

Received September 3, 2019, accepted October 4, 2019, date of publication October 9, 2019, date of current version October 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946419

# CAR: The Clean Air Routing Algorithm for Path Navigation With Minimal PM2.5 Exposure on the Move

**SACHIT MAHAJAN<sup>1,2,3</sup>, YU-SIOU TANG<sup>2</sup>, DONG-YI WU<sup>2</sup>, TZU-CHIEH TSAI<sup>1,3</sup>, AND LING-JYH CHEN<sup>1,2</sup>, (Senior Member, IEEE)**

<sup>1</sup>TIGP SNHCC, Academia Sinica, Taipei 11529, Taiwan

<sup>2</sup>Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan

<sup>3</sup>Department of Computer Science, National Chengchi University, Taipei 116, Taiwan

Corresponding author: Sachit Mahajan (sachitmahajan90@gmail.com)

This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 108-2218-E-001-002 and Grant MOST 105-2221-E-001-016-MY3.

**ABSTRACT** Transport related pollution is becoming a major issue as it adversely affects human health and one way to lower the personal exposure to air pollutants is to choose a health-optimal route to the destination. Current navigation systems include options for the quickest paths (distance, traffic) and least expensive paths (fuel costs, tolls). In this paper, we come up with the CAR (Clean Air Routing) algorithm and use it to build a health-optimal route recommendation system between the origin and the destination. We combine the open source PM2.5 (Fine Particulate Matter with diameter less than 2.5 micrometers) concentration data for Taiwan, with the road network graph obtained through OpenStreetMaps. In addition, spatio-temporal interpolation of PM2.5 is performed to get PM2.5 concentration for the road network intersections. Our algorithm introduces a weight function that assesses how much PM2.5 the user is exposed to at each intersection of the road network and uses it to navigate through intersections with the lowest PM2.5 exposures. The algorithm can help people reduce their overall PM2.5 exposure by offering a healthier alternative route which may be slightly longer than the shortest path in some cases. We evaluate our algorithm for different travel modes, including driving, cycling and walking. An analysis is done for more than 4,000 real-world travel scenarios. The results show that our approach can lead to an average exposure reduction of 17.1% with an average distance increase of 2.4%.

**INDEX TERMS** Routing, smart navigation, urban computing, air quality.

## I. INTRODUCTION

Smart city initiatives have been widely implemented all over the world with the main goal of providing better living conditions and sustainable development. A smart city project integrates Information and Communication Technology (ICT) and Internet of Things (IoT) to provide people with better facilities and promote sustainable development. In the recent years, IoT has revolutionized the smart city initiative and IoT devices have become the technological backbone of smart cities [1]. The rapid boom in technology and industrialization has resulted in some serious environmental challenges, and deteriorating air quality is one of them. Traffic pollution is one of the major sources of

The associate editor coordinating the review of this manuscript and approving it for publication was Hongxiang Li.

outdoor pollution; higher concentrations of pollutants are typically recorded near the roadways. As a result, higher concentration of the pollutants can affect drivers, cyclists and pedestrians. Common pollutants include fine particulate matter and nitrogen oxide ( $NO_x$ ). Among all the major pollutants, PM2.5 is considered to be one of the most dangerous as it can actually penetrate the lungs and cause many health problems like lung cancer, asthma and other dangerous respiratory diseases [2]. Outdoor pollution is one of the major causes of deteriorating health all over the world. A study [3] has found that traffic related pollution, especially PM2.5 has adverse effects on human health and increases the risk of death from breast cancer. This makes it important to figure out ways to reduce PM2.5 exposure while traveling and doing outdoor activities. Most of the time, people do outdoor activities like hiking or cycling to stay healthy, but

the presence of outdoor pollutants can yield the opposite effect.

One way to deal with the problem of lowering PM2.5 exposure is to use less polluted routes for driving, cycling and jogging. The issue with opting for an alternate path is that the path might be a bit longer than the shortest path but with lower overall PM2.5 exposure. There are numerous route-finding applications available to people today such as Google Maps, TomTom, HERE maps etc.. Most of these applications recommend routes to the user based on personal preferences, like fastest route between two locations, shortest route, and least fuel consumption route. Some of them even offer services like calculating the fastest route based on the real-time traffic data. However, people-especially cyclists and joggers find these applications less convincing and are concerned about negative effects of air pollution and would like an application that minimizes the exposure to air pollutants [4].

Having a health optimal navigation system for driving, cycling and walking would help users to avoid polluted routes by suggesting ones that are healthy but slightly longer in some cases. We follow a sub-optimal approach which means that the recommended path would have lower PM2.5 exposure. But there is a trade-off between exposure reduction and journey length. Such cases can be observed when the distance is too long, when the origin and destination is out of bound or for finding alternate routes for pedestrian paths (as there are limited options for short pedestrian routes). Also, there are some cases in which the CAR algorithm and Google API shows different coordinates for the same location. This is because Google API uses a different geocoding technique and this sometimes leads to increase in journey length. For example, if a person has to go from location A to B. The route recommendation system would recommend multiple routes, including the shortest path as well as the path with the lowest exposure to PM2.5. In our approach, we implement A\* path-finding algorithm [5] to find the lowest cost path between two locations based on two weight metrics i.e. PM2.5 and distance. OpenStreetMap (OSM) software is used to get the road network graph for Taiwan. Every road in OSM data is usually tagged with keywords like “footway” and “pedestrian”. We use Graphhopper, a routing library which is based on OSM. There have been some improvements in the navigation systems, but the idea of a healthy route has still not been widely explored. We believe that the route recommendation services should be able to provide alternative routes depending on personal preferences, especially ones that recommend health optimal routes. Most routing algorithms used in the navigation services need weights applied to different segments of the road [6]. In a general scenario, the algorithm for calculating the shortest path between two locations would compute the total weight (based on the length

of each segment). Also, the time duration is used to find the fastest route. Similarly, to compute the healthiest routes, each intersection would be assigned the weight that would be based on the PM2.5 value.

In this paper, we present the CAR algorithm that recommends shortest health-optimal paths from origin to the destination. Open source PM2.5 data are considered as the environmental factor while looking for shortest paths. PM2.5 data are spatially and temporally interpolated on the Taiwan’s road network obtained from OpenStreetMap which is later used to assign weights to different intersections. The spatio-temporally interpolated PM2.5 is used as the cost function in shortest-path finding algorithm. A comparison is made between the health-optimal routes and traditional shortest paths between the origin and destination. An analysis of total PM2.5 exposure across different recommended routes is done.

The rest of the paper is organized as follows: Section II gives an idea about the background and related works. In Section III, we describe the methodology; the data, interpolation techniques, weighting model and shortest-path finding algorithm are also explained. In Section IV, we show the results and analyze them. We also do an evaluation by comparing our proposed model’s results with conventional fastest path finding algorithm results provided by Google Maps. We consider more than 4,000 real-world travel scenarios to analyze the performance of our algorithm. Section V summarizes the paper and provides some directions about the future work.

## II. RELATED WORKS

Many initiatives have been taken for smart city development, and one of the focuses has been towards developing a smart navigation system. In this paper, we consider the issue of route recommending based on the total PM2.5 exposure over the path length. Our work in route-recommendation domain is mainly related to the research in areas related to efficient path-finding in terms of path air quality and the length of the route. Here we review some of the previous research that has been important in carrying out this project.

There have been works which have addressed the problems related to real-time navigation. Studies have shown that by modelling the energy cost, energy aware vehicle routing frameworks can be achieved [7]. The authors in [7] modelled the road network as the directed graph and used the proposed framework for case studies in realistic road networks. In [8], the authors have proposed real-time navigation systems with a focus on energy optimal route calculation. The authors used real-time traffic information and tested the proposed system in real situations. Their algorithm is based on Dijkstra’s algorithm combined with A\* search speed up technique. Eco-routing navigation [9] frameworks have also been proposed which can recommend most eco-friendly routes by considering fuel consumption or overall emission. The route calculation is based on Dijkstra’s algorithm with binary heap priority queue.

In another work [10], the authors performed dynamic routing using PM10 data. They put the PM10 data on 1x1 km grid and used Open Source Routing Machine (OSRM) for routing. An important part of this work was reducing the calculation time when more data were added. Eco-routing problems have been a hot topic everywhere and there have been some important works which have addressed the issues like energy consumption [11], route choice [12] and environmental impacts [13]. In [14], the authors focus on finding a route that involves maximum physical activity, considering environmental and physical factors to compute the optimal route. In some works [15], authors have used online contents, which act as a geo-reference for recommending routes. Based on multiple user-generated location trajectories, the system recommends an efficient and balanced itinerary to the user. The evaluation was performed by testing the model in Beijing, China, where the results were generated by 125 users. The authors of [16] used machine learning techniques to successfully recommend the next tourist destination on the route.

The authors in [17] proposed a traffic routing system that supports cooperative routing services. Their system involved collecting traffic information from participating drivers with which the system could estimate on-road traffic conditions. There have been works that have studied PM2.5 exposure [18] among cyclists [19] and pedestrians and discussed about how natural and built environment [20] effects the PM2.5 exposure. In contrast to the above-mentioned works which focus on energy aware routing or understanding the exposure level based on transport mode, the hereby presented work focuses on recommending a health-optimal route based on overall PM2.5 concentration. We have used a new cost function that gives an overall exposure to PM2.5 particles along the road segments. We analyze the possibility of reducing the amount of inhaled PM2.5 particles and also compare the shortest paths and the health-optimal paths.

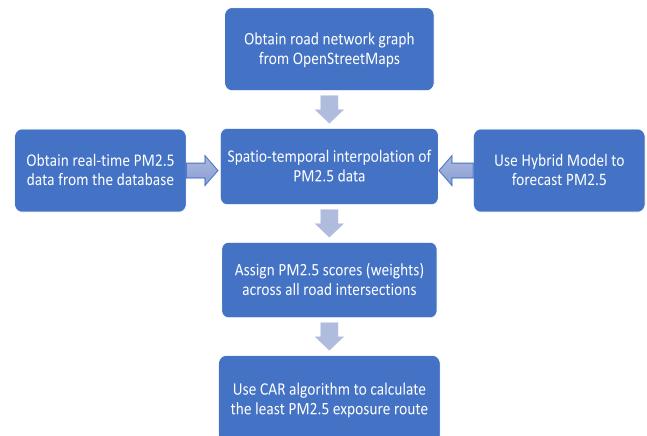
### III. METHODOLOGY

In this section, we will explain step by step the methodology we adopted to achieve our goal: to recommend users a path with minimum PM2.5 exposure. We accomplish it by following the steps shown in Figure 1. We start by obtaining a road network graph and then perform spatio-temporal interpolation of PM2.5 over the study area. This is followed by assigning PM2.5 scores across all the intersections of the road network, which will be used as weights when calculating the shortest path between the origin and destination.

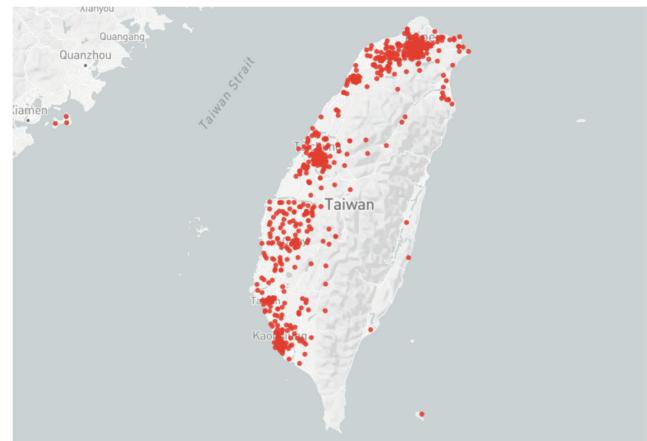
#### A. DATA INTEGRATION

The dataset used for this work contains opensource PM2.5 data. Originally, the PM2.5 dataset in an hourly format. The dataset contains readings from PM2.5 sensors deployed around Taiwan. The locations of the sensors are shown in Figure 2. The dataset has the following attributes: sensor id, day, o'clock, longitude, latitude and predicted

<https://pm25.lass-net.org/en/>



**FIGURE 1.** Flowchart of the proposed system.



**FIGURE 2.** Location of PM2.5 sensors in Taiwan.

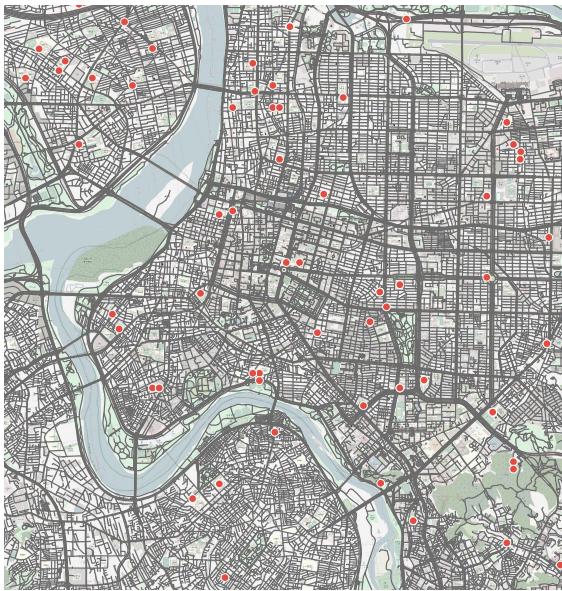
PM2.5 concentration. A road network graph of Taiwan is generated using the data obtained from OpenStreetMap. Figure 3 shows an example of a road network with the locations of the sensors for Taichung region. The boundary data for the region were obtained from a government open data website.

#### B. PM2.5 PREDICTION

Let us suppose that it takes a person two hours to travel from one location to another. So, in order to accurately recommend a path with the lowest total PM2.5 exposure, we first need to know how the PM2.5 varies over the time period of the journey. To do that, we would need to perform PM2.5 prediction [21] for the next hours. The prediction is performed using a Hybrid prediction model [22], which makes use of a Neural Network Autoregression (NNAR) and an Autoregressive Integrated Moving Average (ARIMA) model. The part would create a base to implement a spatio-temporal interpolation method to estimate the PM2.5 concentration values throughout the path.

PM2.5 can be considered as a time-series. A time-series can be composed of linear autocorrelation components

<https://gist-map.motc.gov.tw/>



**FIGURE 3.** Road network of Taipei.

and non-linear components. Combining a linear ARIMA model with non-linear NNAR model can help in understanding complex relationships in time-series data. In our case, PM2.5 time-series  $Z_t$  can be represented as shown in equation (1) where  $X_t$  represents the linear components and  $Y_t$  represents the non-linear components.

$$Z_t = X_t + Y_t \quad (1)$$

Initially, these two components have to be estimated from the PM2.5 data. The next stage is the application of the ARIMA model. In this stage, ARIMA would model the linear components and generates residuals. Let us call  $R_t$  the residuals generated at time  $t$  from the linear model. So it can be written as

$$R_t = Z_t - F_t \quad (2)$$

Here,  $F_t$  is the forecast value for time  $t$ . The residuals are then modelled using neural networks. If there are  $n$  input nodes, then the neural network model for residuals is

$$R_t = f(R_{t-1}, R_{t-2}, \dots, R_{t-n}) + e \quad (3)$$

The non-linear function  $f$  is defined by the neural network and  $e$  is the random error generated. Forecast from the neural network is generated and equation (1) is used to get the final output. Hybrid model uses NNAR (9,5,1) model which uses 9 lagged inputs and the hidden layer consists of 5 nodes. Two methods were tested for weighting the forecasts of the two contributing models. Initially, more weight was assigned to the model with better in-sample performance. Later on, we assigned equal weights to both contributing models. The prediction performance was better when both models were assigned equal weights. Table 1 shows the next 1 hour PM2.5 prediction results for 490 monitoring nodes

**TABLE 1.** Hourly PM2.5 Prediction Results.

Model	Avg. Relative Error	Accuracy (%)	Standard Deviation
Hybrid	0.21	79%	0.24

deployed in Taiwan. It can be observed that the Hybrid prediction model performs forecast with an average accuracy of 79%.

### C. SPATIO-TEMPORAL INTERPOLATION

The existing shortest-path finding algorithms generally use weights on the segments of the road network. Time and distance are usually chosen as the weights for the routing algorithms. The road network of a city is composed of different road segments and intersections. Each road segment has different attributes like length, speed limit and width. As shown in the previous section, the geographical location of the sensors is non-uniform. Also, because of the non-uniform nature of the road structure, it is hard to estimate the pollution level at different road segments. In order to do that we need to perform spatio-temporal data interpolation, which would give us the PM2.5 values for intersections at different segments of the road. Let us consider a road network comprising of nodes and set of edges that connect these nodes. Each edge between two nodes has two weights associated to it. One is road segment length and the other is the expected PM2.5 exposure. The weight function/cost function computes weights for the different intersections going towards the destination. It can be assumed that the travel time between two nodes is directly proportional to the distance. Also, the particles inhaled is proportional to the time of exposure [23]. Therefore, we multiply the particle concentration and the length of road segment to the get the cost function. Based on the time, the computed weight would replace the existing weight and the updated data would be used to calculate the shortest path.

Since the pollutant concentrations are recorded at certain locations at certain times, they have to be interpolated in time and space dimensions. Spatial prediction is used for predicting an unknown value at unsampled locations using known values in a set of observed locations. This process is called spatial interpolation. Many spatial interpolation methods are aimed at providing a suitable prediction for different cases under their own criteria [24]. These methods are commonly used as an integral parts of the Geographic Information System (GIS). The estimation is applied for terrain, temperature, chemical dispersion, pollution, precipitation and soil composition. Basically, all the spatial interpolation methods are based on Tobler's First Law of Geography, that is, they assume a stronger correlation between nearer positions between two components [25], including the two most commonly used methods, IDW (Inverse Data Weighting) [26], Kriging [27] and so on. Many studies have examined and compared IDW and Kriging performances, and the results are controversial [28]. They are similar in the way they weight surrounding measured values, but the major difference is that

**TABLE 2.** Run-time comparison between IDW and Kriging.

Interpolation Method	Run-time ( $\mu s$ )
IDW	1438
Kriging	3564.48

Kriging takes the entire dataset into account to design a fitted model, not only samples that are closer. But IDW depends on the distance of surrounding known samples only. Advantages of the IDW interpolation method are that it is simple and intuitive, and it needs less computation time. We compared the run time for interpolation using 400 monitoring stations using IDW method and Kriging. It was observed that the computation time significantly increased with the number of nodes for Kriging method. The comparison results is shown in Table 2. With this understanding, we designed and implemented IDW interpolation to estimate PM2.5 concentrations over the Taiwan street map. But this also means that the results can be strongly influenced by the outliers and sampling configuration. Many research studies examined the IDW method on pollutants issues [29]. The IDW method is simpler and more efficient when it comes to spatial interpolation [30].

Pollutant concentrations are a continuous space-time function, so our problem turns into a spatio-temporal interpolation problem. Spatio-temporal interpolation estimates the unknown values at unmeasured locations and times [31]. There are some studies that suggest combining temporal and spatial interpolation as an integration model [32]. Some studies suggest that, when spatial and temporal components are processed at the same time, temporal aggregation imposes a bias [33]. So in this paper, we deal with spatial and temporal interpolation separately.

We adapt the traditional spatial IDW method and utilize the following spatio-temporal interpolation formula:

$$Pa_{t+\varepsilon} = Pa_t \times (1 - \varepsilon) + Pa_{t+1} \times \varepsilon, \text{ and } \varepsilon \in [0, 1] \quad (4)$$

given that  $Pa_t$ ,  $Pa_{t+1}$  are the predicted value at location (a) and time  $t$  and  $t + 1$  are the hours.  $\varepsilon$  is the time between 0 and 1 hour, (i.e minute between 0 and 60). A naive way of calculating the value is through fitting linear regression model on PM2.5 data on the hour. After dealing with temporal interpolation, we used the temporal interpolation sampled measurement to calculate the upsampled measurement at every time:

$$Pn_{t+\varepsilon} = \sum_{i=1}^N \frac{\frac{1}{d_i}}{\sum_{j=1}^N \frac{1}{d_j}} P_i_{t+\varepsilon} \quad (5)$$

where  $N$  is the number of nearest neighbors with measured values surrounding the location. Then we choose 10 as the number of nearest neighbors.  $Pn_{t+\varepsilon}$  is the interpolation value at the unmeasured location (a) and time instance  $t + \varepsilon$ .  $d_i$  is the spatio-temporal Euclidean distance between every measured neighbor  $P_i$  and unmeasured  $P_n$ , that is, it is an inverse function that ranges from zero to one. Since it is a

spatio-temporal integrating model, Equations (5) is derived from Equations (4).

### 1) ALGORITHM

The spatio-temporal interpolation algorithm is aimed to generate the PM2.5 measurement for all the time periods and for every location in the study area. This algorithm has two parts as shown in Algorithm 1. The input is the location, time and the hourly predicted PM2.5 data. The results obtained from Algorithm 1 are then fed to the next algorithm. The first part deals with temporal interpolation and the second part is for spatial interpolation. In the temporal interpolation, we have the raw PM2.5 data in hourly format. The raw data are used to interpolate the data and get values for time duration in minutes. This is an important step as some of the journeys might last less than an hour. So to deal with those scenarios, we have to get the PM2.5 data for every minute for different locations. The PM2.5 sensors are randomly located in the test area. Some might be close to the input location and some might be far away. So, first we find the top ten nearest sensors of the input location. Then we generate the inverse distance that is multiplicative inverse of each distance. After that we sum up the inverse distance between these nearest sensors to the input location specified longitude and latitude. The inverse distance is then divided by the sum, and these sum up to 1. These inverse distances divided by the sum are the weights for predicting the input location value. It has to be taken into consideration that with time the location of the user would change. So, we put focus on designing an algorithm which would calculate the PM2.5 values for every intersection of the road network so that depending on the predicted arrival time, we can get an idea of how much PM2.5 concentration would be there.

---

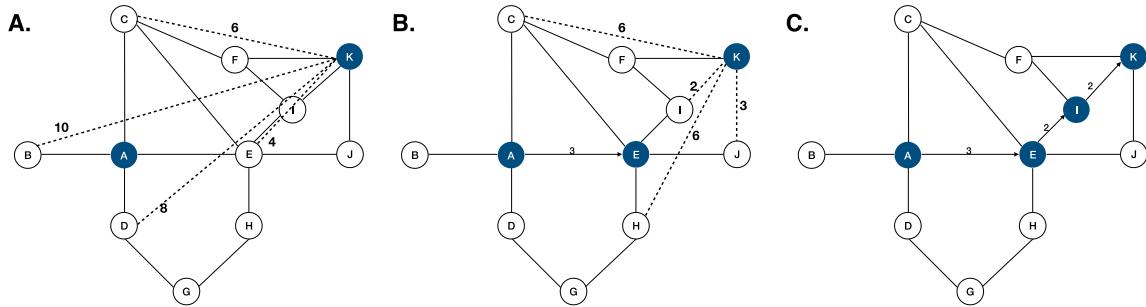
#### Algorithm 1 Spatio-Temporal Interpolation Algorithm

---

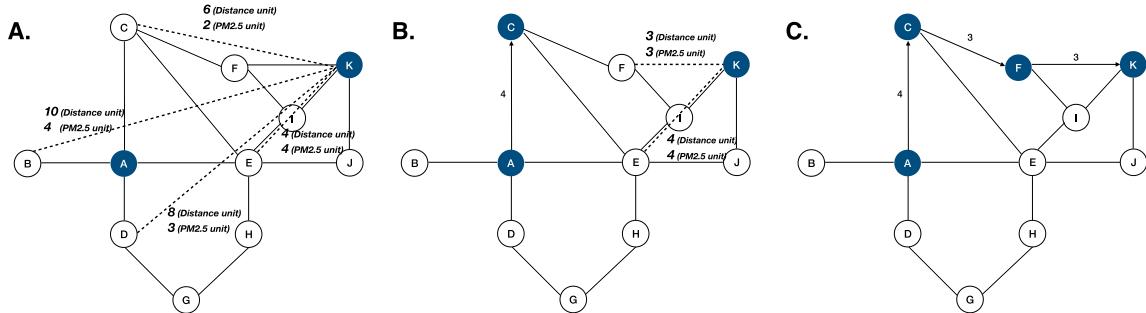
**Input:** location without PM2.5 value, sensor data

**Output:** PM2.5 value at particular time and location

- 1: *TemporalInterpolation* :
  - 2: **if** time is not in the sample **then**
  - 3:     **for** *allstations* **do**
  - 4:         estimation  $\leftarrow$
  - 5:         linear interpolation of the two adjacent sampled data
  - 6:     **end for**
  - 7: **end if**
  - 8:
  - 9: *SpatialInterpolation* :
  - 10: PM2.5 value  $\leftarrow$  the nearest sensor data after temporal interpolation
  - 11:
  - 12: **return** PM2.5 value at particular time and location
  - 13:
-



**FIGURE 4.** Finding the shortest path.



**FIGURE 5.** Finding the health-optimal shortest path.

#### D. PATH SELECTION

In order to take the air pollution exposure concentration into account at the paths scheduling phase, we propose a shortest path algorithm based on PM2.5 concentration. This would inform the (drivers) not only about the current shortest paths to destination but also about paths with best air quality. We take into consideration the critical traffic conditions in city, where the transport networks typically range from some hundreds to many thousands of nodes. We take road networks for Taiwan and the paths data are obtained from OpenStreetMap (OSM), which supports the complete transport networks. In most of the shortest-path finding problems, classical algorithms like Dijkstra's algorithm [34] is used. One disadvantage of Dijkstra's algorithm is that when searching for the shortest-path, it searches uniformly in all directions. Dijkstra's algorithm is optimal but in this work, for path-finding we have used A\* algorithm as it corrects the shortcoming of Dijkstra's algorithm by adding some heuristics based on the knowledge of the geographical location of the target [5]. When dealing with the air quality data, the heuristic would be the product of the distance and PM2.5 data.

The following example is used to illustrate the shortest-path finding algorithm. In Figure 4, the task is to go from the location A to the destination K. To start with, the algorithm will search all the intersections (it is denoted as S set) that location A might meet isotopically, i.e.  $S = \{B, C, D, E\}$  as can be seen in Figure 4-A. After comparing the distance from every intersection to the final destination K,

we will choose intersection with the lowest distance. That is, location E is nearest from destination K i.e. 4 units (units here refer to the distance in kilometers) in this example. In the next step, the algorithm would continue tracking the path towards K from E and at the same time it will also record the distance from location E to starting location A, i.e. 3 units. Repeating the process, it will compare all the distance between isotopic intersections of location E and destination K as shown in the Figure 4, then, it continues at location I and records the distance between location I and last position E. When the intersection of current position is destination, the process will automatically stop as shown in the Figure 4-C.

Using the shortest-path finding approach mentioned previously, our algorithm incorporates the PM2.5 weighting model. The PM2.5 weighting model would assign weights at the intersections which are completely based on PM2.5 value recorded at that particular time. To make the system more efficient, we take into consideration different transportation modes into account. They include car, scooter, bicycle and walking modes. In the next step, we can use the average speed of different transportation mode to calculate the distances of the next potential move and the arriving time. This can be later used to calculate the PM2.5 measurement of the next potential location. Initially the algorithm compares distance between every intersection and the destination. In the next step, the distance is multiplied with the PM2.5 measurement at the intersection at that moment of time as shown in the Figure 5. After choosing location with the lowest value of distance multiplied with the PM2.5 measurement, it continues to

**Algorithm 2** Shortest Health-Optimal Path Algorithm

**Input:** origin, destination, start time, speed, intersections stored in OSM data

**Output:** route, arriving time, total PM2.5 exposure, routing distance

```

1: initial route ← start location
2: repeat
3:   for all intersections of last node in Route do
4:     CalculatingThePM2.5 ofIntersection :
5:     duration ← distance / speed
6:     PM25.value ← call Algorithm 1 by duration
7:
8:     ChooseTheIntersection :
9:     distance ← from intersection to destination
10:    judgment value ← PM2.5 value * distance
11:    #choose the intersection with lowest judgment
12:    value as next node#
13:  end for
14:
15:  PM2.5 exposure ← PM2.5 value * distance
   between last node and chosen intersection
16:
17:  Record the 'distance' and 'PM2.5 exposure'
18:
19:  if the current node is blind alley then
20:    track back and restart from the 2nd low judgment
21:    value
22:  end if
23:
24: until No intersection between destination and last node
   and record last distance and PM2.5.
25:
26: return Total exposure, distance, duration and route
27:

```

find the location of lowest value and records the overall value based on distance, PM2.5 and the time. When the intersection of current position is destination, the process stops as shown in Figure 5.

**1) ALGORITHM**

We feed the output from Algorithm 1 to Algorithm 2. Algorithm 2 deals with finding the shortest health-optimal path from the origin to the destination. Initially, we input the origin and final destination and decide what mode of transport would be used. The choices being walking, riding a scooter, driving a car and cycle. The journey duration is calculated based on the total distance and speed. For speed, we consider the default speed limits for different vehicle classes. According to the transport mode the arrival time to the next intersection can be projected. In this algorithm, we retrieved the road network data from OSM that contains road networks and every intersection. So to begin with, for the current position, we retrieve the intersection from the OSM data. The next step

involves calling the spatio-temporal interpolation algorithm to get the PM2.5 data, and the measurements times the distance between the intersection and destination. According to the result, the intersection with the lowest measurement is selected and the process is repeated again until there is no more intersection between the destination and the last node.

**IV. IMPLEMENTATION**

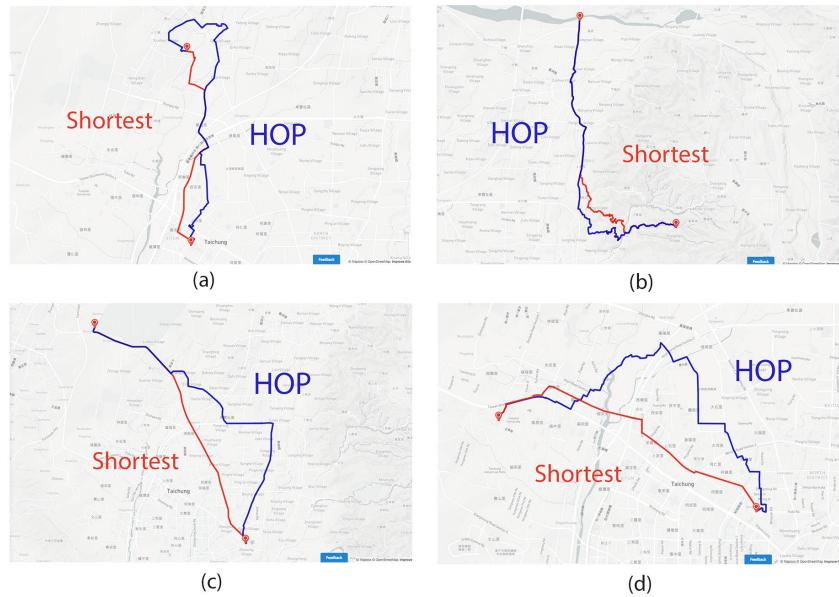
In this section, we describe the implementation of the CAR algorithm and also discuss the results. We would also describe our evaluation approach.

**A. RESULTS**

Here, we would discuss the results which are obtained after the implementation of CAR algorithm for different transport modes. Initially we would show both the paths (shortest as well as the health-optimal path) on a map. It would make it simple for a user to see how the two routes are different (in terms of distance, time, PM2.5 exposure) from each other. Figure 6-(a) shows the shortest path (SP) and the health-optimal path (HOP) from Taichung High Speed Rail station to National Taichung Theater. The mode of transport chosen is a scooter. The total PM2.5 exposure for the shortest path and health-optimal path are  $45 \text{ km} * \mu\text{g}/\text{m}^3$  and  $18.7 \text{ km} * \mu\text{g}/\text{m}^3$  respectively. Figure 6-(b) shows the comparison of bicycle routes from Houfeng Bike Trail Starting Point to Dakeng Scenic Area. The total PM2.5 exposure for the shortest path and health-optimal path are  $44.9 \text{ km} * \mu\text{g}/\text{m}^3$  and  $37.4 \text{ km} * \mu\text{g}/\text{m}^3$  respectively. In Figure 6-(c), the case shows car routes from Taichung Train Station to Taichung Airport(24.252998, 120.598701). The total PM2.5 exposure for the shortest path and health-optimal path are  $63.6 \text{ km} * \mu\text{g}/\text{m}^3$  and  $49.7 \text{ km} * \mu\text{g}/\text{m}^3$  respectively. In Figure 6-(d), walking routes are shown from Luce Chapel to Science Museum, Taichung. The recorded PM2.5 exposure for shortest path and health-optimal path are  $63.5 \text{ km} * \mu\text{g}/\text{m}^3$  and  $26.9 \text{ km} * \mu\text{g}/\text{m}^3$  respectively. These are just a few cases to illustrate how the routes on the map can be visualized for different transport modes.

**B. EVALUATION**

Our aim is to recommend shortest paths that lead to low exposure of PM2.5 during the journey. To determine the effectiveness of our proposed method at meeting the goal, we perform evaluation to answer two main questions. First, does the CAR algorithm recommends routes which have low PM2.5 exposure as compared to other navigation application? Second, are the low PM2.5 exposure paths longer than the shortest paths? To answer these questions, we select 20 points of interest within Greater Taipei Area and simulate trips between all the pairs of these nodes. The results are then compared with that of Google Directions API. The analysis of the results showed us that for most the cases there is a decrease in the overall PM2.5 exposure for the routes suggested by the CAR algorithm. Table 3 illustrates an example of trade-off between exposure reduction and distance increase



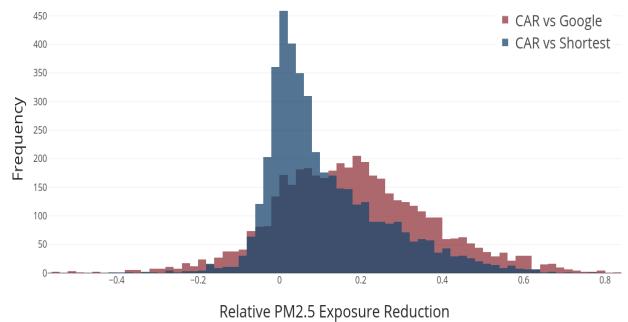
**FIGURE 6.** Map showing shortest path (SP) and health optimal path (HOP) for different transport modes (a) scooter, (b) bicycle, (c) driving and (d) walking.

**TABLE 3.** Exposure Reduction Versus Distance Increase.

Euclidean Distance Group	No. of Simulated Trips	Distance Decrease/Increase	Exposure Reduction	Execution Time (Seconds)
<4km	1,024	-1.8%	8.2%	0.0010
4-6km	1,078	+2.1%	9.5%	0.0241
6-8km	718	+4.4%	19.7%	0.1978
>8km	1,544	+6.0%	22.0%	0.5395
All combined	4,364	+2.4%	17.1%	0.2296

between CAR algorithm and Google's fastest route. On an average, the PM2.5 exposure is lower than relative distance increase for all the distance groups. For 4,364 simulated trips, the average PM2.5 exposure reduction is 17.1% and distance increase is 2.4% on an average. Also, it can be observed that for shorter journeys (less than 4km), the relative journey distance decreases by 1.8%, whereas for the other categories the relative distance increase is less than 6%. It has to be noted that the results are averaged for across different location pairs, transport modes and time frames and trade-off for individual cases might vary. The average execution time is given to showcase the efficiency of our algorithm. The graph of Greater Taipei Area contains 129,403 nodes and 181,669 edges, whereas the graph we used in production, which cover the whole Taiwan, has 640,545 nodes and 886,104 edges.

Figure 7 shows a plot for personal exposure reduction with CAR algorithm when compared to Google API results and Shortest path (Dijkstra's) algorithm. The y-axis shows the number of cases (trips) and the x-axis shows the relative PM2.5 exposure reduction. For most of the cases, the results are towards the positive side showing reduction in PM2.5 exposure. We also compared our work with similar state of the art works related to clean air routing. It won't be fair to say that our system outperforms other



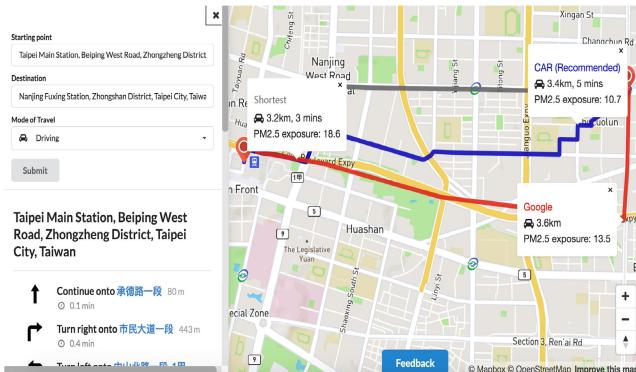
**FIGURE 7.** Plot for distribution of exposure reduction between CAR and Google, CAR and Shortest path algorithm.

state of the art works as all these works were conducted under different settings and with different data sets. Nevertheless, such comparisons give an idea about the proposed method's overall performance. The results are shown in Table 4. Hasenfratz *et al.* [35] proposed health-optimal routing algorithm and considered exposure to ultra fine particles. Stolfi and Alba [36] addressed green house gas emission, fuel consumption and travel times in four European cities. We compared the results for particulate matter in this case. Hatzopoulou *et al.* [37] presented a web-based tool which reduces exposure to traffic pollution (nitrogen dioxide)

**TABLE 4.** Comparison with state of the art works.

Model	Testing Area	Simulated Trips	Pollutants	Exposure Reduction	Distance Increase
CAR	Big Taipei	4,364	PM2.5	17.1%	2.4%
Hasenfratz et al.[35]	Zurich	1,000	UFP	7.1%	6.4%
Stolfi et al.[36]	Malaga, Stockholm, Berlin and Paris	200	PM	4.1%	60.5%
Hatzopoulou et al.[37]	Montreal	2,307	$NO_2$	4%	<1km (Average)*

\* Distance increase percentage not available for this work  
PM: Particulate Matter, UFP: Ultrafine Particles,  $NO_2$ : Nitrogen Dioxide

**FIGURE 8.** Application user interface.

for cyclists. It can be observed from the table that in all the works there is a trade-off between exposure reduction and journey length.

We also follow another way to tackle performance evaluation. 1) We have tested CAR algorithm to design a route recommendation application for Taiwan as shown in Figure 8. The application provides options including route recommended by the CAR algorithm, by Google Directions API (fastest path) and also by the conventional methods like Dijkstra's. Depending on the personal preference, the user can select any route out of three routes. It can be observed from Figure 8 that our algorithm successfully recommends routes with low personal PM2.5 exposure as compared to other options. 2) We have a feedback section on the application through which users can give us a feedback about the application and that constantly helps us to monitor the performance of the application. The application is currently in the testing stage and would be available online once the development phase is done.

## V. CONCLUSION AND FUTURE WORK

In this paper, we demonstrate a model which considers PM2.5 concentration to compute a health-optimal path to the destination. We propose the CAR algorithm that provides path navigation with minimal PM2.5 exposure on the move. PM2.5 data are first spatially and temporally interpolated on the road networks in Taiwan. The model is then implemented for real-world travel scenarios with transport modes like driving, cycling, riding a scooter and walking. A detailed comparative analysis is also done which shows

<https://car.lass-net.org>

the comparison between the path length, journey time, total PM2.5 exposure for the two routes recommended by CAR algorithm, Google Maps and Shortest path algorithm. However, it has to be understood that the difference between the health-optimal route and the shortest route significantly depends on the origin and destination. For some cases the difference between the shortest path and the health-optimal path is not very much but for some cases it significantly differs. In addition, the results suggest that there is a trade-off between PM2.5 exposure reduction and journey length.

The future work would include extending this work to multiple new dimensions. One task would be to add an option that for cars and other vehicles that emit pollution, they could be recommended a route where they do not add to already heavily polluted routes. Another idea is to include tourist hot-spots on the map that would give the user more options like whether to take a more touristy path or stick to the path with best air quality. An important step would be to implement the algorithm to a mobile application and perform user testing and get a constructive feedback.

## REFERENCES

- [1] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 112–121, Apr. 2014.
- [2] G. Kiesewetter, W. Schoepp, C. Heyes, and M. Amann, "Modelling PM<sub>2.5</sub> impact indicators in Europe: Health effects and legal compliance," *Environ. Model. Softw.*, vol. 74, pp. 201–211, Dec. 2015.
- [3] L.-J. Hung, S.-S. Tsai, P.-S. Chen, Y.-H. Yang, S.-H. Liou, T.-N. Wu, and C.-Y. Yang, "Traffic air pollution and risk of death from breast cancer in Taiwan: Fine particulate matter (PM<sub>2.5</sub>) as a proxy marker," *Aerosol Air Qual. Res.*, vol. 12, no. 2, pp. 275–282, 2012.
- [4] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "AirCloud: A cloud-based air-quality monitoring system for everyone," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, New York, NY, USA, 2014, pp. 251–265.
- [5] R.-H. Li, L. Qin, J. X. Yu, and R. Mao, "Optimal multi-meeting-point route search," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 770–784, Mar. 2016.
- [6] F. Kellner, A. Otto, and C. Brabander, "Bringing infrastructure into pricing in road freight transportation—A measuring concept based on navigation service data," *Transp. Res. Procedia*, vol. 25, pp. 794–805, Jan. 2017.
- [7] Z. Yi and P. H. Bauer, "Optimal stochastic eco-routing solutions for electric vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3807–3817, Dec. 2018.
- [8] A. Cela, T. Jurik, R. Hamouche, R. Natowicz, A. Reama, S.-I. Niculescu, and J. Julien, "Energy optimal real-time navigation system," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 3, pp. 66–79, Oct. 2014.
- [9] K. Boriboonsomsin, M. J. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1694–1704, Dec. 2012.

- [10] S. Müller and A. Voisard, "Air quality adjusted routing for cyclists and pedestrians," in *Proc. 1st ACM SIGSPATIAL Int. Workshop GIS Emergency Manage.*, New York, NY, USA, 2015, p. 19.
- [11] G. De Nunzio, L. Thibault, and A. Sciarretta, "Bi-objective eco-routing in large urban road networks," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst.*, Oct. 2017, pp. 1–7.
- [12] Y. M. Nie and Q. Li, "An eco-routing model considering microscopic vehicle operating conditions," *Transp. Res. B, Methodol.*, vol. 55, pp. 154–170, Sep. 2013.
- [13] E. Yao and Y. Song, "Study on eco-route planning algorithm and environmental impact assessment," *J. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 42–53, 2013.
- [14] M. H. Sharker, H. A. Karimi, and J. C. Zgibor, "Health-optimal routing in pedestrian navigation services," in *Proc. First ACM SIGSPATIAL Int. Workshop GIS Public Health*, New York, NY, USA, 2012, pp. 1–10.
- [15] H. Yoon, Y. Zheng, X. Xie, and W. Woo, "Smart itinerary recommendation based on user-generated GPS trajectories," in *Proc. 7th Int. Conf. Ubiquitous Intell. Comput.* Berlin, Germany: Springer, 2010, pp. 19–34.
- [16] R. Baraglia, C. I. Muntean, F. M. Nardini, and F. Silvestri, "LearnNext: Learning to predict tourists movements," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage.*, New York, NY, USA, 2013, pp. 751–756.
- [17] R. Liu, H. Liu, D. Kwak, Y. Xiang, C. Borcea, B. Nath, and L. Iftode, "Balanced traffic routing: Design, implementation, and evaluation," *Ad Hoc Netw.*, vol. 37, pp. 14–28, Feb. 2016.
- [18] I. Y. Hernández-Paniagua, G. L. Andraca-Ayala, U. Diego-Ayala, L. G. Ruiz-Suarez, J. C. Zavala-Reyes, S. Cid-Juárez, L. Torre-Bouscoulet, L. Gochicoa-Rangel, I. Rosas-Pérez, and A. Jazcilevich, "Personal exposure to PM<sub>2.5</sub> in the megacity of mexico: A multi-mode transport study," *Atmosphere*, vol. 9, no. 2, p. 57, 2018.
- [19] V. S. Brand, P. Kumar, A. S. Damascena, J. P. Pritchard, K. T. Geurs, and M. de FatimaAndrade, "Impact of route choice and period of the day on cyclists' exposure to black carbon in London, Rotterdam and São Paulo," *J. Transp. Geography*, vol. 76, pp. 153–165, Apr. 2019.
- [20] J. Gilliland, M. Maltby, X. Xu, I. Luginaah, and T. Shah, "Influence of the natural and built environment on personal exposure to fine particulate Matter (PM<sub>2.5</sub>) in Cyclists using city designated bicycle routes," *Urban Sci.*, vol. 2, no. 4, p. 120, 2018.
- [21] S. Mahajan, L.-J. Chen, and T.-C. Tsai, "Short-term PM<sub>2.5</sub> forecasting using exponential smoothing method: A comparative analysis," *Sensors*, vol. 18, no. 10, p. 3223, 2018.
- [22] S. Mahajan, H.-M. Liu, T.-C. Tsai, and L.-J. Chen, "Improving the accuracy and efficiency of PM<sub>2.5</sub> forecast service using cluster-based hybrid neural network model," *IEEE Access*, vol. 6, pp. 19193–19204, 2018.
- [23] H. Kruize, O. Hänninen, O. Breugelmans, E. Lebret, and M. Jantunen, "Description and demonstration of the EXPOLIS simulation model: Two examples of modeling population exposure to particulate matter," *J. Exposure Sci. Environ. Epidemiol.*, vol. 13, no. 2, pp. 87–99, 2003.
- [24] J. Li and A. D. Heap, "Spatial interpolation methods applied in the environmental sciences: A review," *Environ. Model. Softw.*, vol. 53, pp. 173–189, Mar. 2014.
- [25] T. J.-J. Li, S. Sen, and B. Hecht, "Leveraging advances in natural language processing to better understand Tobler's first law of geography," in *Proc. 22nd ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, New York, NY, USA, 2014, pp. 513–516.
- [26] L. de Mesnard, "Pollution models and inverse distance weighting: Some critical remarks," *Comput. Geosci.*, vol. 52, pp. 459–469, Mar. 2013.
- [27] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY, USA: Springer, 2012.
- [28] G. Gong, S. Matnevada, and S. E. O'Bryant, "Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas," *Environ. Res.*, vol. 130, pp. 59–69, Apr. 2014.
- [29] N. Ya'acob, A. M. Azize, N. M. Adnan, A. L. Yusof, and S. S. Sarnin, "Haze monitoring based on air pollution index (API) and geographic information system (GIS)," in *Proc. IEEE Conf. Syst., Process Control*, Dec. 2016, pp. 7–11.
- [30] M. A. Azpurua and K. D. Ramos, "A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude," *Prog. Electromagn. Res.*, vol. 14, pp. 135–145, Jul. 2010.
- [31] M. Kilibarda, T. Hengl, G. Heuvelink, B. Gräler, E. Pebesma, M. P. Tadić, and B. Bajat, "Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution," *J. Geophys. Res., Atmos.*, vol. 119, no. 5, pp. 2294–2313, 2014.
- [32] L. Li, T. Losser, C. Yorke, and R. Piltner, "Fast inverse distance weighting-based spatiotemporal interpolation: A Web-based application of interpolating daily fine particulate matter PM<sub>2.5</sub> in the contiguous U.S. using parallel programming and k-d tree," *Int. J. Environ. Res. Public Health*, vol. 11, no. 9, pp. 9101–9141, 2014.
- [33] M. Percoco, "Temporal aggregation and spatio-temporal traffic modeling," *J. Transp. Geography*, vol. 46, pp. 244–247, Jun. 2015.
- [34] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.
- [35] D. Hasenfratz, T. Arn, I. de Concini, O. Saukh, and L. Thiele, "Health-optimal routing in urban areas," in *Proc. 14th Int. Conf. Inf. Process. Sensor Netw.*, 2015, pp. 398–399.
- [36] D. H. Stolfi and E. Alba, "Eco-friendly reduction of travel times in European smart cities," in *Proc. Annu. Conf. Genetic Evol. Comput.*, 2014, pp. 1207–1214.
- [37] M. Hatzopoulou, S. Weichenthal, G. Barreau, M. Goldberg, W. Farrell, D. Crouse, and N. Ross, "A web-based route planning tool to reduce cyclists' exposures to traffic pollution: A case study in Montreal, Canada," *Environ. Res.*, vol. 123, pp. 58–61, May 2013.

• • •