace{1cm} (1)

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (STA_t - FK_t)^2}{n}}$$  \hspace{1cm} (2)

$$RE = \frac{|STA_t - FK_t|}{STA_t} \cdot 100\%$$  \hspace{1cm} (3)

Where $STA_t$ is the station data value at hour $t$, $FK_t$ is the corresponding FAirKit data value, $\overline{STA}$ and $\overline{FK}$ is the average value of station data and FAirKit data respectively, $n$ is total number of hours that FairKit and station were co-located.

3.2. Air pollution information and management system - FAirNet

FAirNet is an air pollution information control system which consists of 4 components: sensor node FairKits, server FAirServer, website FAirWeb, and mobile application FAirApp, as shown in Figure 4.
Figure 4. FAirNet System Architecture

FAirServer is a web service based on the architecture of the OGC's Sensor Web Enablement (SWE) standard. The OGC's SWE standards enable developers to make all types of sensors, transducers and sensor data repositories discoverable, accessible and useable via the Web [4]. The main adopted or pending OGC Standards in the SWE framework include:

- Observations & Measurements (O&M) – The general models and XML encodings for observations and measurements.
- PUCK Protocol Standard – Defines a protocol to retrieve a SensorML description, sensor "driver" code, and other information from the device itself, thus enabling automatic sensor installation, configuration and operation.
- Sensor Model Language (SensorML) – Standard models and XML Schema for describing the processes within sensor and observation processing systems.
- Sensor Observation Service (SOS) – Open interface for a web service to obtain observations and sensor and platform descriptions from one or more sensors.
- Sensor Planning Service (SPS) – An open interface for a web service by which a client can 1) determine the feasibility of collecting data from one or more sensors or models and 2) submit collection requests.
- SWE Common Data Model – Defines low-level data models for exchanging sensor related data between nodes of the OGC® Sensor Web Enablement (SWE) framework.
- SWE Service Model – Defines data types for common use across OGC Sensor Web Enablement (SWE) services. Five of these packages define operation request and response types.

FAirServer was implemented based on O&M, SensorML and SOS [5]. FAirServer provides an application development interface (REST API) for FAirKit to collect information from these devices. Submitted data will be stored in the Air Quality Database. After processing and analysing these data, FAirServer enables FAirWeb and FAirApp applications to access air quality monitoring data, manage FAirKit devices in real-time. FAirWeb is a web-based application in order to provide information on air pollution to the public. FAirApp is a mobile application, that has the same features as FAirWeb, developed to provide another channel of information to users.
4. The air pollution monitoring network

Designing air quality monitoring networks involves determining the number of stations and their locations. The number of monitoring stations will depend on the scale and topography of the area, composition of pollution sources and monitoring objectives. Methods of designing the network of monitoring stations include: Geospatial method, statistical analysis method, model method using dispersion model, Multi-objective design method, assuming virtual monitoring station method. Geospatial methods determine the location of monitoring stations based on minimization of estimation [26]–[28]. It has been applied to design networks for Spain [26], Canada and Germany [28]; The method of statistical method analysis is applied to select the location of the monitoring station [29]–[31]. The principle of this method is grouping of same characteristics stations, then use the pollution map to eliminate redundant stations. Principal component analysis methods (PCA) and cluster analysis are used to optimize air quality monitoring network in Portugal [31], Hongkong [32] and Japan [33]; The model method using dispersion model is applied in Argentina [34]. This method uses atmospheric dispersion models to identify the affected areas and therefore the number of residents may be exposed. A process for selecting the minimum number of air quality monitoring stations and their locations needed to detect the presence of background concentrations is greater than the reference concentration values in the metropolitan area; The multi-objective approach is based on simultaneous consideration of environmental, social and economic indicators [35],[36]. A multi-objective optimization model developed in Taiwan is based on the modified bounded implicit enumeration algorithm with the constraint arrangement method [35]. Another study [36] has developed a multi-objective evaluation approach based on GIS model to assess O3 and PM10 monitoring networks in the US in which weights were applied to emphasize important indicators. Recently, the assuming virtual monitoring stations method has been used in a number of studies to minimize monitoring costs [37]–[39]. Artificial neural network (Artificial Neural Network - ANN) is used to simulate virtual stations or rebuild stopped stations by developing nonlinear relationship between PM10 concentration of active stations and stopped stations [39].

The general basis of the observation network design methodology for Hoan Kiem district is illustrated in Figure 5.
4.1. Data collection

The air quality assessment required various kinds of data including historical monitoring data, emissions sources, receptors, topography and meteorology. Based on human activities, emissions sources and pollutants are identified based on indirect information such as population, agriculture, traffic activities. Historical monitoring data reflect status of air pollution in study area for further assessment. Besides, topography and meteorology data are taken into account because they affect directly to air quality.

4.2. Air quality monitoring network design

The design of air quality monitoring network is to determine a reasonable number of monitoring stations and their locations in the study area using statistical method and spatial distribution of pollutant concentration levels, respectively. In the next step, priority for monitoring station locations are considered together, based on Vietnamese air quality standards and network goals. For example, areas with dense traffic and population density will be set at higher priority for monitoring. At the implementation step, field survey is conducted following the network design. The specific location of each station will be determined based on actual conditions and may be modified. Finally, the sensor network is deployed and periodically evaluated.

**Number of station estimation**

The number of monitoring stations for the whole of Hanoi is determined by random sampling method. The monitoring data of air pollutants at ground stations are collected. Then, the mean and standard deviation of air pollutants are calculated. Assuming that the population of measurement stations is following the standard deviation, sample mean will follow Student-t distribution. So, estimating confidence interval will refer to t-score as follows:

\[
CI = (\text{sample mean} - t_{score} \times \text{sample_std}, \text{sample mean} + t_{score} \times \text{sample_std})
\]

(4)

where \(\text{sample mean}, \text{sample_std}\) are estimated from available air pollution data. \(t_{score}\) has a value depending on the desired Confidence Interval (CI) and the degrees of freedom = sample size - 1

From (4), the number of stations is calculated according to the formula:

\[
n = (t_{score} \times 100 / \text{CI})^2 \times (\text{sample_std} / \text{sample_mean})^2
\]

(5)

**Air Quality Assessment**

Cokriging interpolation is used to estimate PM concentration from multiple data sources. Cokriging methods are used to take advantage of the covariance between two or more related variables when the primary variable is sparse but secondary variables are abundant. In this study, PM concentration from monitoring station is the primary variable and other secondary including traffic and population density. The equation for Cokriging is following:

\[
PM_o' = \sum_{i=1}^{n} \alpha_i PM_i + \sum_{j=1}^{m} \beta_j \text{TRAFFIC}_DEN_j + \sum_{k=1}^{p} \gamma_k \text{POP}_DEN_k
\]

(6)

Where \(PM_o'\) is the estimate at the given grid point, \(\alpha_i\) is the weight assigned to the primary variable; \(PM_i\) is observed primary variable at given location, \(\text{TRAFFIC}_DEN_j\) and \(\text{POP}_DEN_k\) are secondary variables including traffic density and population density; \(\beta_j\) and \(\gamma_k\) is weight assigned to secondary variables; \(m,n\) and \(p\) are number of corresponding available PM, traffic density and population density pixels.

**Stations location design**

The monitoring stations is located using Kanaroglou’s method [27]. Firstly, a demand surface over the study area is estimated. A higher value of demand surface increases the need for monitoring. Two criteria for building demand surface are used, that is, a large number of monitors should be located
where the spatial variability of the demand surface and population density are high. The first criteria is implemented as the following equation:

\[
\hat{z}(\vec{x}, h) = \frac{1}{2} \sum_R \left[ \left( z(\vec{x}) - \hat{z}(\vec{x} + \vec{h}) \right)^2 \right]
\]

(7)

Where \( Z(\vec{x}) \) represents PM level at location \( \vec{x} \), \( h \) is the distance between \( \vec{x} \) and other points. The second criteria are implemented by following:

\[
W_R = \frac{P_R / P_T}{\hat{z}_R / \hat{z}_T}
\]

(8)

Here \( P_R \) is population of region \( R \) within study area and \( P_T \) is the population for the entire study area. The numerator is proportion of the total population in study area that resides in region \( R \) meanwhile denominator is proportion of total variability for entire study area that can be attributed to region \( R \) [27].

After calculating demand surface, locating of \( n \) number of stations begin by Location-Allocation procedure. Each pixel in the study area is a candidate whose weight is valued by the demand surface in that location. Attendance Maximizing Problem on ArcGIS toolbox is used to places \( n \) primary stations in a way that sum of weighted distances for all demand locations from their nearest neighbor station is minimized as following equation[27]:

\[
Z = \sum_{i=1}^{k} \sum_{j=1}^{m} w_i (1 - bd_{ij}) x_{ij},
\]

(9)

Where \( k \) is the number of demand locations and \( m \) is the number of candidate locations. The weight \( w_i \) at location \( i \) represents demand surface while \( d_{ij} \) is the distance between location \( i \) and \( j \), \( b \) is the attendance decreasing parameter, \( x_{ij} \) is equal 1 if demand location \( i \) is served by station in \( j \) and equal 0, otherwise [27].

4.3. Implementation

Based on the distribution map of the monitoring stations in theory, the survey will be conducted at specific installation locations to provide information for the deployment. After the field survey, the installation location of the monitoring stations was evaluated and adjusted to suit the actual conditions, based on both technical and safety requirements. They are easy access, prioritize areas of state agencies, public places rather than people's houses, private locations, etc. Finally, implementation plans for each station (description of installation area, location, height, equipment, ...) will be proposed. After a working period, each station data is analyzed and evaluated in order to adjust or reposition if necessary.

5. Results

5.1. Hoan Kiem sensor network

Firstly, relevant data are collected to determine air pollution status in Hanoi. Data includes air pollution observations at available ground stations (e.g. US Embassy, Center for Environmental Monitoring, Vietnam Environmental Administration, DONRE, …), population, road density, meteorological parameters. Data are analyzed and selected to create current air pollution maps for a Hanoi area. Since the lack of ground observation, only PM\(_{10}\) and PM\(_{2.5}\) maps are created in one month. PM\(_{10}\) and PM\(_{2.5}\) maps over Hoan Kiem are presented in Figure 6.
In order to capture the maximal variation of daily air pollution in the study area, the total number of FAirKit devices are proposed as 547 and 644 for PM$_{2.5}$ and PM$_{10}$, respectively, for ten district areas (i.e. Hoan Kiem, Hai Ba Trung, Dong Da, Ba Dinh, Cau Giay, Tay Ho, Bac Tu Liem, Nam Tu Liem, Thanh Xuan, Hoang Mai) in Hanoi, to guarantee the reliability of 90% (CI) and confidence interval of 5% (Ci) following the simple random sampling theory. Then, the number of required stations for each district is also calculated based on area ratio. In Hoan Kiem district area, the proposed number of stations is approximately 17. The distribution of 17 stations at Hoan Kiem district is based on location-allocation analysis with Maximize Attendance type. Figure 7 presents the sensor network for Hoan Kiem district.

Figure 6. Average PM$_{10}$ (a) and PM$_{2.5}$ (b) maps over Hoan Kiem district in October 2017

Figure 7. Air pollution monitoring stations in Hoan Kiem District. The squares refer to designed station location, while the triangles are tentative station after the preliminary survey.
Based on the designed network map, the survey will be conducted at specific sites to provide information for implementation. The survey will be conducted at two levels: the preliminary survey and detailed survey. The preliminary survey is to find potential locations for FAirKit devices. Based on the suggestion from Hoan Kiem’s people committee, governmental offices, schools, public places … should be selected. Based on the designed network, we did a preliminary survey to find candidates for installation FAirKit at each site. The detailed survey focuses on type of station (i.e. residential or traffic sites), installation position, wifi, electricity, maintenance, etc.

Currently, two stations at Cau Dong market and Pho Sach street have been installed and going to operation. The next stations are undergoing detailed surveys.

5.2. FAirKit device and FAirNet system

A FAirKit prototype is showed in Figure 8a. FAirKit was designed to be able to use conventional 220V power through an attached adapter. If the FAirKit is installed at outdoor area, it will be equipped with a solar panel and attached battery to ensure the device can operate stably. A FAirKit box for outdoor environment is shown in Figure 8b. All installed FAirKit devices are sending data online to the FAIRNet system. People can view online station at the link https://fairnet.vn. A detail view of a FAirKit is shown in Figure 9. Besides, the website provides other functions such as map viewer, searching, warning, and user managements.

![Figure 8. Prototype of FAirKit (a) and outdoor box (b)](image-url)
5.3. Evaluation of FAirKit devices

5.3.1. FAirKit calibration. The FAirKit device was co-located with an automatic and continuous station at Hanoi Environmental Protection Department (EPA), at 17 Trung Yen 3, Trung Hoa ward, Cau Giay district, Ha Noi. Measurements were taken over two periods of 18 days. The FAirKit data are evaluated on all six parameters including PM$_{2.5}$, PM$_{10}$, CO, NO$_2$, temperature, relative humidity. The data processing includes pre-processing, correlation analysis and calibration.

In the pre-processing stage, both FAirKit and EPA station data are grouped by hour and aligned together based on their corresponding time. There is a total of 395 data records. After that, coefficient of determination ($R^2$) between FAirKit and station parameters are calculated. Results are shown in Table 1.

| Parameters     | PM$_{2.5}$ | PM$_{10}$ | CO     | NO$_2$ | TEMP | HUD  |
|----------------|------------|-----------|--------|--------|------|------|
| PM$_{2.5}$ F  | 0.41       | 0.4       | 0.21   | 0.32   | 0.23 | 0.03 |
| PM$_{10}$ F   | 0.48       | 0.46      | 0.26   | 0.38   | 0.25 | 0.04 |
| CO F          | 0.57       | 0.6       | 0.47   | 0.57   | 0.66 | 0.13 |
| NO$_2$ F      | 0.44       | 0.47      | 0.38   | 0.47   | 0.66 | 0.22 |
| TEMP F        | 0.25       | 0.27      | 0.17   | 0.28   | 0.94 | 0.1  |
| HUD F         | 0.22       | 0.25      | 0.17   | 0.26   | 0.78 | 0.34 |

FAirKit and EPA station parameters are highly correlated. It is especially high for temperature ($R^2 = 0.94$) and moderate for relative humidity (0.34), PM$_{2.5}$ (0.41), PM$_{10}$ (0.46), CO (0.47), and NO$_2$ (0.47). On the other hand, some FAirKit parameters have a cross-correlation with station’s parameters such as CO, NO$_2$ with temperature, humidity and CO with NO$_2$. It is highlighted the potential of using multivariate linear regression for calibration. However, at present, a simple linear regression model was applied due to short-term data series that may lead to large errors of future data if overfit model was
used. However, further investigation of calibration method will be considered in future. Table 2 presents the $R^2$, RMSE and RE of hourly calibrated FairKit data to EPA station data using simple linear regression model.

### Table 2. Statistical result of calibrated FAirKit data and EPA station data

| Parameter          | PM$_{2.5}$ | PM$_{10}$ | CO   | NO$_2$ | Temp. | RH.  |
|--------------------|------------|-----------|------|--------|-------|------|
| $R^2$              | 0.41       | 0.46      | 0.47 | 0.47   | 0.94  | 0.34 |
| RMSE               | 26.13      | 41.04     | 729.84 | 19.95 | 1.66  | 9.72 |
| PM$_{2.5}$, PM$_{10}$, CO, NO$_2$ ($\mu$g m$^{-3}$) |            |           |       |        |       |      |
| Temperature (°C)   |            |           |       |        |       |      |
| Relative Humidity (%) | 36.51    | 34.38     | 46.48 | 49.45  | 5.41  | 12.91 |

5.3.2. FAirKit in operation. Two FAirKit at Cau Dong Market and Pho Sach Street have been calibrated, installed and going to operation. The first one represented for city background was hung on rooftop of four-story-building (usually 15 m above ground). Another one represented for traffic site that was installed at lamp post about 1.8 m above ground and roughly 0.5 m away Ly Thuong Kiet Street. This section will be presented the preliminary results of these device based on 3 months data (from November 1st, 2018 to January 30th, 2019). Based on type of monitoring site, the PM$_{2.5}$ levels data from FAirKit at Cau Dong market was compare with a nearby beta attenuation in order to investigate of the potential applicability of FAirKit under actual environment. The left one data collection was used for determining the diurnal and weekly variation in concentration of PM$_{2.5}$, PM$_{10}$, CO and NO$_2$.

5.3.2.1. Comparison of PM$_{2.5}$ data between FAirKit and a beta-attenuation monitor. Figure 11 demonstrates the correlation between average hourly and daily PM$_{2.5}$ mass concentration obtained from FAirKit at Cau Dong market and PM$_{2.5}$ data obtained from a Beta Attenuation Monitor (BAM). BAM was installed on the rooftop of US Embassy, located 4.0 km away southwest of the Cau Dong market (see Figure 10). This dataset was obtained from the AirNow website.

![Location of monitoring sites](image.jpg)
Figure 11. Correlation between average hourly (a), daily (b) PM$_{2.5}$ mass concentration measured with FAirKit at Cau Dong market and PM$_{2.5}$ mass concentration measured with a nearby BAM.

The PM$_{2.5}$ sensor (in FAirKit) has good correlation with BAM, with R$^2$ = 0.73 and 0.50 in daily and hourly PM$_{2.5}$ levels, respectively (Figure 11). This result suggested that the effect of nearby site sources of PM$_{2.5}$ is quite significant sometimes. Therefore, two sub-datasets were divided by month to compare.

Table 3. Correlation between PM$_{2.5}$ levels measured with FAirKit and a nearby BAM by month

|          | Hourly PM$_{2.5}$ mass concentration | Daily PM$_{2.5}$ mass concentration |
|----------|--------------------------------------|------------------------------------|
|          | y = 1.1476x - 5.0281                  | y = 1.7644x - 36.78                 |
|          | n = 677                                | N = 30                              |
|          | R$^2$ = 0.55                          | R$^2$ = 0.84                        |
| Nov-18   | y = 1.1904x - 12.216                  | y = 1.4013x - 20.135                |
|          | n = 726                                | N = 31                              |
|          | R$^2$ = 0.65                          | R$^2$ = 0.91                        |
| Dec-18   | y = -0.8613x + 5.0496                 | y = 1.1917x - 11.502                |
|          | n = 652                                | N = 29                              |
|          | R$^2$ = 0.32                          | R$^2$ = 0.52                        |

Correlation between FAirKit and BAM device by month were summarized in Table 3. The correlation was good agreement in December, 2018 (R$^2$ = 0.65 and 0.91 in hourly and daily data, respectively), moderate in November, 2018 and bad in January, 2019. These results indicated that PM$_{2.5}$ levels variation of both sites was similar in December but different in November and January.

Monthly variation in concentration of PM$_{2.5}$ of both sites are compared in Table 4. Median and mean levels were equal in both sites. This figure was in same range of PM$_{2.5}$ levels in Hanoi during 2001 – 2008 period reported in [40]. However, maximum levels measuring by BAM at US Embassy was much higher than its at Cau Dong market. It is implied that FAirKit seem good at measuring in low values but high levels (larger than 200 μg m$^{-3}$). The calibration period, which was not long enough for FAirKit to capture such high levels, could be a reason. On the other hand, difference of sampling height and location Cau Dong market and US Embassy should be noticed in this comparison.

Table 4. Monthly variation in PM$_{2.5}$ levels (μg m$^{-3}$)

|          | November FAirKit | November BAM | December FAirKit | December BAM | January FAirKit | January BAM |
|----------|------------------|--------------|------------------|--------------|----------------|-------------|
| Min      | 16               | 8            | 13               | 3            | 13             | 1           |
| Median   | 47               | 44           | 34               | 27           | 48             | 38          |
| Mean     | 52               | 52           | 38               | 34           | 52             | 50          |
| Max      | 108              | 163          | 103              | 166          | 108            | 169         |
5.3.2.2. Diurnal and weekly variations of CO, NO₂ mixing ratios and PM mass concentration. Hourly PM$_{2.5}$, PM$_{10}$, CO and NO₂ concentration are obtained from FAirKit installed at Pho Sach between December 19$^{th}$ 2018 and January 30$^{th}$ 2019. Diurnal variation in the concentration of four pollutants are presented in Figure 12. The highest levels of CO and NO₂ are observed in morning rush hours (7:00 ~ 9:00 am) and evening rush hours (4:00 ~ 6:00 pm). Meanwhile, the lowest levels are shown in mid-night and mid-day time. A similar daily trend of CO and NO₂ is also reported by previous studies in Hanoi [41], [42]. The dominant source of NO₂ in Hanoi was concluded as traffic based on the relationship between NO₂ and population density in [43].

However, there is no obviously peak in the levels of PM$_{2.5}$ and PM$_{10}$ (around 8 am and 6 pm). These results suggested that other local, regional sources could be also contributors of PM, besides traffic emission at the observation site. The similar conclusion is found in a previous study on the level of PM$_{2.5}$ in the year-round 2016-2017 by [44].

![Figure 12. Diurnal variation in levels of (a) CO; (b) NO₂; (c) PM$_{2.5}$ và (d) PM$_{10}$ (μg m$^{-3}$)](image-url)
Figure 13. Weekly variation in levels of (a) CO; (b) NO₂; (c) PM₂.₅; và (d) PM₁₀ (μg m⁻³)

Figure 13 compares the variation between working day and weekend of CO, NO₂ mixing ratios and PM mass concentration averaged over the period. It is clearly seen that the peak in the levels of all substances were found in the weekend (Friday and Saturday), in contrast, the lowest levels were found in Wednesday. On the other hand, no clear difference between levels of PM₂.₅ on weekdays and weekends was observed in [44]. They also reported that the results may be partly because of total vehicle numbers in weekdays and weekends in Hanoi are not greatly different [45], [46]. In this study, these trends are different from other observed sites in previous research, which was probably due to higher traffic volume in the weekend than weekdays. High values for PM₂.₅, PM₁₀, CO, NO₂ were expected to be due to large numbers of people travel to the city center through the weekend.

6. Conclusion

In this paper, an air pollution monitoring network, called FAirNet, using low-cost sensor devices for Hoan Kiem district, Hanoi, Vietnam is presented. The network is design to capture air pollution in an urban area with high density of population. The network is planning to have 17 outdoor stations measuring PM₂.₅, PM₁₀, CO, NO₂, temperature, and relative humidity by the FAirKit devices. In operation, the FAirKit are sent data in real-time to FAirServer which is intermediately distributed information to users by a web or a mobile application. The FAirNet provides information of air pollution status, alert citizens and support authorities for decision making of air pollution in the study area.
The calibration procedure based on simple linear regression model guarantees quality of FAirKit data with $R^2 = 0.94, 0.34, 0.41, 0.46, 0.47,$ and 0.47 for temperature, relative humidity, PM$_{2.5}$, PM$_{10}$, CO, and NO$_2$, respectively and relative error from 5.41- 49.45% in comparison with EPA station in the period of 395 hours (~ 18 days). In operation, PM$_{2.5}$ of the FAirKit at Cau Dong market, considered as an ambient site, has shown a high correlation with PM$_{2.5}$ of US embassy’s BAM, with $R^2 = 0.73$ and 0.50 for daily and hourly average from November 2018 to January 2019. Regarding PM$_{2.5}$ ranges, the FAirKit seem capturing well low PM$_{2.5}$ values but high levels (larger than 200 $\mu$g m$^{-3}$). The analysis of diurnal variation of air pollutants based on FAirKit at the Pho Sach street, a traffic site, has shown strong impact of daily traffic during the rush hours (7:00 ~ 9:00 am and 4:00 ~ 6:00 pm) by observing higher CO and NO$_2$ concentrations. The weekly variation points out the highest levels of CO, NO$_2$, PM$_{2.5}$, PM$_{10}$ observed during weekend (Friday and Saturday) and their lowest levels happened on Wednesday.

The study with preliminary results has highlighted potential of using low-cost sensor network for air pollution monitoring in a dense urban area. However, there are still many arguments needed further investigation, especially for sensor data, such as calibration methods, quality of data, usage of sensor measurements, etc. In the future, we focus on improving FAirKit data quality in all aspects mentioned above. Besides, FAirKit installation will be continuous at Hoan Kiem district and extended to other places. Operation of FAirKit network and information distribution to end user by a website and mobile application are taken into account in more detail.

Acknowledgement
This research was funded by Vietnam National University Hanoi (VNU) under grant number QMT 17.03. The authors would like to thank Hanoi Environmental Protection Agency (Hanoi EPA), Hanoi Department of Natural Resources and Environment (Hanoi DONRE), Deutsche Gesellschaft für Internationale Zusammenarbeit Gmb H (GIZ Vietnam), the United States of America in Hanoi for providing air quality and meteorological data.

References
[1] Ministry of Natural Resources and Environment of the Socialist Republic of Vietnam 2016 National Environmental Report: Air Environment
[2] Kumar P, Morawska L, Martani C, Biskos G, Neophytou M, Di Sabatino S, Bell M, Norford L and Britter R 2015 The rise of low-cost sensing for managing air pollution in cities Environ. Int. 75 199-205
[3] AirVisual [https://www.airvisual.com/]
[4] Alphasense [https://www.alphasense.com/]
[5] Airbox [https://www.edimax.com/edimax/merchandise/merchandise_detail/data/edimax/global/air_quality_monitoring_semioutdoor/ai-1001w_v2/]
[6] AirQualityEgg [https://airqualityegg.com/]
[7] SNAQ [http://www.snaq.org/]
[8] Captor [https://www.captor-project.eu/]
[9] Everyaware [http://www.everyaware.eu/]
[10] CityOS [https://cityos.io/air/]
[11] Maag B, Zhou Z, and Thiele L 2018 A Survey on Sensor Calibration in Air Pollution Monitoring Deployments IEEE Internet Things J. 5 4857-70
[12] Spinelle L, Gerboles M, Villani M G, Aleixandre M and Bonavitacola F 2015 Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A : Ozone and nitrogen dioxide Sensors and Actuators Chem. 215 249–57
[13] Spinelle L, Gerboles M, Villani M G, Aleixandre M and Bonavitacola F 2016 Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B : NO, CO and CO2 Sensors and Actuators Chem. 238 706–15
[14] Cheng Y, Li X, Li Z, Jiang S, Li Y, Jia J and Jiang X 2014 AirCloud : A Cloud-based Air-Quality Monitoring System for Everyone Proc. of the 12th ACM Conf. on Embedded Network Sensor Systems (Memphis, Tennessee) 251-65

[15] Jiao W et al. 2016 Community Air Sensor Network (CAIRSENSE) project : evaluation of low-cost sensor performance in a suburban environment in the southeastern United States Atmos. Meas. Tech. 9 5281–92

[16] Pang X, Shaw M D, Lewis A C, Carpenter L J and Batchellier T 2017 Electrochemical ozone sensors : A miniaturised alternative for ozone measurements in laboratory experiments and air-quality monitoring Sensors Actuators Chem. 240 829–37

[17] Hagan D H, Isaacman-vanwertz G, Franklin J P, Wallace L M M, Kocar B D, Heald C L and Kroll J H 2018 Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments Atmos. Meas. Tech. 11 315–28

[18] Masson N, Piedrahita R and Hannigam M Chemical 2015 Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring Sensors Actuators Chem. 208 339–45

[19] Kamionka M, Breuil P and Pijolat C 2006 Calibration of a multivariate gas sensing device for atmospheric pollution measurement Sensors Actuators Chem. 118 323–27

[20] Moltchanov S, Levy I, Etzion Y, Lerner U, Broday D M and Fishbain B 2015 On the feasibility of measuring urban air pollution by wireless distributed sensor networks Sci. Total Environ. 502 537–47

[21] Maag B, Zhou Z, Saukh O, Thiele L and Zurich E T H 2017 SCAN: Multi-Hop Calibration for Mobile Sensor Arrays Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies vol 1 19

[22] Budde M, Masri R El, Riedel T and Beigl M 2013 Enabling Low-Cost Particulate Matter Measurement for Participatory Sensing Scenarios Proc. of the 12th Int. Conf. on Mobile and Ubiquitous Multimedia (Luleå, Sweden) 19

[23] Tian F, Feng J, Li G, Dang L, Yin X and Peng X 2013 On-line Calibration of Semiconductor Gas Sensors Based on Prediction Model Journal of Computers 8 2204-08

[24] Deshmukh S, Kamde K, Jana A, Korde S, Bandyopadhyay R, Sankar R, Bhattacharyya N and Pandey R A 2014 Calibration transfer between electronic nose systems for rapid In situ measurement of pulp and paper industry emissions Anal. Chim. Acta 841 58-67

[25] Fonollosa J, Fernández L, Gutiérrez-gálvez A, Huerta R and Marco S 2016 Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization Sensors Actuators Chem. 236 1044–53

[26] Trujillo-Ventura A and Hugh Ellis J 1991 Multiobjective air pollution monitoring network design Atmos. Environ. Part A, Gen. Top. 25 469–79

[27] Kanaroglou P S, Jerrett M, Morrison J, Beckerman B, Arain M A, Gilbert N L and Brook J R 2005 Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach Atmos. Environ. 39 2399–409

[28] Shahraiyni H T, Sodoudi S and Cubasch U 2014 Determination the optimum number and positions of monitoring stations for proper spatial modeling of mean PM10 concentration in Berlin Geophysical Research Abstracts 16 13042

[29] Pires J C M, Sousa S I V, Pereira M C, Alvim-Ferraz M C M and Martins F G 2008 Management of air quality monitoring using principal component and cluster analysis—Part II: CO, NO2 and O3 Atmos. Environ. 42 1261–74

[30] Pires J C M, Sousa S I V, Pereira M C, Alvim-Ferraz M C M and Martins F G 2008 Management of air quality monitoring using principal component and cluster analysis—Part I: SO2 and PM10 Atmos. Environ. 42 1261–74

[31] Pires J C M, Pereira M C and Martins F G 2009 Identification of redundant air quality measurements through the use of principal component analysis Atmos. Environ. 43 3837–42

[32] Lu W, He H and Dong L 2011 Performance assessment of air quality monitoring networks using...
principal component analysis and cluster analysis. Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Build. Environ.* **46** 577–83

[33] Shirato S, Iizuka A, Mizukoshi A and Noguchi M 2015 Optimized Arrangement of Constant Ambient Air Monitoring Stations in the Kanto Region of Japan *Int. J. Environ. Res. Public Health* **12** 2950–66

[34] Mazzeo N A and Venegas L E 2008 Design of an air-quality surveillance system for Buenos Aires City integrated by a NOx monitoring network and atmospheric dispersion models *Environ. Model. Assess.* **13** 349–56

[35] Chen C H, Liu W L and Chen C H 2006 Development of a multiple objective planning theory and system for sustainable air quality monitoring networks *Sci. Total Environ.* **354** 1–19

[36] Pope R and Wu J 2014 A multi-objective assessment of an air quality monitoring network using environmental, economic, and social indicators and GIS-based models *J. Air Waste Manag. Assoc.* **64** 721–37

[37] Ung A et al. 2010 Air pollution mapping over a city – virtual stations and morphological indicators *Proc. of 10th Int. Sym. on Transport and Air pollution* (Boulder)

[38] Beaulant A L, Perron G, Kleinpeter J, Weber C, Ranchin T and Wald L 2008 Adding virtual measuring stations to a network for urban air pollution mapping *Environ. Int.* **34** 599–605

[39] Shahraiyni H T, Sodoudi S, Kerschbaumer A and Cubasch U 2015 The development of a dense urban air pollution monitoring network *Atmos. Pollut. Res.* **6** 904–15

[40] Cohen D D, Crawford J, Stelcer E and Bac V T 2010 Characterisation and source apportionment of fine particulate sources at Hanoi from 2001 to 2008 *Atmos. Environ.* **44** 320–8

[41] Sakamoto Y, Shoji K, Bui M T, Pham T H, Vu T A, Ly B T and Kajii Y 2018 Air quality study in Hanoi, Vietnam in 2015 – 2016 based on a one-year observation of NOx, O3, CO and a one-week observation of VOCs *Atmos. Pollut. Res.* **8** 544-51

[42] Hien P D, Loc P D and Dao N V 2011 Air pollution episodes associated with East Asian winter monsoons *Sci. Total Environ.* **409** 5063–8

[43] Hien P D, Hangartner M, Fabian S and Tan P M 2014 Concentrations of NO2, SO2 and benzene across Hanoi measured by passive diffusion samplers *Atmos. Environ.* **88** 66–73

[44] Ly B, Matsumi Y, Nakayama T, Sakamoto Y, Kajii Y and Nghiem T D 2018 Characterizing PM2.5 in Hanoi with New High Temporal Resolution Sensor *Aerosol Air Qual. Res.* **18** 2487–97

[45] Phuc N H and Kim Oanh N T 2018 Determining factors for levels of volatile organic compounds measured in different microenvironments of a heavy traffic urban area *Sci. Total Environ.* **627** 290–303

[46] Truc V T Q and Kim Oanh N T 2007 Roadside BTEX and other gaseous air pollutants in relation to emission sources *Atmos. Environ.* **41** 7685-97