

## Research Article

# A Mobile Sensing System for Urban PM<sub>2.5</sub> Monitoring with Adaptive Resolution

**Hongjie Guo, Guojun Dai, Jin Fan, Yifan Wu, Fangyao Shen, and Yidan Hu**

*The Department of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China*

Correspondence should be addressed to Guojun Dai; [daigj@hdu.edu.cn](mailto:daigj@hdu.edu.cn) and Yifan Wu; [yfwu@hdu.edu.cn](mailto:yfwu@hdu.edu.cn)

Received 30 December 2015; Revised 23 March 2016; Accepted 30 March 2016

Academic Editor: Yasuko Y. Maruo

Copyright © 2016 Hongjie Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper develops a mobile sensing system, the first system used in adaptive resolution urban air quality monitoring. In this system, we employ several taxis as sensor carries to collect original PM<sub>2.5</sub> data and collect a variety of datasets, including meteorological data, traffic status data, and geographical data in the city. This paper also presents a novel method AG-PCEM (Adaptive Grid-Probabilistic Concentration Estimation Method) to infer the PM<sub>2.5</sub> concentration for undetected grids using dynamic adaptive grids. We gradually collect the measurements throughout a year using a prototype system in Xiasha District of Hangzhou City, China. Experimental data has verified that the proposed system can achieve good performance in terms of computational cost and accuracy. The computational cost of AG-PCEM is reduced by about 40.2% compared with a static grid method PCEM under the condition of reaching the close accuracy, and the accuracy of AG-PCEM is far superior as widely used artificial neural network (ANN) and Gaussian process (GP), enhanced by 38.8% and 14.6%, respectively. The system can be expanded to wide-range air quality monitor by adjusting the initial grid resolution, and our findings can tell citizens actual air quality and help official management find pollution sources.

## 1. Introduction

Fine particulate matter (PM<sub>2.5</sub>) has been identified as the most health-damaging particles to public health [1]. In particular, urban areas of developing countries such as Beijing and New Delhi are suffering from this threat seriously. Due to the intricateness of structures and diversity of functional areas in urban, traditional monitoring urban particulate matter method based on sparsely fixed stations (e.g., the testing region of our work covering a 64 km<sup>2</sup> area only had one official air quality monitoring station [2]) is far away to tell citizens the actual air quality they breath in.

Recently, inferring fine-grained urban air quality has gained much attention. In AirCloud proposed by Cheng et al. [3], they use ANN and GP to infer PM<sub>2.5</sub> concentration based on amounts of sensors which are built on several points of interest (POIs). U-air [4] proposes a semisupervised learning approach based on air quality data reported by a few official monitor stations and meteorological data, POIs data, road networks, and taxi trajectory to infer air quality information throughout a city. These two systems' inference accuracy and resolution of concentration distribution highly relied

on the number of POIs and selection of POIs. To improve data coverage and distribution resolution, several systems resort to collected urban pollutants data with mobile, low-cost sensors. In [5], authors collected the measurements using mobile sensor nodes installed on top of transport vehicles and develop land-use regression (LUR) models to infer the pollution concentration distribution with a high resolution of 100 m × 100 m. Similarly, in the paper proposed by Hu et al. [6], the authors collect the data samples by mobile sensors and proposed a Probabilistic Concentration Estimation Method (PCEM) to infer regional PM<sub>2.5</sub> concentration distribution with 200 m × 200 m resolution. These methods' scalability can be easily challenged given a large monitoring area since too high resolution produces huge computational cost.

Overall, inferring pollutant concentration distribution with adaptive resolution is of great importance. A large grid size can lead to unacceptable errors for many pollutants formed via nonlinear chemical reactions, while too high resolution can lead to high computational complexity. Adaptive grid method used in air quality modeling has been explored for many years. Tomlin et al. [7] used adaptive unstructured triangular grids to model air pollution transport. They solved

the discretized atmospheric diffusion equation using a finite-volume approach and applied the concentration gradient as refinement criteria. They kept the original grid nodes fixed and refinement of the grids by splitting each triangle into 4 smaller, similar triangles when concentration gradient exceeded a threshold. The method developed by Tomlin et al. was applied to model nuclear contamination dispersion [8] and air pollution formation [9] and used in vertical domain [10, 11] as well. Srivastava et al. [12] propose a new adaptive grid algorithm (mesh movement) for simulating reactive atmospheric pollutants. They employ a constant number of grid nodes to keep the computational time manageable and calculated grids weight (spatial error) by a linear combination of curvatures of different chemical species. The grid nodes are clustered in area of high weight to improve simulating accuracy. Garcia-Menendez et al. [13] developed an adaptive grid version of the Community Multiscale Air Quality Modeling System (AG-CMAQ) using the adaptive grid algorithm proposed by Srivastava et al. Lagzi et al. [9] reported an assessment approach for AG-CMAQ, showing AG-CMAQ has better simulating results compared to fixed grid CMAQ [14]. Constantinescu et al. [15] develop an adaptive resolution system (mesh enrichment) for modeling regional air pollution based on the Sulfur Transport and Deposition Model (STEM) [16]. Refinement is achieved by dividing a block into smaller blocks based on the curvature in concentration fields. The authors claim that the adaptive grid algorithm required only a quarter of the time spent compared to static fine grid under the condition of reaching the close accuracy.

Above adaptive grids methods are all model based approaches used in regional-to-global air pollution modeling. They fail to provide a fine-grained and precision air pollution concentration distribution as the data is collected by sparsely deployed monitoring sites. Furthermore, their concentration gradient or concentration curvature based refinement criteria require inferring each grid air pollution concentration before refining grids, which bring modeling delay and poor efficiency. So far, there is no effective adaptive resolution method used in fine-grained air quality inferring.

To this end, we develop a mobile sensing system for fine-grained urban  $PM_{2.5}$  monitoring with adaptive resolution. We divide the testing area into amounts of initial  $500\text{ m} \times 500\text{ m}$  grids and let mobile sensors randomly collect  $PM_{2.5}$  concentration in the testing area and collect meteorological data, traffic status data, and geographical data in the city. We also propose AG-PCEM: an adaptive grid method to infer the fine-grained urban distribution. We develop novel refinement criteria to refine grids before inferring  $PM_{2.5}$  concentration using (historical and real-time)  $PM_{2.5}$  concentration data collected by several mobile sensors and a variety of datasets we observed in the city; we also define several refinement levels and grid resolutions to adaptively adjust grid resolution. The evaluation results show the computational cost of AG-PCEM is reduced by about 40.2% compared with a static grid method PCEM under the condition of reaching the close accuracy, and the accuracy of AG-PCEM is far superior as widely used artificial neural network (ANN)

and gaussian process (GP), enhanced by 38.8% and 14.6%, respectively.

Our contributions can be summarized as follows:

- (i) The proposed AG-PCEM provides a fine-grained  $PM_{2.5}$  concentration distribution which can be expanded to wide-range air quality monitor by adjusting the initial grid resolution.
- (ii) Novel refinement criteria and refining method are used to refine grids before inferring  $PM_{2.5}$  concentration distribution which assure inferring efficiency.
- (iii) Inferring  $PM_{2.5}$  concentration distribution using dynamic adaptive grids reduces computation cost obviously and ensures algorithm accuracy, and finer-grained  $PM_{2.5}$  concentration distribution tell citizens actual air quality and help official management find pollution sources.
- (iv) We develop a system using mobile, low-cost sensors to collect original data over one year and validate that our method AG-PCEM has a good performance in terms of accuracy and computational cost.

The rest of this paper is organized as follows. Section 2 introduces several definitions used in this work and the framework of our system. In Sections 3–6, the proposed AG-PCEM algorithm is presented in detail. A prototype experimental system is developed in Section 7 to validate AG-PCEM. The performance of the proposed algorithm is evaluated in Section 8. Finally, the conclusions of this paper are discussed in Section 9.

## 2. System Overview

### 2.1. Preliminaries

*Definition 1* (air quality index and individual air quality index). AQI is a number that describes the status of air quality. People are more likely to experience health risks as the AQI increases; AQI is calculated from 6 kinds of main pollutant concentrations; they are fine particulate matter ( $PM_{2.5}$ ), particulate matter ( $PM_{10}$ ), sulfur dioxide ( $SO_2$ ), nitrogen dioxide ( $NO_2$ ), ozone ( $O_3$ ), and carbon monoxide (NO).

IAQI is calculated for each pollutant whose value varies among different pollutants; in the above 6 pollutants,  $PM_{2.5}$  has the biggest impact on human and environment, so  $PM_{2.5}$  IAQI value is the most important index to assess air quality.  $PM_{2.5}$  IAQI values are divided into ranges, and their standards differ in different countries. Considering the real testing environment, in this paper, we use the standard issued by Chinese Environmental Protection Administration [17] as shown in Table 1.

*Definition 2* (POI). A point of interest (POI) is a place (like a school and factory) in the physical world that we are interested in.

TABLE 1: PM<sub>2.5</sub> IAQI range of Chinese standard.

PM <sub>2.5</sub> IAQI	24-hour average (μg/m <sup>3</sup> )	Air quality ranges
0–50	0–35	I (optimal)
50–100	35–75	II (good)
100–150	75–115	III (light pollution)
150–200	115–150	IV (moderate pollution)
200–300	150–250	V (high pollution)
300–500	250–500	VI (severe pollution)

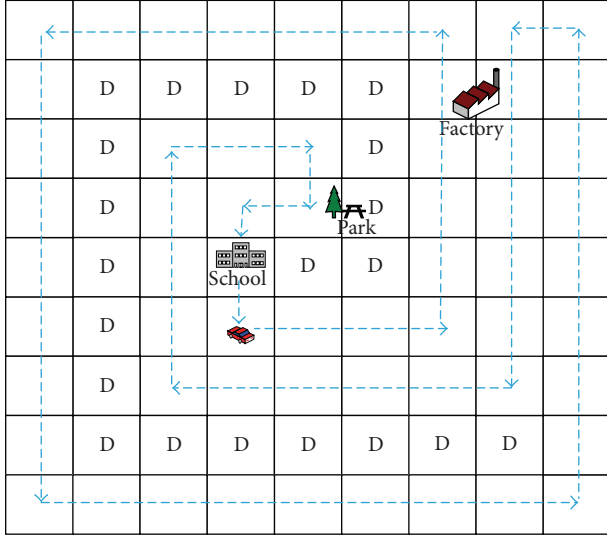


FIGURE 1: Mobile sensing scenario. The grids denoted by D mean that PM<sub>2.5</sub> concentration was not detected for them.

*Definition 3* (grid and mobile sensing). The testing region would be divided into grids with 500 m × 500 m initial resolution. As shown in Figure 1, a testing car with built-in particle sensor collected PM<sub>2.5</sub> data along the route noted by blue arrow. Grids which are not in this route imply that PM<sub>2.5</sub> concentration in these areas is not detected directly. A grid  $s_{ij}$  with a resolution of hundred meters would be regarded as a point in terms of geography and the PM<sub>2.5</sub> concentration  $x_{ij}$  in a grid is uniform, where  $s_{ij}$  means that the location of current grid is in the  $i$ th row and  $j$ th column in the grid map of testing area and  $x_{ij}$  means the PM<sub>2.5</sub> concentration value in the  $i$ th row,  $j$ th column in the grid map.

*Definition 4* (refining and derefining). Refining is a process to divide a grid into 4 or 16 small and same-sized grids, while derefining means merging 4 adjacent grids into one grid, and the process of refining and derefining the grid map of a region is termed as regridding.

*Definition 5* (grid resolution and refinement level). Grid resolution denotes grid size. In this work, the initial grid resolution is 500 m × 500 m with four levels of refinement (i.e., -1, 0, 1, and 2) and four matched grid resolutions (i.e., 1000 m × 1000 m, 500 m × 500 m, 250 m × 250 m, and 125 m × 125 m). As shown in Figure 2, grid with green border, red

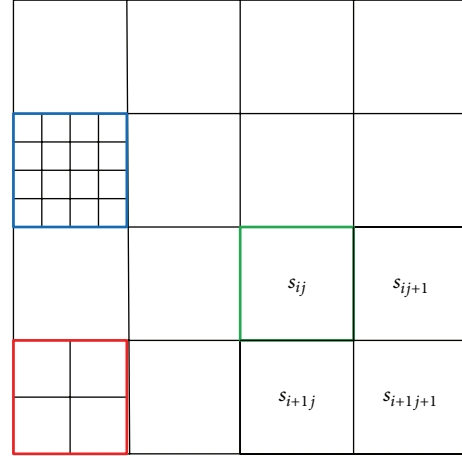


FIGURE 2: Different grid resolution.

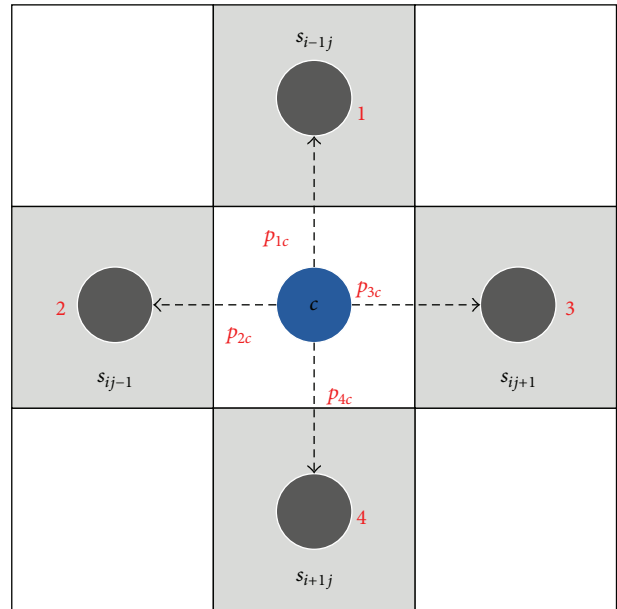


FIGURE 3: PM<sub>2.5</sub> diffusion scenario.

border, and blue border denotes grid at levels -1, 1, and 2, respectively, and the rest of grids denotes grid at level 0. A grid ( $s_{ij}$ ) at level -1 denotes that the grid ( $s_{ij}$ ) and its right grid ( $s_{ij+1}$ ), lower grid ( $s_{i+1j}$ ), and lower right grid ( $s_{i+1j+1}$ ) should be merged into one coarser grid with a 1000 m × 1000 m resolution; a grid at level 0 means the initial grid should keep its initial resolution with a 500 m × 500 m resolution, a grid at level 1 denotes the initial grid should be split into four grids with a 250 m × 250 m resolution, and a grid at level 2 denotes the initial grid should be split into sixteen grids with a 125 m × 125 m resolution.

*Definition 6* (random walk and transition probabilities). The term random walk was first introduced by Pearson in 1905 [18] and is a mathematical formalization of a path that consists of a sequence of successive random steps. In our work, we simulate the particle (PM<sub>2.5</sub>) diffusion referring to the concept of random walk, as shown in Figure 3, and

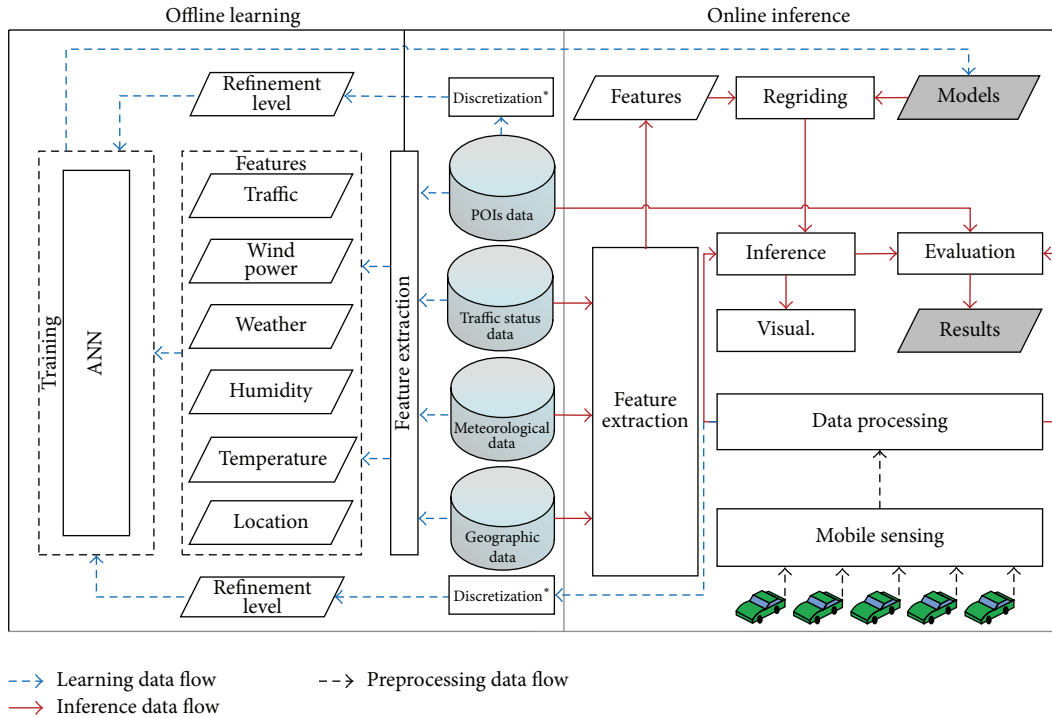


FIGURE 4: Framework of our system.

particles in center grid  $c$  will move to adjoining neighbors  $(1, 2, \dots, n)$  with different transition probabilities influenced by numerous variables, such as wind direction, geographical conditions, and factors, where  $p_{ij}$  represents the probability from location  $i$  to its nearest neighbor  $j$ .

**Definition 7** (false positive rate and false negative rate). In this work, FPR and FNR are used to demonstrate the error of ANN training results. FPR denotes that the rate of grids, supposed to be at level  $-1$ , is instead derefined at  $0, 1$ , or  $2$ , which will increase computational cost. FNR denotes that the rate of grids, supposed to be at level  $1$  or  $2$ , is instead derefined at  $0$  or  $-1$ , which will reduce inferring accuracy.

**2.2. Framework.** As shown in Figure 4, the framework of our system consists of two parts, offline learning and online inference, which generate three kinds of data flows: preprocessing, inference, and learning data flows.

**Preprocessing Data Flow.** In this data flow (denoted by broken black arrows), we employ several mobile sensors to collect the original  $PM_{2.5}$  concentration data in the testing region and store them in server for processing. The processed historical  $PM_{2.5}$  concentration data is used for offline learning, and the processed real-time  $PM_{2.5}$  concentration data is used for online inference and evaluates the inference results.

**Learning Data Flow.** In this data flow (represented by broken blue arrows), we extract features for each grid from a variety of data pieces as the input of ANN training model and calculate refinement level from POIs data and processed

mobile sensing data as output. Among the features, temperature, humidity, weather, and wind power are extracted from meteorological data; traffic and location are extracted from traffic status data and geographic data. To improve the training accuracy, we also set some POIs to extract accurate features and refinement levels to supplement training dataset, as detailed in Section 7.3, POIs data is also used to evaluate the inference results; this process is also performed offline.

**Inference Data Flow.** In this data flow (denoted by red solid arrows), we first calculate the real-time features for each grid from meteorological data, traffic status data, and geographic data and feed the features into ANN model to get output refinement level of each grid. Then the grid map of testing area would be regridding according to each grid's refinement level. After regridding the map, the grid  $PM_{2.5}$  concentration of the testing region would be inferred, detailed in Section 6.

### 3. Refinement Criteria

Increments of  $PM_{2.5}$  concentration with value of  $5 \mu\text{g}/\text{m}^3$  and  $10 \mu\text{g}/\text{m}^3$  are considered as vital boundaries. European Study of Cohorts for Air Pollution Effects (ESCAPE) used data from 17 cohort studies based in nine European countries. Prospective analysis shows that long-term exposure to  $PM_{2.5}$ , even with low concentration, will significantly increase lung cancer risk:  $PM_{2.5}$  concentrations increase every  $5 \mu\text{g}/\text{m}^3$  and lung cancer and lung adenocarcinoma risk increased by 18% and 55%, respectively [19]. The research carried out by American Cancer Society (ACS) shows the relative risk of lung cancer mortality is in correlation with  $10 \mu\text{g}/\text{m}^3$

changing of  $PM_{2.5}$ . The risk increased by 8% using  $PM_{2.5}$  concentration data from 1979 to 1983, by 13% using  $PM_{2.5}$  data collected from 1999 to 2000 [20, 21].

According to the  $PM_{2.5}$  IAQI and the change of  $PM_{2.5}$  concentration, the refinement criteria are as follows.

*Level -1.* A grid and its adjacent grids  $PM_{2.5}$  concentrations are all within  $75 \mu\text{g}/\text{m}^3$  (air quality range is “good”), which is within the range of “good” air quality. Those grids could be merged.

*Level 1.* A grid  $PM_{2.5}$  concentration is more than  $115 \mu\text{g}/\text{m}^3$  (air quality range is “moderate pollution”) and the mean difference value (MDV) about such grid and its surrounding grids is more than  $5 \mu\text{g}/\text{m}^3$  but not more than  $10 \mu\text{g}/\text{m}^3$ ; this grid  $PM_{2.5}$  concentration is unhealthy and there may be a pollution source in the grid; this grid should be refined to the finer level (level 1).

*Level 2.* A grid  $PM_{2.5}$  concentration is more than  $115 \mu\text{g}/\text{m}^3$  (air quality range is “moderate pollution”) and the mean difference value about such grid and its surrounding grids is more than  $10 \mu\text{g}/\text{m}^3$ ; this grid’s  $PM_{2.5}$  concentration is very unhealthy and there is a high possibility that a pollution source is in the grid; this grid should be refined to the finest level (level 2).

*Level 0.* Other grids.

There is the calculation of refinement criteria as the following equations show where  $L$  denotes grid refinement level:

$$\begin{aligned} \text{MDV} &= \frac{(|x_{ij} - x_{i-1j}| + |x_{ij} - x_{i+1j}| + |x_{ij} - x_{ij-1}| + |x_{ij} - x_{ij+1}|)}{4}, \\ L &= \begin{cases} -1, & \text{if } \max [x_{ij}, x_{i+1j}, x_{ij+1}, x_{i+1j+1}] \leq 75 \mu\text{g}/\text{m}^3, \\ 1, & \text{if } x_{ij} > 115 \mu\text{g}/\text{m}^3, 5 \mu\text{g}/\text{m}^3 < \text{MDV} \leq 10 \mu\text{g}/\text{m}^3, \\ 2, & \text{if } x_{ij} > 115 \mu\text{g}/\text{m}^3, \text{MDV} > 10 \mu\text{g}/\text{m}^3, \\ 0, & \text{else.} \end{cases} \end{aligned} \quad (1)$$

#### 4. Features Extraction

In this work, we determined a grid’s refinement level by its  $PM_{2.5}$  concentration feature as depicted in Section 3, while many previous researches have proved that the concentration of air pollutants is influenced by some features like temperature, humidity, and traffic flow [22]; the  $PM_{2.5}$  concentration data and features data we collected span one year and also verified such conclusions, detailed in Section 7.2. That is, grid refinement levels are influenced by those features. Accordingly, we identify six grid features as follows.

- (1) Temperature feature ( $F_t$ ): this feature denotes the temperature of a grid with initial resolution, and it is acquired by public data and POIs data.

- (2) Humidity feature ( $F_h$ ): this feature denotes the humidity of a grid with initial resolution; this feature is acquired by public data and POIs data.
- (3) Weather feature ( $F_w$ ): this feature denotes the weather condition of a grid with initial resolution; it is divided into being sunny (denoted by numeral 1), being cloudy (denoted by numeral 2), light rain (denoted by numeral 3), heavy rain (denoted by numeral 4), and being snowy (denoted by numeral 5); this feature is acquired by public data.
- (4) Wind power feature ( $F_{w\_p}$ ): this feature denotes the wind power of a grid with initial resolution; this feature is acquired by public data and POIs data.
- (5) Traffic feature ( $F_{tr}$ ): this feature denotes the road traffic status in a grid with initial resolution; it is divided into being smooth (denoted by numeral 1), being slow (denoted by numeral 2), being crowded (denoted by numeral 4), and being heavily crowded (denoted by numeral 8); this feature is acquired by public data.
- (6) Location feature ( $F_l$ ): this feature denotes the geographic location of a grid with initial resolution in the grid map of testing area; this feature is acquired by geographic data.

Among the features, location feature is spatially related, extracted from geographical data offline since the feature does not change with time; other features (including temperature, humidity, weather, wind power, and traffic) are temporally related, extracted from POIs data and public data (traffic status data and meteorological data), and updated every hour, detailed in Section 7.3.

#### 5. Offline Learning

We propose a grid refinement level inference model based on ANN, as Figure 5 depicted. Here, the model consists of input, hidden, and output layers, and there are six nodes in input layer while output layer only has one node, where  $F_t^k$ ,  $F_h^k$ ,  $F_w^k$ ,  $F_{w\_p}^k$ ,  $F_{tr}^k$ ,  $F_l^k$ , and  $L^k$  denote the temperature, humidity, weather, wind power, traffic, location, and refinement level of grid  $k$ . The function of the hidden layer is to modify weights in the training procedure for the error minimization [23].

#### 6. Online Inference

In this work, we dynamically adapt grid resolution and infer fine-grained  $PM_{2.5}$  concentration distribution hourly. We first regrid grid map according to the real-time grid features extracted from meteorological data, traffic status data, and geographical data and then infer urban  $PM_{2.5}$  concentration distribution based on the data collected by mobile sensors within an hour.

*6.1. Regriding.* As Figure 6 shows, in the regridding process, the real-time grid features of each grid at initial resolution would be input to the ANN network generated by Section 5,

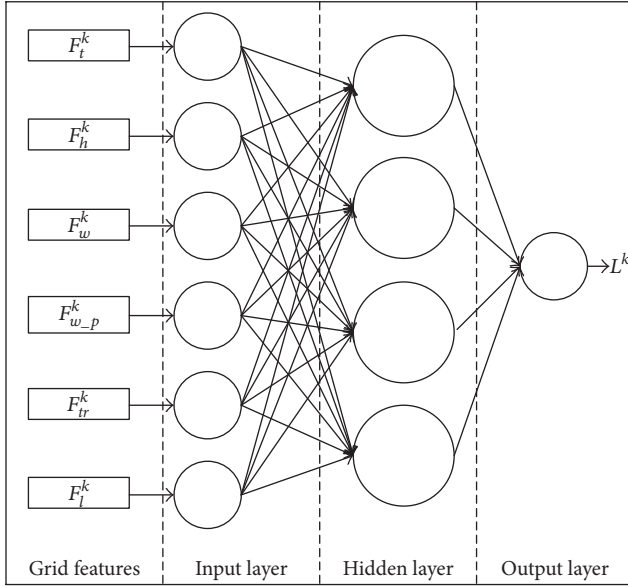


FIGURE 5: Grid refinement level inference model.

and then the grid map of testing area would be regriding according to the output grid refinement levels.

After regriding grid map, the original data collected by several vehicles would be processed and merged into each grid based on their geographic information, while some grids have no data since the vehicles have not passed such areas; the method to infer the  $PM_{2.5}$  concentration in undetected areas is introduced as follows.

**6.2. Calculating Transition Probability Matrix.** We assumed that the  $PM_{2.5}$  concentration of each grid remains the same in a certain interval.  $PM_{2.5}$  concentration of a random grid  $c$  can be identified by the particles transited from surrounding neighbors (affecting region) [4]. In (2), we define the transition probability matrix  $P$ :

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_c \\ \vdots \\ P_N \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1c} & p_{2c} & \cdots & p_{nc} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{nN} \end{bmatrix} \quad (2)$$

$$\sum_{i=1}^n p_{ij} = 1; \quad 0 \leq p_{ij} \leq 1,$$

$$1 \leq i \leq n, \quad 1 \leq j \leq N, \quad c \leq N,$$

where  $N$  is the number of all grids and  $P_c = (p_{1c}, p_{2c}, \dots, p_{nc})$  denotes the particle transport probability between grid  $c$  and its nearest neighbors.

As a result, addressing the problem of particle transport is equal to resolving the transition probabilities for each certain grid. We also assume that the geographic feature is similar and

hourly weather condition is equivalent in the testing region that means different grids in the same testing area follow the same transition probabilities, but for a certain grid  $c$ , different directions have different transition probabilities due to the influence of numerous meteorological factors; in this work, we assume the particle will diffuse in four directions,  $n = 4$ ; that is,  $P_1 = P_2 = \dots = P_N$  and  $p_{1c} \neq p_{2c} \neq p_{3c} \neq p_{4c}$ . Therefore, the problem of modeling particle transport can be simplified as estimating the parameters in  $P_c$ . We calculate the transition probability matrix  $P_c$  using initial resolution grids as (3) shows and then infer the concentration for uncollected area based on it:

$$\begin{aligned} \min \quad & f(P_c) = \left( \sum_{i=1}^4 (p_{ic} \times x_i) - x_c \right)^2 \\ \text{s.t.} \quad & 0 \leq p_{ic} \leq 1, \quad 1 \leq i \leq 4 \end{aligned} \quad (3)$$

$$p_{cc} = 1 - \sum_{i=1}^4 p_{ic}$$

$$0 \leq p_{cc} < 1, \quad c \leq N.$$

**6.3. Inferring.** After regriding process, there are four kinds of grids (coarser, initial, finer, and finest) in the grid map. In this section we introduce the method for estimating the  $PM_{2.5}$  concentration of undetected grids.

The  $PM_{2.5}$  concentration of an undetected initial grid  $c$  can be estimated as follows:

$$x_c = \frac{([p_{1c}^*, p_{2c}^*, p_{3c}^*, p_{4c}^*] \cdot [x_1, x_2, x_3, x_4]^T)}{p_{cc}^*}, \quad (4)$$

where  $x_c$  and  $(x_1, x_2, x_3, x_4)$  denote the  $PM_{2.5}$  concentration of a certain grid  $c$  and its neighbors.

While a certain grid's neighbor may appear, the following four scenarios are as follows as Figure 7 shows.

(1) A center grid's neighbors were split into 4 grids after regriding process, for example, grid 1 in the figure. For this circumstance, we calculate its  $PM_{2.5}$  concentration as follows:

$$x_1 = \frac{(x_{11} + x_{12} + x_{13} + x_{14})}{k}, \quad (5)$$

where  $(x_{11}, x_{12}, x_{13}, x_{14})$  denotes the  $PM_{2.5}$  concentration of four small grids and  $k$  represents nonzero number among  $x_{11}, x_{12}, x_{13}$ , and  $x_{14}$ , that is, the number of detected grids among the four grids.

(2) A center grid's neighbors were split into 16 grids after regriding process, for example, grid 2 in the figure. Similarly, we calculate  $x_2$ ,

$$x_2 = \sum_{i=1}^{16} \frac{x_{2i}}{m}, \quad (6)$$

where  $(x_{21}, \dots, x_{216})$  denotes the  $PM_{2.5}$  concentration of sixteen small grids and  $m$  represents nonzero number among  $(x_{21}, \dots, x_{216})$ .

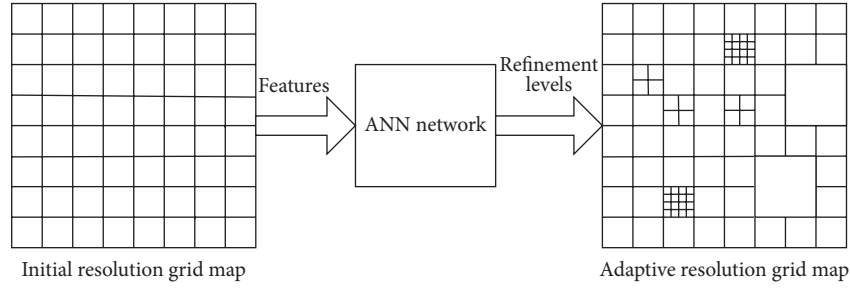


FIGURE 6: Regriding process.

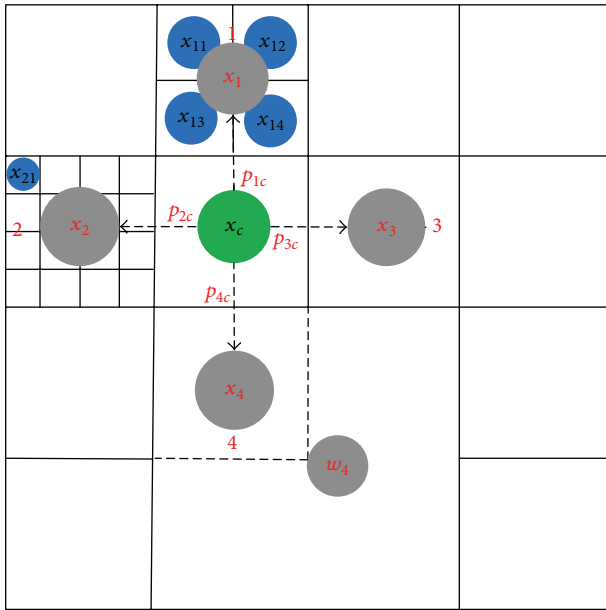


FIGURE 7: Inferring process.

(3) A center grid's neighbor keeps its initial resolution after regriding process, for example, grid 3 in the figure. For this circumstance,  $x_3$  keeps its value.

(4) A center grid's neighbor is redefined to a coarser grid after regriding process, that is, grid 4 in the figure. For this circumstance,  $x_4 = w_4$ , where  $w_4$  denotes the coarser grid's  $PM_{2.5}$  concentration.

The  $PM_{2.5}$  concentration of undetected coarser, finer, and finest grids is estimated by the same method of an undetected initial grid used.

## 7. Experiments

### 7.1. Mobile Data Collection

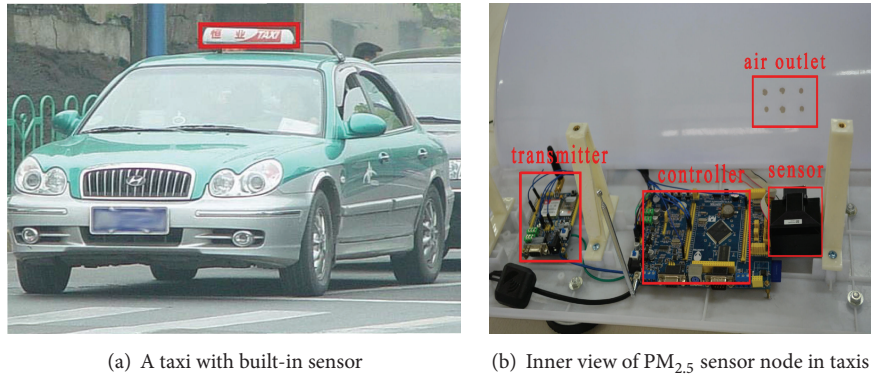
**7.1.1. System Prototype.** We select a low-cost sensor, DN7C3CA006 [24] by SHARP as the built-in sensors; it continually samples the air every 10 ms and provides relative consistent readings. For the sensor calibration and system evaluation, we choose an advanced sensor Lighthouse 3016IAQ [25] as the reference. Lighthouse 3016IAQ is an

advanced portable sensor with  $0.1 \mu\text{g}/\text{m}^3$  estimated error. We employ urban taxis as sensors carriers to collect the mobile data in real-time; it has been verified that mobile sensing can overcome the coverage and granularity problem with its larger coverage and fast movement speed [26]. To cope with complex measuring conditions caused by changeable vehicle speed, sensor nodes are installed on the top of taxis to avoid the physical damaging and keep work even in the worst environmental condition as shown in Figure 8(a); an urban area nearly of  $64 \text{ km}^2$  is covered by 5 taxis in an hour with built-in sensor nodes.

The inner view of a sensor node is shown in Figure 8(b). The sensor nodes are equipped with low-cost  $PM_{2.5}$  sensors DN7C3CA006, GPS (Global Position System), control and transmission modules, and the power interface. This sensor node can be charged directly by vehicle igniter.

**7.1.2. Testing Area.** As shown in Figure 9(a), we choose a local region Xiasha in Hangzhou City of China as the testing region which suffered from air pollution especially  $PM_{2.5}$  for more than 70% days of 15 months (from January 1, 2014, to March 25, 2015); the  $PM_{2.5}$  level of this region was identified as threatened for sensitive groups according to the Chinese standard of  $PM_{2.5}$  level. Apart from the serious hazy days, it also has some comprehensive elements including universities, residential areas, a block of industrial zone, an expressway junction, and a bankside of the Qiantang River. Universities and several parks locate in the northwest. Near the riverbank, there are several residential areas. Industrial zone is in the southeast and the expressway crossing with heavy traffic is in the northeast.  $PM_{2.5}$  concentrations of different locations are collected while taxis are driving randomly in the monitoring area as shown in Figure 9(b), and the prototype system is kept working over a year to monitor local  $PM_{2.5}$  concentration.

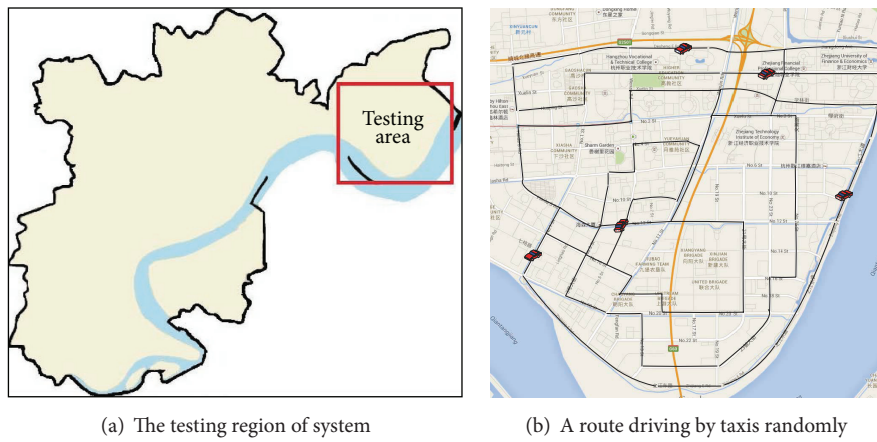
**7.1.3. Sensor Calibration.** Mobile sensing systems normally require high sensor consistency. Therefore, low-cost sensor calibration shall not be carried out only in laboratory with different environmental conditions but also be verified in real world. In our system, the sensor calibration consists of two parts. On one hand, we focus on exploring the relationship between standard values and low-cost sensor reading in the laboratory to eliminate the initial hardware variations. On the other hand, we identify different system compensations through practical experiments. The system



(a) A taxi with built-in sensor

(b) Inner view of PM<sub>2.5</sub> sensor node in taxis

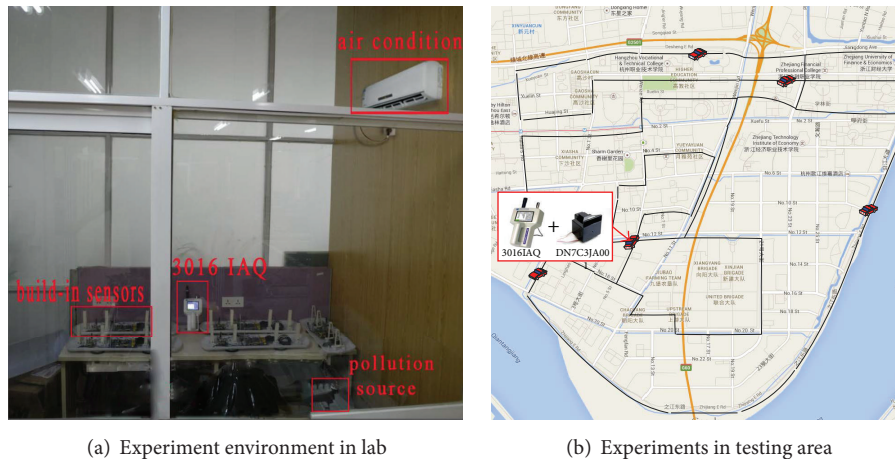
FIGURE 8: Sensors deployment and their inner view.



(a) The testing region of system

(b) A route driving by taxis randomly

FIGURE 9: Testing area and collection schemes.



(a) Experiment environment in lab

(b) Experiments in testing area

FIGURE 10: Experiment for calibration.

compensations mainly consider a possible vibration caused by intermittent moving, high wind, temperature, and complex testing environment. Characteristic of gross errors is also studied based on outdoor samples. Figure 10 shows the detail of testing environment both for laboratory and in real world.

To improve DN7C3CA006 sensors' sensibility for concentration change and stability for same environmental condition, we develop embedded software and design a process for sensor calibration, as shown in Figure 11.

After testing all sensors with different environmental conditions, we find that accidental error follows Gaussian



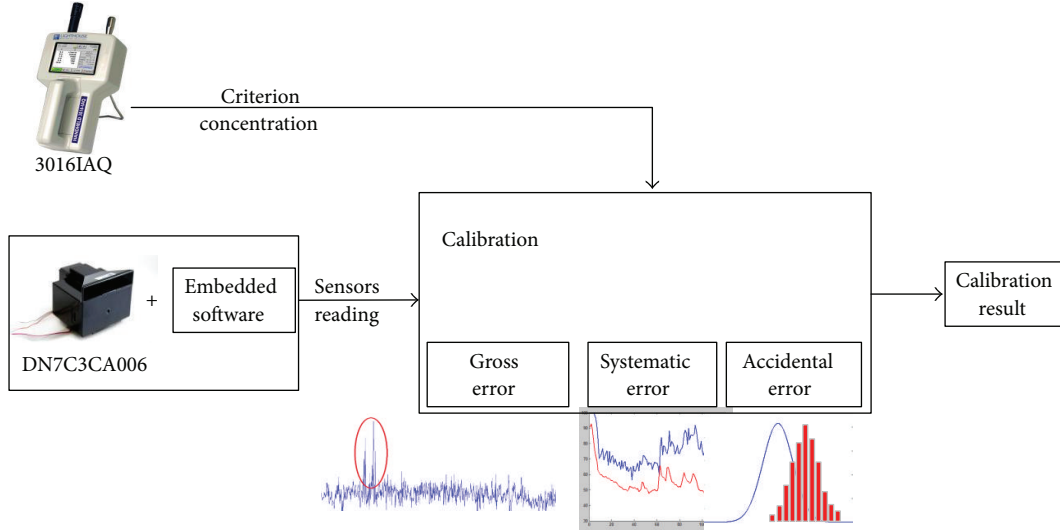


FIGURE 11: Calibration process.

distribution and is independent from sensor to sensor. Systematic error is predictable. It is an inherent bias  $\delta$  in the system [27], which could be caused by testing conditions or the organization of hardware modules. These two errors can be minimized by calibration. The relationship between the reference value and detected concentration is described in (7).  $X^* = [x_1^*, \dots, x_n^*]$  is an observed value of  $\text{PM}_{2.5}$  concentration and  $X = [x_1, \dots, x_n]$  stands for the reference value which is obtained by Lighthouse 3016IAQ. For a sensor  $I$ , accidental deviations are unpredictable and have no expected value [27].  $\varepsilon$  is a set which describes random errors:

$$X^* = X + \delta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \quad (7)$$

In laboratory, we focus on optimizing data redundancy to obtain an acceptable range of random errors. Hardware parameters are also estimated to minimize the systematic errors. To minimize accidental deviations, we change the detection period of DN7C3CA006 from 10 ms to 10 s achieve intensive sampling. This strategy still can be adopted with a certain movement speed. Based on this data piece, we find that the calibrated sample is within  $\alpha = 10 \mu\text{g}/\text{m}^3$  of actual value with 85% confidence level, according to (8) deriving from Chebyshev's Inequality,

$$P\{|X^* - X - \delta| < \alpha\} \geq 1 - \frac{\sigma^2}{\alpha^2}. \quad (8)$$

In mobile sensing scene, gross error detection is performed under real-time experimental condition. For mobile data, gross errors are eliminated once samples are sniffed by sliding window technique which is verified in dynamic systems [28, 29]. We normalize time variable into 10 s interval in each window. The maximum number of windows is  $n_M$  and current number is  $n_C$ ; for a mobile sensor  $i$ , if system sniffs a possible error at the  $t$ th interval, the size of current window will be extended and recalculate the fault tolerant  $\Delta_i(t)$  according to (9). As a result, samples corrupted by gross

error will be eliminated; otherwise, it will be uploaded to database when it turns to left from window:

$$\begin{aligned} \Delta_i(t) &= \frac{\sum_{j=1, j \neq t}^{\min\{n_C, n_M\}} |t - j| \times |x_i(t) - x_i(j)| / \text{mean}(x_i(1), \dots, x_i(n_C))}{\sum_{j=1, j \neq t}^{\min\{n_C, n_M\}} |t - j|} \\ n_C &= \begin{cases} n_C + 1 & \text{if } \Delta_i(t) > \text{threshold,} \\ \lfloor \frac{n_M}{2} \rfloor & \text{else.} \end{cases} \end{aligned} \quad (9)$$

Systematic error is either constant bias or related to the actual value. To explore the relationship between referenced concentration  $Y$  collected by 3016IAQ and corrupted value  $X$  detected by DN7C3CA006, we refer to the linear dependence  $XB \approx Y$  provided by [24]. Then, the singular value decomposition  $\Sigma$  is computed for  $[XY] = U\Sigma V^T$  and the partitioning definition is described as follows:

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \quad (10)$$

According to the total least square [30], the parameter matrix  $B$  can be estimated as  $B_{\text{tls}} = -V_{12}V_{22}^{-1}$ .

In particular, for moving samples, we also consider the possible influences from external conditions, such as temperature, wind speed, wind direction, humidity, and traffic volumes. We denote each possible influence factor as  $f_i$  and the difference between 3016IAQ and DN7C3CA006 as  $\vartheta$ . Then factor analysis ( $\vartheta = w_0 + w_1f_1 + \dots + w_mf_m$ ) is used to estimate different influence factors to minimize systematic error.

After calibration process, the performance under different testing conditions is shown in Figure 12. Compared with the concentration detected by 3016IAQ, Figure 12(a) describes the calibration result of DN7C3CA006 sensors over 1058 samples in laboratory. It shows that two datasets

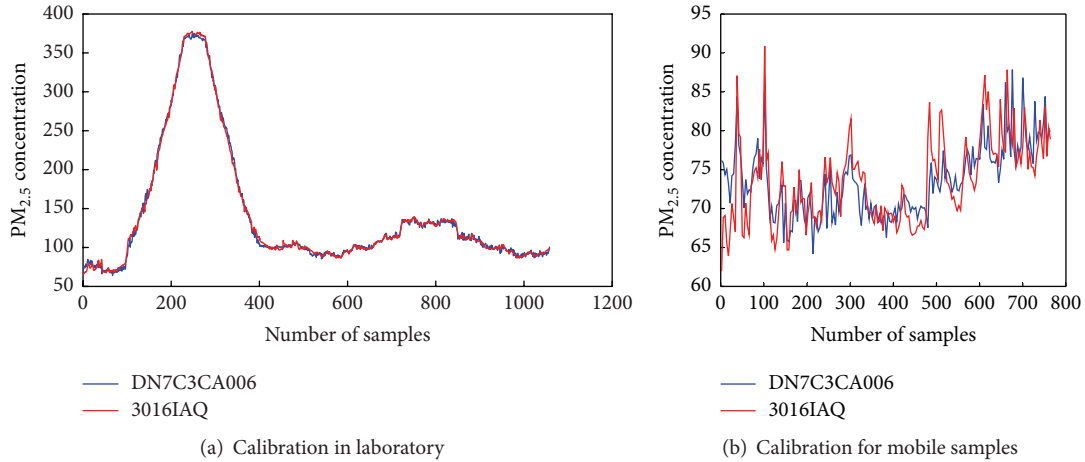


FIGURE 12: Sensor calibration.

TABLE 2: Performance of AG-PCEM with different initial resolution.

Initial resolution	Computational cost(s)	Accuracy (%)
250 m $\times$ 250 m	1.44	0.907
500 m $\times$ 500 m	0.622	0.863
1000 m $\times$ 1000 m	0.215	0.412

have a similar trend. Furthermore, it reflects good Pearson correlations of  $r = 0.864$  between calibrated DN7C3CA006 and 3016IAQ. Figure 12(b) shows that samples collected in practical experimental conditions have a lower correlation of  $r = 0.673$ , but it is still considered as an acceptable one.

**7.1.4. Suitable Resolution.** In the real testing region, there are various functional areas and complex geographic structures, and the region suffered from air pollution badly in most days of a year, which result in great difference of  $PM_{2.5}$  concentration between different locations (in real test, the maximum value is nearly twice minimum value in real test). To find out the most suitable initial resolution for our algorithm, we analyze the performance of AG-PCEM with different initial resolutions. As Table 2 shows, generally, higher initial resolution for AG-PCEM is beneficial for accuracy improvement and brings more computational cost. Taking computational cost and accuracy into account, we adopt an initial resolution with 500 m  $\times$  500 m and a coarser resolution with 1000 m  $\times$  1000 m and two finer resolutions with 250 m  $\times$  250 m and 125 m  $\times$  125 m to dynamically adapt the grid size. The initial resolution can be easily adjusted to fit different application environment.

**7.1.5. Detection Strategy.** Current official monitoring systems take the hourly detection strategy. It means  $PM_{2.5}$  concentration, temperature, wind power, and other meteorological factors will be refreshed once in an hour. All of the analysis in meteorology is built on such basis. Therefore, in this paper, we also assume that influencing features remain stable in an hour. To verify the assumption, we detect the hourly variation

of  $PM_{2.5}$  concentration at certain location, by an advanced sensor 3016IAQ. It shows that the hourly detection strategy can be adopted in this method as it only has around 10% variation in an hour, as Figure 13(a) has shown. We also analyze the impact of vehicle number and find out that 5 testing taxis have 98.8% coverage of all main streets which generate 86.3% accuracy, and the improvement of coverage and accuracy is very little by adding more testing cars, depicted in Figure 13(b).

As a result, hourly samples collected by 5 mobile carriers can be seen as the data acquisition from a large-scale distributed sensor network configured in the area. In this way, we dramatically increase the concentration samples of monitoring area at an extremely low expense; sensing the air every 10 ms provides large amounts of original data under the premise that guarantees the stability of the sensors.

**7.2.  $PM_{2.5}$  Concentration Influenced by Meteorological Features.** We use a dataset including meteorological conditions data and  $PM_{2.5}$  concentration data which span a year (from December 10, 2014, to December 30, 2015) to figure out the correlation between them and results as shown in Table 3.

Through the analysis of experiment results, we find that meteorological features and  $PM_{2.5}$  concentration have significant correlation; in particular, relative humidity has a great positive impact on  $PM_{2.5}$  concentration. Furthermore, the partial correlation coefficient of a meteorological feature (except temperature) and  $PM_{2.5}$  concentration obviously increased by controlling the influence of other variables; it demonstrates that meteorological features have integrated impacts on  $PM_{2.5}$  concentration and they cannot be analyzed one to one.

### 7.3. Dataset

**7.3.1. POIs Data.** We collect the  $PM_{2.5}$  concentration, temperature, humidity, and wind power data of POIs to supplement training datasets and use measured  $PM_{2.5}$  concentration to

TABLE 3: The correlation between meteorological features and PM<sub>2.5</sub> concentration.

Meteorological feature	Correlation coefficient	Partial correlation coefficient	Controlled variable
Temperature	0.153	0.159	Relative humidity, wind power, weather
Relative humidity	0.435	0.509	Temperature, wind power, weather
Wind power	-0.202	-0.267	Relative humidity, temperature, weather
Weather	-0.110	-0.216	Relative humidity, wind power, temperature

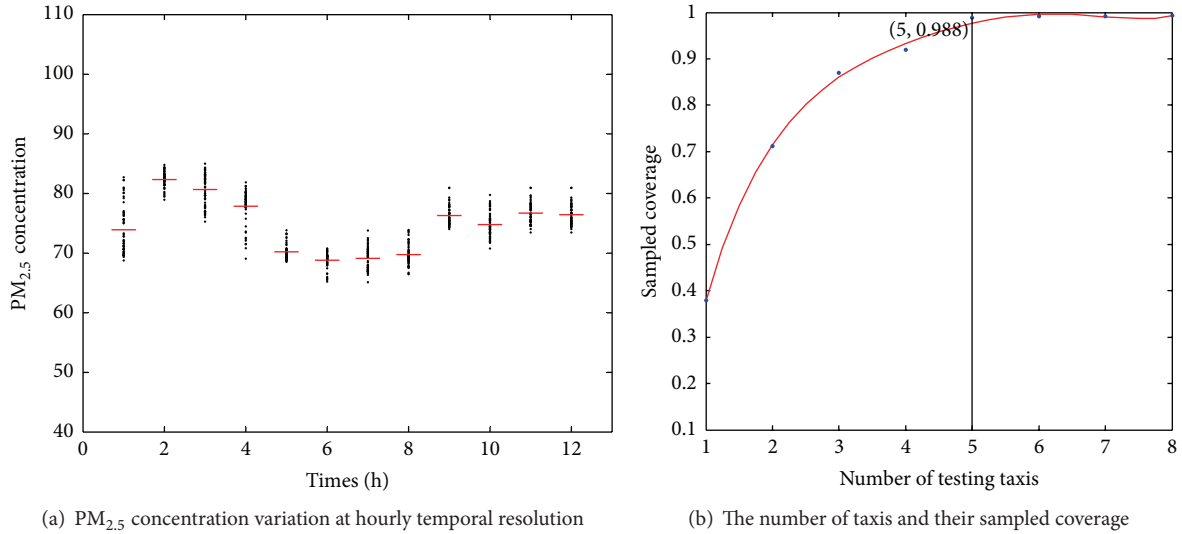


FIGURE 13: The effect of hourly temporal resolution and cars number.

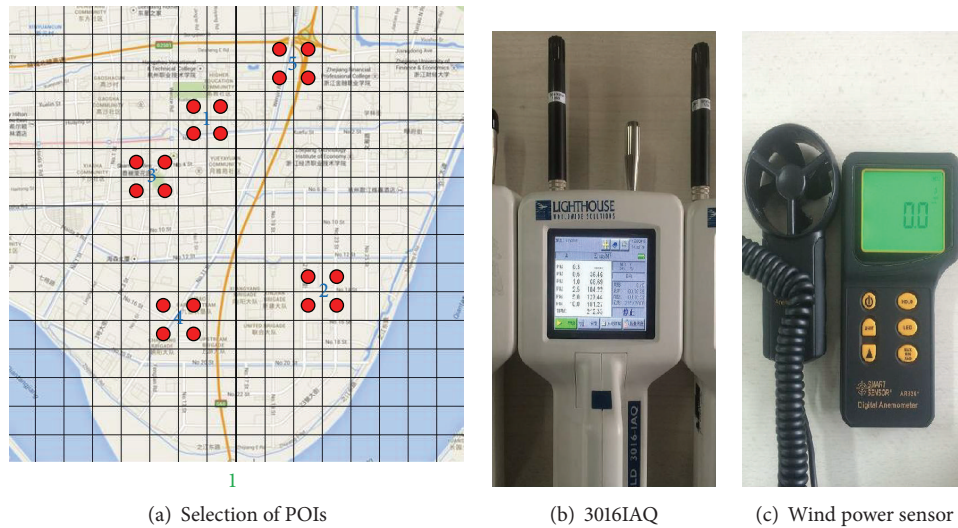


FIGURE 14: POIs data collection.

evaluate the inference accuracy offline. To ensure the diversity of samples, we deliberately select some representative places in the testing region as POIs, including universities (1), residential areas (2), commercial district (3), industrial zone (4), and expressway junction (5) as shown in Figure 14(a). We select four 3016IAQ advanced sensors with temperature/relative humidity probe and four wind power sensors

[31] as Figures 14(b) and 14(c) show and employ them for four places in the same time where their location is four neighboring grids in the grid map (denoted by red dot in Figure 14(a)).

7.3.2. *Traffic Status Data.* We collect traffic status data using web crawler from a public traffic status website.

TABLE 4: Comparison between AG-PCEM and PCEM with different resolution.

Model	Data coverage (%)	Computational cost(s)	Accuracy (%)
AG-PCEM	41.6	0.622	0.863
PCEM with 100 m × 100 m resolution	8.71	8.86	0.815
PCEM with 200 m × 200 m resolution	19.4	2.19	0.911
PCEM with 250 m × 250 m resolution	24.7	1.44	0.876
PCEM with 500 m × 500 m resolution	42.2	0.473	0.725

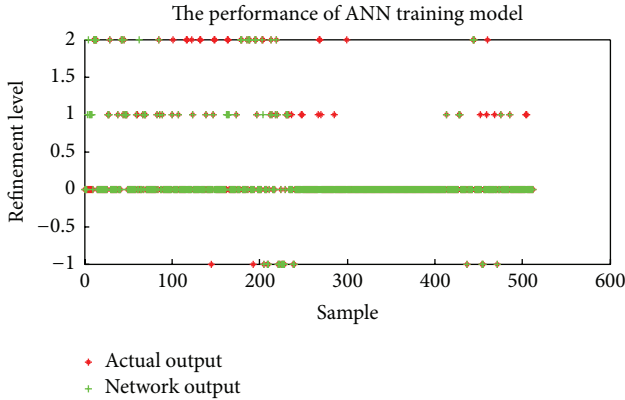


FIGURE 15: ANN performance.

**7.3.3. Meteorological Data.** We collect meteorological data, consisting of temperature, humidity, weather, and wind power, from a public website monitored by nearest official station [32] every hour.

**7.3.4. Geographical Data.** Geographical data is mainly used to map the original  $PM_{2.5}$  concentration data collected by mobile sensors to grids online and calculate location feature offline. We collected geographical data used as a GPS module; this module has the same sampling frequency as  $PM_{2.5}$  sensor. Controller module combines geographical data and  $PM_{2.5}$  concentration data as data for transmission.

## 8. Experimental Results

In this section, we evaluate the performance of AG-PCEM on its offline learning accuracy and online inferring computational cost and accuracy.

**8.1. Performance of Offline Learning.** We randomly choose 512 grid samples from a large amount of historical data; each sample was described by a set of attributes  $A = \{\text{refinement level, temperature, humidity, weather, wind power, traffic, location}\}$ . We use  $\{\text{temperature, humidity, weather, wind power, traffic, location}\}$  as the input of ANN network and compare the output to  $\{\text{refinement level}\}$ .

Figure 15 shows ANN has a good performance in learning grid refinement level. The accuracy of learning

result is 93% and FPR and FNR are 11.1% and 19.5%, respectively.

**8.2. Performance of Online Inference.** To validate the performance of our algorithm, we choose the original  $PM_{2.5}$  concentration data detected by sensors on August 26, 2015, as the evaluating sample and randomly choose 100 monitoring locations as the testing points. Algorithm was tested on a 64-bit server with a Core™ 3.30 G CPU and 4 GB RAM. We adopt two parallel experiments to analyze the performance between AG-PCEM and PCEM with different fixed grid resolutions and the accuracy between AG-PCEM and other widely used methods.

**AG-PCEM and PCEM.** Table 4 resolution shows the performance of AG-PCEM and PCEM with different fixed grid resolution. The results demonstrate AG-PCEM has good performance in terms of accuracy and computational cost. Comparing AG-PCEM and PCEM with the same initial (500 m × 500 m) resolution, AG-PCEM performs much better in inferring accuracy with an acceptable computational cost. Comparing AG-PCEM and PCEM with 250 m × 250 m resolution, the computational cost of AG-PCEM is reduced by about 40.2% under very close accuracy. Though high resolution with 200 m × 200 m of PCEM can improve inferring accuracy, the increment is too little compared to multiple computational cost; we also find that too high resolution of PCEM with 100 m × 100 m reduces the accuracy instead due to bad data coverage (only 8.71% of grids have original data in this situation).

**Accuracy of AG-PCEM.** We also compare our system to some widely used methods, such as classical multivariable linear regression (MLR), artificial neural network (ANN), and Gaussian process (GP). Figure 16 shows the inference accuracy for each of them; result shows that average estimated error of our system can be reduced by about 42.9%, 38.8%, and 14.6% compared with MLR, ANN, and GP respectively.

**Visualization.** Figure 17 shows the heat map of  $PM_{2.5}$  concentration in testing area; it demonstrates that, at the same time,  $PM_{2.5}$  concentration is highly different at different

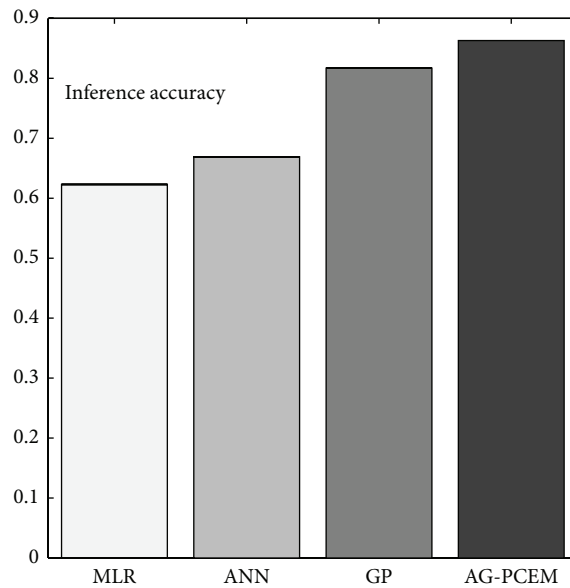
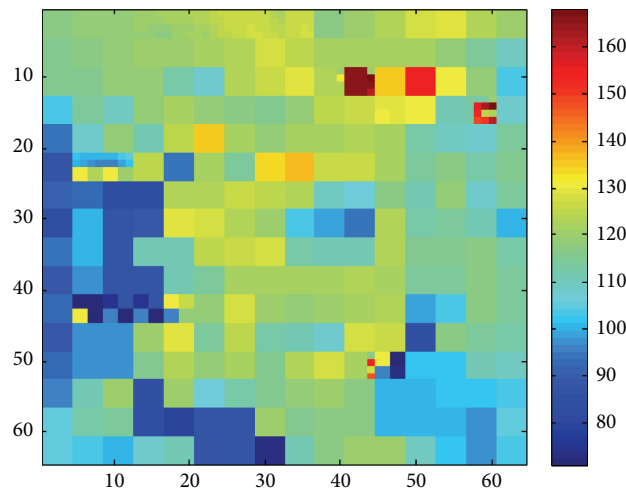


FIGURE 16: Accuracy comparison.

FIGURE 17:  $PM_{2.5}$  concentration heat map.

location in the testing region, and it is valuable for official management of locating pollution sources. Also, citizens can acquire the information of immediate environment conveniently through the applications, as shown in Figure 18.

## 9. Conclusions

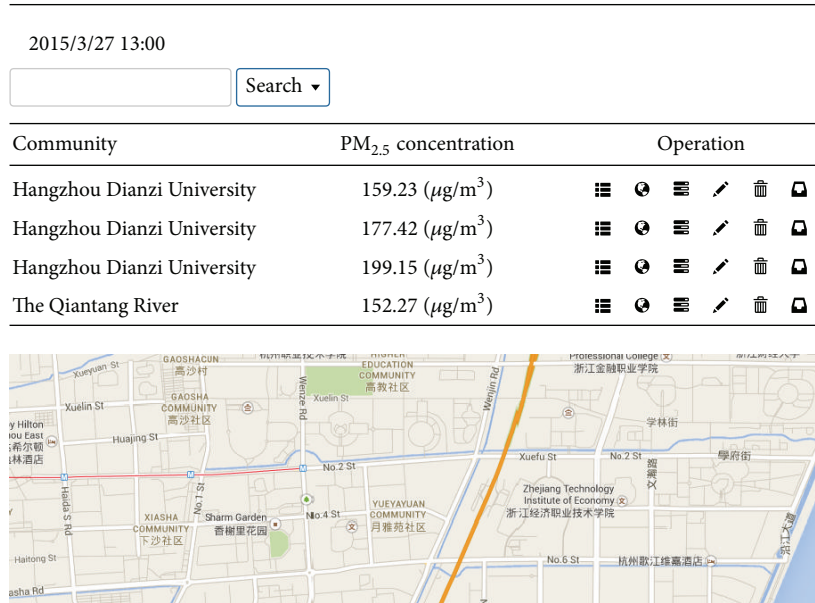
In this paper, we have proposed a mobile sensing system to collect  $PM_{2.5}$  data in the city and present AG-PCEM to infer the  $PM_{2.5}$  concentration for undetected grids using dynamic adaptive grids. Our system can provide a precision  $PM_{2.5}$  concentration distribution for citizens and help official management find pollution sources.

A prototype system has been prepared and implemented in real world over a year and has been tested by employing 5 taxis from October 11, 2014, to November 25, 2015. The results show that the proposed system presents low computational cost and high accuracy.

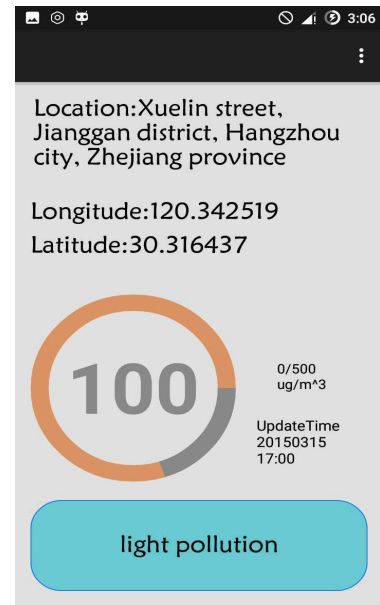
As the first system providing urban air quality monitoring with adaptive resolution, our system can provide deeper understanding of  $PM_{2.5}$  concentration, and it can be easily expanded to wide-range air quality monitor.

## Competing Interests

The authors declare that they have no competing interests.

PM<sub>2.5</sub> distribution in Hangzhou

(a) Immediate information on website



(b) Mobile termination

FIGURE 18: Application.

## Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant nos. 61190113, 61401135, and 61471150.

## References

- [1] J. LePeule, F. Laden, D. Dockery, and J. Schwartz, "Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard six cities study from 1974 to 2009," *Environmental Health Perspectives*, vol. 120, no. 7, pp. 965–970, 2012.
- [2] Hangzhou, <http://dwz.cn/ft4lx>.
- [3] Y. Cheng, X. Li, Z. Li et al., "Aircloud: a cloud-based air-quality monitoring system for everyone," in *Proceedings of the 12th ACM Conference on Embedded Networked Sensor Systems*, Memphis, Tenn, USA, November 2014.
- [4] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: when urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, Ill, USA, August 2013.
- [5] D. Hasenfratz, O. Saukh, C. Walser et al., "Deriving high-resolution urban air pollution maps using mobile sensor nodes," *Pervasive and Mobile Computing*, vol. 16, pp. 268–285, 2015.
- [6] Y. Hu, J. Fan, H. Zhang, X. Chen, and G. Dai, "An estimated method of urban PM<sub>2.5</sub> concentration distribution for a mobile sensing system," *Pervasive and Mobile Computing*, pp. 88–103, 2015.
- [7] A. Tomlin, M. Berzins, J. Ware, J. Smith, and M. J. Pilling, "On the use of adaptive gridding methods for modelling chemical transport from multi-scale sources," *Atmospheric Environment*, vol. 31, no. 18, pp. 2945–2959, 1997.
- [8] I. Lagzi, D. Kármán, T. Turányi, A. S. Tomlin, and L. Haszpra, "Simulation of the dispersion of nuclear contamination using an adaptive Eulerian grid model," *Journal of Environmental Radioactivity*, vol. 75, no. 1, pp. 59–82, 2004.
- [9] I. Lagzi, T. Turányi, A. S. Tomlin, and L. Haszpra, "Modelling photochemical air pollutant formation in Hungary using an adaptive grid technique," *International Journal of Environment and Pollution*, vol. 36, no. 1–3, pp. 44–58, 2009.
- [10] A. S. Tomlin, S. Ghorai, G. Hart, and M. Berzins, "3-D multi-scale air pollution modelling using adaptive unstructured meshes," *Environmental Modelling and Software*, vol. 15, no. 6–7, pp. 681–692, 2000.
- [11] S. Ghorai, A. S. Tomlin, and M. Berzins, "Resolution of pollutant concentrations in the boundary layer using a fully 3D adaptive gridding technique," *Atmospheric Environment*, vol. 34, no. 18, pp. 2851–2863, 2000.
- [12] R. K. Srivastava, D. S. McRae, and M. T. Odman, "An adaptive grid algorithm for air-quality modeling," *Journal of Computational Physics*, vol. 165, no. 2, pp. 437–472, 2000.
- [13] F. Garcia-Menendez, A. Yano, Y. Hu, and M. Talat Odman, "An adaptive grid version of CMAQ for improving the resolution of plumes," *Atmospheric Pollution Research*, vol. 1, no. 4, pp. 239–249, 2010.
- [14] United States Environmental Protection Agency (USEPA), <https://www.cmascenter.org/cmaq>.
- [15] E. M. Constantinescu, A. Sandu, and G. R. Carmichael, "Modeling atmospheric chemistry and transport with dynamic adaptive resolution," *Computational Geosciences*, vol. 12, no. 2, pp. 133–151, 2008.
- [16] G. R. Carmichael, L. K. Peters, and R. D. Saylor, "The STEM-II regional scale acid deposition and photochemical oxidant model—I. An overview of model development and applications," *Atmospheric Environment Part A: General Topics*, vol. 25, no. 10, pp. 2077–2090, 1991.

- [17] Ministry of Environmental Protection of the People's Republic of China, "Technical regulation on ambient air quality index," HJ 633-2012, 2012.
- [18] K. Pearson, "The problem of random walk," *Nature*, vol. 72, p. 294, 1905.
- [19] O. Raaschou-Nielsen, Z. J. Andersen, B. Beelen et al., "Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)," *The Lancet Oncology*, vol. 14, no. 9, pp. 799–805, 2013.
- [20] C. A. Pope III, R. T. Burnett, M. J. Thun et al., "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *The Journal of the American Medical Association*, vol. 287, no. 9, pp. 1132–1141, 2002.
- [21] R. M. Harrison, D. J. T. Smith, and A. J. Kibble, "What is responsible for the carcinogenicity of PM 2.5?" *Occupational and Environmental Medicine*, vol. 61, no. 10, pp. 799–805, 2004.
- [22] S. Vardoulakis, B. E. A. Fisher, K. Pericleous, and N. Gonzalez-Flesca, "Modelling air quality in street canyons: a review," *Atmospheric Environment*, vol. 37, no. 2, pp. 155–182, 2003.
- [23] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14–15, pp. 2627–2636, 1998.
- [24] Sharp, *Device Specification for PM2.5 Sensor Module*, Electronic Components and Devices Division, Sharp Corporation, 2014.
- [25] Lighthouse, <http://www.golighthouse.nl/en/indoor-air-quality>.
- [26] X. Xu, P. Zhang, and L. Zhang, "Gotcha: a mobile urban sensing system," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (SenSys '14)*, pp. 316–317, Memphis, Tenn, USA, November 2014.
- [27] K. Schäfer, "Introduction to the theory of error, von Yardley Beers. Addison-Wesley Publish. Comp. INC., Cambridge 42 Mass. 1953. 1. Aufl. VI, 65 S., brosch. \$ 1.25," *Angewandte Chemie*, vol. 67, no. 16, pp. 432–432, 1955.
- [28] P. Vachhani, R. Rengaswamy, and V. Venkatasubramanian, "A framework for integrating diagnostic knowledge with nonlinear optimization for data reconciliation and parameter estimation in dynamic systems," *Chemical Engineering Science*, vol. 56, no. 6, pp. 2133–2148, 2001.
- [29] D. M. Prata, M. Schwaab, E. L. Lima, and J. C. Pinto, "Simultaneous robust data reconciliation and gross error detection through particle swarm optimization for an industrial polypropylene reactor," *Chemical Engineering Science*, vol. 65, no. 17, pp. 4943–4954, 2010.
- [30] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Processing*, vol. 87, no. 10, pp. 2283–2302, 2007.
- [31] Wind Power Sensor, <http://www.smartsensor.cn>.
- [32] China Weather Reuters, <http://www.weather.com.cn>.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

