



## A short-distance healthy route planning approach

Li-Na Gao<sup>a</sup>, Fei Tao<sup>a,b,c</sup>, Pei-Long Ma<sup>a</sup>, Chen-Yi Wang<sup>a</sup>, Wei Kong<sup>a</sup>,  
Wen-Kai Chen<sup>a</sup>, Tong Zhou<sup>a,\*</sup>

<sup>a</sup> School of Geographical Sciences, Nantong University, Nantong, 226007, China

<sup>b</sup> Department of Geography, University of Wisconsin-Madison, Madison, WI, 53706, USA

<sup>c</sup> Key Laboratory of Virtual Geographical Environment, MOE, Nanjing Normal University, Nanjing, 210046, China



### ARTICLE INFO

**Keywords:**

Short-distance  
Healthy route planning  
PM2.5 concentration  
BPNN  
Dynamic dijkstra algorithm

### ABSTRACT

With the development of the economy and the accumulation of social wealth, urban residents have begun to give more attention to quality of life than to material needs. Consequently, environmental factors that affect human health, such as air quality, have become a new focus when traveling. A travel scheme with relatively low pollutant exposure to travelers can not only improve their health and satisfy their goals but also benefit social stability and sustained progress. However, low spatiotemporal resolution and coarse spatial details of the distribution of PM2.5 (particles with an aerodynamic diameter of 2.5  $\mu\text{m}$  or less) educe the success rate of short-distance healthy travel route planning. This paper proposes a short-distance healthy route planning approach that is based on PM2.5 retrieval with high spatiotemporal resolution and a dynamic Dijkstra algorithm. First, fine spatial resolution images, meteorological data, and socioeconomic data are used to retrieve the spatial distribution of PM2.5 concentration in hourly intervals via a back-propagation neural network (BPNN). Second, a PM2.5 concentration value is obtained for each road section, and the harm degree to the human body is calculated as the weight of each road section. Then, the healthiest route is obtained based on the Dijkstra algorithm. Finally, the route planning effectiveness is verified by comparing the PM2.5 potential dose descending rate between the healthy route and the shortest route. The results show that the coefficient of determination ( $R^2$ ) of the PM2.5 retrieval approach that is based on multisource data and BPNN is 0.85, which can ensure the accuracy of the PM2.5 data at the street level. On this basis, the potential dose reduction rate of the healthy route can reach up to 20%, which proves that our approach can perform well. It can effectively improve the safety of travel and alleviate the anxiety that is caused by air pollution. In addition, it provides an easy implementation strategy for software for health management.

### 1. Introduction

With the enrichment of material life, people pursue a better and healthy lifestyle and give more attention to air pollution problems during travel and other occasions in life (Orru et al., 2016; Zuurbier et al., 2011). Exposure to air pollutants endangers human health, and its harm is more severe than previously thought (Brook et al., 2010; Kim et al., 2015). PM2.5, which a typical air pollutant, has become the fourth leading cause of death in China (Lim et al., 2012). The components of PM2.5 have small diameters but large surface

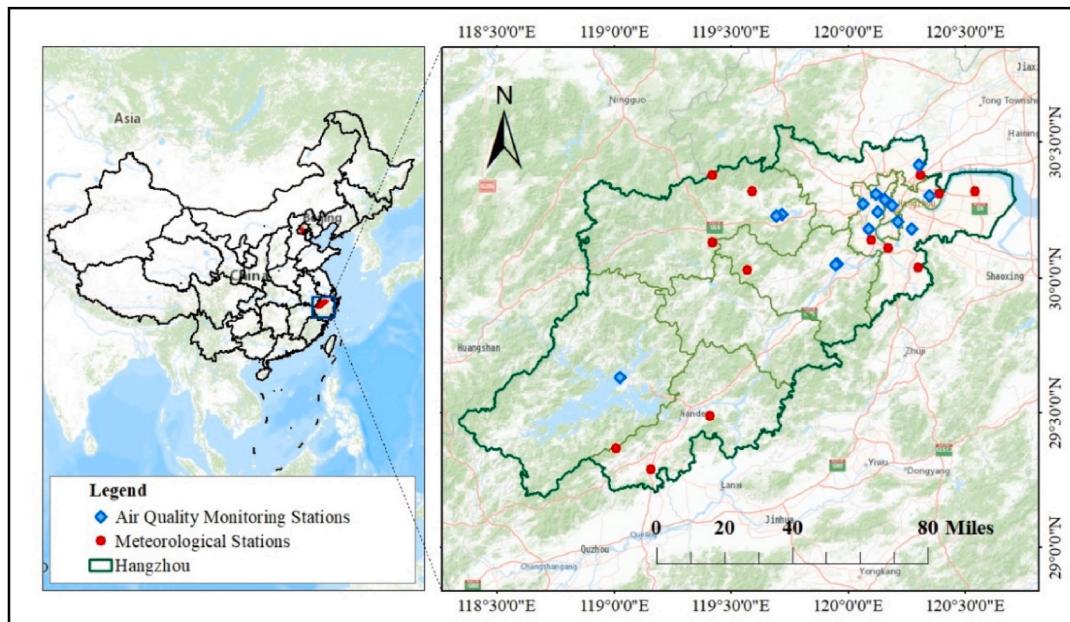
\* Corresponding author.

E-mail address: [zhoutong@ntu.edu.cn](mailto:zhoutong@ntu.edu.cn) (T. Zhou).

areas and, therefore, may be capable of carrying various toxic particles that can pass through the nose, reach the lungs and damage other organs of the body (Xing et al., 2016). Therefore, route planning that aims at reducing PM2.5 hazards can increase travel quality (Hertel et al., 2008), increase people's life expectancy and satisfy people's desire to enjoy a healthy life. Available healthy route planning methods focus on the improvement of route planning algorithms and ignore the significance of basic pollutant concentration data, which renders short-distance healthy route analysis challenging. The length of a short-distance route is approximately 10 km or less because commuting distances range from 9.1 to 18.1 km for all ten megacities in China (Zhang et al., 2021). PM2.5 distribution data with detailed spatial heterogeneity and a route planning algorithm collectively constitute short-distance healthy route planning.

Previous studies on healthy route planning can be divided into two main categories. The methods in the first categories classify roads and use pollutant monitors to measure pollutant exposure on various types of roads. By comparing exposures levels, roads with low values are determined to be healthy routes (Jarjour et al., 2013). These methods do not provide specific travel routes and are only suitable for people who are familiar with the city. From a regional perspective, these methods have difficulty fully covering the entire study area. The methods in the other category realize healthy route planning based on pollution distribution data that are obtained by using multi-data (Luo et al., 2018), in combination with a land use regression model (LUR) (Hatzopoulou et al., 2013; B. Zou et al., 2020), operational street pollution Model (OSPM) (Hertel et al., 2008) or interpolation method (Zahmatkesh et al., 2015; Z. Zou et al., 2020). Then, the Dijkstra or dynamic Dijkstra route planning algorithm is used to plan healthy routes (Mahajan et al., 2019). Indicators such as traffic volume, potential dose (Wang et al., 2018) that can directly or indirectly characterize the degree of harm to the human body are used as road network weights. These collectively lead to a healthy travel route. However, these methods do not take full advantage of the current advanced pollutant retrieval technology, hence, the accuracy of the pollutant distribution is poor and the credibility of the planned health route is low. Consequently it is difficult to plan healthy routes when the starting and ending points are close.

High spatiotemporal resolution of PM2.5 distribution data, which can be used to precisely distinguish the PM2.5 concentration on close roads and track PM2.5 distribution changes in time, are vital to healthy travel route planning. Traditional methods usually rely on the interpolation of fixed monitoring station values to assess the spatial distribution of PM2.5 concentration. However, these methods are limited by the numbers and locations of the stations, hence, the accuracy of the interpolation results is low (Yang et al., 2020). To solve this problem, scholars have installed substantial amounts of monitoring equipment in cities, but such equipment is often too costly (Higgins et al., 2019). Recently, the retrieval model of PM2.5 has centered around the use of aerosol optical depth (AOD) information that is obtained from remote sensing images as a key predictor of fine particulate matter levels (Lin et al., 2015). The result are affected by input variables and modeling methods. As input variables, the AOD products (Chen et al., 2018; Unnithan and Gnannappazham, 2020), precipitation, wind speed (Xu et al., 2020), relative humidity (Yu et al., 2019), temperature, and vegetation coverage (Jin et al., 2020; Yan et al., 2021) in relation to PM2.5 all form a multi-source data set. The more comprehensive the set of input variables is, the higher the accuracy and robustness are. As modeling methods, multiple linear regression (MLR) (Stadlober et al., 2008), support vector machine (Drucker et al., 1997; Zhu et al., 2018), random forest (Hu et al., 2017), and neural network (Voukantis et al., 2011) are commonly used in related research. However, the spatial resolution of AOD products is at the kilometer level, which makes it difficult to reflect the spatial distribution of PM2.5 at the street level. In addition, the AOD products are processed by the atmospheric transmission model, and the errors that are generated during the processing will be transmitted to the PM2.5 retrieval



**Fig. 1.** Study area and distributions of stations.

results.

In general, poor timeliness and low resolution of pollutant distribution results make planning healthy routes for short-distance travel impossible; hence, meeting the practical demand for healthy travel is difficult. This paper contributes to research in this area in three aspects:

First, a low-cost short-distance healthy route planning method that can be implemented in cities based on PM2.5 distribution data with a high spatiotemporal resolution is developed, which meets the needs of people for short-distance healthy travel. Second, hourly PM2.5 concentration data are integrated with the dynamic Dijkstra algorithm to update real-time healthy routes, which can help people avoid air pollution and maximally protect their health. Finally, this paper explores factors that influence the accuracy of the short-distance healthy route.

The remainder of this paper is organized as follows. In Section 2, the study area and datasets are briefly introduced, and the methodology of the proposed model is described. The experimental results are presented and discussed in Section 3, and the conclusions of this study are presented in Section 4.

## 2. Data and method

### 2.1. Study area

Hangzhou (29.25°–30.5° N, 118.34°–120.75° E), which is the capital city of Zhejiang Province in eastern China, is selected as the study area. There are 14 air quality monitoring stations and 13 meteorological stations. Stations are more abundant than in other cities; hence, more data can be used for model training and testing, which improves the accuracy of PM2.5 retrieval and, thus, the accuracy of

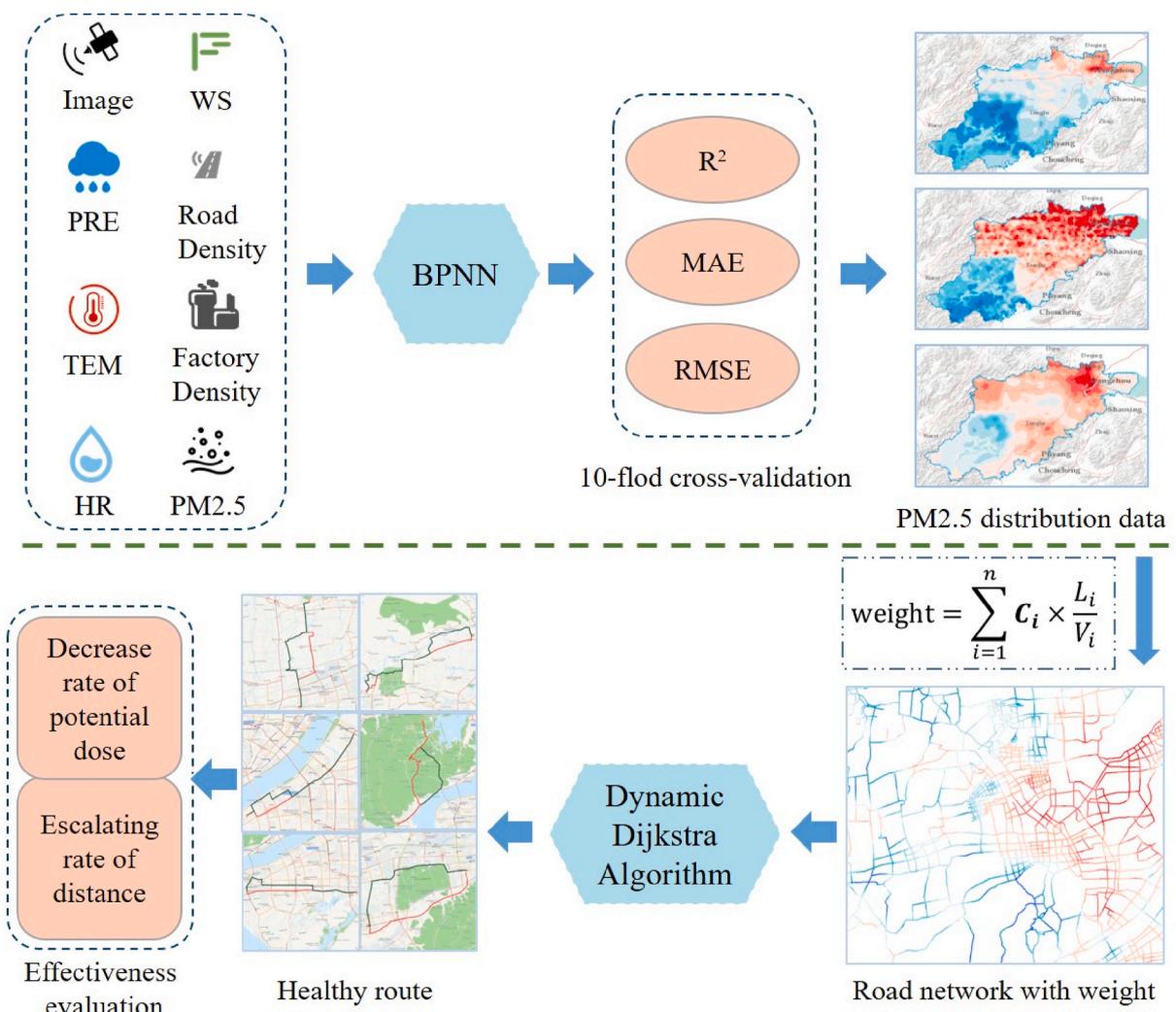


Fig. 2. Workflow of the proposed approach.

healthy path planning. The distributions of these stations are shown in Fig. 1.

## 2.2. Dataset and data processing

Daily PM2.5 data and meteorological data from 2015 to 2019 are used in this experiment. The PM2.5 concentration data are obtained from the national urban air quality real-time platform (<http://106.37.208.233:20035/>). The meteorological data including temperature, precipitation, relative humidity, and wind speed are obtained from the China Meteorological Administration (<http://www.cma.gov.cn>). Because the numbers and geographic locations do not match, the relationships between air quality monitoring stations and meteorological stations are linked via a Voronoi polygon. It is widely used in spatial adjacency analysis and is established by Russian mathematician Georgy Fedoseevich Voronoi (1908).

The Landsat 8 OLI C1 Level-1 product covers 2015 to 2019 and is obtained from the NASA official website (<https://earthexplorer.usgs.gov/>).

The preprocessing includes radiometric correction and atmospheric correction and is implemented on the PIE-Basic platform. Then, the normalized difference vegetation index (NDVI) of Hangzhou NDVI is obtained. NIR and R refer to the near-infrared reflectance and red reflectance, respectively.

$$NDVI = \frac{NIR - R}{NIR + R} \quad (1)$$

The considered socioeconomic data consist of the factory density and road density. The point data of the factories are acquired by Baidu Map API and the road network data are obtained from the OpenStreetMap official website (<https://www.openstreetmap.org/>).

The factory density and the road density of Hangzhou are calculated as follows:

$$D_f = \frac{Count}{\pi r^2} \quad (2)$$

$$D_r = \frac{length}{\pi r^2} \quad (3)$$

where  $D_r$  and  $D_f$  denote the road density and factory density, respectively;  $r$  is the self-defined neighborhood radius;  $length$  is the total length of roads in a circle that is drawn around each raster cell center with radius  $r$ ; and  $Count$  is the number of factory points in the circle that is drawn around each raster cell center with radius  $r$ .

## 2.3. Methodology

The workflow of the proposed approach is composed of four steps, as illustrated in Fig. 2. First, multisource data such as remote sensing images, precipitation, humidity, and temperature data are integrated to build a fine-grained retrieval model of PM2.5 concentration using a back-propagation neural network (BPNN). Second, the accuracy of the trained model is verified using three indices: the coefficient of determination ( $R^2$ ), the absolute deviation (MAE) and the root mean square error (RMSE). Finally, a road network with weights is constructed based on PM2.5 distribution data and the healthy route is obtained by using the Dijkstra algorithm.

### 2.3.1. The retrieval of PM2.5

The distribution of PM2.5 is obtained from a BPNN. It is a multilayer feedforward network that is trained by an error back-propagation algorithm, which is widely used in nonlinear modeling, function approximation, and system identification, among other applications. Its minimum unit is a neuron, which is also known as a single-layer perceptron. The BPNN has a three-layer structure, and it consists of an input layer, a hidden layer, and an output layer. Its learning process consists of forward propagation of signal and backward propagation of error. Samples are introduced from the input layer, processed by the hidden layer, and transmitted to the output layer. If the actual result of the output layer differ from the expected value, the error is propagated backward. The reverse propagation of the error involves of transfer of the error to the input layer in a suitable form and allocation the error to all the units in each layer. The error of each unit is the basis for correcting the weight.

### 2.3.2. Model verification method

Cross-validation is a classic machine-learning method that is often used to assess the predictive validity of a regression analysis or classification problems. K-fold cross-validation divides the training set into  $k$  subsamples: one subsample is reserved for validation of the model, and the other  $k-1$  samples are used for training; this process is repeated  $k$  times. The average result of the  $k$  repeats is used as the final estimation result. This paper calculated  $R^2$ , MAE, and RMSE based on k-fold cross-validation. The formulas of  $R^2$ , MAE, and RMSE are as follows:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(\hat{y}_i - y_i)| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (6)$$

where  $y_i$ ,  $\bar{y}_i$  and  $\hat{y}_i$  refer to the true, average, and predicted values, respectively, of PM2.5 and n is the quantity of testing data.

### 2.3.3. Dynamic Dijkstra algorithm

The Dijkstra algorithm was proposed by Edsger Wybe Dijkstra in 1959. It aims to calculate the shortest route from one vertex to the remaining vertices based on a weighted road network. The basic strategy is to construct a road network using the road length as the weight and search for the route with the smallest cumulative weight from the starting point to another vertex. The vertices are divided into two sets: S, which contains all the vertices for which the shortest route is found, and T, which contains all the vertices for which the shortest route is not found. According to the order in terms of the cumulative weight, the vertices in T are added to S one by one until the terminal vertex has been added to S. Traditional route planning by using the Dijkstra algorithm takes the shortest distance as the target, and it is used to achieve healthy route planning. The purpose of dynamic healthy route planning is to minimize the accumulation of PM2.5 exposure. The core algorithm is the greedy algorithm. Its dynamics are such that as the spatial distribution data of PM2.5 change, the corresponding road network weights will also change. Once these changes are detected, the current point will be regarded as a new starting point, and the above algorithm will be recalled to achieve real-time updating of healthy routes.

### 2.3.4. Weighting of the road network

The PM2.5 distribution of the road network is accessed from the hourly data of PM2.5 distribution. The detailed illustrated in detail in Fig. 3. The road network is overlaid on the spatial distribution data of PM2.5, then the original continuous road is truncated into multiple segments, and the PM2.5 concentration of each segment is the corresponding grid value. Finally, a road network with weight is obtained.

The exposure dose is a comprehensive indicator that reflects the concentration of pollutants and the time of exposure to pollutants, which is used to accurately assess the damage level of pollution to humans. Hence, it is used to weight the road network. The calculation formula is as follows:

$$E = \int_{t_1}^{t_2} C(t) \times dt \quad (7)$$

where  $E$  is the exposure dose,  $C$  is the pollutant concentration and  $t$  is the time of exposure to the pollutants.

Since the exposure time is difficult to determine, the formula is simplified. The exposure time of each road is equal to the ratio of distance  $L$  to velocity  $V$ . The exposure of a road is obtained by summing the exposure of all sections, as expressed in Formula 8:

$$E_r = \sum_{i=1}^n C_i \times \frac{L_i}{V_i} \quad (8)$$

where  $E_r$  denotes the exposure of the road, which consists of  $n$  sections;  $C_i$  and  $L_i$  denote the PM2.5 concentration and length, respectively, of section  $i$ ; and  $V_i$  is the velocity when traversing section  $i$ .

The potential dose reflects the number of pollutants that are inhaled by people when exposed to air pollutants. The formula is as follows:

$$P = C \times IR \times t \quad (9)$$

where  $P$  is the potential dose,  $C$  is the pollutant concentration,  $IR$  is the respiration rate and  $t$  is the time of exposure to the pollutants.

The exposure time  $t$  is replaced with distance-velocity ratio. Then, for a person who traverses a road that consists of  $n$  sections, the calculation formula of the potential of this road is expressed as Formula 10:

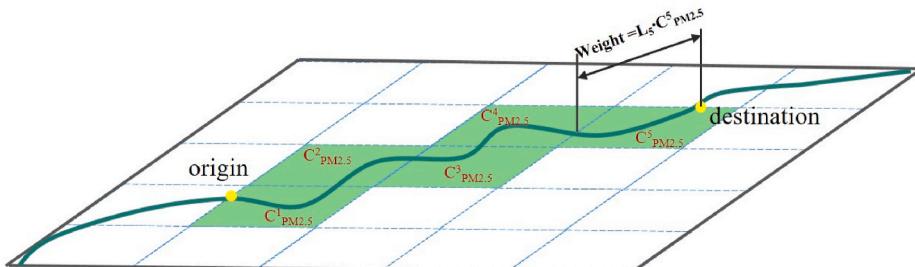


Fig. 3. Road weights.

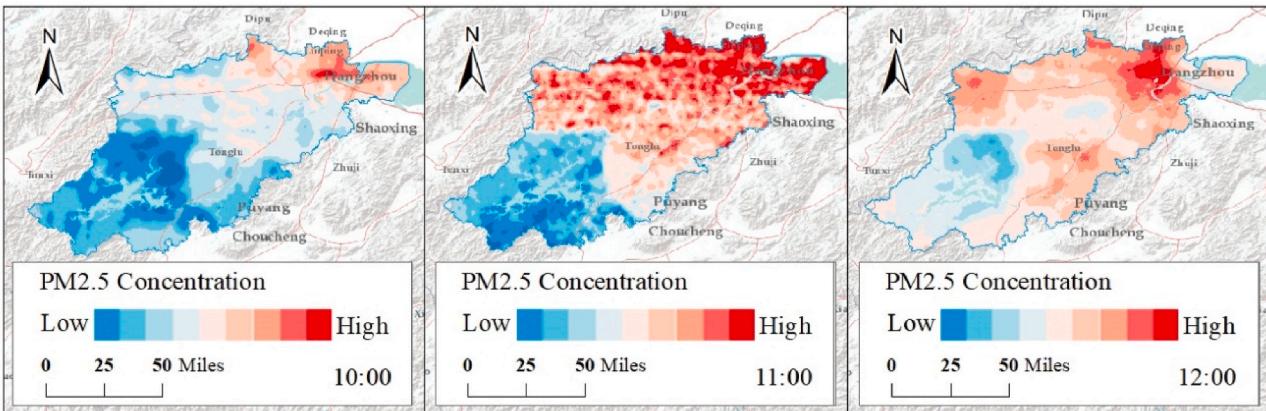


Fig. 4. PM2.5 distribution data that were retrieved by the BPNN.

$$P_r = \sum_{i=1}^n C_i \times L_i \times \frac{IR_i}{V_i} \quad (10)$$

where  $P_r$  is the potential dose of a road which is composed of  $n$  sections and  $IR_i$  is respiration when traversing section  $i$  of the road.  $C_i$ ,  $L_i$ , and  $V_i$  have the same meaning as in [Formula 8](#).

### 3. Results and discussion

#### 3.1. Analysis of the retrieval results

Remote sensing image and their derived data, meteorological data, and socioeconomic data were combined with PM2.5 concentration data to construct and train the BPNN model. The spatial resolution of the PM2.5 concentration distribution data was 30 m. Since the renewal cycles of the meteorological and PM2.5 ground monitoring data were at an hourly level, the time granularity of the retrieval results could also reach the hourly level.

The spatial distribution data of PM2.5 at 10:00, 11:00, and 12:00 on May 18, 2019, in Hangzhou, was achieved by using this method as shown in [Fig. 4](#). The distribution result of PM2.5 showed significant discrepancies in the temporal distributions. The concentration of PM2.5 in the whole area was low, and there was a small area of high concentration in the northeast at 10 a.m. At 11 a.m., the PM2.5 concentration in the southwest was higher than that in the northeast and central regions, where the high-value distribution was scattered. High-value and low-value regions, which were of small area, were located in northeastern and southwestern of Hangzhou, respectively.

##### 3.1.1. Verification of the model accuracy

Spatial distribution data of PM2.5 concentration that are retrieved by the MLR model are commonly used in healthy route planning. In this study,  $R^2$ , MAE, and RMSE were calculated to evaluate the accuracies of the MLR model and BPNN model. As presented in [Table 1](#), the fitting degree of BPNN was higher than that of MLR, and  $R^2$  of BPNN was 0.85, which was greater than that of MLR, namely, 0.51. The MAEs of BPNN and MLR were  $7.14 \mu\text{g}/\text{m}^3$  and  $13.29 \mu\text{g}/\text{m}^3$ , respectively, and the latter was almost double the former. The RMSEs of the two models was  $10.38 \mu\text{g}/\text{m}^3$  and  $18.78 \mu\text{g}/\text{m}^3$ , respectively. Compared to the blue trend line which indicates that the predicted value is equal to the true value, the scatter points of the BPNN are concentrated near the line, while those of the MLR model deviate from the line gradually as the value increase, as shown in [Fig. 5](#). This shows that the BPNN model has great advantages in high pollution prediction.

The overall accuracy of the BPNN model was better than that of the MLR model, as shown in [Fig. 6](#). The BPNN model not only produced satisfactory prediction results for high values but also had a good fitting effect for low values without negative abnormal values. It is shown that BPNN, which is a nonlinear model, performs better in dealing with complex fuzzy mapping relations without knowledge of the distribution of the data or the relationship between variables ([Wang and Elhag, 2007](#)).

##### 3.1.2. Effect of spatial heterogeneity

The spatial distribution of pollutants that is interpolated from monitoring station data is commonly used in traditional healthy travel route planning. To compare the spatial heterogeneity of the PM2.5 concentration among models, we used Kriging interpolation and the BPNN model to obtain PM2.5 distribution data with a resolution of 30 m at 10:00 on April 15, 2019. [Fig. 7](#) shows partially enlarged detail images of two results at the same location. The left panel is the PM2.5 distribution map of the retrieval method that is proposed in this study and the right one is a map of interpolation. According to A2 and B2 in [Fig. 7](#), the interpolation results of monitoring stations can only reflect regional variations in PM2.5 in cities. The PM2.5 values in a specified area around the center of a monitoring station were essentially the same, and the extracted values for the in this area were also the same. Therefore the interpolation method is not suitable for planning healthy travel routes. Instead, A1 and B2 can reflect the spatial distribution of PM2.5 in more detail. Therefore, the method that is proposed in this study provides more detailed information about the changes in PM2.5 on roads and even on sections of the same road. The results indicate that the proposed method is more suitable for healthy route travel planning. Although the spatial resolution of the interpolation results was consistent with that of the retrieval results, due to fewer monitoring stations and uneven distribution, it was still unable to reflect accurately and in detail the spatial distribution of PM2.5 in the city.

#### 3.2. Healthy route validation

The spatial distribution data of PM2.5 with higher spatial-temporal accuracy for 10 a.m. on April 15th, 2019, were obtained by

**Table 1**  
Accuracy comparison between BPNN and MLR.

Evaluation Index	BPNN	MLR
R-square	0.85	0.51
MAE	7.14	13.29
RMSE	10.38	18.78

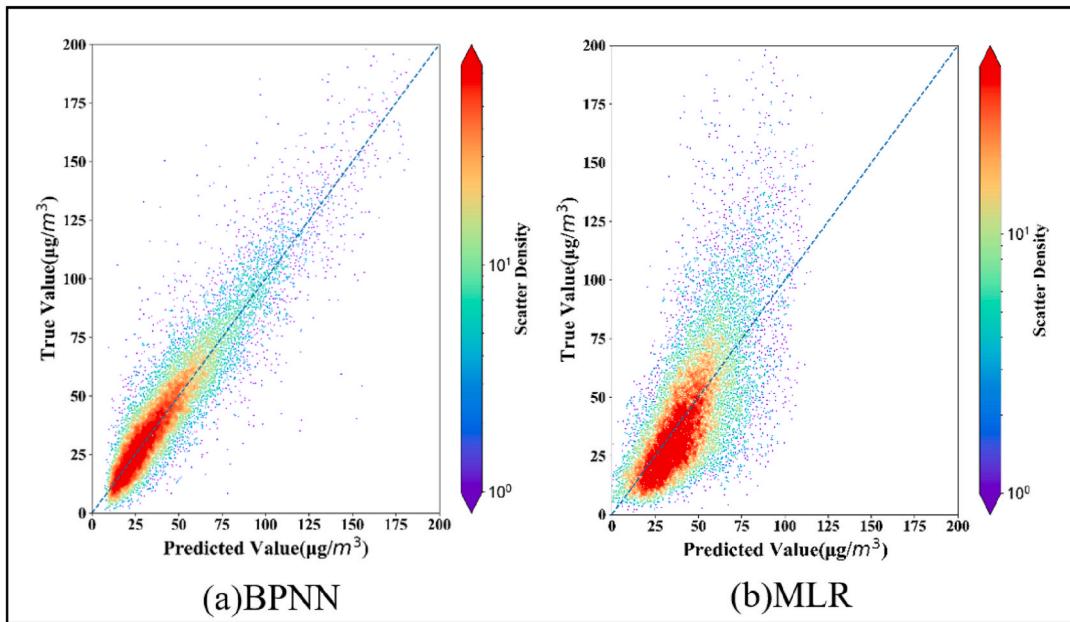


Fig. 5. Scatter plots of true and predicted values.

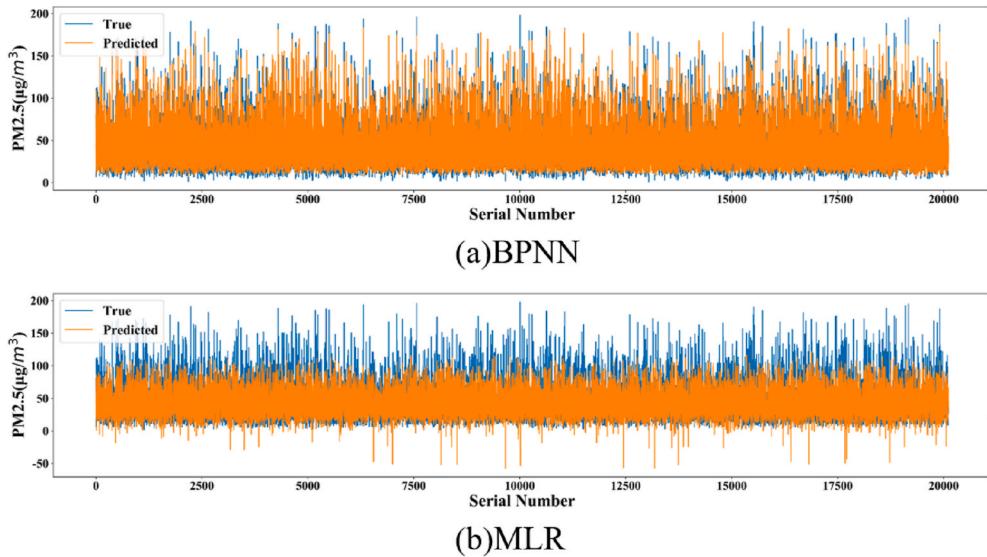
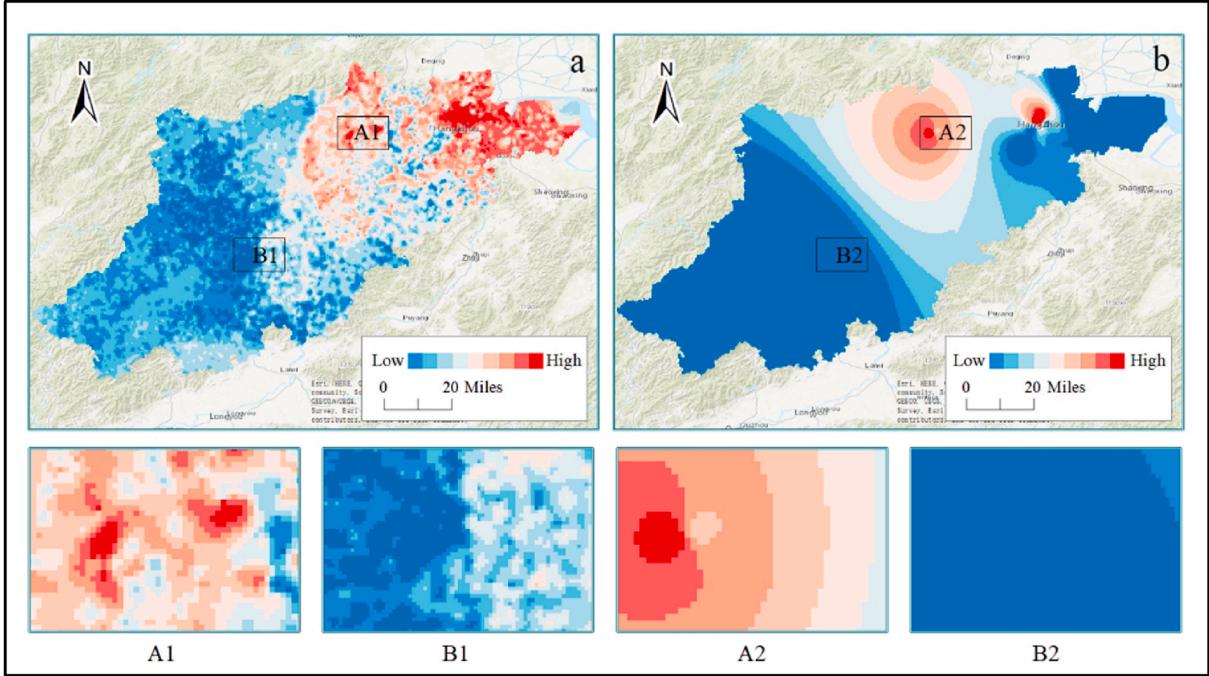


Fig. 6. Accuracy verification chart.

using multisource data and the BPNN model. Then, the PM2.5 distribution results of the road were extracted using overlay analysis of GIS. Due to the spatial heterogeneity of the PM2.5 concentration, the exposure risk of PM2.5 varied among routes. Consequently, the hazard degree of PM2.5 to people was selected as the weight of the road network. Through the Dijkstra algorithm, the travel route with the minimum exposure risk of PM2.5 was identified, as shown in Fig. 8, and the healthy route was not equivalent to the shortest route. Although the healthy route corresponded to a long journey, the exposure risk to PM2.5 was minimal.

### 3.2.1. Quantitative analysis of the effectiveness of health routes

The effectiveness of taking a healthy running route that avoids pollution was determined by comparing the potential dose of the healthy running route and the shortest route between two points in the city. Moving velocity and respiratory rate are variables in the potential dose formula, thus different people may have different potential doses of PM2.5 even when exposed to the same concentration of PM2.5. In this paper, choosing different speed for different classes of roads, and an average respiratory rate of  $1.350 \text{ m}^3/\text{h}$  for men and women under heavy physical activity, as reported in the “Handbook of Exposure Parameters for Chinese Population”, to



**Fig. 7.** Comparison of BPNN retrieval results and interpolation results under the same spatial resolution. a and b show the results of the retrieval method and the interpolation method, respectively. A1 and B1 are the local enlarged results of a, and A2 and B2 are the local enlarged results of b.

calculate the exposure dose and potential dose of two routes by Formula. 8 and [Formula 10](#).

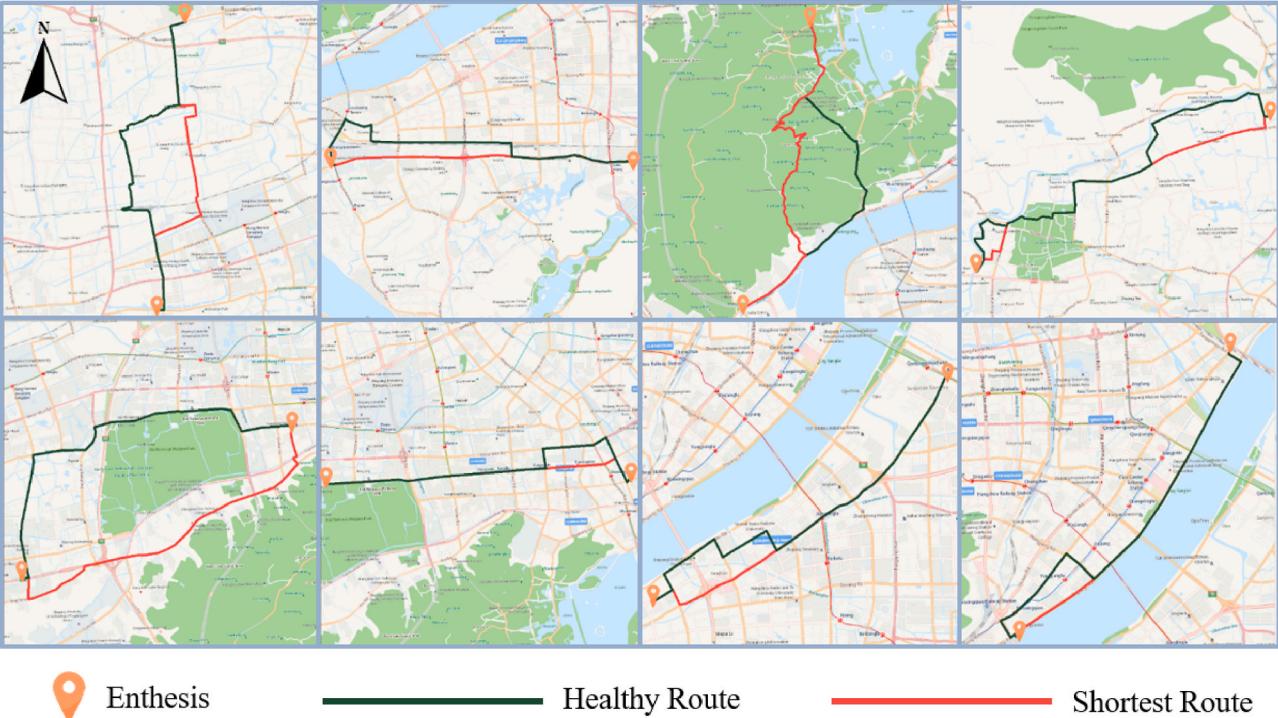
Taking Hangzhou as an example, the potential doses of the shortest routes and healthy routes that corresponded to 8 pairs of starting and ending points (SEPs) were calculated. The decrease rate of the potential dose (DP) and the escalation rate of the distance (ED) of healthy running routes relative to the shortest route were also calculated, as presented in [Table 2](#). The healthy route was longer than the shortest route, but its potential dose was significantly lower, and DP was between 0.55% and 27.37%, which demonstrates that healthy travel routes can significantly reduce the harm of pollutants to people during travel. In addition, the distances between the eight groups SEP were less than 12 km, which demonstrates that this method can improve the performance in solving the short-distance healthy route planning problem.

Healthy route planning reduces the harm of pollutants to people by identifying low-value areas of PM2.5 for travel while sacrificing time. Simply taking the amount of inhaled pollutants as the evaluation index cannot fully characterize the effectiveness. Therefore, both the distance and PM2.5 potential dose were considered in the evaluation of the healthy route in this study. The escalation rate of distance (ED), DP of each healthy route relative to the shortest route, and their ratios (ERP) are presented in [Table 2](#). Healthy routes with ERP values of greater than one and less than one are called high-benefit healthy routes and low-benefit healthy routes, respectively. The DP of the high-efficiency healthy route was greater than the ED, thereby indicating that a small amount of bypass can reduce pollutant inhalation more, and the greater the ERP is, the more effective the healthy route is in avoiding pollution. Low-efficiency healthy routes can reduce the harm of pollutants to humans, but they require longer travel distances. Moreover, the smaller the ERP is, the worse the healthy route. Although both types of a healthy routes can reduce PM2.5 inhalation, high-efficiency healthy routes are preferable.

High-efficiency and low-efficiency routes are presented in [Fig. 9](#). Points that correspond to the shortest and the healthy routes are plotted in a coordinate system with distance as the abscissa and potential dose as the ordinate, and each pair of points connected two routes between the same SEP. In each pair of points, the point with a higher potential dose represents the shortest route, and the other point represents the healthy route. The slope of the line between the two points correspond to the ERP, and the dotted line in the graph represents the ratio of one. The slope of the line that corresponds to the high-efficiency healthy route is greater than that of the dotted line, and the greater the slope is, the better the effect of the healthy route is. 1-Sep and 3-Sep in [Fig. 9](#) are typical examples of high-efficiency healthy routes. Similarly, the slope of the low-efficiency route is smaller than that of the dotted line, and the smaller the slope is, the worse the effect is. Typical representatives of the low-efficiency healthy route are 4-Sep and 6-Sep. Nevertheless, our results show that both high-efficiency routes and low-efficiency routes can effectively reduce the risk of PM2.5.

### 3.2.2. Factors that affect healthy short-distance route planning

The quality of pollutant distribution data can directly affect healthy route planning. Spatial resolution is one of the indicators that is commonly used to describe data quality and reflect the accuracy of data products, but sometimes, it does not reflect the degree of detail for an attribute. For example, the interpolation method and retrieval method can both be used to obtain pollutant data with a



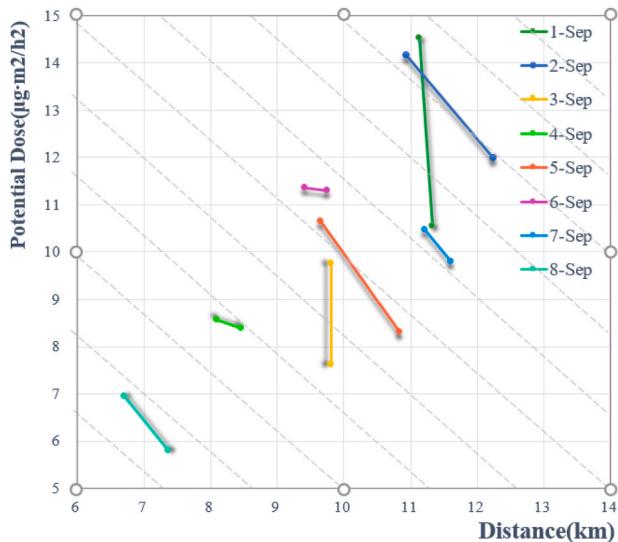
**Fig. 8.** Healthy routes and shortest routes.

**Table 2**

Escalation rate of distance and decrease rate of potential dose of healthy routes.

Starting and ending points (SEP)	Route	Distance (km)	Escalation rate of distance (ED)	Potential dose ( $\mu\text{g}$ )	Decrease rate of potential dose (DP)	ED/DP (ERP)
1	shortest route	11.137	1.73%	14.517	27.37%	15.79
	healthy route	11.330		10.543		
2	shortest route	10.930	12.03%	14.171	15.45%	1.28
	healthy route	12.245		11.982		
3	shortest route	9.807	0.02%	9.760	21.97%	1077.49
	healthy route	9.809		7.616		
4	shortest route	8.099	4.42%	8.558	1.92%	0.44
	healthy route	8.457		8.393		
5	shortest route	9.652	12.25%	10.647	21.96%	1.79
	healthy route	10.834		8.308		
6	shortest route	9.417	3.56%	11.356	0.55%	0.15
	healthy route	9.752		11.294		
7	shortest route	11.216	3.54%	10.468	6.41%	1.81
	healthy route	11.613		9.796		
8	shortest route	6.704	9.93%	6.947	16.51%	1.66
	healthy route	7.370		5.799		

SEP refers to the starting and ending points. ED equals the difference between the length of the healthy route and that of the shortest route divided by the length of the shortest route. DP equals the difference between the potential dose of the shortest route and the potential dose of the healthy route divided by the potential dose of the shortest route. ERP equals ED divided by DP.



**Fig. 9.** Relationship between distance and potential dose. The slope of the line between the two points represents the ERP and the dotted line in the graph represents an ERP value of one.

resolution of 30 m, but the former cannot reflect the spatial heterogeneity of the PM2.5 distribution, which hinders the realization of healthy route planning.

Spatial resolution and spatial heterogeneity are both indicators that should be considered. To investigate this, we considered eight groups of SEPs that were specified in the previous chapter, and identified the healthy travel routes and the shortest routes based on the PM2.5 distribution data with spatial resolutions of 30, 500, and 1000 m that were obtained by the retrieval method and the PM2.5 distribution data with a spatial resolution of 30 m that were obtained by the interpolation method. The DP values of the healthy routes were calculated and compared. The results are presented in [Table 3](#).

[Fig. 10](#) further illustrates why the spatial resolution affects healthy route planning. The PM2.5 concentrations of Route 1 and Route 2 in [Fig. 10](#) are the same with a resolution of 5000 m, but the air quality of the two routes differs with a resolution of 30 m. This further demonstrates that pollutant distribution data with high spatial resolution are important for the efficient identification of healthy routes.

### 3.3. Discussion

Currently, residents give more attention to health with a higher degree of urbanization. Choosing a healthy route for travel can effectively reduce the harm of air pollution to the human body. Therefore, it is essential to study and develop methods for healthy route planning. However, pollutant distribution data without sufficient accuracy or details limit the accuracy of short-distance healthy route planning. In this paper, a PM2.5 retrieval method with high spatiotemporal resolution and the dynamic Dijkstra algorithm collectively overcome the above challenges.

Studies can achieve short-distance healthy route planning based on many ground monitoring instruments, but not all cities or regions can afford the high cost of the sensors. In this paper, low-cost data with wide coverage, such as easily available remote sensing images, meteorological data, and socioeconomic data, are used, which makes the healthy path planning method more feasible and easy to popularize.

Only the road category was considered when assigning a speed to the road. However, road conditions are a constantly changing factor that should be considered in future research. In addition, only PM2.5 is considered in this paper. The main pollutants differ among regions, and consideration of all pollutants is more scientific ([Tian et al., 2020](#)).

This paper improves the established healthy route planning method to make it more suitable for the daily travel of residents and reduces the cost of implementation in cities.

## 4. Conclusions

To realize short-distance healthy route planning, this paper proposes a retrieval method of PM2.5 with high spatiotemporal resolution and solves the problems of difficult implementation and high cost in short-distance healthy route planning. This contributes to the large-scale implementation of a healthy route planning approach that is suitable for the daily travel of inhabitants. This study has found that a suitable representation of the spatial heterogeneity of pollutants is critical for scientific healthy route planning. In various cities or regions of the same city, determining the spatial resolution of pollutant data that maximally reduces the data redundancy on the basis of healthy path planning will be the focus of future research.

### Author statement

Li Na-Gao: Conceptualization, Methodology, Software, Data Curation.

Tao Fei: Project administration, Funding acquisition.

Pei-Long Ma: Writing Original Draft, Data Curation.

Cheng Yi-Wang: Software, Data Curation.

Wei Kong, Wen Kai-Chen: Visualization.

Zhou Tong: Conceptualization, Methodology, Supervision, Writing Review & Editing, Funding acquisition, Investigation.

All authors have read and agreed to the published version of the manuscript.

### Funding

This research was funded by the Major project of the National Social Science Fund (19ZDA189), in part by National Natural Science Foundation of China under Grant 41301514 and Grant 41401456, in part by the Natural Science and Technology Project of Nantong (MS12020075 and MS12021082), in part by 2020–2022 Transportation Education Scientific Research Projects of the China Institute of Communications Education (JTYB20-45), in part by National College Students Innovation and Entrepreneurship Training Program (202110304041Z).

### Declaration of competing interest

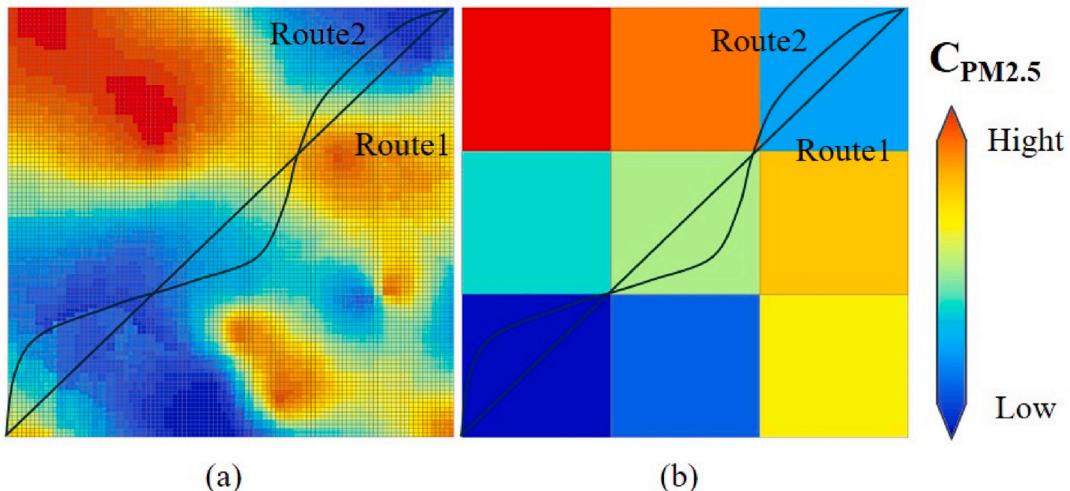
There are no financial conflicts of interest to disclose.

**Table 3**

DP values of healthy routes that were obtained based on PM2.5 distribution data with spatial resolutions of 30, 500, and 1000 m that were obtained by the retrieval method and with a spatial resolution of 30 m that were obtained by the interpolation method.

SEP	retrieval			interpolation DP (30)
	DP (30 m)	DP (500 m)	DP (1000 m)	
1	27.3%	27.3%	0%	0%
2	15.45%	15.45%	0%	15.45%
3	21.97%	21.97%	0%	0%
4	1.92%	0%	0%	0%
5	21.96%	21.96%	21.96%	0%
6	0.55%	0.55%	0%	0%
7	6.41%	6.41%	6.41%	0%
8	16.51%	16.51%	16.51%	0%

A DP value of zero indicates that the planned healthy route is the shortest route, namely, the real healthy route was not found. Five DP values are zero when using PM2.5 distribution data with a resolution of 1000 m, and one DP value is zero when using PM2.5 distribution data with a resolution of 500. In addition, seven DPs are zero when using PM2.5 distribution data with a spatial resolution of 30 m that are acquired by interpolation. These results significantly indicate that the spatial resolution and the details of the spatial heterogeneity are both factors that affect the accuracy of healthy route planning. In general, higher resolution and more detailed data help improve the rationality of healthy route planning.



**Fig. 10.** Comparison of resolutions. (a) and (b) are PM2.5 distribution data with resolutions of 30 m and 5000 m, respectively.  $C_{PM2.5}$  represents the concentration of PM2.5.

## Acknowledgments

The authors would like to thank the editor and the anonymous reviewers for insightful comments on improving this paper.

## References

- Brook, R.D., Rajagopalan, S., Pope III, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., Holguin, F., Hong, Y., Luepker, R.V., Mittleman, M.A., 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation* 121, 2331–2378.
- Chen, H., Li, Q., Wang, Z., Sun, Y., Mao, H., Cheng, B., 2018. Utilization of MERSI and MODIS data to monitor PM2.5 concentration in Beijing-Tianjin-Hebei and its surrounding areas. *Yaogan Xuebao/Journal Remote Sens.* 22, 822–832. <https://doi.org/10.11834/jrs.20187123>.
- Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. In: *Advances in Neural Information Processing Systems*, pp. 155–161.
- Hatzopoulou, M., Weichenthal, S., Barreau, G., Goldberg, M., Farrell, W., Crouse, D., Ross, N., 2013. A web-based route planning tool to reduce cyclists' exposures to traffic pollution: a case study in Montreal, Canada. *Environ. Res.* 123 <https://doi.org/10.1016/j.envres.2013.03.004>.
- Hertel, O., Hvidberg, M., Ketzler, M., Storm, L., Stausgaard, L., 2008. A proper choice of route significantly reduces air pollution exposure—a study on bicycle and bus trips in urban streets. *Sci. Total Environ.* 389, 58–70.
- Higgins, C.D., Adams, M.D., Réquie, W.J., Mohamed, M., 2019. Accessibility, air pollution, and congestion: capturing spatial trade-offs from agglomeration in the property market. *Land Use Pol.* 84, 177–191.
- Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y., 2017. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. Department of Environmental Health , Rollins School of Public Health , Emory University , Department of Biostatistics & Bioinformatics , Rollins School of Environ. Sci. Technol., vols. 1–29.
- Jarjour, S., Jerrett, M., Westerdahl, D., De Nazelle, A., Hanning, C., Daly, L., Lipsitt, J., Balmes, J., 2013. Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study. *Environ. Heal. A Glob. Access Sci. Source* 12, 1–12. <https://doi.org/10.1186/1476-069X-12-14>.

- Jin, N., Li, J., Jin, M., Zhang, X., 2020. Spatiotemporal variation and determinants of population's PM2. 5 exposure risk in China, 1998–2017: a case study of the Beijing-Tianjin-Hebei region. *Environ. Sci. Pollut. Res. Int.* 27, 31767–31777.
- Kim, K.-H., Kabir, E., Kabir, S., 2015. A review on the human health impact of airborne particulate matter. *Environ. Int.* 74, 136–143.
- Lim, S.S., et al., 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380 (9859), 2224–2260.
- Lin, C., Li, Y., Yuan, Z., Lau, A.K.H., Li, C., Fung, J.C.H., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2. 5. *Remote Sens. Environ.* 156, 117–128.
- Luo, J., Boriboonsomsin, K., Barth, M., 2018. Reducing pedestrians' inhalation of traffic-related air pollution through route choices: case study in California suburb. *J. Transp. Heal.* 10, 111–123.
- Mahajan, S., Tang, Y.-S., Wu, D.-Y., Tsai, T.-C., Chen, L.-J., 2019. CAR: the clean air routing algorithm for path navigation with minimal PM2. 5 exposure on the move. *IEEE Access* 7, 147373–147382.
- Orru, K., Orru, H., Maasikmets, M., Hendrikson, R., Ainsaar, M., 2016. Well-being and environmental quality: does pollution affect life satisfaction? *Qual. Life Res.* 25, 699–705.
- Stadlober, E., Hörmann, S., Pfeiler, B., 2008. Quality and performance of a PM10 daily forecasting model. *Atmos. Environ.* 42, 1098–1109.
- Tian, J., Fang, C., Qiu, J., Wang, J., 2020. Analysis of pollution characteristics and influencing factors of main pollutants in the atmosphere of Shenyang city. *Atmosphere* 11, 766.
- Unnithan, S.L.K., Gnanappazham, L., 2020. Spatiotemporal mixed effects modeling for the estimation of PM2. 5 from MODIS AOD over the Indian subcontinent. *GIScience Remote Sens.* 57, 159–173.
- Voronoi, G., 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. *J. für die Reine Angewandte Math. (Crelle's J.)* 97–102. <https://doi.org/10.1515/crll.1908.133.97>, 1908.
- Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2. 5 concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Sci. Total Environ.* 409, 1266–1276.
- Wang, J.Y.T., Dirks, K.N., Ehrgott, M., Pearce, J., Cheung, A.K.L., 2018. Supporting healthy route choice for commuter cyclists: the trade-off between travel time and pollutant dose. *Oper. Res. Heal. Care* 19, 156–164. <https://doi.org/10.1016/j.orhc.2018.04.001>.
- Wang, Y.-M., Elhag, T.M.S., 2007. A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks. *Expert Syst. Appl.* 32, 336–348.
- Xing, Y.-F., Xu, Y.-H., Shi, M.-H., Lian, Y.-X., 2016. The impact of PM2. 5 on the human respiratory system. *J. Thorac. Dis.* 8, E69.
- Xu, G., Ren, X., Xiong, K., Li, L., Bi, X., Wu, Q., 2020. Analysis of the driving factors of PM2. 5 concentration in the air: a case study of the Yangtze River Delta, China. *Ecol. Indicat.* 110, 105889.
- Yan, J.-W., Tao, F., Zhang, S.-Q., Lin, S., Zhou, T., 2021. Spatiotemporal distribution characteristics and driving forces of PM2. 5 in three urban agglomerations of the Yangtze river economic belt. *Int. J. Environ. Res. Publ. Health* 18, 2222.
- Yang, J., Wong, M.S., Ho, H.C., Krayenhoff, E.S., Chan, P.W., Abbas, S., Menenti, M., 2020. A semi-empirical method for estimating complete surface temperature from radiometric surface temperature, a study in Hong Kong city. *Remote Sens. Environ.* 237, 111540.
- Yu, Y., Wang, J., Yu, J., Chen, H., Liu, Vm, 2019. Spatial and temporal distribution characteristics of PM 2.5 and PM 10 in the urban agglomeration of China's Yangtze river Delta, China. *Pol. J. Environ. Stud.* 28.
- Zahmatkesh, H., Saber, M., Malekpour, M., 2015. A new method for urban travel rout planning based on air pollution sensor data. *Curr. World Environ.* 10, 699–704.
- Zhang, H., Xu, S., Liu, X., Liu, C., 2021. Near "real-time" estimation of excess commuting from open-source data: Evidence from China's megacities. *J. Transport Geogr.* 91, 102929.
- Zhu, S., Lian, X., Wei, L., Che, J., Shen, X., Yang, L., Qiu, X., Liu, X., Gao, W., Ren, X., 2018. PM2. 5 forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. *Atmos. Environ.* 183, 20–32.
- Zou, B., Li, S., Zheng, Z., Zhan, B.F., Yang, Z., Wan, N., 2020a. Healthier routes planning: a new method and online implementation for minimizing air pollution exposure risk. *Comput. Environ. Urban Syst.* 80, 101456.
- Zou, Z., Cai, T., Cao, K., 2020b. An urban big data-based air quality index prediction: a case study of routes planning for outdoor activities in Beijing. *Environ. Plan. B Urban Anal. City Sci.* 47, 948–963.
- Zuurbier, M., Hoek, G., Oldenwening, M., Meliefste, K., van den Hazel, P., Brunekreef, B., 2011. Respiratory effects of commuters' exposure to air pollution in traffic. *Epidemiology* 219–227.