



# Strategies of Method Selection for Fine Scale PM<sub>2.5</sub> Mapping in Intra-Urban Area Under Crowdsourcing Monitoring

Shan Xu<sup>1</sup>, Bin Zou<sup>1</sup>, Yan Lin<sup>2</sup>, Xiuge Zhao<sup>3</sup>, Shenxin Li<sup>1</sup>, Chengxia Hu<sup>1</sup>

<sup>1</sup>School of Geosciences and Info-Physics, Central South University, Changsha, Hunan, 410083, China

5 <sup>2</sup>Department of Geography & Environmental Studies, University of New Mexico, Albuquerque, New Mexico, 87131, United States

<sup>3</sup>Chinese Research Academy of Environmental Sciences, Beijing, 100012, China

Correspondence to: Bin Zou (210010@csu.edu.cn)

Abstract. Fine particulate matters (PM<sub>2.5</sub>) are of great concern to public due to their significant risk to human health. Numerous

- 10 methods have been developed to estimate spatial PM<sub>2.5</sub> concentrations at unobserved locations due to the sparse fixed monitoring stations. On the other hand, as the rising of low-cost sensing for air pollution monitoring, crowdsourcing activities has been gradually introduced into fine exposure control in cities. However, the optimal mapping method for conventional sparse fixed measurements may not suit this new high-density monitoring way. This study therefore for the first time presents a crowdsourcing sampling campaign and strategies of method selection for hundred meter-scale level PM<sub>2.5</sub> mapping in intra-
- 15 urban area of China. In this process, the crowdsourcing sampling campaign was developed through a group of volunteers and their smart phone applications; the best performed mapping approach was chosen by comparing three widely used modelling method (ordinary kriging (OK), land use regression (LUR), and universal kriging combined OK and LUR (UK)) with increasing training sites. Results show that crowdsourcing based PM<sub>2.5</sub> measurements varied significantly by sites (i.e. urban microenvironments) (Period 1: 28–136 µg m<sup>-3</sup>; Period 2: 115–266 µg m<sup>-3</sup>) and clearly differed from those at national
- 20 monitoring sites (Period 1: 20–58  $\mu$ g m<sup>-3</sup>; Period 2: 146–219  $\mu$ g m<sup>-3</sup>). Despite the performance of the three models in estimating PM<sub>2.5</sub> concentrations all improved as the number of training sites increase, OK interpolation performed best under conditions with non-peak traffic (9:00-11:00) in Period 1 (i.e. light-polluted period) with the hold-out validation R<sup>2</sup> ranging from 0.47 to 0.82. Meanwhile, the accuracy of UK was the highest for 8:00 and 12:00 with less than 70% training sites (0.40–0.69) and all five hours of Period 2 (i.e. heavy-polluted period) (0.32–0.68). Comparatively, LUR demonstrated limited ability in PM<sub>2.5</sub>
- 25 concentration simulations (0.04–0.55). Moreover, spatial distributions of PM<sub>2.5</sub> concentrations based on the selected model with crowdsourcing data clearly illustrated their hourly intra urban variations which are generally concealed by the results from national air quality monitoring sites. This method selection strategy provides solid experimental evidence for method selection of PM<sub>2.5</sub> mapping under crowdsourcing monitoring and a promising access to the prevention of exposure risks for individuals in their daily life.





### **1** Introduction

Fine particulate matters (PM<sub>2.5</sub>) have been associated with increased risk of morbidity and mortality for both long-term and short-term (Beverland et al., 2012; Cohen et al., 2017; Di et al., 2017; Lelieveld et al., 2017). Nevertheless, it comes down to the persistent cumulative effects from the exposure in daily activities, especially in daily traveling (Kingham et al., 2013;

5 Hankey et al., 2017). It would be very helpful for health protection if individuals could consciously choose the location and time for outdoor activities based on detailed knowledge about the spatiotemporal variation of PM<sub>2.5</sub> concentration, and thus exposure.

In situ measurement is the most reliable way to capture the  $PM_{2.5}$  concentrations across every corner of a city in real time. However, fixed monitoring stations under conventional air quality monitoring networks are sparse. As a result, it has been

- 10 difficult for site-based observations to capture spatiotemporal variations of air pollutants, especially in intra-urban areas with unevenly-distributed emission sources and dispersion conditions (Kumar et al., 2015; Zou et al., 2016; Apte et al. 2017). For this, spatial mapping methods including air dispersion modelling, spatial interpolation, satellite remote sensing (RS), and empirical models have been increasingly employed to estimate concentrations of PM<sub>2.5</sub> at unobserved locations over the past two decades (Jerrett et al., 2005; Henderson et al., 2007; El-Harbawi, 2013; Kim et al., 2014; Rice et al., 2015; Fang et al.,
- 15 2016; Zou et al., 2017; Zhai et al., 2018; Xu et al., 2018). Among them, the outputs of dispersion model largely depend on detailed emission inventories and meteorological information which are not usually available for many cities. The coarse spatial resolution (≥1-10 km) of satellite instruments and data missing problem due to the cloud cover prohibit the widespread use of RS in PM<sub>2.5</sub> concentration mapping in urban environments (Zou et al., 2015; Apte et al., 2017).
- In contrast, geostatistical and empirical models can estimate concentrations at high spatial resolution with a rather low
  requirement for data. The most commonly used ones are ordinary kriging (OK) interpolation and land use regression (LUR) modelling. Moreover, some studies have improved the estimating accuracy through a universal kriging (UK) interpolation by combining these two technologies (Beelen et al., 2009; Mercer et al., 2011; De Hoogh et al., 2018). While they have been successfully applied to map the spatial variability of PM<sub>2.5</sub> concentrations for various geographic areas, their accuracy varies as concentration levels and sample sizes change (Wang et al., 2012; Mercer et al., 2011; Lee et al., 2014; Zou et al. 2015;
  Gillespie et al., 2016; Choi et al., 2017; De Hoogh et al., 2018).

As the rising of low-cost sensing for air pollution monitoring, the real-time strategies for fine exposure control in cities have been further developed (Kumar et al., 2015). Crowdsourcing activities based on informal social networks and web 2.0 technologies that allowed citizens themselves to produce geospatial data among others (Heipke, 2010). Unlike the traditional fixed monitoring stations usually mount on the roofs (i.e. 3 to 20 meters above the ground) for the sake of instruments

30 protection, crowdsourcing activities provide real-time  $PM_{2.5}$  monitoring that reflect the real exposure for individuals living and working on the ground. Although crowdsourcing activities tend to produce observations with questionable quality, they enable us to obtain measurements of ambient air pollution in dense networks at relatively low cost. Some studies have employed these data to display the air pollution concentration and investigate the exposure risks (Thompson, 2016; Miskell et al., 2017; Jerrett





et al., 2017). Nevertheless, these observations are still point measurements which are only representative of the limited area around the site and cannot meet the demand of obtaining air pollution concentration whenever and wherever we want. One way to address the aforementioned challenge is to combine high-density crowdsourcing observations with spatial mapping methods. One of the important investigations was carried out by Schneider et al. (2017) in Oslo, Norway. They presented a

- 5 universal kriging technique for urban NO<sub>2</sub> concentration mapping combining near real-time crowdsourcing observations of urban air quality with output from an air pollution dispersion model. However, the high-density crowdsourcing measurements may vary among urban microenvironments with different human daily activities and differed from the sparsely distributed conventional in situ measurements. Using the elected mapping methods from previous studies to depict variation of air pollution at very fine spatial and temporal scale under new monitoring ways may lead to the misclassification of exposure and
- 10 an underestimate of risk. Moreover, as the number of valid crowdsourcing observations may change significantly due to the instruments fault, human error, and other quality issues, the applicability of mapping methods to different sampling sizes still needs sound scientific evidence.

In this study, we therefore presented strategies of method selection for  $PM_{2.5}$  concentration mapping based on crowdsourcing datasets with varying sizes. The intra-urban crowdsourcing sampling campaign was conducted in the city of Changsha, China,

15 over two periods under different pollution scenarios. The performance of OK, LUR and UK in estimating PM<sub>2.5</sub> pollution was evaluated and compared with increasing training sites. The best performed one was then employed to plot the variation of hourly PM<sub>2.5</sub> concentration and identify pollution hotspots in the intra-urban area. Results from this study will provide evidence for method selection of PM<sub>2.5</sub> mapping under crowdsourcing monitoring and contribute significantly to efficiently and effectively air pollution mapping and exposure assessment in intra-urban areas.

#### 20 2 Data and methods

#### 2.1 PM<sub>2.5</sub> sampling

#### 2.1.1 Measurement instrument

Measurements were performed with 86 portable laser air quality monitors. They are convenient to carry with the overall size of  $25 \times 34 \times 14$  cm (Fig. 1a). And the monitor has the advantages of fast response, rather high resolution ( $0.1 \mu g/m^3$ ) and data

- 25 consistency. The concentration of particulate matter was measured using the light-scattering method. The monitor contains a special laser module, the signals were recorded by the photoelectric receptor when particulate matters passing through the laser light. The count and size of particulate matters were then analysed by the microcomputer after the signals were amplified and converted. Finally, their concentrations were calculated based on the conversion factor (K-Factor) for airborne dust concentration (Fig. 1b).
- 30 In order to ensure the quality of observations from those monitors, we randomly selected 30 monitors and continuously observed two days next to three national monitoring instruments. The scatter plots of hourly measurements for two instruments





were presented in Fig. 1c. The rather high correlation coefficient  $R^2$  of 0.89, and low root-mean-square-errors (RMSE, 5.63) and mean relative error (MRE, 4.48%) has guaranteed the reliability of sampling data, more details can be found in the Supporting Information (Table S1).

# 2.1.2 Sampling design

- 5 The sampling area is located in the Changsha metropolitan area (112 <sup>4</sup>9'-113 <sup>1</sup>4'E, 27 <sup>5</sup>8'-28 <sup>2</sup>4'N), which covers an area of approximately 920 km<sup>2</sup> consisting of seven districts (see Fig. 2). Changsha is the capital of Hunan Province with a population of greater than 7 million people. The area experienced high-level exposure to air pollutants due to the growing anthropogenic activities and intensive energy consumption.
- To make sure the sampling sites distributed typical for different urban microenvironments (i.e. residential community, building site, school, park, etc.) and relatively even. A series of rules were designed to set up the PM<sub>2.5</sub> sampling network based on air quality related local geographical and meteorological conditions as well as the distribution of potential emission sources (see Table 1). The data supporting sampling design consist of important points of interest (POI), dust surfaces, and main road network. All these data were collected from the Information Center of Department of Land and Resources of Hunan Province. A total of 208 PM<sub>2.5</sub> sampling sites were selected.

#### 15 2.1.3 Sampling and data processing

Sampling was carried out in two time periods in the winter of 2015 to examine the effect of air quality grades on mapping results. The first period was between 8:00 and 12:00, representing a light-polluted period (Period 1). The second period was between 14:00 and 18:00, when Orange (i.e. the pollution level was high) warning signals of haze were released by the Changsha Meteorology Bureau, representing a heavy-polluted period (Period 2). At each monitoring site, samples were

- 20 uploaded twice-to-three times hourly through a smart phone application (App) by a group of volunteers. Geographic coordinates of sampling sites have been uploaded as well. To represent the hourly pollution level of the sampling area, we averaged valid PM<sub>2.5</sub> concentrations measured at all sites. In total, 179-208 samples were successfully collected at each hour over Period 1 and 105-118 samples over Period 2. The official observations at 10 national monitoring sites in the study area were also obtained (China Environmental Monitoring Center, CEMC: http://106.37.208.233:20035/) and averaged for
- 25 comparison purposes.

#### 2.2 Mapping method selection

We divided sampling data into a training set and a validation set (hold-out validation) for each hour to evaluate the performance of OK, LUR and UK with increasing training sites. The training data sets were divided into groups based on the percentages of 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the total number of monitoring sites. Therefore, a series of groups of

30 training samples (36–42, 54–62, 72–83, 90–104, 107–125, 125–146, 143–166, 161–189 sites in Period 1 and 31–35, 42–47, 52–59, 63–71, 73–83, 84–94, 94–106 in Period 2) were extracted using the Subset Features Tool of ArcGIS (version 10.0).





We repeated this process 100 times for each training set size for Period 1 and 50 times for Period 2, respectively. Statistics including the coefficient  $R^2$ , RMSE and MRE between predicted and observed concentrations of  $PM_{2.5}$  in the independent validation set were employed to evaluate and compare their performance.

# 2.2.1 Ordinary kriging

5 Ordinary kriging (OK) estimates the target variable at unsampled location as a linear combination of neighboring observations. It relies on a weighting scheme where closer observations have greater impact on the final prediction. The weighting scheme is dictated by the variogram (Pang et al., 2010; Zou et al., 2015). It can be described as follows:

$$Z^*(X_0) = \sum_i^n \omega_i Z(X_i)$$

$$\sum_i^n \omega_i = 1$$
(1)

where  $Z^*(X_0)$  is the estimation of an unknown sample point,  $Z(X_i)$  and  $\omega_i$  are the value of the *i*<sup>th</sup> known sample point 10 surrounding the unknown sample point and its corresponding weight, respectively, and n is the number of known sample points.

#### 2.2.2 Land use regression

Land use regression (LUR) modeling predicts air pollution concentration by linking measurements of monitoring sites and geographic elements around them using the least square method. It is composed of predictor variable extraction and selection as well as regression modelling and validation.

- 15 Geographic factors including pollution sources (dust surface and pollution industries), road networks, and land use/cover were employed to indirectly characterize the  $PM_{2.5}$  emissions in this study. These data were generated using multiple ring buffers with different radii (50–1000m) at each monitoring site. Meteorological data with a spatial resolution of roughly 0.4 sites per 100 km<sup>2</sup> (wind speed, atmospheric pressure, relative humidity, temperature) that might affect the dispersion of  $PM_{2.5}$  were obtained as well. Geographic factors were made available by the Information Centre of Department of Land and Resources of
- 20 Hunan Province. Meteorological data were released by the Hunan Meteorology Bureau. All variables (Table 2) were extracted using ArcGIS (version 10.0). The optimal buffer radius for the percentage of dust surfaces and land use, pollution industries density, and road density were defined based on the maximum Pearson correlation coefficients.

An automatic forward-backward stepwise regression procedure was used to select the best fitting LUR models based on the screened-out predictors. The final LUR models in this study were determined based on the criteria of the lowest Akaike information criterion (AIC) value and the highest fitting R<sup>2</sup>. The model structure can be expressed as follows:

$$PM_{2.5,s} = a_0 + a_1 X_{1,s} + a_2 X_{2,s} + \dots + a_n X_{n,s} + \mu,$$
<sup>(2)</sup>

where  $PM_{2.5,s}$  is the estimation of hourly-averaged PM<sub>2.5</sub> concentration of site s,  $X_{i,s}$  (i=1,2,...,*n*) are independent variables,  $a_0$  is a constant,  $a_i$  (i=1,2,...,n) are regression coefficients, and  $\mu$  is the random error estimated using the least squares method. This process was conducted in R statistical software (version 3.3.2) (Fox and Weisberg 2011, R Core Team 2016).





# 2.2.3 Universal Kriging

Universal Kriging (UK) in this study is a two-stage statistical procedure. Firstly, sperate standard LUR models were developed based on crowdsourcing observations in training dataset for each hour. Secondly, the residuals for LUR models was calculated and interpolated for each hour using OK technology. Finally, the estimations of residuals at validation sites were extracted and

5 added to LUR estimations.

In this study, OK was performed with Ordinary Kriging using the Geostatistical Analyst Tool of ArcGIS (version 10.0), and interpolated residuals were obtained using the Extract Values to Point Tool. The whole process was implemented with Python scripts.

# 2.3 PM<sub>2.5</sub> concentration mapping

10 Based on the best modelling method with 90% training sites, maps of the spatial distribution of  $PM_{2.5}$  concentration for each hour were estimated. In this study, nearest-neighbor distances range from 15 to 60 meters for Period 1 and 54 to 98 meters for Period 2 between sampling sites. Considering the resolutions of potential predictors, 100 meters was therefore used as the mapping grid size.

#### **3 Results**

# 15 **3.1 Descriptive statistics of PM<sub>2.5</sub> concentrations**

Table 3 shows descriptive statistics of  $PM_{2.5}$  concentration at each sampling hour. We found that the  $PM_{2.5}$  concentrations for each sampling hour in this study were significantly different from those at national monitoring sites. For Period 1,  $PM_{2.5}$  concentrations at 8:00, 9:00, 10:00, 11:00, and 12:00 were (69.67±18.81), (72.97±17.04), (73.08±15.57), (74.12±13.96), and (76.45±14.55) µg m<sup>-3</sup>, respectively. The highest and lowest values of  $PM_{2.5}$  concentration was 136 µg m<sup>-3</sup> (12:00) and 28 µg

- 20 m<sup>-3</sup> (10:00), respectively. These were clearly higher than those of national monitoring sites (i.e. (36.9±10.97) (41.2±8.68) μg m<sup>-3</sup>; 58 and 20 μg m<sup>-3</sup>). Correspondingly, the PM<sub>2.5</sub> concentrations for Period 2 ranged from 115 μg/m3 (16:00) to 266 μg m<sup>-3</sup> (17:00). Their averages and standard deviations were (167.91±14.48), (165.75±17.06), (162.72±16.91), (167.69±20.49) and (171.89±22.4) μg m<sup>-3</sup> at 14:00, 15:00, 16:00, 17:00, and 18:00, respectively. The mean, minimum, and standard deviation of PM<sub>2.5</sub> from crowdsourcing sampling were slightly lower than those from national monitoring. Clearly, the average PM<sub>2.5</sub>
- concentrations of Period 2 were two times higher than those of Period 1 and the highest ones occurred when traffic had peaked (i.e. 12:00 and 18:00) for both periods.

Figure 3 demonstrates the spatial variation of  $PM_{2.5}$  measurements over the two periods in the study area, and strong spatial variations can be found between different sampling sites and two periods. For Period 1, the  $PM_{2.5}$  concentrations decreased gradually from north to south and from west to east. Higher concentrations of  $PM_{2.5}$  (> 75 µg m<sup>-3</sup>) were observed on sampling

30 sites in the northwest corner of the study area. The sampling sites in Changsha County with high levels of green vegetation





5

cover had lower  $PM_{2.5}$  concentrations in comparison with the sites in the inner city. For Period 2, in contrast, sampling sites in the central and eastern parts of study area had higher  $PM_{2.5}$  concentrations than ones in the western part. Monitoring sites had  $PM_{2.5}$  concentrations higher than 150 µg m<sup>-3</sup> in most areas, except for the western Yuelu district. Particularly, sampling sites in areas along the Xiangjiang River, especially in the higher education mega center had experienced extreme  $PM_{2.5}$  pollution (> 210 µg m<sup>-3</sup>).

# 3.2 Model performance for OK, LUR and UK

The box plots of Fig. 4 show the variation of hold-out validation  $R^2$  for the three mapping approaches in relation to the number of training sites. Overall, the variability ranges and average values of  $R^2$  for OK, LUR and UK were positively associated with the increasing training sites. UK outperformed OK significantly for all five sampling hours of Period 2 while only for 8:00 and

10 12:00 of Period 1 when the number of training sites smaller than a certain threshold. LUR demonstrated poorest performance for both periods.

For Period 1, PM<sub>2.5</sub> estimating accuracy were generally highest at 8:00 and lowest at 12:00. The average validation R<sup>2</sup> of OK at 8:00, 9:00, 10:00, 11:00 and 12:00 increased from 0.58, 0.56, 0.51, 0.47, 0.24 (20%,  $n \le 42$ ) to 0.72, 0.78, 0.82, 0.71, 0.48 (90%,  $n \ge 161$ ), respectively. Comparing with OK, the accuracy of LUR demonstrated a significant and steady decrease. The

- 15 lowest and highest average R<sup>2</sup> were 0.26, 0.29, 0.16, 0.16, 0.13 and 0.55, 0.54, 0.40, 0.36, 0.34, respectively. As for UK, the average R<sup>2</sup> ranged from 0.64, 0.59, 0.50, 0.48, 0.40 to 0.71, 0.69, 0.66, 0.60, 0.46. They were weakly smaller than OK at 9:00, 10:00, and 11:00, and the difference between them were positively associated with the increasing training site. The average R<sup>2</sup> of UK at 8:00 and 12:00 were higher than OK when less than 70% sampling sites were divided into training datasets (8:00: 0.65–0.69 vs. 0.58–0.68; 12:00: 0.40–0.44 vs. 0.24–0.41). For Period 2, the validation R<sup>2</sup> from high to low clearly followed a
- sequence of UK > OK > LUR. The average validation R2 of OK at 14:00, 15:00, 16:00, 17:00 and 18:00 increased from 0.25, 0.34, 0.40, 0.27, 0.18 (30%,  $n \le 35$ ) to 0.49, 0.50, 0.59, 0.39, 0.27 (90%,  $n \ge 95$ ), respectively. They were significantly lower than those in Period 1. The average R<sup>2</sup> of LUR were even lower, the lowest ones were found to be 0.08, 0.07, 0.15, 0.06, and 0.04, and the highest ones were 0.22, 0.25, 0.42, 0.22, and 0.16, respectively. Combined OK and LUR, the performance of UK clearly improved with average R<sup>2</sup> ranging from 0.43, 0.44, 0.43, 0.36, and 0.32 to 0.60, 0.68, 0.52, 0.54, and 0.57.
- 25 Average and standard deviation of RMSE and MRE between the observed and predicted concentrations of PM<sub>2.5</sub> in hold-out validation presented in Supporting Information (Table S4–S5) further demonstrated the better performance yet larger variation of these three methods with larger training data sets in PM<sub>2.5</sub> concentration estimation. The average RMSE and MRE of OK and UK were close and significantly smaller than LUR. Meanwhile, those of OK were generally smaller than UK in Period 1 while cases in Period 2 were the opposite.
- Figure 5 shows scatterplots for the repeated holdout-validation results with 90% training sites. For Period 1, the overall R<sup>2</sup> of OK were 0.71, 0.79, 0.82, 0.71, and 0.46 for 8:00, 9:00, 10:00, 11:00, and 12:00, respectively (Fig. 5a). They were slightly larger than those of UK (0.68, 0.67, 0.66, 0.60, 0.44, Fig. 5c) and significantly larger than those of LUR (0.53, 0.52, 0.38, 0.34, 0.29, Fig. 5b). Correspondingly, their RMSE and MRE from low to high were OK (5.95–10.36; 6.80%–9.91%) < UK (8.23–</p>





5

10.92; 9.80%-11.91%) < LUR (10.68-13.16; 12.91%-14.97%). For Period 2, however, the UK presented the highest accuracy with overall R<sup>2</sup> ranging from 0.45 (17:00) to 0.66 (14:00). The OK took the second place with lowest and highest R<sup>2</sup> for five sampling hours being 0.27 (17:00) and 0.54 (16:00) while LUR performed poorest (0.06-0.36). Meanwhile, the values of RMSE and MRE went from LUR (13.61-23.13; 6.60%-10.25%) to OK (10.90-17.31; 4.58%-6.24%) then down to UK (8.98-15.38; 4.11%-5.78%).

3.3 Spatial patterns of crowdsourcing PM<sub>2.5</sub> concentration

Figure 6 reveals the spatial distributions  $PM_{2.5}$  concentrations estimated by the best performed mapping method with 90% training sites over the two periods (i.e. OK for Period 1 and UK for Period 2) in the study area from both crowdsourcing sampling sites and national monitoring sites. Significant difference can be found between two sources, and the crowdsourcing-

10 based hourly PM<sub>2.5</sub> concentration maps clearly demonstrate more detailed intra urban variations than the national monitoringbased ones, especially during the Period 1.

For Period 1, OK interpolated  $PM_{2.5}$  concentration from national monitoring sites (Fig. 6b) for almost all of the study area was less than 35 µg m<sup>-3</sup>. While that from crowdsourcing sampling sites (Fig. 6a) generally shows a three-step growth from the south-east to the north-west and multiple hot spots of  $PM_{2.5}$  concentrations. During each zone of the growth, area with more

- 15 factories and major roads were experiencing relatively higher  $PM_{2.5}$  concentration than other areas. For Period 2, with the exception of 14:00, the national monitoring-based  $PM_{2.5}$  concentration maps estimated by UK (Fig. 6d) showed high-east and low-west patterns, lower concentrations of  $PM_{2.5}$  (<175 µg m<sup>-3</sup>) centered in the center of Yuelu district. However, crowdsourcing based  $PM_{2.5}$  concentrations revealed by UK (Fig. 6c) demonstrate extensive cold spots of  $PM_{2.5}$  concentrations in not only the southern Changsha County but also the southern Kaifu district. While the southern Yuelu and the west Tianxin
- 20 with high-density of factories were hot spots of PM<sub>2.5</sub> concentration.

#### **4** Discussion

Aiming at efficiently and effectively mapping  $PM_{2.5}$  concentration in intra-urban area at fine scale under crowdsourcing monitoring, a high-density crowdsourcing sampling campaign and strategies of popular mapping method selection with increasing training sites were presented in China for the first time.

- 25 The numbers of sampling sites were 18 and 10 per 100 km<sup>2</sup> for Period 1 and Period 2, respectively. This is a tremendous improvement in comparison with the density of about 0.015 sites per 100 km<sup>2</sup> in the national air quality monitoring network in China. As expected, crowdsourcing based  $PM_{2.5}$  measurements varied significantly by urban microenvironments. Meanwhile, compared with observations obtained from 10 national  $PM_{2.5}$  monitoring sites, the average and variability range of  $PM_{2.5}$  observations for each sampling hour in this study were found to be significantly different from those at national
- 30 monitoring sites. These findings suggest that the national air quality monitoring sites are relatively inadequate and inaccurate for exposure risk assessment of PM<sub>2.5</sub> on the ground in urban environment. Crowdsourcing based sampling, on the other hand,





could reduce the monetary cost of sampling activity and guarantee the data collection occur on the real-ground simultaneously, thus could be an effective alternative for high-density  $PM_{2.5}$  monitoring and be strongly supportive to the short-term air pollution exposure assessment for epidemiologic studies.

However, the difference of PM<sub>2.5</sub> observations between crowdsourcing sampling sites and national monitoring sites vary with

- 5 air quality grades. The crowdsourcing observations were significantly larger than national monitoring ones during Period 1 (light-polluted), while that was clearly opposite during Period 2 (heavy-polluted). This suggests the inconvenient truth that the exposure risk remains relatively high for the public when official air pollution levels are "Good" (i.e. No health implications) and "Moderate" (i.e. Members of sensitive groups should reduce outdoor activities) and this risk is significantly different across the urban area. Meanwhile, it further confirms the necessity of developing real-ground high-density PM<sub>2.5</sub> monitoring
- 10 networks. The variation of PM<sub>2.5</sub> difference between two periods may possibly because the change of the major pollution sources in study area; the major contribution of local sources especially the vehicle emission during light-polluted period may lead to the accumulation of PM<sub>2.5</sub> near the ground; the sources of long-range transport of regional pollution during heavypolluted period could increase the concentration of PM<sub>2.5</sub> on the upper layer.
- Unlike previous studies that conducted comparisons of performance and the corresponding exposure assessments of OK, LUR and UK in estimating air pollution concentration at annual and seasonal scale based on measurements from static and sparse fixed stations (Beelen et al., 2009; Mercer et al., 2011; Lee et al., 2014; Zou et al. 2015; Choi et al., 2017; De Hoogh et al., 2018), this is the first study to evaluate and compare their performance with increasing training sites at the hourly scale for different air quality grades under crowdsourcing monitoring.

As expected, the performance of three methods all improved with the increasing training sites. Compared with former studies

- 20 normally developed in other fields (e.g. spatial variability analysis of soil components in the environmental sciences) (Li and Heap, 2014), this study further confirmed the better performance of OK interpolation with larger training data sets in air pollution estimation. Meanwhile, we substantiated the findings of Johnson et al. (2010) that LUR models developed with fewer sampling sites may perform poorly using real-ground  $PM_{2.5}$  measurements. But average hold-out validation R<sup>2</sup> (0.04–0.55) between observed and predicted concentrations of  $PM_{2.5}$  in this study were smaller than those in Johnson et al. (2010) (0.29–
- 0.67) and results in similar studies of NO<sub>2</sub> presented by Wang et al. (2012) and Gillespie et al. (2016) (0.44–0.85). The main reason is probably because the variations of hourly average PM<sub>2.5</sub> concentration between monitoring sites were generally sharper in comparison to those of annual average. Moreover, meteorological conditions played a more sensitive role in shortterm transmission and diffusion of PM<sub>2.5</sub> than long-term processes. These findings suggest that the most effective way to improve the accuracy of mapping method would still be increasing the number of sampling sites. They confirm, again, the
- 30 necessity of developing high-density crowdsourcing sampling for  $PM_{2.5}$  monitoring. However, the increased variability ranges of  $R^2$  and standard deviation of RMSE and MRE with increasing training sites also suggest that the performance of these methods was affected by more than sampling size. The spatial distribution of samples, for example, may influence their estimating accuracy too (Li and Heap 2014).





Contrary to the findings of Zou et al. (2015) and Choi et al. (2017) conduced at the annual scale, OK interpolation surprisingly showed better performance in estimating  $PM_{2.5}$  concentrations compared to the LUR modelling with clearly higher average  $R^2$  and lower RMSE and MRE. UK combined LUR and OK performed clearly better than LUR too (0.32–0.71 vs. 0.04–0.55). This is consistent with the findings of Mercer et al. (2011) (0.67–0.75 vs. 0.48–0.74) and De Hoogh et al. (2018) (0.66 vs.

- 5 0.59). Moreover, it was the best performed method for 8:00 and 12:00 of Period 1 with less than 70% training sites and for all five hours of Period 2. These results suggest that OK interpolation based on crowdsourcing sampling is the best strategy for PM<sub>2.5</sub> mapping in intra urban area when official air pollution levels are "Good" and "Moderate" for non-peak traffic conditions in this study, while UK is the best one when pollution levels are "Heavy-polluted". These findings challenge the traditional point on LUR model's good performance in air pollution mapping and verify that the applicability of mapping methods varies
- as the monitoring technology and sampling density change. Meanwhile, it further implies the major contribution of vehicle emission on PM<sub>2.5</sub> concentrations during light-polluted period considering the peak traffic condition of 8:00 and 12:00. In addition, the accuracy of OK and LUR were obviously higher for Period 1 (0.24–0.82; 0.13–0.55) than for Period 2 (0.18– 0.59; 0.04–0.42), while that for UK was rather stable (0.40–0.71 vs. 0.32–0.68). This indicates the robustness and generalization capability of UK in estimating PM<sub>2.5</sub> concentration.
- 15 Using the selected mapping method, the spatial distributions of hourly PM<sub>2.5</sub> concentration based on crowdsourcing sampling data and national air quality observations were successfully plotted and compared. Clearly, the former one provides more information on intra urban PM<sub>2.5</sub> variations than the latter one. The nearest-neighbor distances ranging from 15 to 60 m between crowdsourcing sampling sites make it real for PM<sub>2.5</sub> concentration mapping to reach the hundred meter-scale level. Meanwhile, in the light-polluted period, this phenomenon was more pronounced and the crowdsourcing estimations were clearly higher
- 20 than the national monitoring ones. These findings above not only suggest the superiority of crowdsourcing activities in  $PM_{2.5}$ monitoring at fine scale, but also prompt us to pay more attention to the scenarios with low-level air pollution. This is critical to the long-term future of air pollution prevention and control and public health protection for China, given that its main emphasis has gradually shifted from the control of heavy pollution to the prevention of exposure risks.
- However, despite a technology course and a set of standard quality parameters before crowdsourcing sampling and data cleaning before data processing have been provided, the possible measurement bias caused by the lack of professional services and the relatively low sampling frequency (twice/ three times per hour) during the process is still unavoidable. Future developments in the workflow of crowdsourcing system, the automatic processing technique of crowdsourced data, and the rational uniform quality standard could improve the efficiency of PM<sub>2.5</sub> concentration sampling and lower the measurement bias (Heipke 2010). Additionally, as the urban air quality grade has an important effect on spatial distribution of samples
- 30 (spatial autocorrelation, heterogeneity etc.), which may also be affected by sample size, the mechanism for this influence is somewhat equivocal and needs further research.





### **5** Conclusions

This study presented strategies of method selection for efficiently and effectively  $PM_{2.5}$  concentration mapping with increasing training sites based on a crowdsourcing sampling campaign. According to the results, it can be confirmed that  $PM_{2.5}$  concentrations in microenvironments varied significantly across the intra urban area in China's city. And these variations could be clearly disclosed by the crowdsourcing based  $PM_{2.5}$  sampling rather than the national air quality monitoring sites.

- 5 be clearly disclosed by the crowdsourcing based  $PM_{2.5}$  sampling rather than the national air quality monitoring sites. Meanwhile, the selection of models for fine scale  $PM_{2.5}$  concentration mapping should be adjusted with the changing sampling and pollution circumstances. Generally, OK interpolation performs best under conditions with non-peak traffic situation in light-polluted period, while the UK modeling can perform better for conditions with the peak traffic and relatively few sampling sites in heavy-polluted period. Additionally, it has to be noticed that the LUR model demonstrates limited ability in
- 10 estimating  $PM_{2.5}$  concentrations at very fine scale in this study. In short, this method selection strategy provides solid experimental evidence for method selection of  $PM_{2.5}$  mapping under crowdsourcing monitoring and a promising access to the prevention of exposure risks for individuals in their daily life.

*Author contribution*. SX performed the experiments and wrote the manuscript text. BZ supervised and designed the research work, as well as helped with the manuscript. YL and XZ helped with discussion and revisions. SL and CH participated in data processing.

Competing interests. The authors declare that they have no conflict of interest.

*Acknowledgments*. This work was supported by the National Key Research and Development Program of China (No. 2016YFC0206201/05), the National Nature Science Foundation of China (No. 41871317), and the Innovation Driven Program of Central South University (No. 2018CX016).

#### 20 References

15

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C.J., Vermeulen,
R. C., and Hamburg, S. P.: High-resolution air pollution mapping with google street view cars: exploiting big data, Environ.
Sci. Technol., 51, 6999–7009, doi:10.1021/acs.est.7b00891, 2017.

Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., Hoogh, K. D., and Briggs, D. J.: Mapping of background air pollution at a fine spatial scale across the European Union, Sci. Total Environ., 407, 1852–1867, doi:10.1016/j.scitotenv.2008.11.048, 2009.





5

Beverland, I. J., Cohen, G. R., Heal, M. R., Carder, M., Yap, C., Robertson, C., Hart, C. L., and Agius, R. M.: A comparison of short-term and long-term air pollution exposure associations with mortality in two cohorts in Scotland, Environ. Health Perspect., 120, 1280–1285, doi:10.1289/ehp.1104509, 2012.

Choi, G., Bell, M. L., and Lee, J. T.: A study on modeling nitrogen dioxide concentrations using land-use regression and conventionally used exposure assessment methods, Environ. Res. Lett., 12, 044003, doi:10.1088/1748-9326/aa6057, 2017.

- Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., YangLiu, Y., Martin, R., Morawska, L., PopeIII, A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., Donkelaar, A., Vos, T., DPhile, C., and Forouzanfar, M. H.: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an
- analysis of data from the global burden of diseases study 2015, Lancet, 389, 1907–1918, doi:10.1016/S0140-6736(17)30505-6, 2017.

De Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U. A., Katsouyanni, K., Klompmaker, J., Martin, R.V., Samoli, E., Schwartz, P. E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., and Hoek, G.: Spatial PM<sub>2.5</sub>, NO<sub>2</sub>, O<sub>3</sub> and BC models for Western Europe-Evaluation of spatiotemporal stability, Environ. Int., 120, 81–92, doi:10.1016/j.envint.2018.07.036, 2018.

- spatiotemporal stability, Environ. Int., 120, 81–92, doi:10.1016/j.envint.2018.07.036, 2018.
  Di, Q., Dai, L., Wang, Y., Zanobetti, A., Choirat, C., Schwartz, J. D., and Dominici, F.: Association of short-term exposure to air pollution with mortality in older adults, Jama, 318, 2446–2456, doi:10.1001/jama.2017.17923, 2017.
  El-Harbawi, M.: Air quality modelling, simulation, and computational methods: a review, Environ. Rev., 21, 149–179, doi:10.1139/er-2012-0056, 2013.
- Fang, X., Zou, B., Liu, X., Sternberg, T., and Zhai, L.: Satellite-based ground PM<sub>2.5</sub> estimation using timely structure adaptive modeling, Remote Sens. Environ., 186, 152–163, doi:10.1139/er-2012-0056, 2016.
   Fox, J., and Weisberg, S.: An R companion to applied regression, 2nd ed., SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks, California 91320, the United States of America, 2011.
   Gillespie, J., Beverland, I., Hamilton, S., and Padmanabhan, S.: Development, evaluation, and comparison of land use
- 25 regression modeling methods to estimate residential exposure to nitrogen dioxide in a cohort study, Environ. Sci. Technol., 50, 11085–11093, doi:10.1021/acs.est.6b02089, 2016. Hankey, S., Lindsey, G., and Marshall, J. D.: Population-level exposure to particulate air pollution during active travel:

planning for low-exposure, health-promoting cities, Environ. Health Perspect.,125, 527–534, doi:10.1289/EHP442, 2017. Heipke, C.: Crowdsourcing geospatial data, ISPRS J. Photogramm., 65, 550–557, doi: 10.1016/j.isprsjprs.2010.06.005, 2010.

30 Henderson, S. B., Beckerman, B., Jerrett, M., and Brauer, M.: Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter, Environ. Sci. Technol., 41, 2422–2428, doi:10.1021/es0606780, 2007.





Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, and J., Giovis, C.: A review and evaluation of intraurban air pollution exposure models, J. Expo. Anal. Env. Epid., 15, 185–204, doi:10.1038/sj.jea.7500388, 2005.

Jerrett, M., Donaire-Gonzalez, D., Popoola, O., Jones, R., Cohen, R. C., Almanza, E., de Nazelle, A., Mead, I., Carrasco-

5 Turigas, G., Cole-Hunter, T., Triguero-Mas, M., Seto, E., and Nieuwenhuijsen, M.: Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science, Environ. Res., 158, 286–294, doi:10.1016/j.envres.2017.04.023, 2017.

Johnson, M., Isakov, V., Touma, J. S., Mukerjee, S., and Özkaynak, H.: Evaluation of land-use regression models used to predict air quality concentrations in an urban area, Atmos. Environ., 44, 3660–3668, doi: 10.1016/j.atmosenv.2010.06.041,

10 2010.

Kim, S. Y., Yi, S. J., Eum, Y. S., Choi, H. J., Shin, H., Ryou, H. G., and Kim, H.: Ordinary kriging approach to predicting long-term particulate matter concentrations in seven major Korean cities, Environ. Health Toxicol., 29, e2014012, doi:10.5620/eht.e2014012, 2014.

Kingham, S., Longley, I., Salmond, J., Pattinson, W., and Shrestha, K.: Variations in exposure to traffic pollution while

15 travelling by different modes in a low density, less congested city, Environ. Pollut., 181, 211–218, doi:10.1016/j.envpol.2013.06.030, 2013.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di, S. S., Bell M, Norford, L., and Britter, R.: The rise of low-cost sensing for managing air pollution in cities, Environ. Int., 75, 199–205, doi:10.1016/j.envint.2014.11.019, 2015.

Lee, J. H., Wu, C. F., Hoek, G., De, H. K., Beelen, R., Brunekreef, B., and Chan, C. C.: Land use regression models for
estimating individual NO<sub>x</sub> and NO<sub>2</sub> exposures in a metropolis with a high density of traffic roads and population, Sci. Total
Environ., 472, 1163–1171, doi:10.1016/j.scitotenv.2013.11.064, 2014.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367–371, doi:10.1038/nature15371, 2015.

Li, J., and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: a review, Environ. Modell. Softw.,
53, 173–189. doi:10.1016/j.envsoft.2013.12.008, 2014.

- Mercer, L. D., Szpiro, A. A., Sheppard, L., Lindström, J., Adar, S. D., Allen, R. W., Avol, EL., Oron, A. P., Larson, T., Liu, L. J., and Kaufman, J. D.: Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NO<sub>x</sub>) for the multi-ethnic study of atherosclerosis and air pollution (MESA Air), Atmos Environ, 45, 4412–4420, doi:10.1016/j.atmosenv.2011.05.043, 2011.
- Miskell, G., Salmond, J., and Williams, D. E.: Low-cost sensors and crowd-sourced data: observations of siting impacts on a network of air-quality instruments, Sci. Total Environ., 575, 1119–1129, doi:10.1016/j.scitotenv.2016.09.177, 2017.
   Pang, W., Christakos, G., and Wang, J. F.: Comparative spatiotemporal analysis of fine particulate matter pollution, Environmetrics, 21, 305–317, doi:10.1002/env.1007, 2010.





Rice, M. B., Ljungman, P. L., Wilker, E. H., Dorans, K. S., Gold, D. R., Schwartz, J., Koutrakis, P., Washko, G. R., O'Connor, G. T., and Mittleman, M. A.: Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the framingham heart study, Am. J. Respir. Crit. Care. Med., 191, 656–64, doi:10.1164/rccm.201410-1875OC, 2015. Saraswat, A., Apte, J. S., Kandlikar, M., Brauer, M., Henderson, S. B., and Marshall, J. D.: Spatiotemporal land use regression

5 models of fine, ultrafine, and black carbon particulate matter in new Delhi, India, Environ. Sci. Technol., 47, 12903–12911, doi:10.1021/es401489h, 2013.

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., and Bartonova, A.: Mapping urban air quality in near realtime using observations from low-cost sensors and model information, Environ. Int., 106, 234–247, doi:10.1016/j.envint.2017.05.005, 2017.

Thompson, J. E.: Crowd-sourced air quality studies: a review of the literature & portable sensors, Trends in Environmental Analytical Chemistry, 11, 23–34, doi:10.1016/j.teac.2016.06.001, 2016.
 Team, R. D. C.: R: a language and environment for statistical computing. R foundation for statistical computing, R foundation for statistical computing, Vienna, Austria, Computing, 14, 12–21, 2009.
 Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., and Brunekreef, B. Systematic evaluation of land use regression

- models for NO<sub>2</sub>, Environ. Sci. Technol., 46, 4481–4489, doi:10.1021/es204183v, 2012.
   Xu, S., Zou, B., Shafi, S., and Sternberg, T.: A hybrid Grey-Markov/ LUR model for PM<sub>10</sub> concentration prediction under future urban scenarios, Atmos. Environ., 187, 401–409, doi:10.1016/j.atmosenv.2018.06.014, 2018.
   Zhai, L., Li, S., Zou, B., Sang, H., Fang, X., and Xu, S.: An improved geographically weighted regression model for PM<sub>2.5</sub> concentration in large areas, Atmos. Environ., 181, 145–154, doi:10.1016/j.atmosenv.2018.03.017, 2018.
- Zou, B., Luo, Y., Wan, N., Zheng, Z., Sternberg, T., and Liao, Y.: Performance comparison of LUR and ok in PM<sub>2.5</sub> concentration mapping: a multidimensional perspective, Sci. Rep., 5, 8698, doi:10.1038/srep08698, 2015.
   Zou, B., Pu, Q., Bilal, M., Weng, Q., Zhai, L., and Nichol, J. E.: High-resolution satellite mapping of fine particulates based on geographically weighted regression, IEEE Geosci. Remote. S., 13, 495–499, doi:10.1109/LGRS.2016.2520480, 2017.





| Code | Туре                | N  | Rules  |
|------|---------------------|----|--|
| 1    | Vertex point        | 5  | U1 = $\{X \mid X \in (Vertex point of sampling area \cap Landmark)\}.$   |
| 2    | Industrial park     | 28 | A2 = {X   X $\in$ ((Industrial park $\cup$ (Metal & cement & power industrial  |
|      |                     |    | agglomeration)) – High-tech industrial park)};   |
|      |                     |    | $U2 = \{X \mid X \text{ has the largest number of factories within its 100m buffer zone AND}$  |
|      |                     |    | $X \in A2$ .   |
| 3    | Dust surface        | 13 | A3 = {X   $X \in (POI \cap Dust surface)$ AND area of dust surface ranks in the top 4 of   |
|      |                     |    | each district};  |
|      |                     |    | U3 = {X   Distance between X $\geq$ 200m AND X $\in$ A3}.  |
| 4    | Depot               | 16 | $U4 = \{X \mid X \in (Coach \text{ station} \cap Railway \text{ station})\}.$  |
| 5    | Scenic area         | 27 | $A5 = \{X \mid X \in ((Park - Neighborhood park) \cap well-known scenic area)\};\$   |
|      |                     |    | U5 = {X   Distance between X>200m AND X $\in$ A5}.   |
| 6    | Hospital            | 11 | A6 = {X   X $\in$ (Hospital ranks in the top 3 of each district $\cup$ Children's hospital   |
|      |                     |    | $\cup$ Respiratory special hospital)};   |
|      |                     |    | U6 = {X   Distance between X>200m AND X $\in$ A6}.   |
| 7    | Residential area    | 12 | A7 = {X   Distance between X and U1 $\leq$ 200m OR Distance between X and  |
|      |                     |    | $U3 \le 200m, X \in \text{Residential area };$   |
|      |                     |    | U7 = {X   Distance between X>200m AND X $\in$ A7}.   |
| 8    | School              | 15 | $U8 = \{X \mid Distance between X and U1 \le 200 m OR$ |
|      |                     |    | U3 $\leq$ 200m, X $\in$ School, in order of priority:  |
|      |                     |    | Kindergarten>Primary>Secondary>Universities)}.   |
| 9    | Commercial area     | 9  | $U9 = \{X \mid X \text{ is the building with the highest Population density, } X \in Commercial$   |
|      |                     |    | area}.   |
| 10   | Other important POI | 8  | U10 = {X   X $\in$ (Corresponding sampling site of national monitoring site $\cup$   |
|      |                     |    | Background site $\cup$ Museum)}.   |
| 11   | Road                | 56 | A11 = {X   X $\in$ (Junction of (Expressway $\cup$ Main road))};   |
|      |                     |    | U11 = {X   X is $50/100$ meters away from A11 OR X $\in$ A11}.   |
| 12   | Supplementary point | 3  | U12 = {X   $X \in POI$ where four neighboring grids have no site}.   |





# Table 2. Description of potential predictor variables for LUR.

| GIS dataset   | Predictor Variables                      | Unit              | Buffer size (radius in meters) |  |  |
|---------------|--|-------------------|--------------------------------|--|--|
|               | Piling surface                           | %                 |                                |  |  |
|               | Construction surface                     | %                 |                                |  |  |
| Dust surfaces | Rolling trample surfaces                 | %                 | 50, 100, 200, 300, 500, 1000   |  |  |
|               | Bare surfaces                            | %                 |                                |  |  |
|               | Total                                    | %                 |                                |  |  |
| Pollution     | Inverse distance to nearest industries   |                   | NA                             |  |  |
| industries    | Industries density                       |                   | 50, 100, 200, 300, 500, 1000   |  |  |
|               | High-density residential area            | %                 |                                |  |  |
|               | Low-density residential area             | %                 |                                |  |  |
|               | Urban green land                         | %                 |                                |  |  |
| Land use      | Other built-up area                      | %                 | 50, 100, 200, 300, 500, 1000   |  |  |
|               | High-density forest                      | %                 |                                |  |  |
|               | Low-density forest                       | %                 |                                |  |  |
|               | Agricultural land                        | %                 |                                |  |  |
| Troffic       | Inverse distance to a nearest major road | NA                |                                |  |  |
| Traffic       | Road density                             |                   | 50, 100, 200, 300, 500, 1000   |  |  |
|               | Average wind speed                       | Meter/s           | NA                             |  |  |
| Mataaralagu   | Atmospheric pressure                     | Ра                | NA                             |  |  |
| Meteorology   | Relative humidity                        | %                 | NA                             |  |  |
|               | Temperature                              | Fahrenheit degree | NA                             |  |  |

# **Table 3.** Descriptive statistics of $PM_{2.5}$ concentration (µg m<sup>-3</sup>).

|          | •     | Mean              |                  | Max  |     | Min  |     | Standard deviation |       |
|----------|-------|-------------------|------------------|------|-----|------|-----|--------------------|-------|
|          |       | SAMP <sup>a</sup> | NAT <sup>b</sup> | SAMP | NAT | SAMP | NAT | SAMP               | NAT   |
|          | 8:00  | 69.67             | 39.8             | 128  | 58  | 36   | 27  | 18.81              | 10.46 |
|          | 9:00  | 72.97             | 36.9             | 132  | 54  | 30   | 20  | 17.04              | 10.97 |
| Period 1 | 10:00 | 73.08             | 38.5             | 113  | 58  | 28   | 21  | 15.57              | 11.57 |
|          | 11:00 | 74.12             | 39.4             | 106  | 54  | 30   | 27  | 13.96              | 8.78  |
|          | 12:00 | 76.45             | 41.2             | 136  | 53  | 44   | 29  | 14.55              | 8.68  |
|          | 14:00 | 167.91            | 188.3            | 220  | 207 | 145  | 165 | 14.43              | 14.48 |
|          | 15:00 | 165.75            | 182              | 227  | 206 | 133  | 153 | 16.68              | 17.06 |
| Period 2 | 16:00 | 162.72            | 178.7            | 212  | 201 | 115  | 149 | 15.96              | 16.91 |
|          | 17:00 | 167.69            | 177.8            | 266  | 209 | 136  | 146 | 18.92              | 20.49 |
|          | 18:00 | 171.89            | 182.1            | 250  | 219 | 132  | 149 | 21.5               | 22.4  |

a: sampling sites of the crowdsourcing sampling campaign.

5 b: national monitoring sites.







Figure 1: Principle and accuracy of measurement instrument.







Figure 2: Sampling area and PM<sub>2.5</sub> sampling sites.







Figure 3: Spatial variation of PM<sub>2.5</sub> concentration of sampling sites: (a) Period 1, (b) Period 2.



5 **Figure 4:** Box plots of hold-out validation R<sup>2</sup> between the observed and predicted concentrations of PM<sub>2.5</sub> for OK, LUR and UK with increasing training sites: (a) Period 1; (b) Period 2. The boundaries of the boxes indicate the 75th and 25th percentiles (Q3 and Q1). The line within the box marks the median (Q2), and the crosses are averages. The errors bars above and below indicate the highest datum (Q3+1.5IQR, IQR is the interquartile range, IQR=Q3-Q1) and the lowest datum (Q1-1.5IQR). Dots above and below the error bars indicate the outliers.







**Figure 5:** Scatterplots of repeated validating results with 90% training sites for (a) OK, Period 1; (b) LUR, Period 1; (c) UK, Period 1; (d) OK, Period 2; (e) LUR, Period 2; (f) UK, Period 2. The solid line is the 1:1 line as a reference.







**Figure 6:** Spatial distributions of PM<sub>2.5</sub> concentrations from crowdsourcing sampling sites and national monitoring sites. (a) Period 1, crowdsourcing; (b) Period 1, national monitoring; (c) Period 2, crowdsourcing; (d) Period 2, national monitoring.