



Identifying low-PM_{2.5} exposure commuting routes for cyclists through modeling with the random forest algorithm based on low-cost sensor measurements in three Asian cities



Tzong-Gang Wu ^{a,b}, Yan-Da Chen ^{a,c}, Bang-Hua Chen ^d, Kouji H. Harada ^c, Kiyoung Lee ^e, Furong Deng ^f, Mark J. Rood ^g, Chu-Chih Chen ^h, Cong-Thanh Tran ^{i,j}, Kuo-Liong Chien ^j, Tzai-Hung Wen ^k, Chang-Fu Wu ^{a,b,*}

^a Institute of Environmental and Occupational Health Sciences, College of Public Health, National Taiwan University, No. 17, Xuzhou Rd, Taipei, 10055, Taiwan

^b Innovation and Policy Center for Population Health and Sustainable Environment, College of Public Health, National Taiwan University, No. 17, Xuzhou Rd, Taipei, 10055, Taiwan

^c Department of Health and Environmental Sciences, Kyoto University Graduate School of Medicine, Kyoto University, Yoshida-konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan

^d Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, No. 17, Xuzhou Rd, Taipei, 10055, Taiwan

^e Department of Environmental Health Sciences, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, South Korea

^f Department of Occupational and Environmental Health Sciences, School of Public Health, Peking University, No. 38 Xueyuan Road, Beijing, 100191, China

^g Department of Civil and Environmental Engineering, University of Illinois, 205 N. Mathews Ave., Urbana, IL, 61801, USA

^h Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County, 35053, Taiwan

ⁱ University of Science, Vietnam National University Ho Chi Minh City, 227 Nguyen Van Cu Street, Dist. 5, Ho Chi Minh City, Viet Nam

^j Institute of Epidemiology and Preventive Medicine, National Taiwan University, No. 17, Xuzhou Rd, Taipei, 10055, Taiwan

^k Department of Geography, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan

ARTICLE INFO

ABSTRACT

Keywords:

PM_{2.5}
Cyclist
Routing
Asian city
Random forest
Spatial autocorrelation

Cyclists can be easily exposed to traffic-related pollutants due to riding on or close to the road during commuting in cities. PM_{2.5} has been identified as one of the major pollutants emitted by vehicles and associated with cardiopulmonary and respiratory diseases. As routing has been suggested to reduce the exposures for cyclists, in this study, PM_{2.5} was monitored with low-cost sensors during commuting periods to develop models for identifying low exposure routes in three Asian cities: Taipei, Osaka, and Seoul. The models for mapping the PM_{2.5} in the cities were developed by employing the random forest algorithm in a two-stage modeling approach. The land use features to explain spatial variation of PM_{2.5} were obtained from the open-source land use database, OpenStreetMap. The total length of the monitoring routes ranged from 101.36 to 148.22 km and the average PM_{2.5} ranged from 13.51 to 15.40 µg/m³ among the cities. The two-stage models had the standard k-fold cross-validation (CV) R² of 0.93, 0.74, and 0.84 in Taipei, Osaka, and Seoul, respectively. To address spatial autocorrelation, a spatial cross-validation approach applying a distance restriction of 100 m between the model training and testing data was employed. The over-optimistic estimates on the predictions were thus prevented, showing model CV-R² of 0.91, 0.67, and 0.78 respectively in Taipei, Osaka, and Seoul. The comparisons between the shortest-distance and lowest-exposure routes showed that the largest percentage of reduced averaged PM_{2.5} exposure could reach 32.1% with the distance increases by 37.8%. Given the findings in this study, routing behavior should be encouraged. With the daily commuting trips expanded, the cumulative effect may become significant on the chronic exposures over time. Therefore, a route planning tool for reducing the exposures shall be developed and promoted to the public.

* Corresponding author. Institute of Environmental and Occupational Health Sciences, College of Public Health, National Taiwan University, No. 17 Xuzhou Road, Room 717, Taipei, 10055, Taiwan.

E-mail address: changfu@ntu.edu.tw (C.-F. Wu).

1. Introduction

Particulate matter with a diameter less than or equal to 2.5 μm ($\text{PM}_{2.5}$) is associated with increased morbidity and mortality for cardiopulmonary diseases and lung cancer (Lo et al., 2017; Pope et al., 2009; Pope et al., 2020). For example, long-term exposure to $\text{PM}_{2.5}$ may increase the risk of mortality affecting arrhythmias, ischemic heart disease, and heart failure, while short-term exposure to $\text{PM}_{2.5}$ is found to have an association with the development of arrhythmias, myocardial infarction, thrombosis, and stroke (Zhao et al., 2021). In addition, nasal inflammation (Sun et al., 2021), alternation in lung function (Edginton et al., 2021; Sun et al., 2021), oxidative stress, inflammation, damage to human respiratory system tissue and DNA (Niu et al., 2020) are also related to $\text{PM}_{2.5}$ exposures. One major source of $\text{PM}_{2.5}$ in urban cities is vehicular emissions (Jaiprakash and Habib, 2017; Requia et al., 2018). To reduce these emissions, cycling facilities have been developed, and cycling has become a popular mode of commuting in American, European, and Asian cities. However, compared with commuters using other transportation modes, cyclists are exposed to a higher dose of inhaled air pollutants as a result of cycling on or close to roadways, with an increased respiratory rate during their commute. This can lead to increased health risks (Apparicio et al., 2016; Cole et al., 2018; Lee and Sener, 2019).

To decrease commuter exposure to $\text{PM}_{2.5}$, routing has been suggested in some studies (Bigazzi et al., 2016; Li et al., 2017; Mörter and Lindley, 2015; Ma et al., 2020a). Thus, exploring $\text{PM}_{2.5}$ spatiotemporal variation in cities is necessary. To describe the spatial variability of pollutants, the land use model based on the regression technique (LUR) is the widely used empirical approach (Chen et al., 2019; Hankey and Marshall, 2015b). Conventional LUR studies have conducted measurement at a limited number of fixed sites, although other studies have developed models using mobile monitoring platforms to map pollutant exposure with high spatial and temporal resolution (Hankey and Marshall, 2015b; Hankey et al., 2019; Lim et al., 2019; Minet et al., 2017; Shi et al., 2016).

In terms of modeling techniques, the LUR model is based on the multivariate linear regression algorithm because of the algorithm's simplicity, but the model relies on several assumptions, such as non-collinearity among the independent variables and the predetermined relationship between the independent and dependent variables (Xu et al., 2018). Machine learning (ML) algorithms, such as the random forest (RF) algorithm, do not rely on the aforementioned assumptions but can still provide accurate predictions. The RF algorithm modifies bagging techniques to construct a large collection of decorrelated regression or decision trees. Each tree is constructed with several nodes, and among a subset of randomly chosen predictors, the best-determined split for each node is used. The final predictions are the averaged results of all decision trees (Breiman, 2001; Hastie et al., 2009; Xu et al., 2018). The RF algorithm also provides normalized scores for ranking the effectiveness of predictors through an evaluation of the importance of variables for reducing prediction errors (Chen et al., 2019; Hu et al., 2017; Rahman et al., 2020). This additional information provides users with an alternative aspect for model selection. Thus, the RF algorithm has become widely used and has demonstrated the ability to generate accurate predictions for $\text{PM}_{2.5}$ (Chen et al., 2019; Lim et al., 2019; Rahman et al., 2020). For example, Lim et al. (2019) compared the prediction ability of LUR and land use models based on RF (LURF) with mobile measurements of $\text{PM}_{2.5}$ exposure on the streets of Seoul, South Korea. Their results indicated that the LURF model could provide more accurate predictions of $\text{PM}_{2.5}$ than the LUR model. However, ML algorithms, such as that of RF, cannot address spatial autocorrelation (SAC) in the model building process (Chen et al., 2019). SAC is the relation of each data point to another based on their distance, which may lead to an optimistically biased prediction when evaluating performance through standard cross-validation (CV) (Pohjankukka et al., 2017). Hence, SAC cannot be ignored in studies aiming to map targets.

This study identified alternative routes for cyclists to reduce their

exposure to $\text{PM}_{2.5}$ during commuting. Because the urban air quality tends to be lower in Asian cities than those in European and American cities (Kumar et al., 2018; Wu et al., 2021), the monitoring campaigns in this study were conducted in the Asian cities of Taipei, Osaka, and Seoul. To account for SAC, the models were evaluated using a spatial k-fold CV approach. As described by Pohjankukka et al. (2017), this approach restricts the distance between model building and data validation to minimize SAC (Pohjankukka et al., 2017). In coordinated field campaigns, measurements were conducted using monitoring devices carried by assistants who cycled on predetermined monitoring routes. As the low-cost sensors have a lower price (<\$300) than the reference instrumentation, to increase the applicability and accessibility for different cities, low-cost sensors were employed to provide reliable measurements with quality control procedures (Lim et al., 2019; McKercher and Vanos, 2018; Popoola et al., 2018). Finally, model-based exposures along the shortest distance and alternative routes were simulated and compared in each city.

2. Materials and methods

2.1. Study areas

The monitoring campaigns were conducted in the following three Asian cities (Fig. 1): Taipei (area = 271.8 km^2 ; population density = 9724/ km^2), Osaka (area = 225.2 km^2 ; population density = 12,097/ km^2), and Seoul (area = 605.2 km^2 ; population density = 16,239/ km^2) (City of Osaka, 2018; Directorate-General of Budget Accounting and Statistics Taiwan, 2020; Seoul Metropolitan Government, 2021). Seoul has the highest population density and Taipei has the lowest. The downtown area of each city, ranging between 26.5 and 46.0 km^2 , served as the study area. In Taipei, the sources of $\text{PM}_{2.5}$ included vehicular emissions, transported regional, secondary pollution sources, industrial emissions, biomass burning, oil combustion, and soil resuspension (Kuo et al., 2014; Liao et al., 2021). Secondary pollution, vehicular emissions, industrial emissions, heavy oil combustion, soil and dust resuspension, as well as incinerators, were identified as the fine particle sources in Osaka (Funasaka et al., 2020). In Seoul, several $\text{PM}_{2.5}$ sources were identified, such as secondary pollution sources, vehicular emissions, biomass burning, soil, roadway emissions, as well as coal and oil combustions (Park et al., 2020).

2.2. Data collection and integration

All monitoring campaigns were completed within two weeks in Taipei (June 12 to 19, 2018), Osaka (October 22 to November 1, 2018), and Seoul (October 24 to 29, 2019). The measurements were conducted during "rush hours" (i.e., 07:00–09:00 and 16:00–18:00) on weekdays. In each time period, the on-road $\text{PM}_{2.5}$ of two routes was monitored, with a total of 20 monitoring network routes covering both major and minor roads. The total monitoring distances were 101.36, 148.22, and 134.77 km in Taipei, Osaka, and Seoul, respectively. In addition, the monitoring network routes were also predetermined across network areas of ambient air quality monitoring stations (AQMSs; Fig. 2).

The on-road $\text{PM}_{2.5}$, temperature, and relative humidity (RH) measurements with 10-s intervals were conducted using low-cost sensors (AirVisual Node [AN], IQAir) carried by the study cyclists. The cycling speed varied with traffic conditions (at around 15 km/h). Every on-road $\text{PM}_{2.5}$ measurement was adjusted for RH effects (Wu et al., 2005) and calibrated using a gravimetric correction equation. To retrieve the gravimetric calibration parameters, a total of 208 24-h $\text{PM}_{2.5}$ filter samples were collected using three Harvard impactors (Air Diagnostics and Engineering Inc.) colocated with three ANs in Taipei. The detailed procedure for the RH effect adjustment, gravimetric correction, and other quality control of AN $\text{PM}_{2.5}$ is described in the Supplementary data (Supplementary Sections 1 and 2). Feenstra et al. (2019) evaluated AirVisual sensor performance against hourly $\text{PM}_{2.5}$ measurements

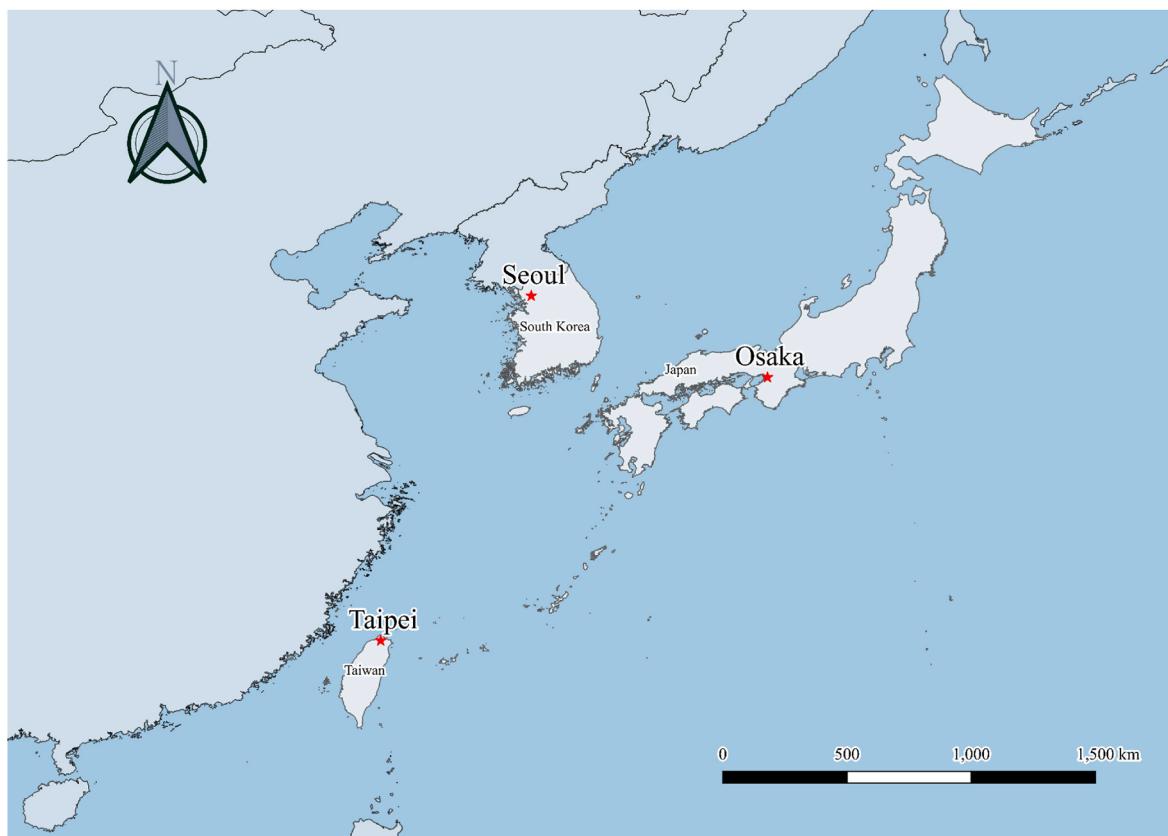


Fig. 1. Location of study cities. Map source: Global Administrative Areas (2020), University of California, Berkeley (<http://www.gadm.org>).

obtained using the Met One Beta-Attenuation Mass Monitor based on the US EPA designated Federal Equivalent Method (EQPM-0308-170). The results revealed that R^2 , slopes, and intercepts ranged from 0.69 to 0.72, 0.70 to 1.30, and -4.0 to 4.0, respectively; the AirVisual sensors demonstrated a moderate correlation with the reference instruments. Moreover, the sensors had an accuracy of approximately 86% in reference to filter measurements (Zamora et al., 2020). Against daily filter measurements, the sensors in this study had an R^2 of 0.92, a slope of 0.71, and an intercept of 3.01, which were comparable to the results reported in the aforementioned studies.

The GPS coordinates were collected using smartphone applications (Green Tracks for the Android system [Wang, 2021] and GPS Status—Record Your Track for the iOS system [Beijing Zhiyue Information Technology Co.Ltd, 2021]). To reduce data noise but to still obtain an adequate number of measurements in every city, the adjusted AN measurements were averaged every 1 min and then 100-m road segment. A spatiotemporal resolution similar to that of Lim et al. (2019) was applied in this study. To adjust for temporal variation, hourly ambient PM_{2.5} measurements were obtained from four governmental AQMSs in Taipei (Taiwan Environmental Protection Administration, 2019), nine in Osaka (Osaka Prefectural Government, 2019), and five in Seoul (AirKorea, 2018) (Fig. 2).

Land use features were retrieved at the centroid of every 100-m road segment in different buffer sizes; the buffer radiiuses were 25, 50, 100, 200, and 500 m. The land use features were downloaded from the OpenStreetMap (OSM) online database (GEOFABRIK, OpenStreetMap Contributors, 2021). OSM is a crowdsourced geographic information system database maintained by numerous volunteers around the world (Lautenschlager et al., 2020; Lim et al., 2019; OpenStreetMap contributors, 2021). The database has been utilized for multiple purposes, e.g., the LUR studies (Hatzopoulou et al., 2013a; Lim et al., 2019; Liu et al., 2019; Shi et al., 2016). Based on the commonality of features among the cities (Table S1), 13 OSM feature types were applied in this study,

namely the surface areas of residential (Residential), industrial (Industrial), school (School), commercial (Commercial), leisure (Leisure), temple (Temples), public amenity (PublicAmenity), natural and open (NaturalOpen), and urban green and wide space areas (UrbanGreen); the lengths of all roads (ARL) and major roads (MRL); and the number of counted traffic signs and facilities (Traffic_C), and shipping ports (Port_C). Because of a lack of ports within the buffer size of 25 m along the monitoring routes in all cities, Port_C in the 25-m buffer was removed during model development. Consequently, with five buffer sizes, 13 feature types, and one removed feature, a total of 64 land use features were applied to develop the models in this study. Table S2 lists the descriptive statistics of the land use features in the buffer with a radius of 500 m. All the GIS analysis was performed using QGIS software (v3.14.15-Pi).

2.3. Modeling approach

To estimate the short-term PM_{2.5} exposure and to control for temporal variation, a two-stage modeling approach (Chen et al., 2012) that separates temporal trends and the spatial variation of the pollutants within the study area was adopted. The spatiotemporal variation of PM_{2.5} was explained separately in two stages. The flowchart was shown in Fig. S3. In the first stage, the ambient model is developed based on the averaged ambient PM_{2.5} measurements from the AQMSs in each city to control for temporal variation, and the model can be expressed as follows:

$$\text{On-road PM}_{2.5}(s_{ij}) = \beta_0 + \beta_1 \text{Ambient}(t_i) + e_{\text{Ambient},ij}, \quad (1)$$

Where On-road PM_{2.5}(s_{ij}) denotes the PM_{2.5} measurement collected along the routes at the location of s_{ij} at the hour i and the minute j and Ambient(t_i) denotes the ambient PM_{2.5} measurement at time t_i. β_0 and β_1 are the intercept and slope of the ambient linear model, respectively, and e_{Ambient,ij} is the residual values of the ambient model.

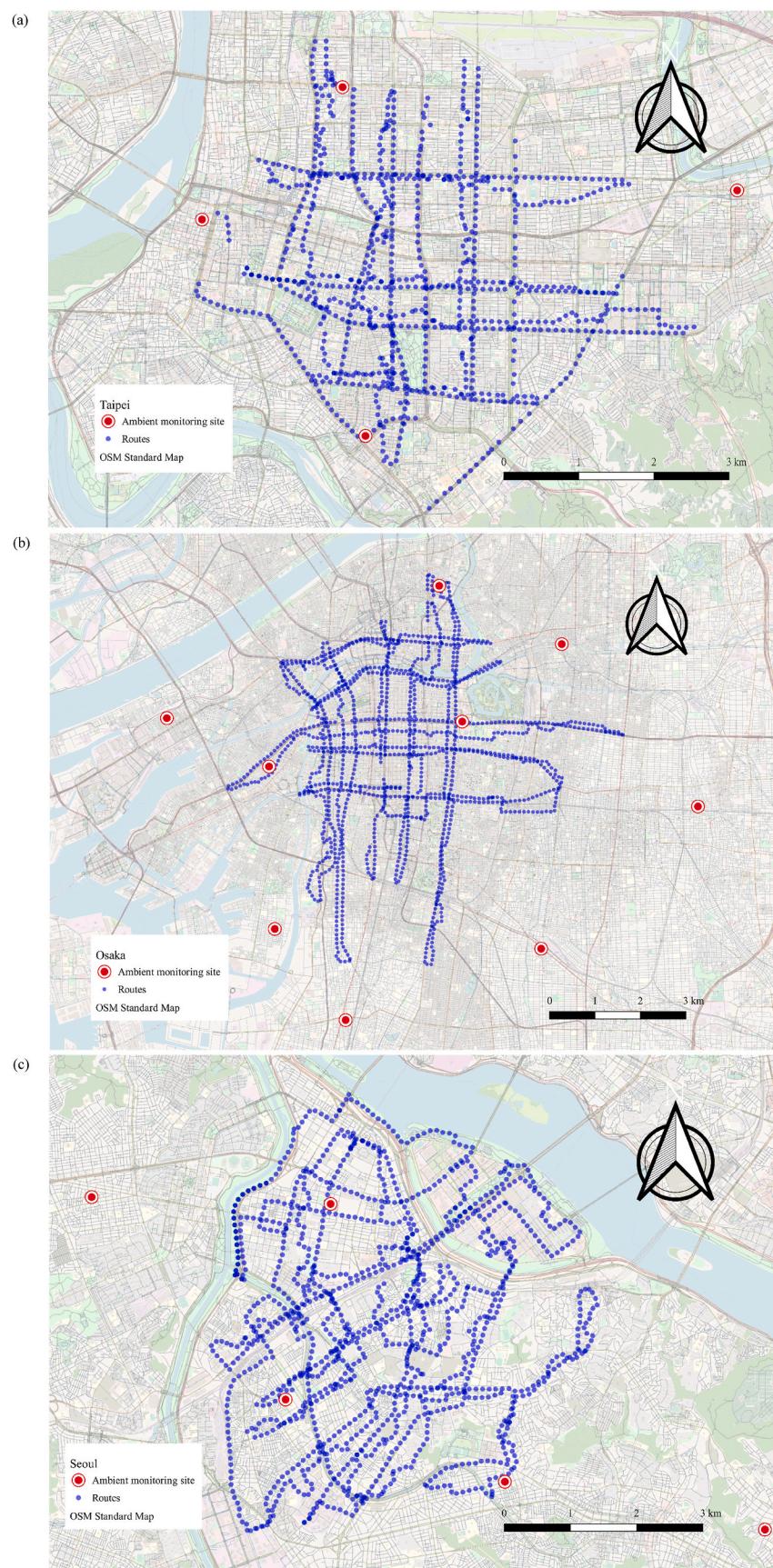


Fig. 2. Network routes in the cities of (a) Taipei, (b) Osaka, and (c) Seoul.

Fig. 2. Network routes in the cities of (a) Taipei, (b) Osaka, and (c) Seoul.

After controlling for temporal variation, in the second stage, a spatial model is applied for $e_{\text{Ambient},ij}$ of the ambient model at time t_{ij} to account for the remaining spatial variation. The spatial model is developed based on the land use features using the RF algorithm and is written as follows:

$$e_{\text{Ambient},ij} = \text{LURF}(s_{ij}) + e_{\text{RF},ij}, \quad (2)$$

Where LURF represents land use random forest model; $\text{LURF}(s_{ij})$ is the predicted $e_{\text{Ambient},ij}$ of the ambient model based on the land use features at location s_{ij} (the centroid of the road segment), and $e_{\text{RF},ij}$ is the residual of the LURF model. Finally, to predict on-road PM_{2.5}, two-stage model prediction is performed by summing the predictions in these two stages (Equation (1) + Equation (2)). The importance scores of the land use features are also calculated in the second stage (LURF model) to identify the major features influencing on-road PM_{2.5} in these cities.

2.4. Model evaluation

R^2 and root mean square error (RMSE) were computed to evaluate model performance based on the model's predicted values against the true measurements. Two types of CVs were conducted to evaluate the generalization and prediction power of the models. First, standard k-fold CV was applied. The dataset was randomly separated into k subsets. The $k - 1$ of the subsets was used to develop the model, and the remaining dataset was used in the testing stage of the developed model. Second, to address potential SAC, based on the method of Poljankukka et al. (2017), a spatial k-fold CV approach was employed. In this approach, training data points that were less than the designated distance from the testing point were omitted, thus minimizing the SAC between the training and testing samples in the CV process. All analyses were performed using Python (v3.85), and the RF models were developed using the *RandomForestRegressor* functions in the *scikit-learn* package.

2.5. Simulation for identifying the lowest-exposure routes

To identify the lowest-exposure routes, the two-stage model was developed using the ambient PM_{2.5} median value (for Equation (1)) in each city to create concentration contour maps. The origins and destinations of the routes were determined following the identification of selected PM_{2.5} hot zones in the city, ensuring that the shortest-distance route crossed into the hot zone. To evaluate the cumulative exposure along the routes, path integrated concentration (PIC) was calculated as the sum of the PM_{2.5} of each road segment multiplied by 100 m (= 0.1 km) with a unit of $\mu\text{g}/\text{m}^3\text{-km}$. A similar method has been described in Hatzopoulou et al. (2013a). Pairs of the shortest-distance and lowest-exposure routes were simulated using the "shortest path (point to point)" tool in QGIS. To generate the lowest-exposure route, a cost factor calculated as 1/concentration was utilized as the proxy for "speed" along the roads using the tool's "fastest" option.

3. Results and discussion

3.1. On-road PM_{2.5} measurements

The total number of on-road PM_{2.5} measurements was 972, 1,458, and 1,314, and the average PM_{2.5} (standard deviation) was 13.51 (8.37), 12.87 (4.21), and 15.40 (5.41) $\mu\text{g}/\text{m}^3$ in Taipei, Osaka, and Seoul, respectively (Fig. 3). Limited studies have explored the personal PM_{2.5} exposure of cyclists in Asia. One Taipei study revealed that cyclists were exposed to higher levels of PM_{2.5}, with the average level being 23.3 $\mu\text{g}/\text{m}^3$ (Wu et al., 2021), than those observed in the current study, which could be explained by seasonal variation (i.e., winter vs summer, respectively). In addition, no study has previously assessed the personal PM_{2.5} exposure of cyclists in Osaka and Seoul. Studies in Singapore (Tran et al., 2020) and Beijing, China (Kumar et al., 2018) have reported that cyclists had an average PM_{2.5} exposure of 17.7 and 49.3 $\mu\text{g}/\text{m}^3$, respectively. For cyclists in European cities, the mean or median PM_{2.5}

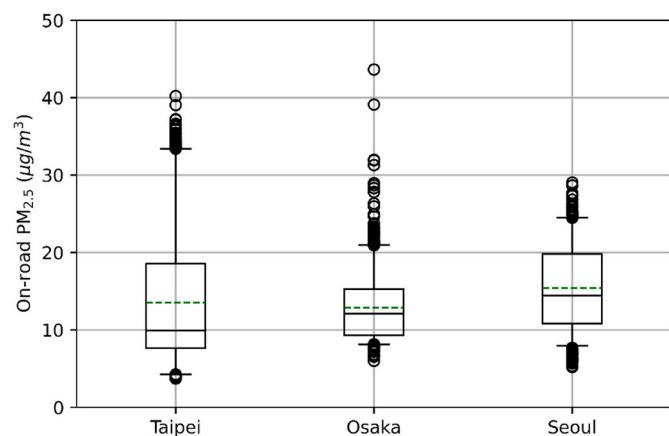


Fig. 3. On-road PM_{2.5} concentration distribution in different cities. The upper, median, and lower solid lines of the box indicate the 25th percentile (Q1), median, and 75th percentile (Q3), respectively. The dashed line in the box represents the mean value. Whiskers (error bars) above and below the box are the 95th and 5th percentiles; the circles represent the outliers that were out of the range between 5th and 95th percentiles.

exposure ranged from 6.3 to 83.4 $\mu\text{g}/\text{m}^3$ in Barcelona (Spain), Lisbon (Portugal), Helsinki (Finland), Rotterdam (the Netherlands), Thessaloniki (Greece), Vienna (Austria), and Milan (Italy) (Cole-Hunter et al., 2016; Correia et al., 2020; de Nazelle et al., 2012; Okonkwo et al., 2017; Ozgen et al., 2016; Ramos et al., 2017; Strasser et al., 2018). The results indicated that over the last decade, PM_{2.5} levels have varied considerably in different European cities. In American cities, the average PM_{2.5} exposure of cyclists ranged from 5 to 17.5 $\mu\text{g}/\text{m}^3$, representing a narrower range and generally lower exposure than those in European and Asian cities (Cole et al., 2018; Good et al., 2016; Ham et al., 2017; Hankey and Marshall, 2015a; Hatzopoulou et al., 2013b; Quiros et al., 2013; Weichenthal et al., 2011).

3.2. Ambient model

Ambient measurements were highly correlated with on-road PM_{2.5}, with correlation coefficients (r) of 0.92, 0.72, and 0.76 for Taipei, Osaka, and Seoul, respectively. This high correlation between PM_{2.5} measurements in each city indicated that the ambient PM_{2.5} can be utilized to adjust for temporal variation in the modeling approach. Accordingly, the ambient model as the first-stage model was developed using Equation (1) and evaluated through 5-fold CV in each city. As detailed in Table 1, the 5-fold CV-R² was 0.85, 0.53, and 0.58, and the CV-RMSE was 4.30, 2.95, and 3.53 $\mu\text{g}/\text{m}^3$ in Taipei, Osaka, and Seoul,

Table 1

CV performance of model in different stages with or without a distance restriction of 100 m. All indicators were calculated based on the predicted values against measured values.

	R ²			RMSE		
	Taipei	Osaka	Seoul	Taipei	Osaka	Seoul
Ambient Model (First stage)						
5-fold CV ^a	0.85	0.53	0.58	4.30	2.95	3.53
LURF Model (Second stage)						
5-fold CV	0.55	0.42	0.61	2.37	2.26	2.28
50-fold CV ^b	0.56	0.46	0.64	2.21	2.19	2.19
50-fold CV-D100M ^c	0.43	0.31	0.50	2.50	2.46	2.54
Two-stage Model						
50-fold CV	0.93	0.74	0.84	2.21	2.19	2.19
50-fold CV-D100M	0.91	0.67	0.78	2.50	2.46	2.54

^a 5-fold CV is the 5-fold CV without a distance restriction of 100 m.

^b 50-fold CV is the 50-fold CV without a distance restriction of 100 m.

^c 50-fold CV-D100M is the 50-fold CV with a distance restriction of 100 m.

respectively. As expected, the ambient models could be used to control for temporal variation in the modeling approach in each city. However, because air pollutants are considered spatiotemporal-dependent and the ambient model cannot provide spatial information, the LURF model as the second-stage model was required to explain spatial variation in each city.

3.3. Spatial model

To explore spatial variation in the cities, the LURF models were developed using Equation (2) in the second stage. As presented in Table 1, the 5-fold CV-R² and CV-RMSE of the LURF models ranged from 0.42 to 0.61 and 2.26 to 2.37 $\mu\text{g}/\text{m}^3$, respectively, among the cities. To address the SAC effects on the training and testing samples, spatial k-fold CV was applied. Based on preliminary sensitivity analysis, a k of 50 and a spatial restriction distance of 100 m were chosen to ensure an appropriate number of training samples in the CV procedure. Because some routes in the network intersected, the spatial distance between the centroids in the road segments could be less than the designated 100 m. Fig. S4 depicts an example of a training and testing dataset pair in the CV procedure with and without the 100-m distance restriction. Initially, in the standard 50-fold CV, the average number of samples for the training model was 953, 1,429, and 1288 in Taipei, Osaka, and Seoul,

respectively. In the CV procedure with the distance restriction, an average of 51, 81, and 71 training samples were omitted, resulting in average sample sizes of 902, 1,348, and 1217 for the training models in Taipei, Osaka, and Seoul, respectively. The average number of testing samples in Taipei, Osaka, and Seoul was 19, 29, and 26, respectively, and remained the same in the two CV procedures.

As described in Table 1, for the LURF model without the distance restriction (50-fold CV), the CV-R² was 0.56, 0.46, and 0.64, and the CV-RMSE was 2.21, 2.19, and 2.19 $\mu\text{g}/\text{m}^3$ in Taipei, Osaka, and Seoul, respectively. With the distance restriction, such optimistic estimates on the prediction performance were prevented. When the distance restriction was set (50-fold CV-D100M), the CV-R² was 0.43, 0.31, and 0.50, and the CV-RMSE was 2.50, 2.46, and 2.54 $\mu\text{g}/\text{m}^3$ in Taipei, Osaka, and Seoul, respectively. The decrements observed for the CV-R² ranged 0.13 and 0.15, and the increments of the CV-RMSE ranged from 0.27 to 0.35 $\mu\text{g}/\text{m}^3$.

To evaluate the influence of predictors on spatial variation, the variable importance scores were examined and are detailed by buffer size and category in Fig. 4. In the smallest buffer size (25 m), the all-road length (ARL) or major-road length (MRL) in each city were the major land use features that had the highest scores, ranging between 1.1% and 2.8%, compared with the other features. This indicated that the road length features had a dominant effect on on-road PM_{2.5} within this

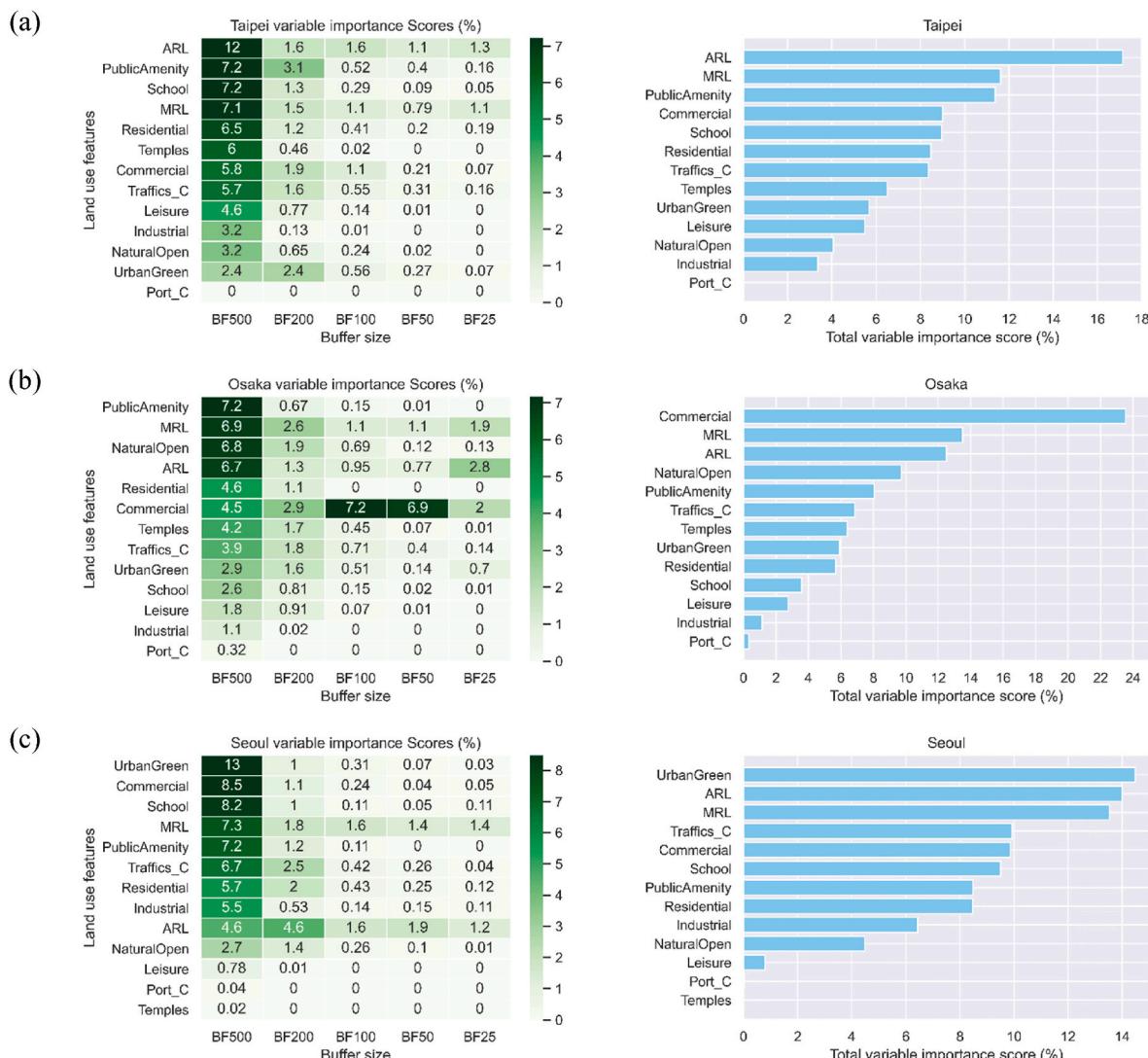


Fig. 4. Variable importance scores and total variable importance scores categorized by land use features in all buffer sizes in the city of (a) Taipei, (b) Osaka, and (c) Seoul.

small-scale area in the city. In general, as the buffer size increased, the importance of the features also increased. With a 500-m buffer size, the road lengths had scores ranging from 4.6% to 12% among the cities. Given the findings, after controlling for temporal variation, the on-road PM_{2.5} was heavily influenced by the land use features in large spaces. Similar buffer sizes have been reported by Shi et al. (2016) and Lim et al. (2019). The bar charts at the right in Fig. 4 presents the total scores classified by land use features in all buffer sizes, and apart from the road length features, the features with total scores over 10% were residential, commercial, and urban green areas. This implied that other than the road length, the major features influencing PM_{2.5} vary in different cities.

3.4. Two-stage model

Following the summing of predictions in the ambient and spatial models, on-road PM_{2.5} predictions were generated in the two-stage modeling approach. As listed in Table 1, the 50-fold CV with the distance restriction of 100 m (50-fold CV-D100M) for the two-stage model generated CV-R² values of 0.91, 0.67, and 0.78 and CV-RMSE values of 2.50, 2.46, and 2.54 µg/m³ in Taipei, Osaka, and Seoul, respectively. Compared with 50-fold CV performance, the 50-fold CV-D100M performance of the two-stage model exhibited decrements of 0.02, 0.07, and 0.06 for the CV-R² and increments of 0.29, 0.27, and 0.35 µg/m³ for the CV-RMSE in Taipei, Osaka, and Seoul, respectively. Use of the two-stage modeling approach with the distance restriction provided less biased performance estimates. ML studies that did not account for SAC in the prediction of air pollutants such as PM_{2.5}, ultrafine particles (UFPs), and black carbon using mobile measurements have reported CV-R² values ranging 0.53 to 0.80 (Krecl et al., 2019; Lim et al., 2019; Weichenthal et al., 2016); these estimates may be optimistic. Regarding a study that addressed SAC, Adams et al. (2020) employed multiple ML algorithms to predict on-road PM_{2.5} using measurements collected by cyclists in the US city of Charlotte, North Carolina. The results indicated a normal 5-fold CV-RMSE of 2.59 µg/m³, with a CV-RMSE that increased to 4.09 µg/m³ after SAC was accounted for.

The averaged on-road and predicted PM_{2.5} measurements on the routes were further calculated, as presented in Fig. 5. The R² was higher than 0.87, and the CV-RMSE was lower than 1.97 µg/m³, representing the reliability in average PM_{2.5} exposure estimation along different routes.

3.5. Comparison of the shortest-distance and lowest-exposure routes

Fig. 6 illustrates five example pairs of shortest-distance (red lines) and lowest-exposure routes (green lines) in the routing simulations based on two-stage models; these models were developed using the ambient PM_{2.5} median values of 10.50, 12.75, and 14.83 µg/m³ in Taipei, Osaka, and Seoul, respectively. The shortest-distance and lowest-exposure routes ranged from 1.65 to 5.32 km and 1.67 km to 5.61 km,

respectively; their average exposure ranged from 9.28 to 17.03 µg/m³ and 7.90 to 15.08 µg/m³, respectively (Table 2). To account for accumulative exposure, the PICs were calculated. The minimum and maximum PICs were 16.06 and 71.46 µg/m³-km on the shortest-distance routes, respectively, and 15.33 and 63.75 µg/m³-km on the lowest-exposure routes, respectively. The largest reduction in average exposure and its related PIC was 3.79 µg/m³ and 3.56 µg/m³-km, representing reduction percentages of 32.1% and 7.9%, respectively. Correspondingly, the largest increment of routing distance was 1.43 km (37.8% increase) with a 7.9% of PIC reduction. These findings support the effectiveness of the routing method for reducing exposure, even with a distance increment of more than 30%. Similarly, another study reported that cycling on alternative routes could significantly reduce the levels of PM_{2.5} exposure from 21.75% to 28.10% (Wu et al., 2021). Li et al. (2017) also stated that cycling on the cleanest designated route instead of the typical daily route could reduce PM_{2.5} dosage by 50%. Nevertheless, their results were estimated using an inverse distance weighting model based on AQMS measurements, which may be challenged to represent commuters' exposure. Elford and Adams (2019) applied a modeling approach to evaluate the simulated UFP dosage for school children on the shortest-distance and lowest-exposure routes. The results revealed a mean UFP dosage reduction of 4.71% with a mean distance increment of 3.08% on the lowest-exposure routes. Another study in school children demonstrated that routing can reduce NO₂ dosage by as much as 36.94%, with a 5.26% increase in commuting distance (Ma et al., 2020b). These findings are consistent with those of a review article that recommended routing despite the increased travel time compared with the shorter route with higher exposure (Tainio et al., 2021).

3.6. Limitations

The strength of this study is the design of the modeling with inputs from low-cost sensors, routine ambient air quality measurements, and the crowdsourced land used features to enhance the applicability of common protocol across cities. However, one of the main limitations in this study is that the measurements were conducted during a short period in a city. All measurements were conducted during morning and afternoon rush hours within two weeks in different countries. The field studies were also conducted in the months with relatively good air quality. The corresponding monthly average PM_{2.5} concentrations at the AQMS for the field campaigns (10.94, 9.31, and 15.48 µg/m³ in Taipei, Osaka, and Seoul, respectively) were lower than the annual means (14.09, 12.77, and 26.05 µg/m³). For a comprehensive long-term exposure assessment for cyclists, seasonal measurements must be obtained in future studies. Second, because the OSM data are maintained voluntarily, more data are available for some sampling locations than for others along the routes (Lim et al., 2019). For example, previous studies have shown that using 3D parameters can improve the model

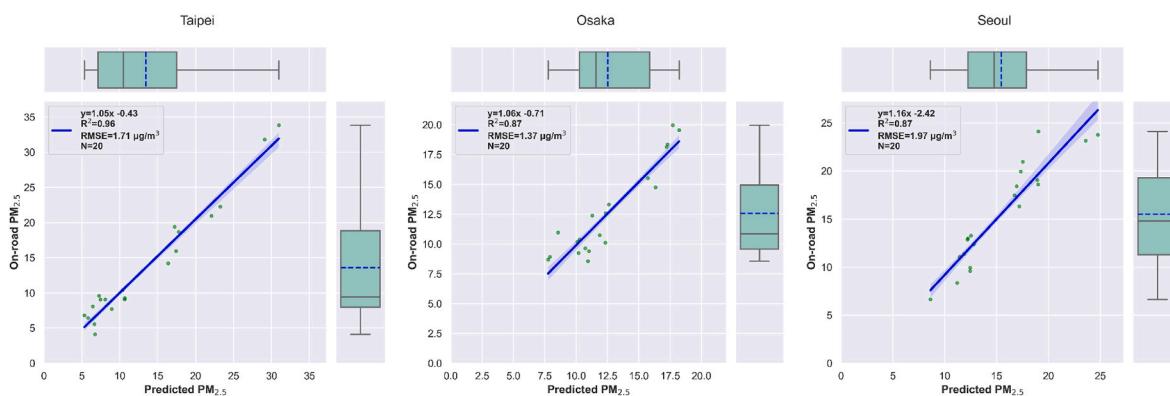


Fig. 5. Two-stage model performance for average route exposure in the city of (a) Taipei, (b) Osaka, and (c) Seoul.

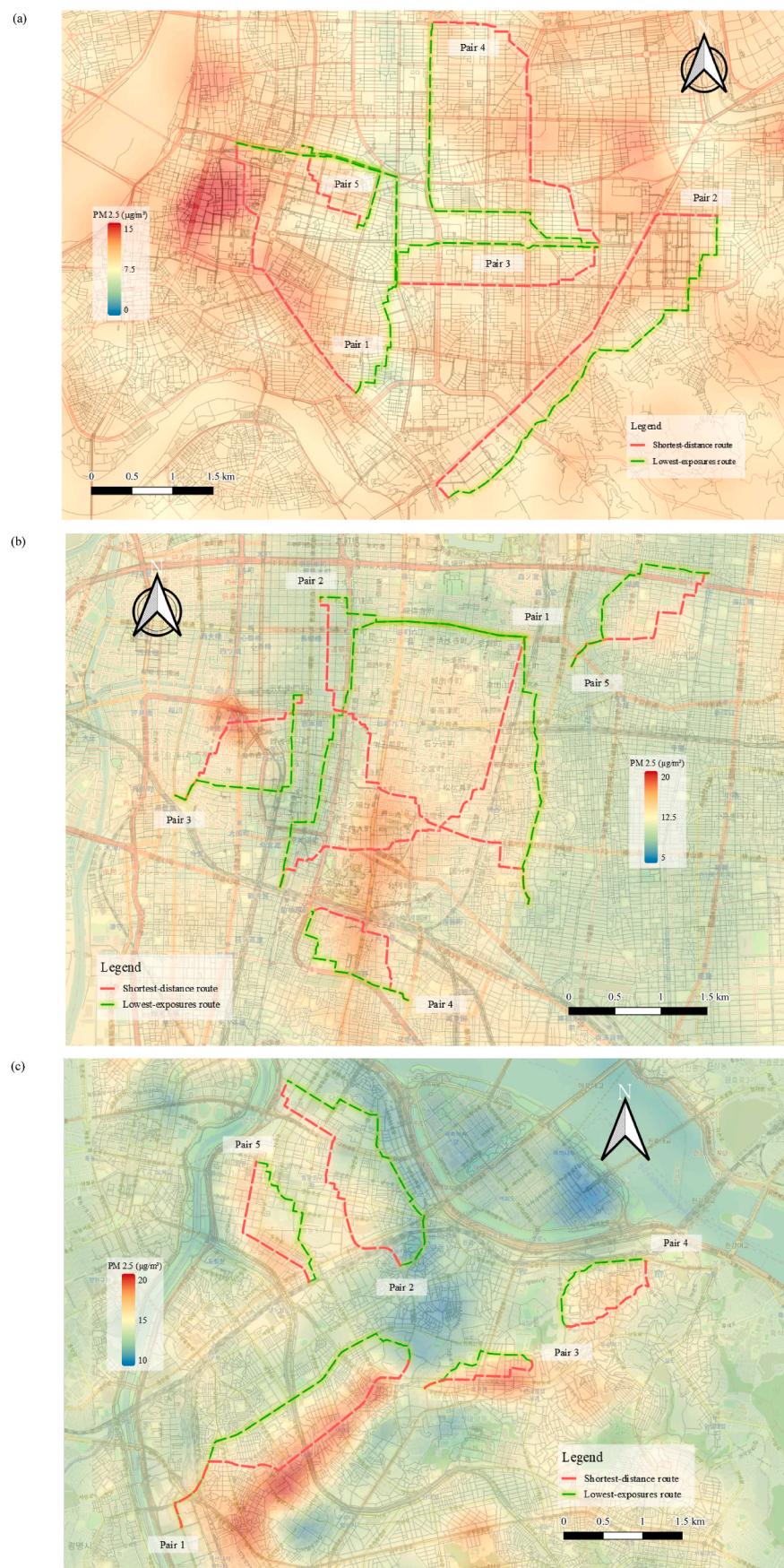


Fig. 6. Pairs of routes in the city of (a) Taipei, (b) Osaka, and (c) Seoul.

Table 2Comparison of PM_{2.5} exposure on the shortest route and cleanest route.

City	Pair	Length (km)		Average Concentration ($\mu\text{g}/\text{m}^3$)		PIC ^a ($\mu\text{g}/\text{m}^3\text{-km}$)		Percentage of increased length (%)	Percentage of reduced average concentration (%)	Percentage of reduced PIC (%)
		Shortest-distance route	Lowest-exposure route	Shortest-distance route	Lowest-exposure route	Shortest-distance route	Lowest-exposure route			
Taipei	1	3.79	5.22	11.81	8.02	45.01	41.45	37.8	-32.1	-7.9
	2	5.32	5.60	10.25	9.30	54.72	52.06	5.3	-9.3	-4.9
	3	2.80	3.00	9.28	8.08	26.11	24.26	7.1	-12.9	-7.1
	4	4.38	4.63	9.57	8.11	41.97	37.41	5.8	-15.2	-10.9
	5	1.70	1.95	9.38	7.90	16.06	15.33	15	-15.8	-4.6
Osaka	1	4.73	5.15	12.80	10.30	60.96	52.85	8.9	-19.5	-13.3
	2	5.24	5.61	12.10	11.10	63.78	62.38	7.1	-8.3	-2.2
	3	2.23	2.54	13.13	11.37	29.56	28.92	14.1	-13.4	-2.2
	4	1.85	1.90	14.33	13.41	26.67	25.53	2.9	-6.4	-4.3
	5	2.26	2.32	11.58	10.71	26.40	24.99	2.7	-7.5	-5.3
Seoul	1	4.25	4.50	16.70	14.10	71.46	63.75	5.9	-15.6	-10.8
	2	3.47	3.70	14.20	12.20	49.68	45.11	6.6	-14.1	-9.2
	3	1.65	1.67	17.03	15.08	28.38	25.18	1.1	-11.4	-11.3
	4	1.68	1.75	15.45	14.26	26.03	24.92	4.4	-7.7	-4.3
	5	2.00	2.13	15.92	14.89	32.01	31.68	6.5	-6.5	-1.0

^a Path integrated concentration (PIC): the sum of the PM_{2.5} in each road segment multiplied by 100 m (0.1 km) with a unit of $\mu\text{g}/\text{m}^3\text{-km}$.

performance (Eeftens et al., 2013; Ghassoun and Löwner, 2017; Su et al., 2008; Tang et al., 2013). However, the preliminary test in this study showed that using aspect ratio (i.e., the ratio of building height to road width) as a predictor variable did not improve the model performance, probably due to the insufficient quality of the building layer in OSM (Supplementary Material Section 6). This limitation can be addressed in future studies as the quality and balance of data in the region are continually improving. Third, the predictor variables in the LURF models were limited to those provided in OSM. To further improve model performance, additional variables associated with emissions (e.g., emission inventory or traffic counts) and dispersion (e.g., meteorological parameters) should be considered in future studies (Eeftens et al., 2013; Ghassoun et al., 2015; Ghassoun and Löwner, 2017; Tang et al., 2013). Fourth, this study focused on the results from the random forest algorithm. As a sensitivity analysis, different algorithms including Nu Support Vector Regression (Nu-SVR), General Regression Neural Network (GRNN), and Extreme Gradient Boosting (XGboost) were also tested and the algorithms based on the regression tree techniques (i.e., RF and XGboost) had a better performance as compared to the others (Table S3). The field of machine learning is developing rapidly and adopting cutting-edge algorithms is recommended. Despite these limitations, this study demonstrated the capability of this modeling approach with the open-sourced data for predicting on-road PM_{2.5} in different cities.

4. Conclusion

This study indicated that routing can reduce the exposure of cyclist commuters in Asian cities. When the traveled distance increased by 37.8% (from 3.79 to 5.22 km), the reduced accumulative PM_{2.5} exposure still reached 7.9%. Although low-cost sensors can be easily obtained and distributed in various cities, appropriate calibration is essential for ensuring data quality. Overall, the study findings suggested that routing behavior must be encouraged to reduce the exposure of cyclist commuters. Such changes in daily commuting would have a beneficial, cumulative effect in relation to chronic exposure over time. A route planning tool based on the exposure modeling approach should be developed and promoted to the public.

CRediT author statement

Tzong-Gang Wu was responsible for conceptualization, methodology, data analysis, data collection, manuscript writing, and revision.

Yan-Da Chen and Bang-Hua Chen were responsible for data collection. Kouji H. Harada and Kiyoung Lee were responsible for methodology, reviewing and editing, and the supervision of data collection. Furong Deng, Cong-Thanh Tran, Kuo-Liong Chien, and Mark J. Rood were responsible for methodology, reviewing, and editing. Tzai-Hung Wen was responsible for Geomatic methodology. Chu-Chih Chen was responsible for modeling methodology. Chang-Fu Wu was responsible for conceptualization, methodology, supervision, project administration, reviewing, and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Greatly appreciate Prof. Yu-Pin Lin for improving the study design and Jung-Chi Chang for assisting the data collection in Taipei. This project is supported by the Office of International Affairs, National Taiwan University; and the Innovation and Policy Center for Population Health and Sustainable Environment (Population Health Research Center, PHRC), College of Public Health, National Taiwan University. It is also partially supported by the Korea Environment Industry & Technology Institute (KEITI) through the Environmental Health Action Program, funded by the Korea Ministry of Environment (Project No. 2018001350001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envpol.2021.118597>.

References

- Adams, M.D., Massey, F., Chastko, K., Cupini, C., 2020. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. *Atmos. Environ.* 230 <https://doi.org/10.1016/j.atmosenv.2020.117479>.
- AirKorea, 2018. AirKorea. <http://www.airkorea.or.kr/eng>. (Accessed 23 September 2020).

- Apparicio, P., Carrier, M., Gelb, J., Séguin, A.-M., Kingham, S., 2016. Cyclists' exposure to air pollution and road traffic noise in central city neighbourhoods of Montreal. *J. Transport Geogr.* 57, 63–69. <https://doi.org/10.1016/j.jtrangeo.2016.09.014>.
- Beijing Zhiyue Information Technology Co Ltd, 2021. GPS status - record your track. Available: <https://apps.apple.com/us/app/gps-status-record-your-track/id831152887>. (Accessed 20 September 2020).
- Bigazzi, A.Y., Broach, J., Dill, J., 2016. Bicycle route preference and pollution inhalation dose: comparing exposure and distance trade-offs. *Journal of Transport & Health* 3, 107–113. <https://doi.org/10.1016/j.jth.2015.12.002>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Chen, C.-C., Wu, C.-F., Yu, H.-L., Chan, C.-C., Cheng, T.-J., 2012. Spatiotemporal modeling with temporal-invariant variogram subgroups to estimate fine particulate matter PM_{2.5} concentrations. *Atmos. Environ.* 54, 1–8. <https://doi.org/10.1016/j.atmosenv.2012.02.015>.
- Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., et al., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environ. Int.* 130 <https://doi.org/10.1016/j.envint.2019.104934>.
- City of Osaka, 2018. Profile of osaka city. https://www.city.osaka.lg.jp/contents/wdu020/enjoy/en/overview/content_CityProfile.html. (Accessed 9 April 2020).
- Cole, C.A., Carlsten, C., Koehle, M., Brauer, M., 2018. Particulate matter exposure and health impacts of urban cyclists: a randomized crossover study. *Environ. Health* 17, 78. <https://doi.org/10.1186/s12940-018-0424-8>.
- Cole-Hunter, T., Weichenthal, S., Kubesch, N., Foraster, M., Carrasco-Turigas, G., Bouso, L., et al., 2016. Impact of traffic-related air pollution on acute changes in cardiac autonomic modulation during rest and physical activity: a cross-over study. *J. Expo. Sci. Environ. Epidemiol.* 26, 133–140. <https://doi.org/10.1038/jes.2015.66>.
- Correia, C., Martins, V., Cunha-Lopes, I., Faria, T., Diapouli, E., Eleftheriadis, K., et al., 2020. Particle exposure and inhaled dose while commuting in Lisbon. *Environ. Pollut.* 257, 113547. <https://doi.org/10.1016/j.envpol.2019.113547>.
- de Nazelle, A., Fruin, S., Westerdahl, D., Martinez, D., Ripoll, A., Kubesch, N., et al., 2012. A travel mode comparison of commuters' exposures to air pollutants in Barcelona. *Atmos. Environ.* 59, 151–159. <https://doi.org/10.1016/j.atmosenv.2012.05.013>.
- Directorate-General of Budget Accounting and Statistics Taiwan, 2020. National Statistics. <http://statdb.dgbas.gov.tw/pxweb/Dialog/statfile9.asp>. (Accessed 9 April 2020).
- Edginton, S., O'Sullivan, D.E., King, W.D., Lougheed, M.D., 2021. The effect of acute outdoor air pollution on peak expiratory flow in individuals with asthma: a systematic review and meta-analysis. *Environ. Res.* 192, 110296. <https://doi.org/10.1016/j.enres.2020.110296>.
- Eeftens, M., Beekhuizen, J., Beelen, R., Wang, M., Vermeulen, R., Brunekreef, B., et al., 2013. Quantifying urban street configuration for improvements in air pollution models. *Atmos. Environ.* 72, 1–9. <https://doi.org/10.1016/j.atmosenv.2013.02.007>.
- Elford, S., Adams, M.D., 2019. Exposure to ultrafine particulate air pollution in the school commute: examining low-dose route optimization with terrain-enforced dosage modelling. *Environ. Res.* 178, 108674. <https://doi.org/10.1016/j.enres.2019.108674>.
- Feeenstra, B., Papastopoulou, V., Hasheminassab, S., Zhang, H., Boghossian, B.D., Cocker, D., et al., 2019. Performance evaluation of twelve low-cost PM_{2.5} sensors at an ambient air monitoring site. *Atmos. Environ.* 216 <https://doi.org/10.1016/j.atmosenv.2019.116946>.
- Funasaka, K., Masumoto, K., Asakawa, D., Kaneko, S., 2020. Size distribution of atmospheric particles: 40-year trends and 20-year comparisons of chemical constituents between residential and roadside areas in Osaka city, Japan. *Asian J. Atmos. Environ.* 14, 345–366. <https://doi.org/10.5572/ajae.2020.14.4.345>.
- GEOFABRIK, OpenStreetMap Contributors, 2021. Openstreetmap data extracts. <http://download.geofabrik.de/asia.html>.
- Ghassoun, Y., Löwner, M.-O., 2017. Land use regression models for total particle number concentrations using 2D, 3D and semantic parameters. *Atmos. Environ.* 166, 362–373. <https://doi.org/10.1016/j.atmosenv.2017.07.042>.
- Ghassoun, Y., Ruths, M., Lowner, M.O., Weber, S., 2015. Intra-urban variation of ultrafine particles as evaluated by process related land use and pollutant driven regression modelling. *Sci. Total Environ.* 536, 150–160. <https://doi.org/10.1016/j.scitotenv.2015.07.051>.
- Good, N., Möller, A., Ackerson, C., Bachand, A., Carpenter, T., Clark, M.L., et al., 2016. The Fort Collins Commuter Study: Impact of route type and transport mode on personal exposure to multiple air pollutants. *J. Expo. Sci. Environ. Epidemiol.* 26, 397–404. <https://doi.org/10.1038/jes.2015.68>.
- Han, W., Vijayan, A., Schulte, N., Herner, J.D., 2017. Commuter exposure to PM_{2.5}, BC, and UFP in six common transport microenvironments in Sacramento, California. *Atmos. Environ.* 167, 335–345. <https://doi.org/10.1016/j.atmosenv.2017.08.024>.
- Hankey, S., Marshall, J.D., 2015a. On-bicycle exposure to particulate air pollution: particle number, black carbon, PM_{2.5}, and particle size. *Atmos. Environ.* 122, 65–73. <https://doi.org/10.1016/j.atmosenv.2015.09.025>.
- Hankey, S., Marshall, J.D., 2015b. Land use regression models of on-road particulate air pollution (particle number, black carbon, PM_{2.5}, particle size) using mobile monitoring. *Environ. Sci. Technol.* 49, 9194–9202. <https://doi.org/10.1021/acs.est.5b01209>.
- Hankey, S., Sforza, P., Pierson, M., 2019. Using mobile monitoring to develop hourly empirical models of particulate air pollution in a rural appalachian community. *Environ. Sci. Technol.* 53, 4305–4315. <https://doi.org/10.1021/acs.est.8b05249>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.
- Hatzopoulou, M., Weichenthal, S., Barreau, G., Goldberg, M., Farrell, W., Crouse, D., et al., 2013a. A web-based route planning tool to reduce cyclists' exposures to traffic pollution: a case study in Montreal, Canada. *Environ. Res.* 123, 58–61. <https://doi.org/10.1016/j.envres.2013.03.004>.
- Hatzopoulou, M., Weichenthal, S., Dugum, H., Pickett, G., Miranda-Moreno, L., Kulka, R., et al., 2013b. The impact of traffic volume, composition, and road geometry on personal air pollution exposures among cyclists in Montreal, Canada. *J. Expo. Sci. Environ. Epidemiol.* 23, 46–51. <https://doi.org/10.1038/jes.2012.85>.
- Hu, X.F., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., et al., 2017. Estimating PM_{2.5} concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* 51, 6936–6944. <https://doi.org/10.1021/acs.est.7b01210>.
- Jaiprakash, Habib G., 2017. Chemical and optical properties of PM_{2.5} from on-road operation of light duty vehicles in Delhi city. *Sci. Total Environ.* 586, 900–916. <https://doi.org/10.1016/j.scitotenv.2017.02.070>.
- Krecl, P., Cipoli, Y.A., Targino, A.C., Toloto, M.D., Segersson, D., Parra, A., et al., 2019. Modelling urban cyclists' exposure to black carbon particles using high spatiotemporal data: a statistical approach. *Sci. Total Environ.* 679, 115–125. <https://doi.org/10.1016/j.scitotenv.2019.05.043>.
- Kumar, P., Patton, A.P., Durant, J.L., Frey, H.C., 2018. A review of factors impacting exposure to PM_{2.5}, ultrafine particles and black carbon in Asian transport microenvironments. *Atmos. Environ.* 187, 301–316. <https://doi.org/10.1016/j.atmosenv.2018.05.046>.
- Kuo, C.-P., Liao, H.-T., Chou, C.-C.-K., Wu, C.-F., 2014. Source apportionment of particulate matter and selected volatile organic compounds with multiple time resolution data. *Sci. Total Environ.* 472, 880–887. <https://doi.org/10.1016/j.scitotenv.2013.11.114>.
- Lautenschlager, F., Becker, M., Kobs, K., Steininger, M., Davidson, P., Krause, A., et al., 2020. OpenLUR: Off-the-shelf air pollution modeling with open features and machine learning. *Atmos. Environ.* 233 <https://doi.org/10.1016/j.atmosenv.2020.117535>.
- Lee, K., Sener, I.N., 2019. Understanding potential exposure of bicyclists on roadways to traffic-related air pollution: findings from El Paso, Texas, using Strava Metro data. *Int. J. Environ. Res. Publ. Health* 16. <https://doi.org/10.3390/ijerph16030371>.
- Li, H.-C., Chiueh, P.-T., Liu, S.-P., Huang, Y.-Y., 2017. Assessment of different route choice on commuters' exposure to air pollution in Taipei, Taiwan. *Environ. Sci. Pollut. Control Ser.* 24, 3163–3171. <https://doi.org/10.1007/s11356-016-8000-7>.
- Liao, H.T., Lee, C.L., Tsai, W.C., Yu, J.Z., Tsai, S.W., Chou, C.C.K., et al., 2021. Source apportionment of urban PM_{2.5} using positive matrix factorization with vertically distributed measurements of trace elements and nonpolar organic compounds. *Atmos. Pollut. Res.* 12, 200–207. <https://doi.org/10.1016/j.apr.2021.03.007>.
- Lim, C.C., Kim, H., Vilcassim, M.J.R., Thurston, G.D., Gordon, T., Chen, L.-C., et al., 2019. Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ. Int.* 131, 105022. <https://doi.org/10.1016/j.envint.2019.105022>.
- Liu, M., Peng, X., Meng, Z., Zhou, T., Long, L., She, Q., 2019. Spatial characteristics and determinants of in-traffic black carbon in Shanghai, China: combination of mobile monitoring and land use regression model. *Sci. Total Environ.* 658, 51–61. <https://doi.org/10.1016/j.scitotenv.2018.12.135>.
- Lo, W.-C., Shie, R.-H., Chan, C.-C., Lin, H.-H., 2017. Burden of disease attributable to ambient fine particulate matter exposure in Taiwan. *J. Formos. Med. Assoc.* 116, 32–40. <https://doi.org/10.1016/j.jfma.2015.12.007>.
- Ma, X., Longley, I., Gao, J., Salmond, J., 2020a. Assessing schoolchildren's exposure to air pollution during the daily commute - a systematic review. *Sci. Total Environ.* 737, 140389. <https://doi.org/10.1016/j.scitotenv.2020.140389>.
- Ma, X., Longley, I., Gao, J., Salmond, J., 2020b. Evaluating the effect of ambient concentrations, route choices, and environmental (in)justice on students' dose of ambient NO₂ while walking to school at population scales. *Environ. Sci. Technol.* 54, 12908–12919. <https://doi.org/10.1021/acs.est.0c05241>.
- McKercher, G.R., Vanos, J.K., 2018. Low-cost mobile air pollution monitoring in urban environments: a pilot study in Lubbock, Texas. *Environ. Technol.* 39, 1505–1514. <https://doi.org/10.1080/09593330.2017.1332106>.
- Minet, L., Gehr, R., Hatzopoulou, M., 2017. Capturing the sensitivity of land-use regression models to short-term mobile monitoring campaigns using air pollution micro-sensors. *Environ. Pollut.* 230, 280–290. <https://doi.org/10.1016/j.envpol.2017.06.071>.
- Möller, A., Lindley, S., 2015. Influence of walking route choice on primary school children's exposure to air pollution-a proof of concept study using simulation. *Sci. Total Environ.* 530–531, 257–262. <https://doi.org/10.1016/j.scitotenv.2015.05.118>.
- Niu, X., Chuang, H.C., Wang, X., Ho, S.S.H., Li, L., Qu, L., et al., 2020. Cytotoxicity of PM_{2.5} vehicular emissions in the shing mun tunnel, Hong Kong. *Environ. Pollut.* 263, 114386. <https://doi.org/10.1016/j.envpol.2020.114386>.
- Okonon, E.O., Yli-Tuomi, T., Turunen, A.W., Taimisto, P., Pennanen, A., Voutsis, I., et al., 2017. Particulates and noise exposure during bicycle, bus and car commuting: a study in three European cities. *Environ. Res.* 154, 181–189. <https://doi.org/10.1016/j.envres.2016.12.012>.
- OpenStreetMap contributors, 2021. Planet Dump. <https://planet.osm.org>. <https://www.openstreetmap.org>.
- Osaka Prefectural Government, 2019. Data download>download>1 hour value data (translated from Japanese). http://taiki.kankyo.pref.osaka.jp/taikiWebApp/bin/form3000/Form_3010.jsp. (Accessed 21 May 2019).
- Ozgen, S., Ripamonti, G., Malandrini, A., Ragettli, M., S., Lonati, G., 2016. Particle number and mass exposure concentrations by commuter transport modes in Milan, Italy. *AIMS Environ. Sci.* 3, 168–184. <https://doi.org/10.3934/environsci.2016.2.168>.

- Park, E.H., Heo, J., Kim, H., Yi, S.M., 2020. Long term trends of chemical constituents and source contributions of PM_{2.5} in Seoul. Chemosphere 251, 126371. <https://doi.org/10.1016/j.chemosphere.2020.126371>.
- Pohjankukka, J., Pahikkala, T., Nevalainen, P., Heikkonen, J., 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. Int. J. Geogr. Inf. Sci. 31, 2001–2019. <https://doi.org/10.1080/13658816.2017.1346255>.
- Pope 3rd, C.A., Ezzati, M., Dockery, D.W., 2009. Fine-particulate air pollution and life expectancy in the United States. N. Engl. J. Med. 360, 376–386. <https://doi.org/10.1056/NEJMsa0805646>.
- Pope 3rd, C.A., Coleman, N., Pond, Z.A., Burnett, R.T., 2020. Fine particulate air pollution and human mortality: 25+ years of cohort studies. Environ. Res. 183, 108924. <https://doi.org/10.1016/j.envres.2019.108924>.
- Popoola, O.A.M., Carruthers, D., Lad, C., Bright, V.B., Mead, M.I., Stettler, M.E.J., et al., 2018. Use of networks of low cost air quality sensors to quantify air quality in urban settings. Atmos. Environ. 194, 58–70. <https://doi.org/10.1016/j.atmosenv.2018.09.030>.
- Quiros, D.C., Lee, E.S., Wang, R., Zhu, Y., 2013. Ultrafine particle exposures while walking, cycling, and driving along an urban residential roadway. Atmos. Environ. 73, 185–194. <https://doi.org/10.1016/j.atmosenv.2013.03.027>.
- Rahman, M.M., Karunasinghe, J., Clifford, S., Knibbs, L.D., Morawska, L., 2020. New insights into the spatial distribution of particle number concentrations by applying non-parametric land use regression modelling. Sci. Total Environ. 702 <https://doi.org/10.1016/j.scitotenv.2019.134708>.
- Ramos, C.A., Silva, J.R., Faria, T., Wolterbeek, T.H., Almeida, S.M., 2017. Exposure assessment of a cyclist to particles and chemical elements. Environ. Sci. Pollut. Control Ser. 24, 11879–11889. <https://doi.org/10.1007/s11356-016-6365-2>.
- Requia, W.J., Higgins, C.D., Adams, M.D., Mohamed, M., Kourtrakis, P., 2018. The health impacts of weekday traffic: a health risk assessment of PM_{2.5} emissions during congested periods. Environ. Int. 111, 164–176. <https://doi.org/10.1016/j.envint.2017.11.025>.
- Seoul Metropolitan Government, 2021. Statistics of Seoul's Resident Registered Population (By Category). <http://data.seoul.go.kr/dataList/419/S/2/datasetView.do>. (Accessed 21 May 2021).
- Shi, Y., Lau, K.K., Ng, E., 2016. Developing street-level PM_{2.5} and PM₁₀ land use regression models in high-density Hong Kong with urban morphological factors. Environ. Sci. Technol. 50, 8178–8187. <https://doi.org/10.1021/acs.est.6b01807>.
- Strasser, G., Hiebaum, S., Neuberger, M., 2018. Commuter exposure to fine and ultrafine particulate matter in Vienna. Wien Klin. Wochenschr. 130, 62–69. <https://doi.org/10.1007/s00508-017-1274-z>.
- Su, J.G., Brauer, M., Buzzelli, M., 2008. Estimating urban morphometry at the neighborhood scale for improvement in modeling long-term average air pollution concentrations. Atmos. Environ. 42, 7884–7893. <https://doi.org/10.1016/j.atmosenv.2008.07.023>.
- Sun, B., Song, J., Wang, Y., Jiang, J., An, Z., Li, J., et al., 2021. Associations of short-term PM_{2.5} exposures with nasal oxidative stress, inflammation and lung function impairment and modification by GSTT1-null genotype: a panel study of the retired adults. Environ. Pollut. 285, 117215. <https://doi.org/10.1016/j.envpol.2021.117215>.
- Tainio, M., Jovanovic Andersen, Z., Nieuwenhuijsen, M.J., Hu, L., de Nazelle, A., An, R., et al., 2021. Air pollution, physical activity and health: a mapping review of the evidence. Environ. Int. 147, 105954. <https://doi.org/10.1016/j.envint.2020.105954>.
- Taiwan Environmental Protection Administration, 2019. Taiwan air quality monitoring network. <https://airtw.epa.gov.tw/ENG/default.aspx>. (Accessed 25 September 2019).
- Tang, R., Blangiardo, M., Gulliver, J., 2013. Using building heights and street configuration to enhance intraurban PM₁₀, NO_x, and NO₂ land use regression models. Environ. Sci. Technol. 47, 11643–11650. <https://doi.org/10.1021/es402156g>.
- Tran, P.T.M., Zhao, M., Yamamoto, K., Minet, L., Nguyen, T., Balasubramanian, R., 2020. Cyclists' personal exposure to traffic-related air pollution and its influence on bikeability. Transport. Res. Transport Environ. 88 <https://doi.org/10.1016/j.trd.2020.102563>.
- Wang, S., 2021. Green tracks - hiking partner. <https://play.google.com/store/apps/details?id=com.mountain.tracks&hl=en&gl=US>. (Accessed 20 September 2020).
- Weichenthal, S., Kulka, R., Dubeau, A., Martin, C., Wang, D., Dales, R., 2011. Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. Environ. Health Perspect. 119, 1373–1378. <https://doi.org/10.1289/ehp.1003321>.
- Weichenthal, S., Van Ryswyk, K., Goldstein, A., Bagg, S., Shekharizfard, M., Hatzopoulou, M., 2016. A land use regression model for ambient ultrafine particles in Montreal, Canada: a comparison of linear regression and a machine learning approach. Environ. Res. 146, 65–72. <https://doi.org/10.1016/j.envrres.2015.12.016>.
- Wu, C.-F., Delfino, R.J., Floro, J.N., Samimi, B.S., Quintana, P.J.E., Kleinman, M.T., et al., 2005. Evaluation and quality control of personal nephelometers in indoor, outdoor and personal environments. J. Expo. Sci. Environ. Epidemiol. 15, 99–110. <https://doi.org/10.1038/sj.jea.7500351>.
- Xu, T.-G., Chang, J.-C., Huang, S.-H., Lin, W.-Y., Chan, C.-C., Wu, C.-F., 2021. Exposures and health impact for bicycle and electric scooter commuters in Taipei. Transport. Res. Transport Environ. 91 <https://doi.org/10.1016/j.trd.2021.102696>.
- Xu, Y., Ho, H.-C., Wong, M.-S., Deng, C., Shi, Y., Chan, T.-C., et al., 2018. Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM_{2.5}. Environ. Pollut. 242, 1417–1426. <https://doi.org/10.1016/j.envpol.2018.08.029>.
- Zamora, M.L., Rice, J., Koehler, K., 2020. One year evaluation of three low-cost PM_{2.5} monitors. Atmos. Environ. 235 <https://doi.org/10.1016/j.atmosenv.2020.117615>.
- Zhao, T., Qi, W., Yang, P., Yang, L., Shi, Y., Zhou, L., et al., 2021. Mechanisms of cardiovascular toxicity induced by PM_{2.5}: a review. Environ. Sci. Pollut. Control Ser. <https://doi.org/10.1007/s11356-021-16735-9>.