



Real-time Route Planning to Reduce Pedestrian Pollution Exposure in Urban Settings

Siqi Zhong

The University of Melbourne
Melbourne, Victoria, Australia
siqzhong@student.unimelb.edu.au

Richard O. Sinnott

The University of Melbourne
Melbourne, Victoria, Australia
rsinnott@unimelb.edu.au

ABSTRACT

PM2.5 refers to fine particulate matter less than 2.5 micrometers in diameter. PM2.5 is a common air pollutant. It is capable of entering the respiratory system, and is associated with a variety of health issues such as asthma and other diseases. Pedestrians are at risk of exposure to traffic-related PM2.5 due in part to increased numbers of vehicles in city settings and their associated exhaust fumes - a key contributor to PM2.5. In this paper, we present a framework to minimise PM2.5 exposure for pedestrians by helping them avoid areas with high PM2.5 concentration levels. Specifically we predict the concentration levels through an XGBoost model and background concentration levels from official air quality monitoring stations around Melbourne. We factor in real-time, portable, air quality monitoring devices, weather conditions and real-time traffic flow information. The coefficient of determination (R^2), root mean squared error (RMSE) and the mean average error (MAE) for the XGBoost model achieves 0.71, 1.98 and 1.1 respectively. The Dijkstra algorithm is then applied to generate the minimum PM2.5 exposure of routes with alternative routes suggested trading off distance and PM2.5 exposure. Compared with the shortest route, experiments show that PM2.5 exposure can be decreased by 11 - 15% with only a marginal increase in route length.

KEYWORDS

PM2.5, Health route planning, Pollution exposure, XGBoost

ACM Reference Format:

Siqi Zhong and Richard O. Sinnott. 2023. Real-time Route Planning to Reduce Pedestrian Pollution Exposure in Urban Settings. In *IEEE/ACM 10th International Conference on Big Data Computing, Applications and Technologies (BDCAT '23), December 4–7, 2023, Taormina (Messina), Italy*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3632366.3632381>

1 INTRODUCTION

Fine particulate matter (PM2.5) is a term for particles that are 2.5 micrometers in diameter or less. Such particles can infiltrate the gas-exchange region of the lungs, but they also have the capacity to breach the respiratory barrier and enter the circulatory system. This

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BDCAT '23, December 4–7, 2023, Taormina (Messina), Italy

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0473-4/23/12...\$15.00
<https://doi.org/10.1145/3632366.3632381>

can result in adverse health effects. Air quality is now recognised as a global health and societal problem [34].

Nowadays, an increasing number of studies have examined the negative health impacts caused by long-term exposure to PM2.5. Exposure to increased levels of PM2.5 has been associated with negative impacts on the respiratory system, including increased incidence and exacerbation of asthma, airway inflammation, and overall decline in lung function. Recent studies also identified an association between exposure to high PM2.5 levels and increased cardiovascular disease mortality risk [26]. Other studies have revealed that long-term exposure to even low PM2.5 levels can increase the risk associated with diabetes, hypertension, and hyperlipidemia [9, 27].

With the increase in vehicle ownership in modern society, vehicle exhaust emissions have been identified as one of the main sources of PM2.5 [25]. Active travel modes such as biking and walking are one solution to reduce traffic-related PM2.5 [42]. However, compared with other transportation modes, pedestrians are at higher risk of exposure to traffic-related PM2.5 due to their potentially increased respiratory rate and their proximity to vehicle exhaust emissions [30]. Hence, pedestrians can face higher health risks related to PM2.5. This is especially the case in urban settings.

To reduce pedestrian exposure to PM2.5, one approach is to actively avoid areas with high concentration levels of PM2.5 and identify alternative routes. This is the focus of this paper. We summarize the limitations of previous works and propose a new low PM2.5 exposure route planning framework aimed at identifying the least polluted routes for pedestrians and proposing alternative routes.

The remainder of this paper is organized as follows. Section 2 provides a literature review of previous related work. We then present the methodology for low PM2.5 exposure route planning in Section 3. Section 4 presents and discusses the results. Section 5 draws conclusions on the work as a whole and identifies possible future extensions to the work.

2 LITERATURE REVIEW

Air pollution impacts on health, hence the demand for optimising minimal exposure routes is of interest [32]. Vehicle emissions are one of the main sources of PM2.5 due to burning of fossil fuels [40]. Vamshi & Prasad [41] proposed a framework to reduce commuter exposure to traffic-related air pollutants by recognizing and avoiding highly congested intersections. Other research separated cyclists from vehicle traffic to reduce cyclists' exposure to pollutants [8]. However, these methods did not consider the actual distribution of the pollutant. In addition, Bigazzi & Figliozzi [8] pointed out that, due to weather changes, the exposure of road

cyclists may not always be higher than those on off-street paths. Therefore, using traffic-related factors as the sole determinant of pollution may result in inaccurate estimations. Meteorological conditions, such as wind speed, relative humidity and temperature, are also important factors that influence pollution levels. For example, PM2.5 concentrations decrease with increasing precipitation [10]. Falling raindrops combine with air particles forcing PM2.5 particles to the ground thereby reducing the concentration of PM2.5 in the air. This process is called "wet deposition" [47]. In addition, the wind can decrease the PM2.5 concentration at one location but increase it in downwind locations [49]. The diffusion of PM2.5 thus depends on wind speed and direction. This results in the different local concentrations of PM2.5 [13, 48].

Many studies plan low air pollutants exposure routes based on the distribution of PM2.5 concentrations. However, obtaining the precision distribution of air pollutant concentration for each street is challenging. Traditionally, the distribution of the PM2.5 concentration relied on interpolation methods associated with fixed monitoring stations, such as inverse distance weighted interpolation, original Kriging interpolation and collaborative Kriging interpolation [15, 32, 50]. However, these methods cannot accurately represent street-level PM2.5 concentrations since they can underestimate PM2.5 concentrations near the pollution source due to the sampling approach [4, 24]. For example, traffic-related PM2.5 can drop rapidly compared to the background pollution level when the distance from the road increases [28, 35]. Therefore, in this situation, an interpolation method based on fixed monitoring station data is likely to underestimate the PM2.5 exposure of pedestrians near the roadside [22]. To solve this problem, some studies utilize taxi-based on-road mobile monitors to collect street level PM2.5 concentrations [43]. However, this method is too expensive to obtain real time PM2.5 concentration data for every street all of the time.

Machine learning and deep learning models have been used to predict PM2.5 based on fixed monitoring station data. Examples of such approaches include artificial neural networks (ANN) [29], XGBoost [36] and Support Vector Regression [31]. These methods have been used to predict the distribution of street level PM2.5 concentration data whilst aiming to achieve acceptable accuracy and low cost. [46] used a land use random forest model to model the street PM2.5 concentration distribution based on official PM2.5 data from fixed monitoring stations, meteorological data, and street data measured by citizen scientists together with land features. [18] used a back propagation neural network to obtain the PM2.5 distribution based on meteorological data, remote sensing images, road density and factory density. However, these methods ignore the impact of traffic volume and flows on PM2.5 concentration levels. It has been shown that PM2.5 concentration on low-traffic roads is lower than that of congested road and associated intersections [5, 23].

From a health perspective, different age groups and genders can inhale different amounts of pollution even in the same situational context [38]. Studies such as [46] that focus on the pollutant concentration only do not consider the impact on the different demographic groups. Further studies such as [14, 18] consider other factors and include human exposure information to weight routes to improve the route planning and associated benefits from a health

perspective. However, these methods only provide the shortest distance and/or minimum exposure route, and do not consider the real time travel needs or dealing with live information updates.

To address the limitations associated with previous studies and accurately identify real-time low PM2.5 exposure routes, this paper proposes a method that utilizes XGBoost to map fixed monitoring station data to street level PM2.5 concentration using the non-linear relationship and complex patterns associated with street PM2.5 concentration levels (as recorded by citizen scientists using Internet-of-Things based devices), traffic flows, weather conditions and official PM2.5 concentration information. The exposure models then used the respiratory minute volume and walking velocity associated with different age groups and genders to accurately quantify PM2.5 exposure dose. Finally, the Dijkstra algorithm is applied to generate the minimum PM2.5 exposure routes and propose alternative routes for different travel purposes - trading off the weight of distance and PM2.5 exposure.

3 METHODOLOGY

3.1 Study Area

The study area for this work based on the city of Melbourne Central Business District (CBD). The CBD is the central hub and commercial heart of Melbourne, Australia as shown in Figure 1. It is a vibrant and bustling urban area that serves as the main financial, cultural, and shopping district of the city. Pedestrians and vehicles are highly concentrated in this area. It has the classic congestion challenges facing many cities combining shopping and bustling pavements brimming with pedestrians, and at the same time many vehicles, traffic lights and associated rush-hour challenges.

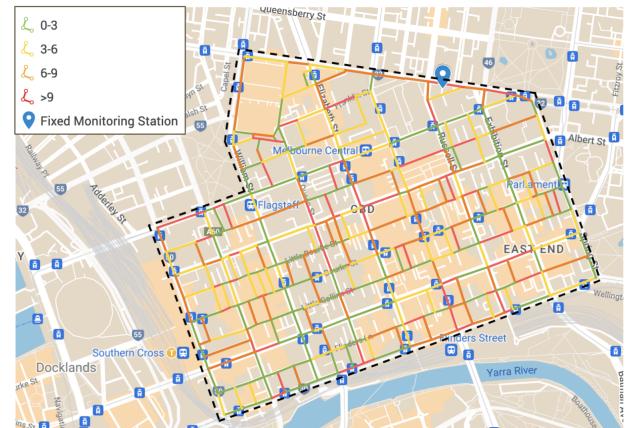


Figure 1: The PM2.5 Distribution for the Melbourne CBD

The PM2.5 distribution as shown in Figure 1 is based on recordings measured by low-cost portable air quality monitoring devices through a given individual navigating the street network over a period of two days. In this work the device used is the AirBeam technology from HabitatMap (<https://www.habitatmap.org/airbeam>). As illustrated in Figure 1, the PM2.5 concentration values have high spatial variation, i.e., a street with high PM2.5 concentration levels can have a nearby street with low PM2.5 concentration levels due to the different traffic patterns, wind direction and the impact

of the building facades (noting that many of the buildings in the Melbourne CBD are skyscrapers). In addition to this citizen science data, the more official hourly PM2.5 concentration value from the Environmental Protection Agency (EPA - www.epa.vic.gov.au) is also used. However the EPA monitoring stations are located at several, fixed (static) monitoring stations and hence they are unable to reflect the actual distribution of the PM2.5 concentration in the CBD. Nevertheless, the hourly PM2.5 concentration value from the fixed monitoring station represents an essential reference for the background concentration levels of the air quality.

3.2 Overall Workflow

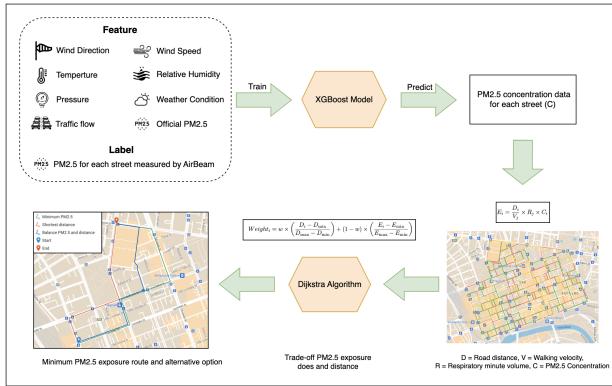


Figure 2: Overall workflow of the PM2.5 exposure route planning framework

The overall workflow of the proposed PM2.5 exposure route planning framework is shown in Figure 2. There are five main steps in this framework that are to identify the minimum PM2.5 exposure routes and potentially to propose alternative routes. Firstly, using the training data and features of the meteorological data, traffic flow, official PM2.5 data from the EPA, and the label of actual street level PM2.5 data from the AirBeam devices, we construct a street level PM2.5 concentration prediction model based on the machine learning method XGBoost. Secondly, the performance of the model is evaluated based on three evaluation metrics including the coefficient of determination (R^2), the mean absolute error (MAE) and the root of mean square error (RMSE) utilising k-fold cross-validation. Thirdly, the human PM2.5 exposure levels for each street section is calculated based on the PM2.5 concentration, route distances, respiratory (lung) volume and the average walking speed. The minimum PM2.5 exposure route is then calculated through Dijkstra's algorithm. Finally, alternative routes for different travel purposes are calculated, trading off the distance and the PM2.5 exposure levels.

3.3 PM2.5 Concentration Prediction Model

3.3.1 Dataset and Data Pre-processing.

3.3.1.1 Official PM2.5 data. As noted, official air quality and hence PM2.5 data is recorded by the EPA in Victoria. Since the study area is the Melbourne CBD, only fixed monitoring station in the vicinity of the Melbourne CBD are considered as shown in Figure 3 [17]. This data provides reliable hourly PM2.5 concentration data

for the Melbourne CBD and surrounding districts. We used the official PM2.5 data as the background PM2.5 concentration and map this background concentration to each street based on several key variables including weather conditions and real-time traffic flows.

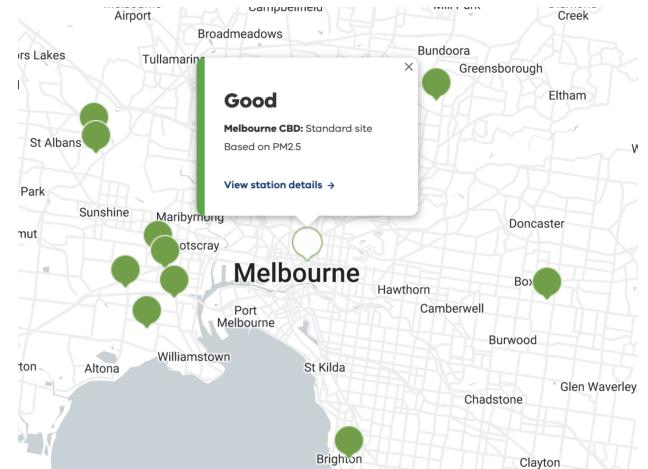


Figure 3: Location of Fixed Monitoring Stations around Melbourne

The data itself can have issues. For example, the fixed monitoring stations can provide null values or even negative PM2.5 values in the case of maintenance, or due to calibration errors or malfunctions of the monitoring equipment. During the training process, if the training instances contain missing or negative PM2.5 values from the fixed monitoring station, these training instances are removed. However, during the prediction process, the null values of negative PM2.5 values from the fixed monitoring stations should be handled to improve the robustness of the model and make sure it can work when issues with the fixed monitoring station arise. One method to achieve this is using the regression Kriging interpolation method to estimate the PM2.5 value of the Melbourne CBD fixed monitoring stations based on other nearby monitoring stations. Regression Kriging interpolation is a spatial prediction method that combines regression analysis with Kriging. Firstly, a regression model is established using spatial covariates as predictors and using the PM2.5 value of fixed monitoring station as the dependent variables. This generates predicted values and residuals for each location. The Kriging is then applied on the regression residuals to learn the spatial correlation for predicting values at unsampled locations using the weighted averages of known values from nearby locations and using the weight assigned based on spatial auto-correlation. Finally, the regression-predicted values and Kriging-predicted residuals are combined for the final prediction. This is shown in Equation 1 [21].

$$\hat{z}(S_0) = \hat{m}(S_0) + \hat{e}(S_0) = \sum_{k=0}^p \hat{\beta}_k \cdot q_k(S_0) + \sum_{i=1}^n \lambda_i \cdot e(S_i) \quad (1)$$

Here the first term $\hat{m}(S_0)$ is used to fit the deterministic part and the second term $\hat{e}(S_0)$ is the residual of the interpolation. $\hat{\beta}_k$ refers to the regression coefficients that can be estimated by generalized least squares, λ_i refers to the Kriging weight and $e(S_i)$ refers to the residual value at the location of other fixed monitoring stations.

This method can account for spatial auto-correlation from the model residuals to effectively explain the spatial variability between fixed monitoring stations. The quality of the estimated result is highly dependent on the availability and quality of nearby fixed monitoring station data.

3.3.1.2 Localized PM2.5 Concentration Data from AirBeam devices. Since the PM2.5 concentration distribution from fixed EPA monitors has strong spatial variability, the localized street-level PM2.5 data is measured using AirBeams. AirBeam provides a low-cost, portable air quality monitoring instrument that can be used to measure concentrations of PM2.5 [2]. It provides the accurate and reliable measurement of the particulate matter by pulling in ambient air and evaluating the quantity and size of these particles using a technique based on light scattering. The location information (latitude and longitude) is captured using the GPS on the phone. The data itself is captured in real time on the phone and once a route is completed, the data is submitted to a centralized air pollution map. AirBeam can measure PM1, PM2.5 and PM10, however only the PM2.5 air quality data was used. The data also includes the coordinates and the time of the recordings.

3.3.1.3 Real-time Traffic Information. The Distance Matrix API in the Google map platform provides live traffic information by specifying the departure_time as now [19]. The Distance Matrix API uses the latitude and longitude of the origins and destination of vehicles. It returns a JSON file that contains the time to cross a given road section and the associated traffic volume during that time window. This live traffic data is based collected from the users of Google who share their geographic location with the Google app on Android devices. To unify the data, hourly live traffic flows are calculated based on the time to cross each road section and the volume of traffic at that time. This is based on Equation 2.

$$\text{Traffic}_i = \frac{V_i}{T_i} \times 60 \quad (2)$$

Here Traffic_i refers to the hourly traffic volume for road section i , V_i refers to the traffic volume during that time and T_i refers to the time in minutes to cross road section i .

3.3.1.4 Weather Data. The Australian Bureau of Meteorology (2023) provides accurate weather data for Australia. This includes the temperature, relative humidity, weather conditions, wind speed, wind direction and atmospheric pressure. Such weather features are used as inputs to the XGBoost model. In the XGBoost model, every feature forms a decision node within a tree, with those near the top of the tree having a greater influence on the final prediction decision. The weather condition is a text description of weather patterns observed at the monitoring site. This can be sunny, partly cloudy, cloudy, mostly cloudy, rain, light rain, rain shower, showers in the vicinity and so on. These are important features needed to predict PM2.5 concentration levels.

3.3.2 XGBoost model. The XGBoost model [12] is used to map the official PM2.5 value from EPA to the street level. XGBoost refers to Extreme Gradient Boosting, which is an improvement over the Gradient-boosted decision tree (GBDT) algorithm. It operates on the principles of gradient tree boosting. In the prediction process,

it predicts the value based on the sum of the predicted result of a set of regression trees as shown in Equation 3 .

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F} \quad (3)$$

Here K refers to the number of regression trees, f_k refers to the k -th regression tree and \mathcal{F} refers to the space of the regression tree.

In the training process, the predictions are improved by iteratively combining a set of weak decision-tree models in an ensemble, where each subsequent decision-tree model attempts to correct the errors of existing ensembles based on the gradient loss function as shown in Equation 4 [12]. This kind of additive model can allow the XGBoost model to add a new tree at each stage to minimize the loss function, using gradient descent to find the optimal parameters as shown in Equation 6 and Equation 7[12]. Through this approach, XGBoost uses a collection of weak prediction models to collectively form a robust predictive model.

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t) \quad (4)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (5)$$

Here the first term is the loss function and second term $\Omega(f_t)$ is the regularization term which is used to penalize the complexity of the regression tree. f_t refers to the t -th regression tree and T is the number of leaves in the tree.

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \quad (6)$$

$$h_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \quad (7)$$

Here g_i and h_i are the parameters to be learned, l refers to the loss function used to measure the difference between the predicted value \hat{y} and the target value y .

There are some advantages of the XGBoost model. Firstly, as shown in Equation 5, it adds an additional regularization term, including both L1 (Lasso) and L2 (Ridge) regularization in the objective function to avoid over-fitting which is a common problem in tree-based models. In addition, it utilizes a column block structure in memory that supports efficient parallelization during the training process. Furthermore, it supports incremental training, which means that it can update the already trained XGBoost model with new data without retraining the entire model. This gives the opportunity to continue to improve the model by adding more training data. XGBoost is widely recognised as having outstanding performance in regression tasks both in terms of accuracy and training efficiency.

XGBoost is a supervisor learning model which is trained on a labeled dataset. Therefore, to map the background PM2.5 concentration data from the fixed monitoring station to the street distribution, the features used in the XGBoost model include the street id, the real time traffic flow of each street, the background PM2.5 concentration from the fixed monitoring station and the associated meteorological data including the relative humidity, wind speed, temperature, wind direction, weather conditions and the atmospheric pressure. The label used in the XGBoost is based on the real street PM2.5 concentration measured by AirBeam devices.

3.3.3 Model Evaluation Metrics. K-fold cross-validation was used to verify the XGBoost performance. This is a common statistical method used in machine learning to assess the predictive performance of models and helps to guard against over-fitting. This technique involves partitioning the dataset into subsets, called “folds”. K-fold cross-validation separates the data into k subsets. The model is then trained k times, each time using one subset as the test set and remaining k-1 subset as the training set [45]. The averaged result over the k iterations is then used as the final result to estimate the model performance on the unseen data. This process ensures the effectiveness and robustness of the model by providing a reliable evaluation of the ability of the model to generalize to new data. As noted, the R^2 , RMSE and MAE are used and these are based on 10-fold cross-validation.

3.3.3.1 Coefficient of Determination. As shown in Equation 8, R^2 represents the percentage of the variation in the dependent variable that can be predicted by independent variables in the regression model [37]. An R^2 closer to 1 indicates that the model better fits the data.

3.3.3.2 Root of Mean Square Error. The RMSE is used to measure how spread-out prediction errors are. A lower RMSE suggests a model is more accurate in its predictions. It is calculated based on the square root of the average squared differences between the actual values and the predictions. It is shown in Equation 9 [1].

3.3.3.3 Mean absolute error. The MAE is the average of the absolute difference between the predictions and the true values. A smaller MAE also means a lower average prediction error. Compared with RMSE, MAE does not square the errors before averaging, which means it gives less weight to large errors. This can make it a more straightforward measure of average error. It is shown in Equation 10 [18].

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(\hat{y}_i - y_i)| \quad (10)$$

Here n is the quantity of training data, \hat{y}_i is the predicted PM2.5 value, y_i is the true PM2.5 value and the \bar{y}_i is the average PM2.5 value.

3.3.4 Validation of XGBoost Model Accuracy. A Multiple Linear Regression (MLR) model with polynomials was used as the baseline model for the performance comparison with the XGBoost model. MLR is a common baseline used in PM2.5 prediction[3, 6, 11]. In order to handle the non-linear relationship between the features and labels, a polynomial is applied to map the features from low-dimensions to higher-dimensions.

As illustrated in Table 1, the overall performance of the XGBoost model is better than that of the MLR. The R^2 for the MLR model was 0.42, demonstrating that it explained 42% of the variability in the dependent variable. In contrast, the XGBoost model delivered a higher R^2 value of 0.71, demonstrating that it could explain 71% of the variability, which indicates a higher fitting degree to the data.

The RMSE of the MLR model was 3.4 which is significantly higher than the XGBoost model (1.98). In addition, the MAE of MLR and XGBoost were 2.7 and 1.1 respectively. Both the RMSE and MAE metrics imply that the XGBoost model’s predictions had smaller average deviations from the actual values compared to the MLR model.

Figure 4 also shows a similar result. The trend line indicates that the true value is equal to the predicted value. The XGBoost model is concentrated near the trend line which shows a good prediction in both low and high PM2.5 values. However, the MLR model deviates from the trend line which shows a poor prediction of PM2.5 values. As a result, it can be seen that the XGBoost performs better in dealing with complex mapping relations which are a key characteristic of ensemble learning and dealing with interactions between features.

Table 1: Performance of XGBoost and MLR

Model	RMSE	MAE	R^2
MLR	3.4	2.7	0.42
XGBoost	1.98	1.1	0.71

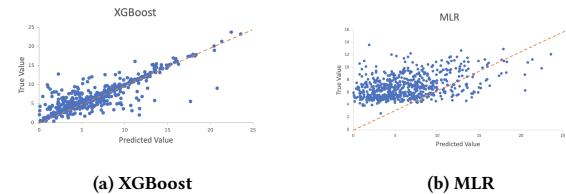


Figure 4: Scatter Plot of XGBoost and MLR

3.4 Human Exposure to PM2.5

The degree of pollutant exposure to an individual is clearly influenced by the concentration levels of PM2.5 and the amount of time they spend walking in higher PM2.5 exposed routes [43]. As shown in Equation 11, the degree of PM2.5 exposure for each street can be mathematically represented as the function of the duration of exposure, individual respiratory volume and the associated PM2.5 concentration levels.

$$E_i = \frac{D_i}{V_j} \times R_j \times C_i \quad (11)$$

Here E_i is the personal PM2.5 exposure level (measured in ug), C_i is the PM2.5 concentration (measured in ug/m³), D_i is the distance of street i. V_j is the speed/velocity of walking (measured in metres/minute) and R_j is the respiratory volume (measured in litres/minute) of demographic group j.

The PM2.5 exposure is mainly dependent on three variables: the PM2.5 concentration, the walking speed and the respiratory volume. The PM2.5 concentration is obtained from the XGBoost model introduced above. The respiratory volume and walking velocity are influenced by age, gender, activity level and physical health. In order to understand the influence of different demographics, the age group is divided into young (18-40), middle-aged (40-65) and older (>65), which account for 31.1%, 30.2%, and 18.2% of the total

Melbourne population respectively based on data from the Australian Bureau of Statistics [7]. The assumption of walking velocity and respiratory volumes for different age groups in different gender are shown in Table 2 [33, 44].

Table 2: Assumption of walking speed and respiratory volumes

Gender	Age group	Walking speed (m/min)	Respiratory volume (L/m ³)
Male	Youth	111.1	9.2
	Middle aged	82.2	9.2
	older	48.6	8.9
Female	Youth	101.4	8.2
	Middle aged	75.6	8.1
	older	42.6	7.8

The aim is to minimise the total PM2.5 exposure dose for the whole path across the CBD. This is calculated based on Equation 12.

$$E = \sum_{i=1}^n C_i \quad (12)$$

Where C_i refers to exposure level on street section i and where n refers to the total number of street sections in the given route.

3.5 PM2.5 Exposure Route Planning Model

The Dijkstra algorithm is primarily used to find the shortest path between two nodes in a weighted graph [16]. This is used as the basic algorithm to establish the PM2.5 levels for the different routes that might be taken by pedestrians. This algorithm begins from a start node and iteratively explores its neighbouring nodes, progressively constructing the shortest-path tree. For each iteration, the node with the minimum distance from the current node will be selected and its neighbouring nodes will be examined. The Dijkstra algorithm maintains a distance list for each node. At the beginning, the distance of the start node is set to zero and all other nodes are set as infinity. When examining the neighbouring nodes, if the shorter path is found, it will update the distance list. If all nodes are visited or the target node is reached, this algorithm will be terminated. As a result, this algorithm guarantees that the shortest path between two nodes can be found.

To plan the minimum PM2.5 exposure route, the weight of each street used in the Dijkstra algorithm is the human PM2.5 exposure level shown in Equation 11. However, to trade off the distance and the PM2.5 exposure, the weight of each street used in the Dijkstra algorithm is based on Equation 13.

$$Weight_i = w \times \left(\frac{D_i - D_{\min}}{D_{\max} - D_{\min}} \right) + (1 - w) \times \left(\frac{E_i - E_{\min}}{E_{\max} - E_{\min}} \right) \quad (13)$$

Here D_i is the distance of street section i , D_{\max} is the maximum distance over all street sections, D_{\min} is the minimum distance over all street sections, E_i is the PM2.5 exposure level of street section i , E_{\max} is the maximum exposure level over all street sections, E_{\min} is the minimum exposure level over all street sections. The min-max normalization strategy is used to transform the value of both distance and PM2.5 exposure levels to the range between 0 and 1. In this way, it can avoid higher numeric ranges that will have a higher impact on the weight than that of the lower numeric ranges. It is worth noting that if $w = 0$, the route is equal to the minimum

PM2.5 exposure route and if $w = 1$, the route is equal to the shortest distance route.

4 RESULT AND DISCUSSION

4.1 Effectiveness of Minimum PM2.5 Route

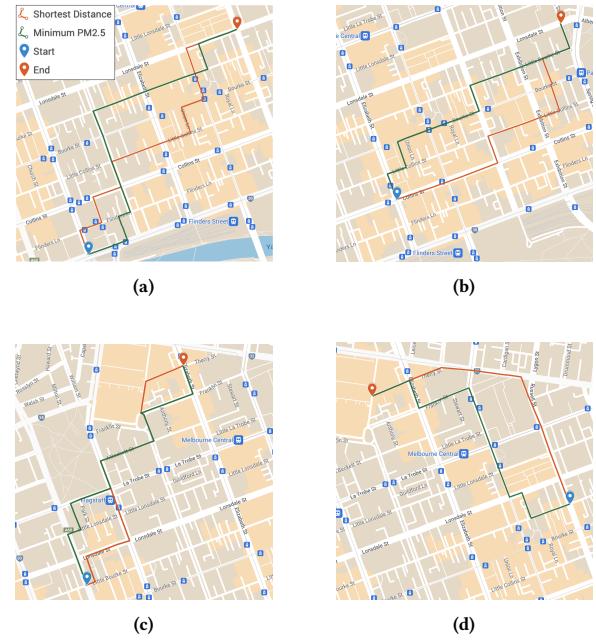


Figure 5: Routes Showing Minimum PM2.5 Exposure and Shortest Distances

Table 3: Comparison between Shortest Distances and Minimum PM2.5 Levels

Graph ID	Route type	Distance(m)	Distance Change	PM2.5(ug)	PM2.5 Change
a	shortest distance	1599.11	-	0.98	-
	minimum PM2.5	1601.27	0.14%	0.87	-11.22%
b	shortest distance	1373.44	-	0.88	-
	minimum PM2.5	1374.93	0.11%	0.78	-11.36%
c	shortest distance	1374.95	-	0.86	-
	minimum PM2.5	1421.1	3.36%	0.74	-13.95%
d	shortest distance	1272.23	-	0.81	-
	minimum PM2.5	1466.2	15.25%	0.69	-14.81%

The effectiveness of the minimum PM2.5 route is seen by comparing it with the shortest distance route. Figure 5 shows the different routes for the shortest distance and the minimum PM2.5 exposure level routes. Table 3 shows the detailed information of Figure 5. Overall, compared with the shortest distance route, although the distance of the minimum PM2.5 exposure route is slightly longer, the PM2.5 exposure decreases by 11.22-14.81%.

The Melbourne CBD is based on a grid pattern in which roads from south to north and from west to east are approximately parallel. This means that there are many routes with similar distances from the same origin to the same destination. Therefore, this design characteristic can be utilized to help pedestrians choose the minimum PM2.5 exposure routes with almost no increase in walking

distance. As shown in Table 3, the minimum PM2.5 exposure route of Figure 5a and Figure 5b decreased by 11.22% and 11.36% respectively, while the distances were almost the same as the shortest distance route, i.e., the PM2.5 exposure was reduced.

However, finding the minimum PM2.5 exposure routes can increase the route distances. Therefore, the efficiency of the proposed routes should also be considered. Ideally an efficient route should only increase the distance marginally (or not at all), but at the same time reduce PM2.5 values by a large amount. One example is the routes shown in Figure 5c. The minimum PM2.5 exposure route reduced by 13.95% while the distance only increased by 3.36%. However the minimum PM2.5 exposure route of Figure 5d shows that the PM2.5 exposure level decreased by 14.81% while the distance increased by 15.25%, indicating that this is a less inefficient route. However, for people who are sensitive to PM2.5 concentration levels, e.g., asthmatics, they may well consider it worthwhile to take longer detours if it means that they can avoid routes with higher PM2.5 concentration levels.

4.2 Alternative Low PM2.5 Route

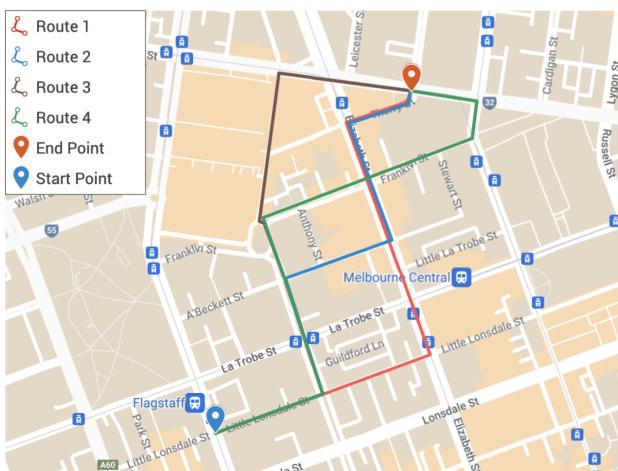


Figure 6: Alternative Low PM2.5 Routes

Table 4: Comparison of Different Alternative Routes

Route ID	Distance(m)	Distance Change	PM2.5(ug)	PM2.5 Change
Route 1	1109.84	-	0.83	-
Route 2	1110.79	0.09%	0.76	-8.43%
Route 3	1171.29	5.54%	0.73	-12.05%
Route 4	1219.83	9.91%	0.67	-19.28%

Since the minimum PM2.5 exposure route usually increases the distance in reducing the PM2.5 exposure levels, there is a trade off between the PM2.5 exposure levels and the route distances. For those walking to work, although PM2.5 exposure levels is important, the distance and the walking duration are also important factors to be considered in route selection, e.g. to arrive in time. Only providing the shortest distance route and minimum PM2.5 exposure route in route planning does not meet all needs of people. For example, for commuters, the detour distance of the minimum PM2.5

exposure route Route 4 is 110 meters, which may be unacceptable. In addition, although Route 1 is the shortest distance route, it has a significantly higher level of PM2.5 exposure compared to Route 2 but it has almost the same distance as Route 2. Therefore, compared to the shortest distance route Route 1 and the minimum PM2.5 exposure route Route 4, the alternative route Route 2 may be more suitable route for people in a hurry.

In addition to the shortest distance routes and minimum PM2.5 exposure routes, alternative routes are also provided for different travel purposes by adjusting the weight between the distance and PM2.5 exposure levels as presented in Figure 6 and Table 4. Individuals can choose Route 2 or Route 3 according to the time needs. If the purpose of pedestrians in the CBD is to undertake exercise or if pedestrians are sensitive to PM2.5, a longer travel distance may be accepted provided there is less PM2.5 exposure. Therefore, Route 4 will be the most suitable route. The correlation between increased distance and reductions in PM2.5 is non-linear, so people can choose the most cost effective route based on their own criteria. For example, by taking Route 2, pedestrians only need to walk an additional 1 meter to avoid 0.04ug of PM2.5 exposure levels which accounts for 8.43% of the shortest distance route, i.e. it is the most cost effective route. However, by taking Route 4, pedestrians need to walk an additional 110 meters to reduce their PM2.5 exposure by 19.28%, which is the least cost effective route.

4.3 Impact on PM2.5 router of different demographics

Table 5: Route Selection and PM2.5 Exposure Levels of Different Demographic Groups

Gender	Age group	Route selection	PM2.5 exposure levels
Male	Youth	Route 4	0.51
	Middle aged	Route 3	0.74
	older	Route 2	1.32
Female	Youth	Route 3	0.53
	Middle aged	Route 3	0.71
	older	Route 2	1.33

Although demographic characteristics impact the walking speed and respiratory volumes of individuals, the overall distribution of PM2.5 exposure will not be affected once the PM2.5 concentration for each street is determined. However, the different walking velocities and respiratory volumes in the different age groups and genders can influence the PM2.5 exposure levels and route selection under the same constraints.

For example, if pedestrians want to limit their detour time to one minute, people of different genders and age groups will choose different routes dependent on the different longest detour distances if they are limited by their walking speed. According to Table 5, the different route selections are shown and compared with the detour time. In this case, younger people may choose the lower PM2.5 exposure routes due to the higher walking speeds compared to older people.

Although some of the different age groups and genders may choose the same route, the exposure rates are also different due to different walking speeds and respiratory volumes. Although men may have a slightly higher walking speed than women and hence are able to walk through routes faster, the larger respiratory volume

of men means that they may be exposed to higher PM2.5 levels than women. In addition, when the age increases, the walking speed may become slower, but the respiratory volume may not have a significant change [39]. Therefore, older people may have higher PM2.5 exposure levels on the same route due to the lower walking speed compared to younger people. For example, both the youth women and middle aged women may choose Route 3 when the largest detour time is within one minute. In this case, middle aged women breath 0.18ug more PM2.5 than that of the younger women group.

5 CONCLUSION AND FUTURE DIRECTIONS

This research proposes a novel framework to assess PM2.5 exposure routes for pedestrians in urban areas and, where necessary it provides alternative routes that may serve the different travel demands. In order to overcome the challenge of obtaining the street level PM2.5 concentration data, XGBoost was used to map the fixed monitoring station PM2.5 data from the EPA to the street level data by learning the non-linear relationship and the complex pattern between the real street PM2.5 concentration data from the AirBeam IoT devices, traffic flow, weather conditions and the background PM2.5 concentration values from the fixed monitoring station. The XGBoost showed an improved accuracy for prediction when compared with the baseline MLR model. Furthermore, the Dijkstra algorithm was used to generate minimum PM2.5 routes. Compared with the shortest distance routes, although the distance of minimum PM2.5 exposure routes could be slightly longer, the minimum PM2.5 exposure routes could effectively reduce the PM2.5 exposure levels of pedestrians by 11–15%. This reduction can have huge benefits for those with respiratory and cardiovascular diseases.

To meet the actual requirement of travel needs of pedestrians, alternative routes are suggested for different travel purposes by adjusting the weights between the distances and the PM2.5 exposure levels. Individuals may choose the most cost-effective route from the different routes based on their specific travel needs. In addition to the travel purposes, the different route selections and exposure levels of different demographic groups was explored factoring in different walking speeds and respiratory capacity of individuals. Although older people might choose the same low PM2.5 exposure routes as younger people, it was illustrated that significantly lower walking speeds may lead to higher PM2.5 exposure. Overall, for the public, this low PM2.5 exposure route planner can effectively reduce the exposure levels of PM2.5 by avoiding areas with high PM2.5 concentration levels. Such a solution may encourage the public to choose active travel modes, such as walking, and help reduce traffic-related PM2.5 emissions and help contribute to environmental sustainability.

However, there are some limitations to this work. Firstly, although emissions from vehicle exhausts are a major contributor to PM2.5 pollution in urban areas, if the range of the route planner is expanded to the whole of Melbourne or Victoria, then smoke from fires/bushfires and emissions from factories are also key contributors of PM2.5 and hence cannot be ignored [5]. Such other factors might also be considered in future extensions to this work.

It is also noted that during the measurement of real-time street PM2.5 data using the AirBeam, the intersections, especially those

with traffic lights and those with heavy traffic, have significantly higher PM2.5 values than nearby non-intersection streets. One primary factor is that vehicles frequently stopping at intersections idle their engines, which leads to the release of higher levels of exhaust emissions. Vehicles also accelerate when traffic signals turn green, which generally results in incomplete combustion within the engine and a higher emission rate compared to driving at a steady speed [20]. Therefore, for future improvement, the number of traffic lights and associated intersections might also be taken into consideration when considering PM2.5 route planning.

Another extension of this work is to deliver the route planning solution through a mobile application (app) that can be accessed and used by individuals interested in air quality or those with respiratory health issues. Once established, this could then be trialled with real citizens to identify whether or not they would accept and use such a pollution minimising app. Such an app could also help to inform policy, e.g., providing information to drivers on the amount of pollution that they release in their driving behaviour and hence encouraging them to use public transport, as well as potentially offering evidence of the need for further disaggregated data collection of air quality monitoring solutions. The EPA in Victoria are considering air quality monitoring on major thoroughfares through monitoring in streetlights, however as with the current approach, these are at fixed/static locations and hence will not cover all areas and intersections.

ACKNOWLEDGMENTS

This work is closely aligned with the Australian Urban Research Infrastructure Network (AURIN - www.aurin.org.au) high impact project focused on establishment of a national air quality platform (<https://naqd.eresearch.unimelb.edu.au>), and the Australian Research Data Commons (ARDC - www.ardc.edu.au) funded bushfire data commons (<https://bdcpoc.eresearch.unimelb.edu.au>). The authors thank the two funders for the investments made in these platforms.

REFERENCES

- [1] Ibrahim Said Ahmad, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub, and Shamsudeen Hassan Muhammad. 2020. A survey on machine learning techniques in movie revenue prediction. *SN Computer Science* 1 (2020), 1–14. <https://doi.org/10.1007/s42979-020-00249-1>
- [2] AirCasting. 2023. Taking Matter into Your Own Hands:About AirBeam. <https://www.habitatmap.org/airbeam>
- [3] M.S. Alam, H. Perugu, and A. McNabola. 2018. A comparison of route-choice navigation across air pollution exposure, CO₂ emission and traditional travel cost factors. *Transportation Research Part D: Transport and Environment* 65 (2018), 82–100. <https://doi.org/10.1016/j.trd.2018.08.007>
- [4] Dimitra Alexiou and Stefanos Katsavounis. 2015. A multi-objective transportation routing problem. *Operational Research* 15 (2015), 199–211. <https://doi.org/10.1007/s12351-015-0173-1>
- [5] Mohammad Hashem Askariyah, Madhusudhan Venugopal, Haneen Khreis, Andrew Birt, and Josias Zietsman. 2020. Near-road traffic-related air pollution: Resuspended PM_{2.5} from highways and arterials. *International journal of environmental research and public health* 17, 8 (2020), 2851. <https://doi.org/10.3390/ijerph17082851>
- [6] Shadi Ausati and Jamil Amanollahi. 2016. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM2.5. *Atmospheric Environment* 142 (2016), 465–474. <https://doi.org/10.1016/j.atmosenv.2016.08.007>
- [7] Australian Bureau of Statistics. 2021. Regional population by age and sex. <https://www.abs.gov.au/statistics/people/population/regional-population-age-and-sex/latest-release>
- [8] Alexander Y. Bigazzi and Miguel A. Figliozzi. 2015. Roadway determinants of bicyclist exposure to volatile organic compounds and carbon monoxide.

- Transportation Research Part D: Transport and Environment* 41 (2015), 13–23. <https://doi.org/10.1016/j.trd.2015.09.008>
- [9] Benjamin Bowe, Yan Xie, Tingting Li, Yan Yan, Hong Xian, and Ziyad Al-Aly. 2018. The 2016 global and national burden of diabetes mellitus attributable to PM_{2.5} air pollution. *The Lancet Planetary Health* 2, 7 (2018), e301–e312. [https://doi.org/10.1016/S2542-5196\(18\)30140-2](https://doi.org/10.1016/S2542-5196(18)30140-2)
- [10] Chate. 2005. Study of scavenging of submicron-sized aerosol particles by thunderstorm rain events. *Atmospheric Environment* 39, 35 (2005), 6608–6619. <https://doi.org/10.1016/j.atmosenv.2005.07.063>
- [11] Asha B. Chelani. 2019. Estimating PM_{2.5} concentration from satellite derived aerosol optical depth and meteorological variables using a combination model. *Atmospheric Pollution Research* 10, 3 (2019), 847–857. <https://doi.org/10.1016/j.apr.2018.12.013>
- [12] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [13] Tao Chen, Jun He, Xiaowei Lu, Jiangfeng She, and Zhongqing Guan. 2016. Spatial and temporal variations of PM_{2.5} and its relation to meteorological factors in the urban area of Nanjing, China. *International Journal of environmental research and public health* 13, 9 (2016), 921. <https://doi.org/10.3390/ijerph13090921>
- [14] Gemma Davies and J. Duncan Whyatt. 2014. A network-based approach for estimating pedestrian journey-time exposure to air pollution. *Science of The Total Environment* 485–486 (2014), 62–70. <https://doi.org/10.1016/j.scitotenv.2014.03.038>
- [15] Lirong Deng. 2015. *Estimation of PM_{2.5} spatial distribution based on kriging interpolation*. 1791–1794 pages. <https://doi.org/10.2991/icismme-15.2015.370>
- [16] EW Dijkstra. 1959. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* 1 (1959), 269–271. <https://doi.org/10.1007/BF01386390>
- [17] EPA. 2023. EPA AirWatch | Environment Protection Authority Victoria. <https://www.epa.vic.gov.au/for-community/airwatch>
- [18] LiNa Gao, Fei Tao, PeiLong Ma, ChenYi Wang, Wei Kong, WenKai Chen, and Tong Zhou. 2022. A short-distance healthy route planning approach. *Journal of Transport & Health* 24 (2022), 101314. <https://doi.org/10.1016/j.jth.2021.101314>
- [19] Google. 2023. Custom Map Tools & Products. <https://mapsplatform.google.com/maps-products/#directions>
- [20] Hongdi He and H Oliver Gao. 2021. Particulate matter exposure at a densely populated urban traffic intersection and crosswalk. *Environmental Pollution* 268 (2021), 115931. <https://doi.org/10.1016/j.envpol.2020.115931>
- [21] Tomislav Hengl, Gerard B.M. Heuvelink, and David G. Rossiter. 2007. About regression-kriging: From equations to case studies. *Computers & Geosciences* 33, 10 (2007), 1301–1315. <https://doi.org/10.1016/j.cageo.2007.05.001>
- [22] Jing Huang, Furong Deng, Shaowei Wu, and Xinbiao Guo. 2012. Comparisons of personal exposure to PM_{2.5} and CO by different commuting modes in Beijing, China. *Science of The Total Environment* 425 (2012), 52–59. <https://doi.org/10.1016/j.scitotenv.2012.03.007>
- [23] Sarah Jarjour, Michael Jerrett, Dane Westerdahl, Audrey de Nazelle, Cooper Hanning, Laura Daly, Jonah Lipsitt, and John Balmes. 2013. Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study. *Environmental Health* 12 (2013), 1–12. <https://doi.org/10.1186/1476-069X-12-14>
- [24] Michael Jerrett, Richard T Burnett, Renjun Ma, C Arden Pope III, Daniel Krewski, K Bruce Newbold, George Thurston, Yuanli Shi, Norm Finkelstein, Eugenia E Calle, et al. 2005. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* (2005), 727–736. <https://doi.org/10.1097/01.ede.0000181630.15826.7d>
- [25] Patrick L. Kinney, Michael Gatari Gichuru, Nicole Volavka-Close, Nicole Ngo, Peter K. Ndiba, Anna Law, Anthony Gachanja, Samuel Mwaniki Gaita, Steven N. Chillrud, and Elliott Sclar. 2011. Traffic impacts on PM_{2.5} air quality in Nairobi, Kenya. *Environmental Science & Policy* 14, 4 (2011), 369–378. <https://doi.org/10.1016/j.envsci.2011.02.005>
- [26] Chayakrit Krittawong, Yusuf Kamran Qadeer, Richard B. Hayes, Zhen Wang, Salim Virani, George D. Thurston, and Carl J. Lavie. 2023. PM_{2.5} and Cardiovascular Health Risks. *Current Problems in Cardiology* 48, 6 (2023), 101670. <https://doi.org/10.1016/j.cpcardiol.2023.101670>
- [27] Guoao Li, Wanying Su, Qi Zhong, Mingjun Hu, Jialiu He, Huanhuan Lu, Wenlei Hu, Jianjun Liu, Xue Li, Jiahua Hao, and Fen Huang. 2022. Individual PM_{2.5} component exposure model, elevated blood pressure and hypertension in middle-aged and older adults: A nationwide cohort study from 125 cities in China. *Environmental Research* 215 (2022), 114360. <https://doi.org/10.1016/j.envres.2022.114360>
- [28] Donghai Liang, Rachel Golan, Jennifer L. Moutinho, Howard H. Chang, Roby Greenwald, Stefanie E. Sarnat, Armistead G. Russell, and Jeremy A. Sarnat. 2018. Errors associated with the use of roadside monitoring in the estimation of acute traffic pollutant-related health effects. *Environmental Research* 165 (2018), 210–219. <https://doi.org/10.1016/j.envres.2018.04.013>
- [29] Xinwei Liu, Muchuan Qin, Yue He, Xiwei Mi, and Chengqing Yu. 2021. A new multi-data-driven spatiotemporal PM_{2.5} forecasting model based on an ensemble graph reinforcement learning convolutional network. *Atmospheric Pollution Research* 12, 10 (2021), 101197. <https://doi.org/10.1016/j.apr.2021.101197>
- [30] Ji Luo, Kanok Boriboonsomsin, and Matthew Barth. 2018. Reducing pedestrians' inhalation of traffic-related air pollution through route choices: Case study in California suburb. *Journal of Transport & Health* 10 (2018), 111–123. <https://doi.org/10.1016/j.jth.2018.06.008>
- [31] Jinghui Ma, Zhongqi Yu, Yuanhao Qu, Jianming Xu, Yu Cao, et al. 2020. Application of the XGBoost machine learning method in PM_{2.5} prediction: A case study of Shanghai. *Aerosol and Air Quality Research* 20, 1 (2020), 128–138. <https://doi.org/10.4209/aaqr.2019.08.0408>
- [32] Sachin Mahajan, YuSiou Tang, DongYi Wu, TzuChieh Tsai, and LingJyh Chen. 2019. CAR: The Clean Air Routing Algorithm for Path Navigation With Minimal PM_{2.5} Exposure on the Move. *IEEE Access* 7 (2019), 147373–147382. <https://doi.org/10.1109/ACCESS.2019.2946419>
- [33] Elaine M Murtagh, Jacqueline L Mair, Elroy Aguiar, Catrine Tudor-Locke, and Marie H Murphy. 2021. Outdoor walking speeds of apparently healthy adults: A systematic review and meta-analysis. *Sports Medicine* 51 (2021), 125–141. <https://doi.org/10.1007/s40279-020-01351-3>
- [34] Timothy D. Nelin, Allan M. Joseph, Matthew W. Gorr, and Loren E. Wold. 2012. Direct and indirect effects of particulate matter on the cardiovascular system. *Toxicology Letters* 208, 3 (2012), 293–299. <https://doi.org/10.1016/j.toxlet.2011.11.008>
- [35] Halil Özkan, Lisa K Baxter, Kathrin L Dionisio, and Janet Burke. 2013. Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of exposure science & environmental epidemiology* 23, 6 (2013), 566–572. <https://doi.org/10.1038/jes.2013.15>
- [36] Jian Peng, Haisheng Han, Yong Yi, Huimin Huang, and Le Xie. 2022. Machine learning and deep learning modeling and simulation for predicting PM_{2.5} concentrations. *Chemosphere* 308 (2022), 136353. <https://doi.org/10.1016/j.chemosphere.2022.136353>
- [37] Luke Plonsky and Hessameddin Ghanbar. 2018. Multiple regression in L2 research: A methodological synthesis and guide to interpreting R² values. *The Modern Language Journal* 102, 4 (2018), 713–731. <https://doi.org/10.1111/modl.12509>
- [38] Carla A Ramos, Humbert T Wolterbeek, and Susana M Almeida. 2016. Air pollutant exposure and inhaled dose during urban commuting: a comparison between cycling and motorized modes. *Air Quality, Atmosphere & Health* 9 (2016), 867–879. <https://doi.org/10.1007/s11869-015-0389-5>
- [39] Gulshan Sharma and James Goodwin. 2006. Effect of aging on respiratory system physiology and immunology. *Clinical interventions in aging* 1, 3 (2006), 253–260. <https://doi.org/10.2147/cia.2006.1.3.253>
- [40] Admír Crésio Targino, Mark David Gibson, Patricia Krecl, Marcos Vinicius Costa Rodrigues, Mauricio Moreira dos Santos, and Marcelo de Paula Corrêa. 2016. Hotspots of black carbon and PM_{2.5} in an urban area and relationships to traffic characteristics. *Environmental Pollution* 218 (2016), 475–486. <https://doi.org/10.1016/j.envpol.2016.07.027>
- [41] B Vamshi and Raja Vara Prasad. 2018. Dynamic route planning framework for minimal air pollution exposure in urban road transportation systems. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE, 540–545. <https://doi.org/10.1109/WF-IoT.2018.8355209>
- [42] Wang, Kim Dirks, Matthias Ehrhart, Jon Pearce, and Alan Cheung. 2018. Supporting healthy route choice for commuter cyclists: The trade-off between travel time and pollutant dose. *Operations Research for Health Care* 19 (2018), 156–164. <https://doi.org/10.1016/j.orhc.2018.04.001>
- [43] Wang, Yizheng Wu, Zhenyu Li, Kai Liao, Chao Li, and Guohua Song. 2022. Route planning for active travel considering air pollution exposure. *Transportation Research Part D: Transport and Environment* 103 (2022), 103176. <https://doi.org/10.1016/j.trd.2022.103176>
- [44] Beibei Wang, Zongshuang Wang, Yongjie Wei, Feifei Wang, and Xiaoli Duan. 2015. 2 - Inhalation Rates. In *Highlights of the Chinese Exposure Factors Handbook (Adults)*, Xiaoli Duan, Xiuge Zhao, Beibei Wang, Yiting Chen, and Suzhen Cao (Eds.). Academic Press, 15–21. <https://doi.org/10.1016/B978-0-12-803125-4.00012-2>
- [45] Tsu-Tsung Wong and Po-Yang Yeh. 2019. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering* 32, 8 (2019), 1586–1594. <https://doi.org/10.1109/TKDE.2019.2912815>
- [46] TzongGang Wu, Yan-Da Chen, BangHua Chen, Kouji H. Harada, Kiyoung Lee, Furong Deng, Mark J. Rood, ChuChih Chen, CongThanh Tran, KuoLiang Chien, TsaiHung Wen, and ChangFu Wu. 2022. 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. *Environmental Pollution* 294 (2022), 118597. <https://doi.org/10.1016/j.envpol.2021.118597>
- [47] Zhao Xin, Sun Yue, Zhao Chuanfeng, and Jiang Huifei. 2020. Impact of precipitation with different intensity on PM_{2.5} over typical regions of China. *Atmosphere* 11, 9 (2020), 906. <https://doi.org/10.3390/atmos11090906>
- [48] ChenXi Zhao, YunQi Wang, YuJie Wang, HuiLan Zhang, and BingQing Zhao. 2014. Temporal and spatial distribution of PM_{2.5} and PM₁₀ pollution status and the correlation of particulate matters and meteorological factors during winter and spring in Beijing. *Huan jing ke xue= Huanjing kexue* 35, 2 (2014), 418–427.
- [49] Liu Zhen, Shen Luming, Yan Chengyu, Du Jianshuang, Li Yang, and Zhao Hui. 2020. Analysis of the Influence of Precipitation and Wind on PM_{2.5} and PM₁₀

- in the Atmosphere. *Advances in Meteorology* 2020 (2020), 1–13. <https://doi.org/10.1155/2020/5039613>
- [50] Bin Zou, Yanqing Luo, Neng Wan, Zhong Zheng, Troy Sternberg, and Yilan Liao. 2015. Performance comparison of LUR and OK in PM_{2.5} concentration mapping: A multidimensional perspective. *Scientific reports* 5, 1 (2015), 1–7. <https://doi.org/10.1038/srep08698>