IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 28, 2021, accepted April 15, 2021, date of publication April 20, 2021, date of current version April 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074303

Analyzing the Effects of Road Type and Rainy Weather on Fuel Consumption and Emissions: A Mesoscopic Model Based on Big Traffic Data

RUI SHANG[1], YI ZHANG[1,2], (Member, IEEE), ZUO-JUN MAX SHEN[1,3], (Member, IEEE), AND YI ZHANG[4]

[1]Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, China
[2]Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Automation, Tsinghua University, Beijing 100084, China
[3]Department of Industrial Engineering and Operations Research, University of California at Berkeley, Berkeley, CA 94720, USA
[4]Institute of Future Human Habitats, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

Corresponding author: Yi Zhang (zy1214@sz.tsinghua.edu.cn)

This research was supported in part by the National 13-5 Key Research and Development Program Projects (Grant No. 2018YFB1600600), Basic Research Program of Shenzhen Science and Technology Innovation Committee (JCYJ20180307123910003), and Scientific Research Start-up Funds of Tsinghua Shenzhen International Graduate School (QD2021007N).

ABSTRACT Road transportation accounts for significant percentages of urban energy consumption and carbon emissions. Therefore, it is important to predict and analyze the fuel consumption and emissions for on-road vehicles, which are varied under different conditions. Previous studies have shown that some traffic elements such as road type and weather condition have considerable influence on transportation fuel consumption and emissions. However, limited to the data availability, most of the existing studies focus on specific routes or scenarios, and few of them consider the effects of road type and weather condition systematically at large scale. In this research, a data-driven mesoscopic model was developed to investigate the effects of road type and weather condition on the link-level fuel consumption and emission factors based on big traffic data. This built model utilized the neural network for the prediction algorithm with inputs including road type, weather condition, and link-level aggregated operation data obtained through link-based segregation over trajectory snippets. The investigation was carried out with real-world big traffic data collected from 10,944 taxis over a 2-month period of operation in Shenzhen, and produced reliable predictions for four road types with clear and rainy weather conditions. Both statistical analysis and model prediction results showed that fuel consumption and emission factors are lower in low-speed range for freeway and expressway, and are lower in middle-speed range for main road and secondary road. In addition, rainy weather condition tends to have lower fuel consumption and emission factors than clear weather condition.

INDEX TERMS Fuel consumption and emission factors, mesoscopic model, neural network, road type, weather condition.

I. INTRODUCTION

The urbanization process has become faster worldwide, while intelligent transportation technology has significantly improved. Urban road transportation is now an indispensable part of the daily lives of city dwellers. In some large cities such as Beijing, road transportation is of the highest demand among the trip modes [1]. At present, the motor vehicles in China are responsible for 9% of the nation's total energy consumption, as well as more than 40 million tons of carbon monoxide and hydrocarbon pollutant emissions. In some cities of China, automobile exhaust has become the largest source of air pollution, contributing more than 50% of the PM2.5 emissions [2], [3]. The situation in the U.S. is similar, that the growth rates of the transportation-related energy consumption and emissions were the highest out of all the fields over the last 30 years, and the transportation industry has become the second largest source of greenhouse gas (GHG) emissions, following the power generating industry [4]. Therefore, motor traffic has a significant influence on

The associate editor coordinating the review of this manuscript and approving it for publication was Jin-Liang Wang.

the total energy consumption and pollutant emissions, which has promoted researchers from all around the world to study how to solve or relieve this problem.

It is known that there are many aspects, including road characteristic, weather condition, and traffic condition, that can affect the fuel consumption and emission of vehicles. It is found that different road types have varied emission contributions [5], [6] Faria, *et al.* [7] investigated data samples from 47 drivers, and found that Road grade and aggressive driving impact fuel consumption rates significantly Tong, *et al.* [8] studied the vehicle emissions and fuel consumption under urban driving conditions, and found that different driving modes led to different performances, with a transient driving mode more polluting than steady-speed driving modes. Some researchers have made efforts on the energy consumption performance analysis with considering the influence factors such as road characteristics, and driving behaviors in microscopic, or for some specific vehicles [9]–[12]. Most of these studies used data of limited scale, and the influences of the aforementioned factors in city-level has not been addressed. However, it is not easy to make field studies in real world to analyze the effects of different factors, especially in large scale. Fortunately, fuel consumption and emission models provide a tool to better understand the correlations and features of transportation fuel consumption and emissions from different perspectives.

Over the past thirty years, there have been an increasing number of studies related to transportation energy consumption and emission models using various methods. Based on the scale and purpose of such models, they can be classified into macroscopic models, mesoscopic models, and microscopic models. The macroscopic models use average aggregate network parameters to estimate and evaluate the transportation energy consumption or emissions over a wide area such as a city or district. Typical macroscopic models include MEET, COPERT, ECOTRANSIT, EMFAC, MOBILE, and MOVES [13]. Microscopic models have the goal of providing a precise description of vehicle energy consumption and emissions by relating the fuel consumption and emission rate to vehicle operation over a series of short time steps. Typical microscopic models include IFCM, VSP, VT-Micro, CPFM, PHEM, and CMEM [13], [14]. Compared to the macroscopic ones, these models focus on calculating the instantaneous energy consumption and emission data of a single vehicle based on instantaneous parameters. The microscopic models can be accurate on the micro-scale, but are not applicable for a large number of vehicles at the city scale because of the extensive computation required.

For practical applications with the goal of reducing urban transportation emissions in an ITS, such as dynamic eco-guidance [15]–[17], temporal–spatial traffic emission evaluation [18]–[21], and the real-time low-carbon operation management of urban transportation [22], [23], it is necessary to estimate the energy consumption and emissions of on-road vehicles. Therefore, researchers have introduced mesoscopic models to calculate the link-level vehicle energy consumption and emission factors by aggregating the operation states over trajectory snippets. Some researchers have focused on energy consumption and emission estimations of fixed route or restricted trips. Li, *et al.* [24] collected vehicle history data by driving an experimental vehicle on a fixed route, and then determine the transportation energy consumption factor of a certain road in Beijing by applying the history data to a microscopic model. Huang and Peng [25] developed an eco-routing method based on data-driven fuel consumption model. Xiaolin, *et al.* [26] studied the average fuel consumption model of touring coach based on OBD data. Liu, *et al.* [27] applied the data collected from taxies in Shanghai to MOVES, and assessed the impact of transportation on air quality. Huang, *et al.* [28] evaluated vehicle energy consumption and related emissions of the Guangdong-Hong Kong-Macao greater Bay Area, applying real-time traffic data and MOVES. Li, *et al.* [29] analyzed the emission by combining the taxi trajectory data in Beijing and a microscopic emission model. In terms of freeway research, some researchers have made contributions to energy and emission estimation and control strategies for freeway transportation [30]–[32]. These studies have produced great achievements and laid the foundation of mesoscopic transportation model research.

However, the previous studies of mesoscopic models mainly focused on specific routes of certain road types. The uncertainty in link-level energy consumption and emissions caused by road characteristics of different road types and weather conditions has not been fully evaluated because of the lack of big data containing large and diverse sets of information. According to the previous studies, the effects of these factors are non-negligible.

Based on big traffic data, a data-driven method can be used to investigate the inherent variability of the influential factors, and provide a more efficient way to develop mesoscopic models that consider more related factors. The massive data involved in a data-driven method used to be costly, which led to the use of a small scale for the source data in previous work. New information technology makes it much easier to collect traffic data, and the data volume has become much larger than ever before. In addition, with the development of big data technology, there have been numerous related tools and technologies that can be applied in various of research fields to have new improvements [33]. Artificial intelligence (AI) is capable of dealing with big traffic data and many related applications have emerged [34], especially the application of machine learning and neural networks [35]–[37]. There have been various works that investigate ITS with big data techniques [38]. Some researchers have used the data from mobile devices such as cell phones to estimate the vehicle fuel consumption using a neural network [39] Alomari, *et al.* [40] developed an efficient tool that is used for traffic-related event detection from twitter data, which have broadened the data sources and application mode for big data in traffic field. For traffic prediction with big traffic data, Aqib, *et al.* [41] proposed a traffic prediction model by bring four complementary cutting-edge technologies, which are big data, deep

learning, in-memory computing, and high performance computing, that can predict traffic flow, speed and occupancy with big data of 11 years. The new findings and achievements of algorithms in machine learning and neural networks make them powerful tools for research involving big traffic data for an ITS.

Most of the aforementioned studies on transportation fuel consumption and emissions focus on small-scale modeling, with few considerations on different impact factors. Therefore, there exists a gap that few studies tend to evaluate the effects of different influential factors at large-scale on urban transportation fuel consumption and emissions. To bridge the gap, this study proposed a data-driven methodology with a neural network to estimate the link-level energy consumption and emission factors for on-road vehicles at the city scale, with consideration given to different road types and weather conditions, and the influences of these factors are analyzed. This study makes several contributions to the existing work. 1) A neural network-based mesoscopic transportation energy consumption and emission model was developed with big traffic data to estimate the link-level fuel consumption and emission factors for running vehicles in urban areas with considering road type and weather condition. 2) The model was performed on the real-world big traffic data of Shenzhen, which were collected from 10,944 light-duty vehicles operating for 2 consecutive months covering all the road types throughout the city. 3) The impact of road type and weather condition on predicting fuel consumption and emission factors were analyzed.

The remainder of this paper is organized as follows. The detailed methodology is described in section 2, and the case study of Shenzhen is introduced in section 3. The results and some discussions are provided in section 4. Finally, the conclusions are presented in section 5.

II. METHODOLOGY

A. RESEARCH FRAMEWORK

The framework for this research consisted of three parts: data collection and pre-processing, model development, influence factor and performance analysis. For the data collection and pre-processing, a massive amount of data was collected using vehicle-mounted data collectors. Taxis might be the ideal data acquisition vehicles because they tend to have long and consecutive hours of operation and cover most of an urban area. The operation status data was turned into trajectories with map data through map matching. The invalid data were detected and removed. The model development applied the Comprehensive Modal Emission Model (CMEM) to generate the instantaneous fuel consumption and emissions of vehicles based on the vehicle operation history data. Then, a link-based segregation method was used to obtain the link-based fuel consumption and emission factors from the instantaneous data. The data features are analyzed in terms of the influence of road type and rainy weather condition. Following this, the neural network and nonlinear regression methods were applied to learn and predict the fuel consumption and emissions considering different road types and weather conditions. Lastly, for the performance analysis, the results predicted by the neural network method and conventional nonlinear regression method were compared and evaluated, and the predictions of fuel consumption and emission factors under clear and rainy weather conditions are assessed. In addition, a parametric study of the various road types as well as weather conditions was conducted to demonstrate their impacts on the fuel consumption and emission factors. The research framework is shown in FIGURE 1.

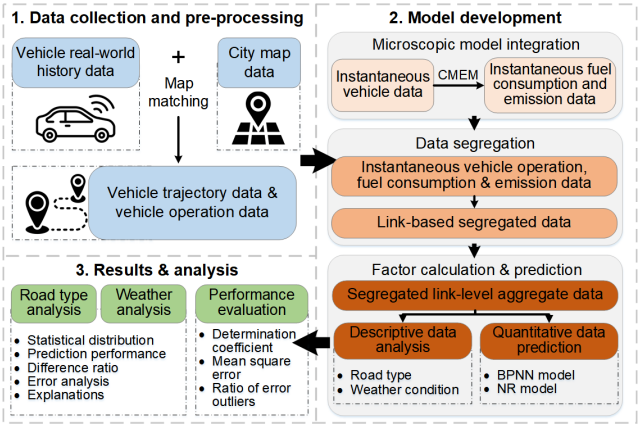


FIGURE 1. Research framework.

B. DATA PREPARATION

The modelling in this research used data-driven methodology with real-world big data. Therefore, the data preparation was important. Because the model was designed to analyze the vehicle fuel consumption and emissions on different roads in an urban area, the data had to cover as many different trip histories as possible. Thus, normal test floating vehicle might not meet the demand for a large database. Taxis make up 8% to 15% of total traffic flows on urban roads during the daytime [27], and the difference of average speed between taxis and private cars is acceptable. Although it is believed that professional drivers have different driving characteristics as non-professional drivers, the velocity features are not so distinguished. Some researchers [42] conducted a comparative study on professional and non-professional drivers, and the results showed that significant differences exist in some aspects, such as driving anger under various conditions, while no significant difference is found in terms of preferred speed on expressway (speed limit of 80km/h) and urban road (speed limit of 60km/h). Similar results were obtained in other two studies [43], [44] from Chen, *et al.*, where data of simulated driving tests of mid-age professional and non-professional drivers are similar in terms of mean value of two indicators, including standard deviation of speed and average speed. Some researchers have tested data collected from taxis, which were very effective for modelling [27]. The data fields involved in this model were the vehicle ID, recording

time, vehicle speed, vehicle acceleration (optional), vehicle moving direction (optional), and vehicle GPS longitude and latitude. To analyze the influence of weather condition, the hourly history data of weather condition were integrated to collected data records.

The map data were needed to match the vehicle data with road links. In the map data, each link had to be defined with the GPS coordinates of nodes on the link, as well as the road type according to the traffic plan and the road's purpose.

C. FUEL CONSUMPTION AND EMISSION MODEL

In this research, the CMEM [45], [46] was applied to calculate the instantaneous fuel consumption and emission data of the vehicles. This model was jointly developed by the University of California Riverside, Center for Environmental Research and Technology, University of Michigan, and Lawrence Berkeley National Laboratory, and was used to predict the real-time energy consumption and emission values of microscopic traffic. With the CMEM, the energy consumption and emission rates of various kinds of vehicles with different features or performances, and operating under different conditions, could be quickly calculated. All the vehicles in the model were divided into 26 categories according to their engine type, power and weight, and total mileage.

It has been shown that for each vehicle category, the fuel consumption and emission rate calculated by the CMEM accurately reflect the fuel consumption and emission values of the corresponding vehicle the in real world, which gives the model a good estimation performance. This research had the goal of studying the energy consumption and emission factors on urban roads with a medium scope. Thus, it was more reasonable to use the parameters for a category of vehicles in the calculations instead of the parameters of a specific vehicle model. The method used for the category selection was based on the CMEM User Guide [47].

D. LINK-BASED SEGREGATION METHOD

In mesoscopic models, the average velocity within a segregated snippet is normally used to estimate the energy consumption and emissions. Various segregation methods are used in mesoscopic models, including trip-based, time-based, distance-based, and link-based methods [24]. According to some studies, a link-based method that segregates the data by road links has the most reliable results. It is also more sensitive to emission effects, and more appropriate and accurate for an ITS [48]–[51]. Therefore, this study applied a link-based segregation method to estimate the link-level fuel consumption and emission factors. In the data processing step, the data were segregated into small snippets based on the link ID found in the GPS record (shown in FIGURE 2). Then, the integral mean values for the velocity, fuel consumption rate, and emission rate within each snippet were calculated as the average velocity, average energy consumption rate, and average emission rate of the road link, respectively. The mathematical expressions (equations 1–3) for this process are

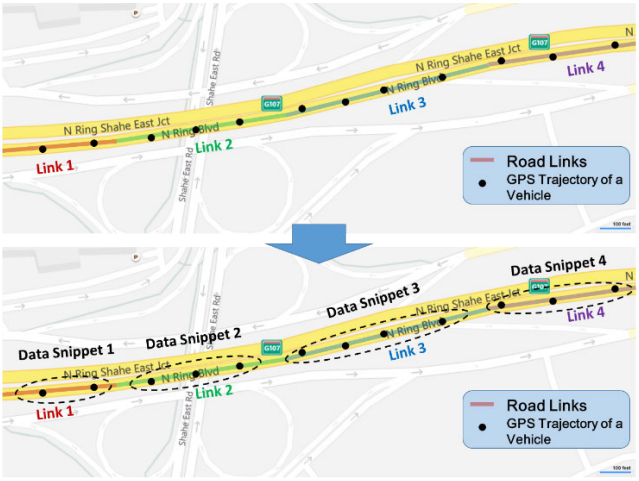


FIGURE 2. Schematic diagram of link-based segregation method.

shown as follows.

$$\bar{v}_{ij} = \frac{\sum_{k=1}^{k_{max}-1}\left[\left(t_{ij,k+1}-t_{ij,k}\right)\left(v_{ij,k+1}+v_{ij,k}\right)/2\right]}{t_{ij,k_{max}}-t_{ij,1}}, \quad (1)$$

$$\overline{FR}_{ij} = \frac{\sum_{k=1}^{k_{max}-1}\left[\left(t_{ij,k+1}-t_{ij,k}\right)\left(FR_{ij,k+1}+FR_{ij,k}\right)/2\right]}{t_{ij,k_{max}}-t_{ij,1}}, \quad (2)$$

$$\overline{ER}_{ij} = \frac{\sum_{k=1}^{k_{max}-1}\left[\left(t_{ij,k+1}-t_{ij,k}\right)\left(ER_{ij,k+1}+ER_{ij,k}\right)/2\right]}{t_{ij,k_{max}}-t_{ij,1}}, \quad (3)$$

where $\bar{v}$ is the average velocity (km/h); t is the recording time (s); v is the instantaneous velocity (km/h); $\overline{FR}$ is the average fuel consumption rate (g/s); FR is the fuel consumption rate (g/s); $\overline{ER}$ is the average emission rate (g/s); and ER is the emission rate (g/s).

The subscript ij represents the data of the jth vehicle on the ith road link; k means the kth record in the data sequence; and k_{max} indicates the number of records in the data sequence.

E. PARAMETER DEFINITION

The previous research showed that some of the external factors have impacts on the fuel consumption and emission values of a vehicle to some extent. These include the traffic conditions, weather conditions, and driver's behavior [52], [53]. However, the velocity and the acceleration are the two main decisive factors. According to the CMEM User Guide [47], if the acceleration data cannot be acquired, valid results may also be obtained using only the velocity data in the calculation.

The IDs of road links are given in the map data file. Road links are the basic units in a road network map, and usually consist of sections of a road between adjacent traffic signal lights, or road sections with uniform features (e.g. direction, gradient, and curvature).

The fuel consumption condition is defined as the fuel consumption factor. This is found by dividing the average fuel

consumption rate by the average velocity (equation 4), and represents the fuel consumption per unit of distance (1 km) for a given vehicle on a given road link. The emission of carbon dioxide was considered to be the key index in this study, because it is the main substance in vehicle exhaust and can represent the emission level. Thus, the emission factor is defined as the average carbon dioxide emission rate divided by the average velocity (equation 5), and represents the emissions per unit of distance (1 km) for a given vehicle on a given road link.

$$\overline{FF}_{ij} = \frac{\overline{FR}_{ij}}{\overline{v}_{ij}/3.6}, \tag{4}$$

$$\overline{EF}_{ij} = \frac{\overline{ER}_{ij}}{\overline{v}_{ij}/3.6}, \tag{5}$$

where $\overline{FF}$ is the average fuel consumption factor (kg/km); $\overline{FR}$ is the average fuel consumption rate (g/s); $\overline{EF}$ is the average emission factor (kg/km); $\overline{ER}$ is the average emission rate (g/s); and $\overline{v}$ is the average velocity (km/h).

The subscript ij indicates the data for the jth vehicle on the ith road link.

F. STATISTICAL ANALYSIS OF AGGREGATED DATA

The aggregated data set of link-level fuel consumption and emission factors generated through link-based data segregation contains those records from different road types and weather conditions. To obtain the effects of road type and weather condition on fuel consumption and emission, statistical features are analyzed within data groups sectionalized according to the corresponding data field.

For road type analysis, each group contains data from a specific road type. The statistical analysis is operated within each group, including speed distribution and fuel consumption and emission factor distribution. All the records within a group are divided into subgroups of 1km/h speed range (i.e. (0km/h,1km/h], (1km/h, 2km/h], …, (n-1km/h, nkm/h]). The descriptive statistical features including mean, median, upper and lower bounds, and percentiles are obtained for each subgroup. Then the comparison analysis among different groups of road types is performed to determine the effects of road type.

For weather condition analysis, a further grouping operation is implemented within each road type group according to data field of weather condition (clear or rainy). Similarly, the speed distribution, as well as the fuel consumption and emission factor distribution are analyzed. The subgrouping strategy is also set as 1km/h speed interval. The comparison analysis between clear and rainy weather for different road types is performed to determine the effects of weather condition.

G. PREDICTION METHODS

1) BACK PROPAGATION NEURAL NETWORK

The data fitting and prediction method used in this research was a back propagation neural network (BPNN), which

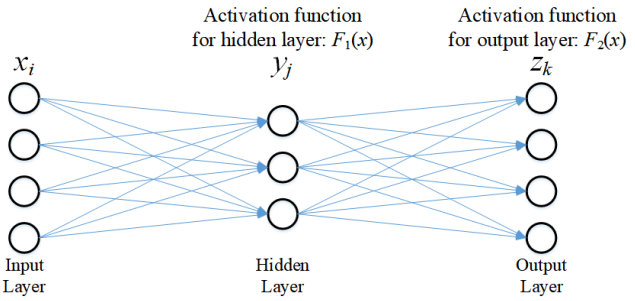


FIGURE 3. Schematic model of neural network [55].

recognized the underlying relationships in a set of data after training the system with the data. Compared to conventional nonlinear regression, the most obvious feature of an artificial neural network is that the function form is not needed in the whole process. The neural network system iteratively trains itself using the given data to learn the rules governing the data, and then predicts the output for a new input that is the closest to the expected value [54]. The basic structure of a neural network is shown as FIGURE 3. It consists of an input layer, a hidden layer, and an output layer. The output of each layer is the input of the next layer. In a BPNN, the error between the output from the network and the true value will be transferred back to each layer to adjust the parameters of the network and achieve better performance.

A 3-layer BPNN was designed for the data training and prediction in this research. Although a BPNN can consist of multiple hidden layers and a number of nodes, the increasing of layers and nodes will make the training process much more complex and time-consuming. It is known that a 3-layer neural network can approximate any continuous functions on compact subset of real numbers [56]. The input layer and output layers had m and n nodes, respectively (depending on the number of parameters to be input and predicted). A sigmoid function ($f(x) = \frac{1}{1+e^{-x}}$) was applied as the activation function in the hidden layer of the BPNN. In addition, the BPNN was used to make continuous prediction instead of being used for data classification or recognition in this research. Therefore, a linear function was used for the activation function in the output layer ($f(x) = x$). Mean square error (*MSE*) of predicted results and labels is applied as the loss function. The basic process for a 3-layer BPNN can be described as follows:

Parameter setting

$$(w, b) = (w_{ij}^{(l)}, b_i^{(l)}), \tag{6}$$

where $w_{ij}^{(l)}$ is the weight value between node j in layer l and node i in layer $l+1$, and $b_i^{(l)}$ is the bias of i in layer l.

Forward propagation

$$O_i^{(1)} = x_i, \tag{7}$$

$$I_i^{(2)} = \sum_{j=1}^{n} (w_{ij}^{(1)} O_j^{(1)} + b_i^{(1)}), \tag{8}$$

$$O_i^{(2)} = f_2(I_i^{(2)}), \tag{9}$$

$$I_i^{(3)} = \sum_{j=1}^{n} (w_{ij}^{(2)} O_j^{(2)} + b_i^{(2)}), \tag{10}$$

$$O_i^{(3)} = f_3(I_i^{(3)}), \tag{11}$$

where: $I_i^{(2)}$ and $I_i^{(3)}$ are the input values of node i in the hidden layer (layer 2) and output layer (layer 3), respectively. $O_i^{(1)}$, $O_i^{(2)}$, and $O_i^{(3)}$ are the output values of node i in the input layer (layer 1), hidden layer (layer 2), and output layer (layer 3), respectively. f_2 and f_3 are the activation functions of the hidden layer (layer 2) and output layer (layer 3), respectively. In this method, $f_2(x) = \frac{1}{1+e^{-x}}, f_3(x) = x$.

Back propagation

Loss function

$$E = \frac{1}{2} \sum_{i=1}^{n} (O_i^{(3)} - y_i)^2, \tag{12}$$

where E is the loss function used to evaluate the output accuracy of the output layer, and y_i is the target value of node i.

For the output layer

$$\frac{\partial E}{\partial w_{ij}^{(2)}} = (O_i^{(3)} - y_i) O_j^{(2)}, \tag{13}$$

$$\frac{\partial E}{\partial b_i^{(2)}} = (O_i^{(3)} - y_i), \tag{14}$$

For the hidden layer

$$\frac{\partial E}{\partial w_{jk}^{(1)}} = O_k^{(1)} O_j^{(2)} (1 - O_j^{(2)}) \sum_{i=1}^{n} (O_i^{(3)} - y_i) w_{ij}^{(2)}, \tag{15}$$

$$\frac{\partial E}{\partial b_j^{(1)}} = O_j^{(2)} (1 - O_j^{(2)}) \sum_{i=1}^{n} (O_i^{(3)} - y_i) w_{ij}^{(2)}, \tag{16}$$

Update process

$$w_{jk}^{(1)+} = w_{jk}^{(1)} - \eta \frac{\partial E}{\partial w_{jk}^{(1)}}, \tag{17}$$

$$b_j^{(1)+} = b_j^{(1)} - \eta \frac{\partial E}{\partial b_j^{(1)}}, \tag{18}$$

$$w_{ij}^{(2)+} = w_{ij}^{(2)} - \eta \frac{\partial E}{\partial w_{ij}^{(2)}}, \tag{19}$$

$$b_i^{(2)+} = b_i^{(2)} - \eta \frac{\partial E}{\partial b_i^{(2)}}, \tag{20}$$

where the superscript+ indicates an updated value.

The stop criterion should be set as maximal iteration reached or required loss value achieved. There have been various studies on application of BPNNs. Please see the related works for more details.

2) NONLINEAR REGRESSION

To evaluate the performance of the BPNN prediction method, the conventional nonlinear regression (NR) method was applied as a control group. In addition, to ensure the application universality of the model, this research applied nonlinear functions with the same form to the data fitting for different road links and different road types. According to a previous study, a logarithmic function has good performance in describing the relationship between the fuel consumption factor and velocity [15], [24]. A test calculation proved that the 6th order logarithmic function was very effective in the data fitting with the given test data. The logarithm was taken for both the fuel consumption factor and emission factor, and then polynomial fitting was used to obtain the best result. The function form used in the fitting process is shown as follows.

$$\log \overline{FF} = a\bar{v}^6 + b\bar{v}^5 + c\bar{v}^4 + d\bar{v}^3 + e\bar{v}^2 + f\bar{v} + g \tag{21}$$

$$\log \overline{EF} = a\bar{v}^6 + b\bar{v}^5 + c\bar{v}^4 + d\bar{v}^3 + e\bar{v}^2 + f\bar{v} + g \tag{22}$$

H. MODEL EVALUATION

To evaluate the performance of the model, coefficient of determination (R^2), *MSE*, and residual analyses were performed with the original test data and data predicted using both the BPNN and NR methods (equations 23–27). The residual is defined as the following function:

$$Re_l = y_l - \hat{y}_l, \tag{23}$$

$$R^2 = 1 - \frac{SSE}{SST}, \tag{24}$$

$$SSE = \sum_l Re_l^2, \tag{25}$$

$$SST = \sum_l (y_l - \bar{y})^2, \tag{26}$$

$$MSE = \frac{SSE}{n}, \tag{27}$$

where Re is the residual; y is the true value of the dependent variable, which was fuel consumption factor or emission factor in this research (kg/km); $\hat{y}$ is the predicted value of the dependent variable, which was the fuel consumption factor or emission factor predicted by the NR or BPNN (kg/km); $\bar{y}$ is the mean value of the dependent variable set (y); *SSE* is the error sum of the squares; *SST* is the total sum of the squares; *MSE* is the mean square error; and n is the length of the data sequence.

The subscript l represents the lth data point in the data sequence. If the model performs well, R^2 will be as close as possible to one, and *MSE* will be as close as possible to zero.

III. CASE STUDY IN SHENZHEN

Shenzhen is located in Guangdong Province and is one of the biggest cities in southern China. It has an area of 1997 km^2 and a population of over 13 million. The road network of Shenzhen has a total length of more than 700 km, including 468 km of freeways and expressways, 1434 km of main roads, and 1033 km of secondary roads. According to the 'Annual Report on Road Network Density in Major Chinese Cities', the road network density in the 181 km^2 built-up area of Shenzhen is 9.50 km/km^2, putting it in 1st place among all the major Chinese cities [57]. Data from the Transport Commission of Shenzhen Municipality show that the total number of motor vehicles in Shenzhen was over 3.3 million at the end of 2018. Thus, it can be inferred that road transportation is quite important in Shenzhen.

FIGURE 4. Road network map of Shenzhen.

A. DATA SOURCE

The big traffic data used in this research were obtained from the Transport Commission of Shenzhen Municipality. The data were collected from 10,944 taxis in Shenzhen that were equipped with data acquisition equipment during a 2-month operational period. The original data file is of csv format, with fifteen data fields for each record and time intervals of 10 seconds between each two consecutive records of a same taxi. And the data volume of a file that includes all records in a single day ranges from 913MB to 2.28GB, containing records of 6,301,381 to 16,216,790 pieces. The total number of records for all the original data files is 780,718,195, and the total size of them is 110GB. The data fields included the record ID, vehicle ID, recording time, vehicle speed, moving direction, and GPS longitude and latitude where the data were recorded and other information. These unordered records spatially cover more than 90% main road network in Shenzhen, which are collected automatically without managing and clustering operation. As previously discussed, the data characteristics of taxis are similar to most private cars, and therefore in this research, it is assumed that citywide operation states can be determined with data from taxis.

The map data included the main road network of Shenzhen City, with the roads grouped into four classes: freeways, expressways, main roads, and secondary roads (as shown in FIGURE 4). The urban branch roads were not considered in the research because of the instability of the traffic flow and complexity of the traffic conditions.

B. DATA PRE-PROCESSING

Some of the original data were invalid because of extensive deviation from the normal data as a results of errors in the data collection process and communication failures during the transmission process, and had to be filtered out. To eliminate the invalid and redundant data, a data filtering process is applied with the main purpose of removing a) duplicated records (which had the same record ID), b) incorrect velocities (outside the normal range), c) data duplicated because of communication failure (data for the same time and vehicle), d) data from vehicles that were not in operation (where the speed was zero or the GPS data remained unchanged for a long time), and e) data from vehicles were outside the designated area (GPS data outside the range of the map data). After deleting the invalid data, the desired parts of the data, including the vehicle ID, recording time, velocity, and GPS coordinates, were retained. The rest of the data were deleted as redundant data. With the aforementioned pre-processing operation, around 30% to 40% records are removed for each day, and only seven data fields are kept for analysis in this research, which reduced the data scale by 65% and made computation more efficient. The map matching method in this research was based on the hidden Markov model (HMM), for which numerous studies have been conducted. For more details about the map matching method based on the HMM, please see the cited references[58]. The vehicle trajectories were obtained through map matching, and the records were connected to the corresponding road links, which allowed the records to show the vehicle operations on the related road links. Each road link is associated with an abbreviation of road type. In this study, four road types are concerned including freeway, expressway, main road and secondary road. Another variable of weather condition was adhered to the data records according to time stamp. The required data of vehicle trajectory, road links of different road type, and weather condition were prepared.

C. MODEL SETUP

In this case study, the model was used to predict the fuel consumption and carbon dioxide emission levels of taxis and similar light-duty vehicles in Shenzhen. Thus, in the CMEM, with the vehicle condition and related parameters of the taxis, all the vehicles were randomly defined as category 12 or 13 according to the CMEM category selection method. The data segregation was based on the link IDs and data of vehicles that were matched to the related links. After segregation operation, 52,430,629 pieces of records were generated for all the four road types.

In descriptive statistical analysis of aggregated fuel consumption and emission factors, data are divided into groups corresponding to velocity interval of 1km/h. Data features of distribution are compared regarding to different road types and weather conditions.

In the data fitting and prediction process, the BPNN was set as a three-layer neural network. With the average velocity, road type and weather condition as the input and fuel consumption or emission factor as the predicted output. The test training and parameter adjustment showed that using fifteen to twenty nodes in the hidden layer was sufficient to obtain a reliable prediction. All the data were divided into four groups according to the road class. Then, pick 1000 samples out from each group as the validation data set, and 70% of the rest part was selected as the training data set, while the remaining 30% was used as the test data set. The stop criteria are set as that the mean loss value of validation data set is smaller than the threshold or the iteration time reaches the given maximum value. After the BPNN was iteratively trained with the training data, the results were generated using the test data.

IV. RESULTS AND DISCUSSIONS

In this section, the influences of road type and weather condition on fuel consumption factor and emission factor

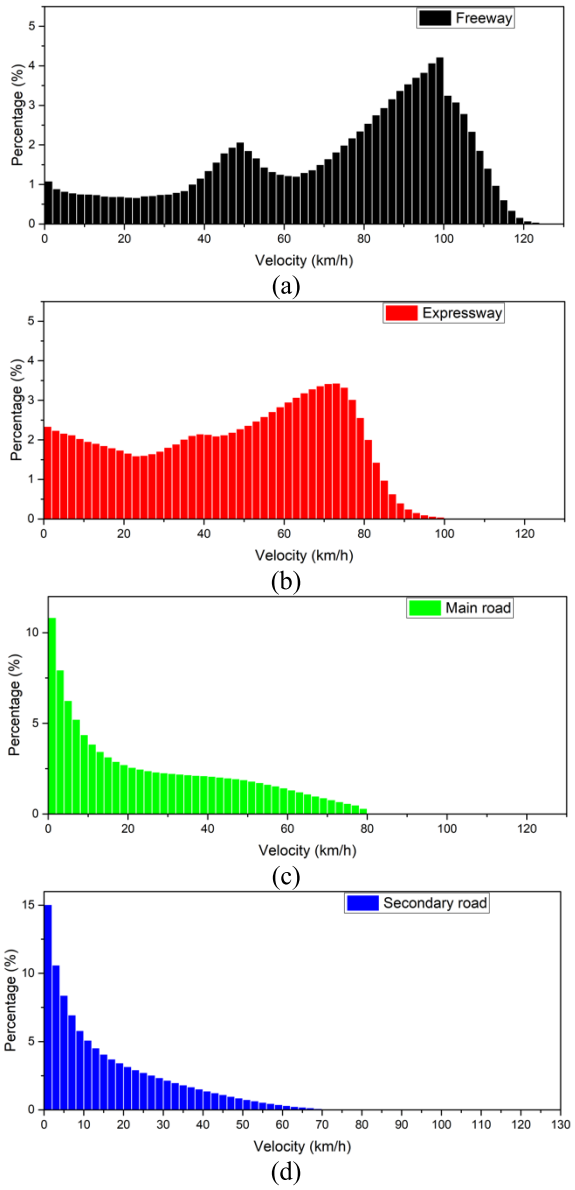


FIGURE 5. Velocity distributions for (a) freeway, (b) expressway, (c) main road, and (d) secondary road.

are analyzed. Firstly, the characteristic of average velocity distribution as well as the fuel consumption and emission factor distributions from different road types under two weather conditions are discussed with aggregated data. Then, the influences of road type and weather condition on fuel consumption and emission factors are evaluated through BPNN model, and the prediction performances of BPNN are evaluated. And last but not least, some ideas for existing problems and further study are presented in the discussion part.

A. DESCRIPTIVE STATISTICAL RESULTS FOR DIFFERENT ROAD TYPES

In this study, as previously mentioned, the original aggregated data were divided into four groups: freeway, expressway, main road, and secondary road based on the road

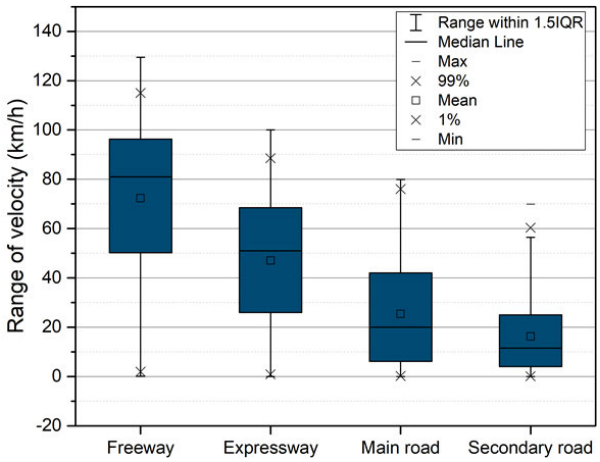


FIGURE 6. Box-plot of average velocities on links of four road types.

classification from the traffic plan and design. The results for the average speed distribution and statistical box plots are shown in FIGURE 5 and FIGURE 6, respectively. Based on the results, there were obvious differences in the average velocity distributions for the four road types. The mean and median values for the four road classes gradually decreased based on the road features. The mean value for freeways is the highest, with two extreme values of data density around 50 km/h and 100 km/h. For expressways, the mean value is relatively high, with data concentrated around average velocity of 75 km/h, which is close to the speed limit. The results for the main roads and secondary roads are similar. There are many records showing a velocity around zero as a result of intersections and traffic signals, and the distribution range for the main roads is mainly below 50 km/h, with much lower values for the secondary roads.

To further analyze the differences in the features of the four road types, density maps of the fuel consumption factor (FIGURE 7) and emission factor (FIGURE 8) were generated. Here, the color of an area indicates the scatters density, with a brighter (light yellow) area indicating a higher scatter density for the area. It can be seen that the density distributions of the four road types have their own special accumulation areas.

The four road types have very unique average velocity distributions and density distributions for the fuel consumption and emission factors. To compare the characteristic of fuel consumption and emission factors for the four road types, the detailed data distribution analyses are presented as follows.

To analyze fuel consumption and emission factor distribution within each road type, the original data is processed with the average velocity interval of 1km/h. In each 1km/h data group, six indexes are obtained with original data, including the mean value, median, upper and lower bound, as well as 25 and 75 percentiles. The fuel consumption factor distribution is shown as FIGURE 9, in which the mean value, upper and lower bound are shown in FIGURE 9-(a), while the median, 25 and 75 percentiles are shown in FIGURE 9-(b).

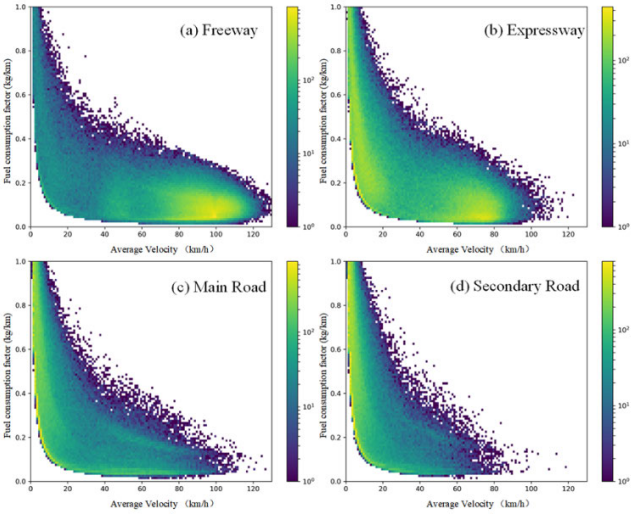


FIGURE 7. Density maps of fuel consumption factors for (a) freeway, (b) expressway, (c) main road, and (d) secondary road.

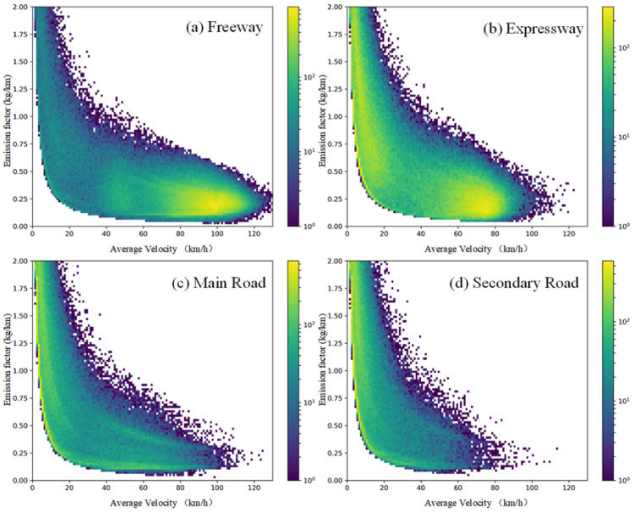


FIGURE 8. Density maps of emission factors for (a) freeway, (b) expressway, (c) main road, and (d) secondary road.

Within each 1km/h interval, all the records have similar average velocity with difference less than 1km/h, while the fuel consumption factors fall in a range, which is because of the different driving behaviors, road link features and other related aspects. Since that the speed and acceleration are the two decisive parameters to estimate fuel consumption rate and emission rate in CMEM, we assume that the main reason for the various fuel consumption factors is different driving behaviors. Normally, the definition of aggressive driving behaviors depends on the frequency of acceleration. While in this research, according to the estimated fuel consumption factors, the non-aggressive, moderate, and aggressive driving behaviors are classified as the records within 0 to 25 percentiles, 25 to 75 percentiles, and over 75 percentiles respectively. It is observed in FIGURE 9 that the lower bounds for the four road types are similar, while within low-speed range (average velocity less than 15km/h) the records of

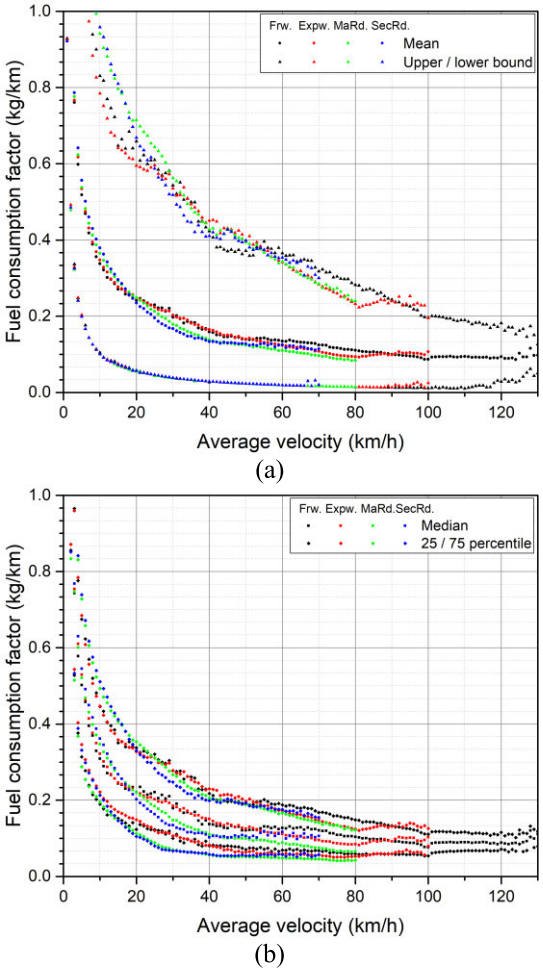
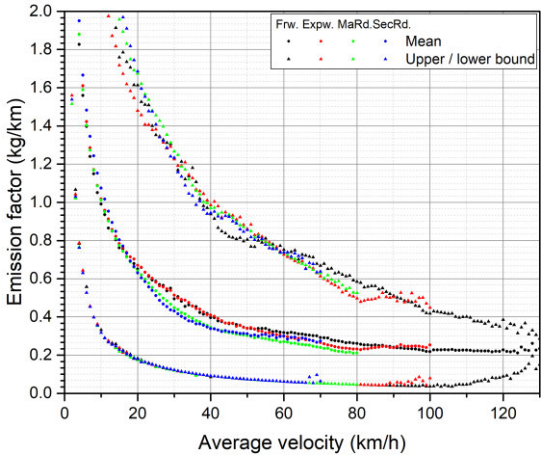
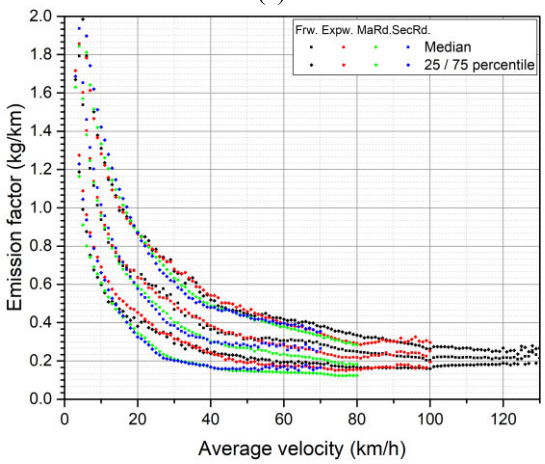


FIGURE 9. Fuel consumption factor distribution of (a) mean and upper/lower bound, (b) median and 25/75 percentiles for four road types.

freeway and expressway tend to accumulate in moderate area, and those records of aggressive driving behaviors are of less aggressiveness compared to main and secondary roads. The records of freeway and expressway in middle-speed range (average velocity from 20 to 80) tend to have higher fuel consumption factors than records of main road and secondary road. The reason of the trend can be explained as follows. In low speed range, the traffic conditions of freeway and expressway are seriously congested, which lead to the slow movement of traffic flows. However, the situation of main road and secondary road are quite different. The existing of intersections and more complex traffic environment that involves non-motor vehicles and pedestrians for main and secondary roads lead to frequent acceleration-deceleration process, and make vehicles consume more fuel than vehicles on freeways and expressways. Then, in middle-speed range, the traffic conditions of main and secondary roads are quite clear, while the situations of freeways and expressways are still with heavy traffic that make drivers push the braking paddle frequently and caused higher fuel consumption. As a result, the different driving behaviors for different road types



(a)



(b)

FIGURE 10. Emission factor distribution of (a) mean and upper/lower bound, (b) median and 25/75 percentiles for four road types.

under similar average velocity cause distinguishing features of fuel consumption factors.

The same statistical analysis is operated with emission factors, and the results are shown as FIGURE 10. The trend of emission factor basically follows the same pattern of fuel consumption factor. Records in middle-speed range show that vehicles on freeways and expressways tend to emit more GHG than those on main and secondary roads of the same average velocity. However, the difference in low-speed range is not as obvious as fuel consumption factor, which might because that the imperfect combustion of engine generates more emissions such as HC and CO other than GHG.

B. DESCRIPTIVE STATISTICAL RESULTS FOR CLEAR AND RAINY WEATHER

Similar as the analysis of road type influences, the influences of clear and rainy weather on fuel consumption and emission factors are discussed with aggregated data. Compared to normal clear weather, rainy weather brings two most challenging conditions for drivers, which are wet road surfaces and affected visibilities. In this study, to analyze the impacts

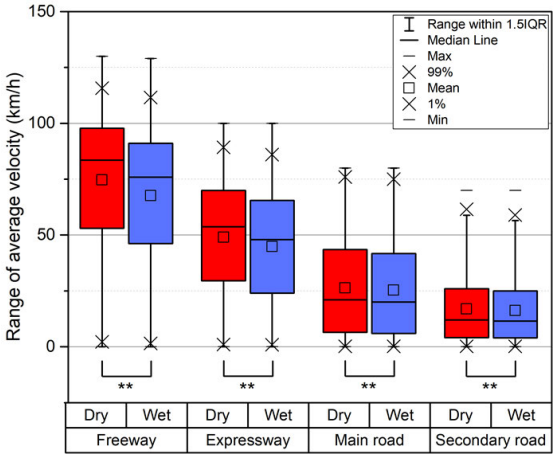


FIGURE 11. Box-plot of average velocities on links of four road types under rainy and clear weather conditions (** means the population mean is significantly different at 0.01 level).

of rainy weather on fuel consumption and emission factors, the historical hourly weather data is integrated to vehicle trajectories, fuel consumption and emission data estimated by CMEM according to timestamp of data records. In the following analysis, the original data and estimated data for rainy weathers are denoted as 'wet', while those from clear weathers are denoted as 'dry'.

The average speed distribution of original data from four road types under rainy and clear weather conditions are shown as FIGURE 11. It is observed that for freeway and expressway, the average speed under rainy weather conditions is lower than clear weather conditions, while the influences on main and secondary road are slightly affected. In addition, through the significance test, the mean value of velocity distribution for dry and wet conditions are significantly different at 0.01 level for all the four road types. It can be explained as follows. For vehicles on freeways and expressways, drivers tend to drive much faster than those on main and secondary roads, which requires drivers to concentrate more on traffic environment. However, the rainy weather lowers the visibility, and the wet road surface condition lowers the tire grip, which make the drivers be more careful and less aggressive, in other words, drivers tend to drive slower and more moderate under rainy condition than normal clear weathers.

For fuel consumption factor under clear and rainy weather conditions, the data distribution analysis is operated within each 1km/h average velocity interval. Eight indexes are concerned, including mean, median, upper and lower bound, 5 and 95 percentiles, 25 and 75 percentiles. The results of four road types are shown as FIGURE 12. For all the four road types, similar trends are observed that the lower bound, 5 percentiles and 25 percentiles are close for clear and rainy weather conditions, while the mean, median, 75 percentiles, 95 percentiles and upper bounds show separation for the two weather conditions. The values of fuel consumption factor of rainy weather condition are lower than clear weather

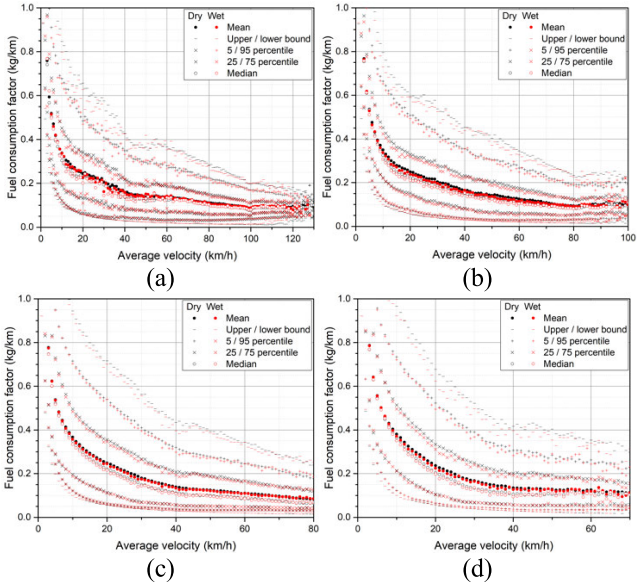


FIGURE 12. Fuel consumption factor distribution under clear and rainy weather conditions of (a) freeway, (b) expressway, (c) main road, (d) secondary road.

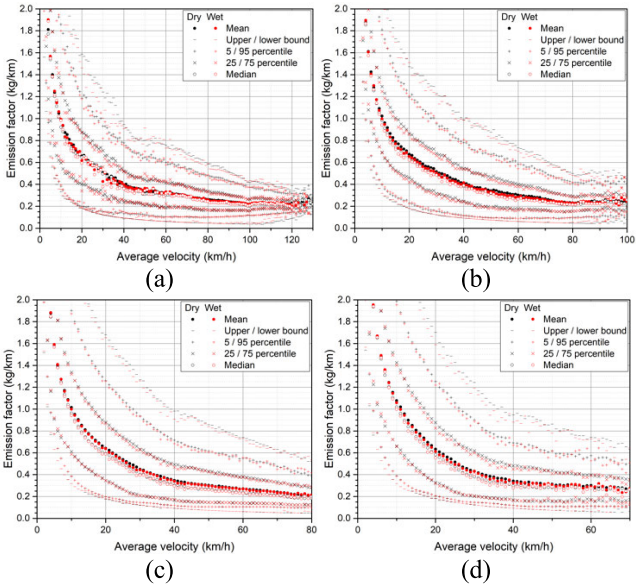


FIGURE 13. Emission factor distribution under clear and rainy weather conditions of (a) freeway, (b) expressway, (c) main road, (d) secondary road.

condition in terms of data records distributed above median within each velocity interval. It means that even drivers with aggressive driving behaviors in clear weathers tend to drive less aggressively under rainy weather conditions. And the results show that the influences of rainy weather conditions on fuel consumption factor are less significant for main road than other three road types.

The results of emission factor distribution in FIGURE 13 show the same trend as fuel consumption factor, in which the emission factors of rainy weather conditions are slightly lower than clear weather conditions.

To further analyze the influence on fuel consumption and emission factors of road type and weather condition, the results of BPNN prediction model are discussed.

C. RESULTS OF BPNN PREDICTION MODEL FOR FOUR ROAD TYPES

To quantitatively evaluate the influence of road type and weather condition on fuel consumption and emission factors, the BPNN model is applied to make predictions. The road type is designed as categorical variable by one-hot code, which introduces four binary variables to represent each road type. Average velocity together with road type variable constitute the five input values. According to the test samples, twenty nodes in hidden layers are enough to produce accurate prediction results. And the output values are fuel consumption factor and emission factor in the BPNN model respectively. To validate the prediction performance, the nonlinear regression (NR) method that are commonly used in previous studies [15], [24] are used as a benchmark. The comparative results of BPNN prediction and NR prediction are shown in FIGURE 14. It is observed that NR results have deviation at initial and end segments of the average velocity.

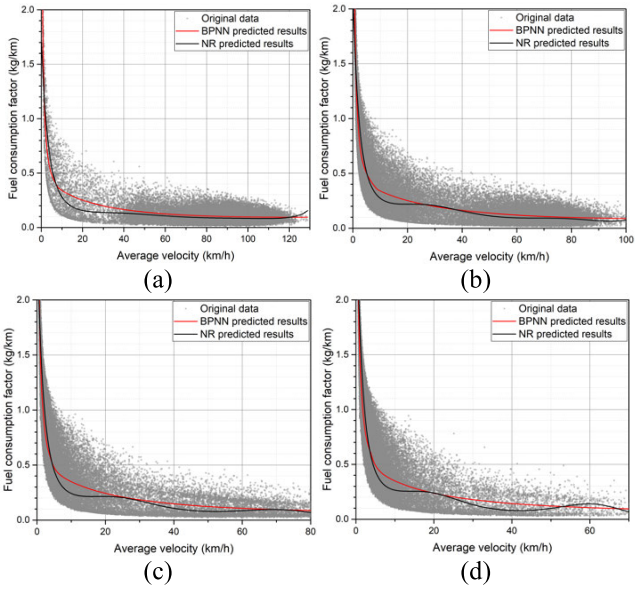


FIGURE 14. Fuel consumption factor prediction results with NR and BPNN model for (a) freeway, (b) expressway, (c) main road, and (d) secondary road.

The coefficients corresponding to the NR models in FIGURE 14 are presented in TABLE 1.

The prediction performance of BPNN and NR model are evaluated with R2 and MSE, and the contrastive error analysis is operated. The outlier ratio of error (EOR) indicates the ratio of error values that fall out of the 1.5 times interquartile range (IQR) of the error population. The key indexes of performance evaluation for BPNN and NR models are listed in TABLE 2. It is obvious that the BPNN model outperforms NR model in terms of all the three indexes, especially for fuel consumption factor prediction of main road and secondary road.

TABLE 1. Coefficients of fuel consumption factor NR model for four road types.

Road type	NR function: $\log\overline{FF} = a\bar{v}^6 + b\bar{v}^5 + c\bar{v}^4 + d\bar{v}^3 + e\bar{v}^2 + f\bar{v} + g$			
	Freeway	Expressway	Main road	Secondary road
Coefficient a	4.894×10^{-11}	3.581×10^{-10}	8.843×10^{-10}	2.739×10^{-9}
Coefficient b	-2.071×10^{-8}	-1.171×10^{-7}	-2.609×10^{-7}	-6.868×10^{-7}
Coefficient c	3.482×10^{-6}	1.481×10^{-5}	2.944×10^{-5}	6.509×10^{-5}
Coefficient d	-2.934×10^{-4}	-9.080×10^{-4}	-1.587×10^{-3}	-2.911×10^{-3}
Coefficient e	1.292×10^{-2}	2.774×10^{-2}	4.203×10^{-2}	6.309×10^{-2}
Coefficient f	-0.287	-0.407	-0.528	-0.646
Coefficient g	0.564	0.768	0.974	1.152

TABLE 2. Key indexes for performance evaluation of BPNN and NR fuel consumption factor prediction model.

Road type	*Model*	R^2	*MSE*	*EOR*
Freeway	BPNN	0.95	6.43×10^{-3}	4.28×10^{-2}
	NR	0.496	6.5×10^{-2}	4.6×10^{-2}
Expressway	BPNN	0.963	2.43×10^{-2}	5.92×10^{-2}
	NR	0.529	0.312	8.46×10^{-2}
Main road	BPNN	0.97	5.09×10^{-2}	1.10×10^{-3}
	NR	0.535	0.781	5.26×10^{-2}
Secondary road	BPNN	0.974	5.96×10^{-2}	2.03×10^{-3}
	NR	0.56	1.014	6.46×10^{-2}

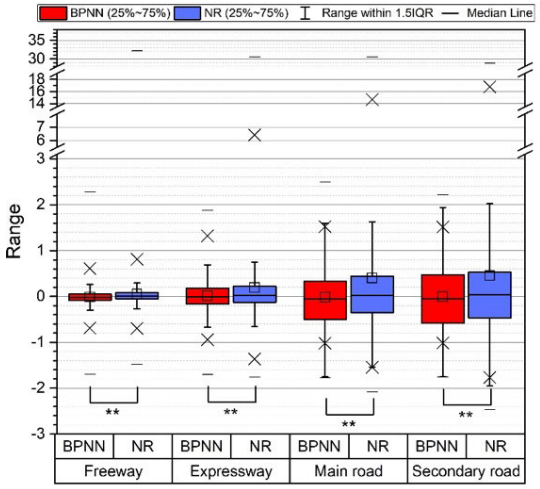


FIGURE 15. Box-plot of BPNN and NR model predicted fuel consumption factor error (means the population mean is significantly different at 0.01 level).**

The error distribution of BPNN and NR models for fuel consumption factor of four road types is shown as FIGURE 15. The mean value of error population generated from the two methods for four road types are all significantly different at 0.01 level. The error distribution of BPNN model fall in a reasonable range, while the upper bound and 99 percentiles of NR model fall extremely far from zero. The error analysis and key indexes including R^2 and *MSE* show that the BPNN model has better prediction performance than conventional NR model.

The comparison of BPNN predicted results for the four road types are shown in FIGURE 16. It is observed that

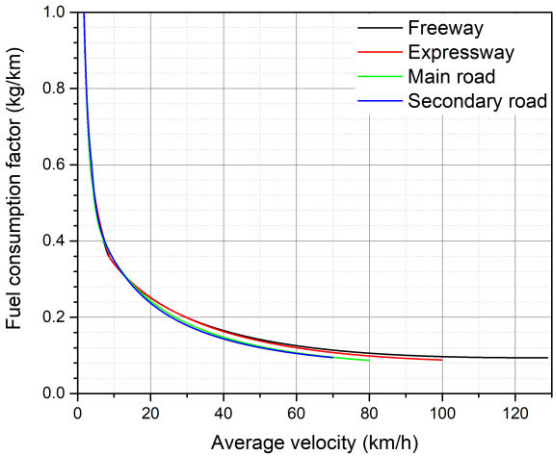


FIGURE 16. Comparison of BPNN predicted fuel consumption factor results for four road types.

the BPNN predicted results follow the characteristics of the aforementioned data distribution analysis. In the low-speed range around 15km/h, freeway and expressway have a lower fuel consumption factor than main and secondary road, while in the middle speed range, the results turn to be just the opposite. The difference between freeway and secondary road around 50km/h is 10%, and up to 17% at 70km/h. However, the results of freeway and expressway are close to each other, while the results of main road and secondary road show the same trends. It can be inferred from the results that freeway and expressway have similar fuel consumption features, and also main road and secondary road have the similar ones. In addition, the fuel consumption factor is mostly determined by the traffic condition. When the road traffic is clear and drivers can drive closely at the speed limit of the corresponding road type, the fuel consumption factor are similarly low. At the middle-speed range, vehicles on main roads and secondary roads are under quite clear traffic condition, while the traffic condition on freeways and expressways are pretty crowded that lead to frequent acceleration and deceleration which cause a higher fuel consumption factor.

For emission factor, the predicted results of NR and BPNN models are shown in FIGURE 17. Similar to the results of fuel consumption factor, NR predicted results have obvious deviation in initial and end segments of the average velocity range.

The coefficients corresponding to the NR models in FIGURE 17 are presented in TABLE 3.

Same as fuel consumption factor analysis, the prediction performance of BPNN and NR model for emission factor are evaluated with R^2, *MSE*, and *EOR*. The key indexes are listed in TABLE 4. It can be seen that BPNN model has better emission factor prediction performance than NR model, especially for main and secondary roads.

The error distribution of BPNN and NR models for emission factor of four road types is shown as FIGURE 18. The mean value of error population generated from the two methods for four road types are all significantly different at

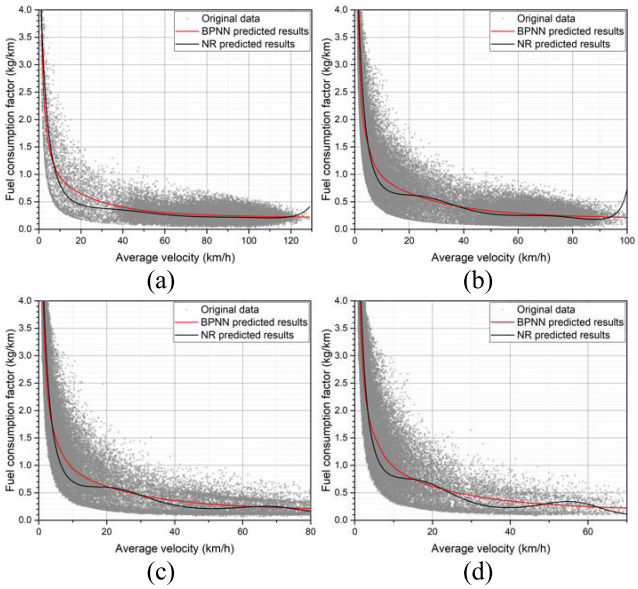


FIGURE 17. Emission factor prediction results with NR and BPNN model for (a) freeway, (b) expressway, (c) main road, and (d) secondary road.

TABLE 3. Coefficients of emission factor NR model for four road types.

NR function: $\log \overline{EF} = a\bar{v}^6 + b\bar{v}^5 + c\bar{v}^4 + d\bar{v}^3 + e\bar{v}^2 + f\bar{v} + g$

	Road type	Freeway	Expressway	Main road	Secondary road
Coefficient	a	5.580×10^{-11}	5.271×10^{-10}	1.052×10^{-9}	4.032×10^{-9}
	b	-2.321×10^{-8}	-1.608×10^{-7}	-2.999×10^{-7}	-9.312×10^{-7}
	c	3.814×10^{-6}	1.898×10^{-5}	3.272×10^{-5}	8.163×10^{-5}
	d	-3.130×10^{-4}	-1.088×10^{-3}	-1.711×10^{-3}	-3.395×10^{-3}
	e	1.342×10^{-2}	3.123×10^{-2}	4.410×10^{-2}	6.897×10^{-2}
	f	-0.294	-0.438	-0.546	-0.674
	g	1.679	1.939	2.114	2.292

TABLE 4. Key indexes for performance evaluation of BPNN and NR emission factor prediction model.

Road type	*Model*	R^2	*MSE*	*EOR*
Freeway	BPNN	0.968	3.15×10^{-2}	5.13×10^{-2}
	NR	0.513	0.539	5.51×10^{-2}
Expressway	BPNN	0.971	0.174	0.115
	NR	0.547	2.748	0.12
Main road	BPNN	0.973	0.4	6.1×10^{-3}
	NR	0.542	6.808	5.12×10^{-2}
Secondary road	BPNN	0.975	0.472	7.01×10^{-5}
	NR	0.574	8.171	5.52×10^{-2}

0.01 level. The error distribution of BPNN model fall in a reasonable range, while the upper bound and 99 percentiles of NR model fall extremely far from zero. The error analysis and key indexes including R^2 and *MSE* show that the BPNN model has better prediction performance than conventional NR model for emission factor estimation.

The comparison of BPNN predicted emission factor results for four road types are shown as FIGURE 19. The results basically follow the features of fuel consumption predictions with some minor differences. In the low-speed range

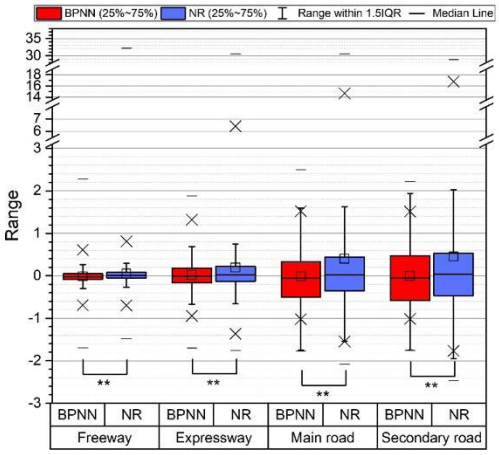


FIGURE 18. Box-plot of BPNN and NR model predicted emission factor error (** means the population mean is significantly different at 0.01 level).

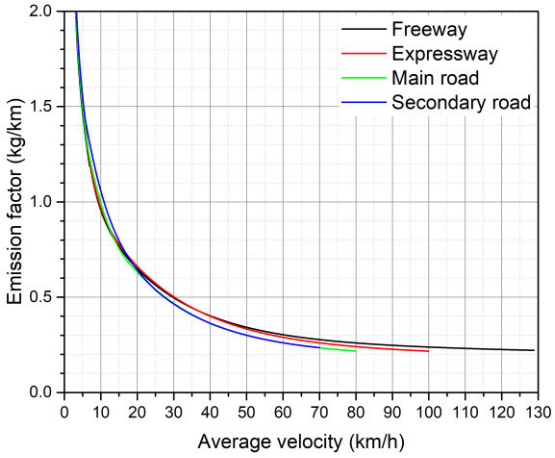


FIGURE 19. Comparison of BPNN predicted emission factor results for four road types.

(5-20km/h), secondary road has the highest emission factor, while freeway and expressway has similarly low ones. And in the middle-speed range (20-80km/h), with the average speed increases, the difference among the four road types are showing more divergences. However, the difference between freeway and expressway, as well as the difference between main and secondary road, are relatively small. In addition, the emission factor of secondary road at 70km/h is 15% lower than freeway, 10% lower than expressway, and similar to main road. The reason of the differences is same to fuel consumption factor, that the smooth traffic on main and secondary road at middle-speed range makes vehicles emit less, while traffic condition on freeway and expressway at the same speed range is congested to some extent. At low-speed range, traffic condition on all four road types are crowded, while traffic on main road and secondary road involves much standing start acceleration at intersections, that lead to a higher emission factor. For smooth traffic condition that vehicles operating around speed limit of the corresponding road type, the emission factors are similarly low.

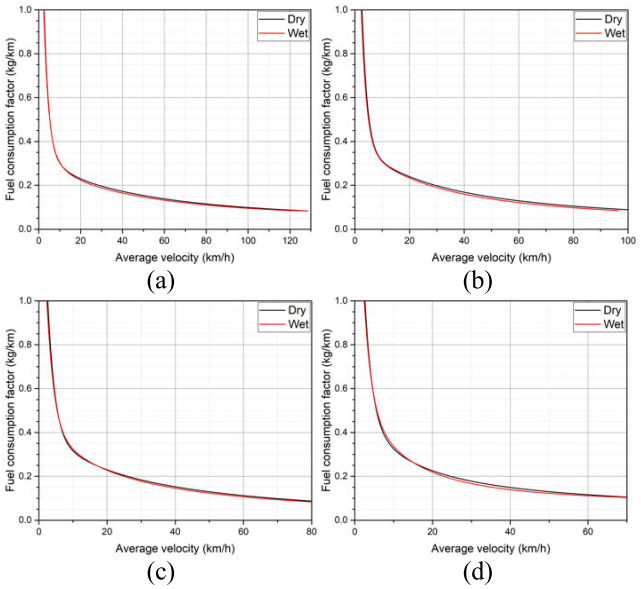


FIGURE 20. BPNN model predicted fuel consumption factors under two weather conditions for (a) freeway, (b) expressway, (c) main road, (d) secondary road.

TABLE 5. Key indexes for BPNN prediction of fuel consumption factors for four road types with weather condition.

Road type	*Weather*	R^2	*MSE*
Freeway	Clear	0.951	6.22×10^{-3}
	Rainy	0.974	6.21×10^{-3}
Expressway	Clear	0.964	2.48×10^{-2}
	Rainy	0.969	2.31×10^{-2}
Main road	Clear	0.97	5.07×10^{-2}
	Rainy	0.973	5.01×10^{-2}
Secondary road	Clear	0.974	5.94×10^{-2}
	Rainy	0.975	5.83×10^{-2}

D. RESULTS OF BPNN PREDICTION MODEL FOR CLEAR AND RAINY WEATHERS

To investigate the influence of rainy weather condition on fuel consumption and emission factors of four road types, another binary variable that indicates the weather condition of rainy or clear is added into the BPNN inputs. The results of BPNN predicted fuel consumption factor under two weather conditions for four road types are presented in FIGURE 20.

It is observed that for each road type, rainy weather conditions have lower fuel consumption factor compared to clear weather conditions. However, the degrees of impact of weather condition for four road types are different. Generally, the BPNN predicted results follow the aforementioned descriptive features of data distribution. The influences on freeway results are mainly in middle-speed range where the difference is up to 4.5%, and in high-speed range the fuel consumption factor for clear and rainy conditions are almost the same. For expressway, the fuel consumption factors in middle- to high- speed range are all lowered in rainy weather, with differences up to 5.9%. Rainy weather has the minimum impact on main roads among all the four road types, that mainly happens in middle-speed range with difference less than 3.9%. And for Secondary road, main difference is on middle-speed range that up to 6.7%. It can be concluded that rainy weather condition has influence on fuel consumption factor in mainly middle-speed range, and expressway and secondary road are most affected road types. From the results of R^2 and *MSE* listed in TABLE 5, it can be seen that compared to the results of the four road types without considering weather conditions, the prediction accuracy for clear weather is quite similar, while for rainy weather, the R^2 increases and *MSE* decreases slightly. It can be inferred that considering the weather condition in BPNN model can increase the prediction accuracy for fuel consumption factor.

TABLE 6. Key indexes for BPNN prediction of emission factors for four road types with weather condition.

Road type	*Weather*	R^2	*MSE*
Freeway	Clear	0.971	3.51×10^{-2}
	Rainy	0.984	3.49×10^{-2}
Expressway	Clear	0.974	0.174
	Rainy	0.977	0.172
Main road	Clear	0.976	0.408
	Rainy	0.978	0.401
Secondary road	Clear	0.978	0.479
	Rainy	0.98	0.471

In the emission factor prediction, weather condition is also considered in BPNN model. It is observed from FIGURE 21 that the impact of rainy weather condition on emission factor is mainly in middle-speed range. The difference ratio of emission factors for the two weather conditions are varied with different road types. Secondary road is the most affected road type with difference up to 5.2%, followed by freeway of 4.7%. Main road and expressway are less affected, with a difference ratio of 3.8% and 3.9% respectively.

The key indexes that evaluate prediction performance of BPNN model for emission factor are listed in TABLE 6. Same trend as fuel consumption factor results exists in the emission factor prediction. Considering weather condition makes emission factor prediction with BPNN model perform slightly better.

In general, from the BPNN model predicted results, it is validated that road type and weather condition do influence the fuel consumption and emission factor. The difference between road types can be up to 17%, while the influence of rainy weather condition for different types of roads ranges from 3.8% to 6.7%.

The results of fuel consumption and emission factors under clear and rainy weather conditions for four road types predicted by NR model are obtained as a comparison. The results of R^2 and *MSE* indicate that the performance of BPNN model is much better than NR model. In addition, the NR model failed to reflect the features shown in the descriptive statistical results of freeway. The detailed information about NR predicted results please see appendix.

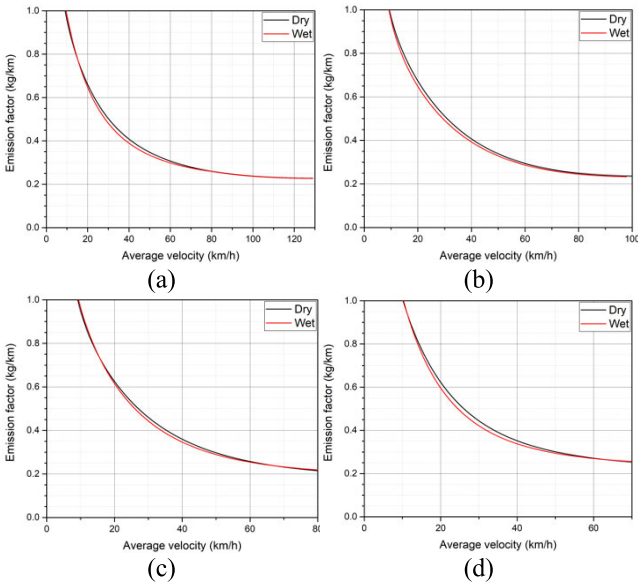


FIGURE 21. BPNN model predicted emission factors under two weather conditions for (a) freeway, (b) expressway, (c) main road, (d) secondary road.

E. DISCUSSIONS

In the aforementioned results and analysis, both descriptive data features and BPNN predicted results show that rainy weather condition mainly have influence on middle-speed range. A possible reason is that traffic condition is quite crowded in low-speed range, which makes drivers pay much attention on road traffic; while many of those records in high-speed range belong to aggressive driving behaviors (drives as fast as clear weather condition) that leads to little difference between two weather conditions. However, since the records of rainy weather condition only accounts for 10% to 15% in total records, and the records in high-speed range are even less. The data sparsity for records in rainy weather condition in high-speed range might affect the data distribution and therefore conceal the inherent characteristics. The impact of weather condition of clear traffic in high-speed range needs to be further investigated with more specific research plan and more collected data.

In this research, the collected data from taxis are applied to generate fuel consumption and emission information. Although some driving behaviors of professional taxi drivers look variant from private car users to some extent, the velocity features are similar. The selected parameters in CMEM match both taxis and most private vehicles, in terms of model year and powertrain characteristics. The speed data were applied as inputs for prediction, and the influence of particularity of data source is reduced to some extent. In addition, we focused on establishing a method to analyze the influence of road type and weather condition on fuel consumption and emissions with big traffic data, it is assumed that there is no significant difference of aforementioned influential factors' effects on different driver population. Also, it is stated in some research [20], [59], [60] that taxi data can be used as floating

TABLE 7. Indicators of prediction performance and regression coefficients of NR fuel consumption factor model for clear and rainy conditions of different road types.

	Road type	Weather condition	NR: $\log\overline{FF} = a\bar{v}^6 + b\bar{v}^5 + c\bar{v}^4 + d\bar{v}^3 + e\bar{v}^2 + f\bar{v} + g$			
			Freeway	Expressway	Main road	Secondary road
Indicator	R^2	Clear	0.490	0.528	0.558	0.583
		Rainy	0.539	0.550	0.564	0.594
	MSE	Clear	6.56×10^{-2}	0.317	0.760	0.944
		Rainy	0.148	0.307	0.775	0.965
Coefficient	a	Clear	4.17×10^{-11}	3.77×10^{-10}	1.47×10^{-9}	4.48×10^{-9}
		Rainy	6.11×10^{-11}	4.25×10^{-10}	1.46×10^{-9}	4.89×10^{-9}
	b	Clear	-1.78×10^{-8}	-1.23×10^{-7}	-3.99×10^{-7}	-1.03×10^{-6}
		Rainy	-2.56×10^{-8}	-1.35×10^{-7}	-3.98×10^{-7}	-1.11×10^{-6}
	c	Clear	3.02×10^{-6}	1.55×10^{-7}	4.15×10^{-5}	8.95×10^{-5}
		Rainy	4.25×10^{-6}	$1.66\text{E}\times10^{-5}$	4.15×10^{-5}	9.51×10^{-5}
	d	Clear	-2.59×10^{-4}	-9.48×10^{-4}	-2.06×10^{-3}	3.69×10^{-3}
		Rainy	-3.50×10^{-4}	-9.96×10^{-4}	-2.06×10^{-3}	-3.86×10^{-3}
	e	Clear	1.17×10^{-2}	2.88×10^{-2}	5.01×10^{-2}	7.37×10^{-2}
		Rainy	1.49×10^{-2}	2.99×10^{-2}	5.04×10^{-2}	7.62×10^{-2}
	f	Clear	-0.268	-0.417	-0.575	-0.697
		Rainy	-0.315	-0.435	-0.583	-0.713
	g	Clear	0.509	0.787	1.037	1.197
		Rainy	0.732	0.838	1.063	1.219

car data (FCD) to infer the traffic condition and analyze the traffic related scientific issues such as emissions and fuel consumption, and data collected from taxis can reflect the traffic condition, which is considered to be a key reason for differences of fuel consumption and emission factors of different road types. Therefore, the results in this research demonstrate the situation of taxis because of the data availability, and it can be a reference for different drivers despite that there are some limitations. Indeed, for many studies performed with data from taxis [29], [61], the researchers believe that the conclusions can be extended to other population with further work. In the future study, more data acquisition devices with higher accuracy are planned to be implemented on floating vehicles which are private cars, and further comparison including the features of driving behaviors of private car owners and taxi drivers would be studied.

It is also concerned that different type of vehicles such as trucks and buses, which only accounts for a small portion in urban road traffic, might have varied reactions in fuel consumption and emission with regard to multiple road types and weather conditions. So it is considerable to collect data from more representative types of vehicles to generate better results in the future.

V. CONCLUSION

In this paper, the influence of road type and rainy weather condition on fuel consumption and emission of urban road transportation is studied by developing a mesoscopic transportation energy consumption and emission model. The modelling method was based on big traffic data and used link-based data segregation strategy and a neural network to estimate the urban link-level fuel consumption and emission factors for on-road vehicles. A mesoscopic transportation fuel consumption and emission model of urban roads in Shenzhen, China was established. The model was based on the

TABLE 8. Indicators of prediction performance and regression coefficients of NR emission factor model for clear and rainy conditions of different road types.

	Road type	Weather condition	NR: $\log\overline{EF} = a\bar{v}^6 + b\bar{v}^5 + c\bar{v}^4 + d\bar{v}^3 + e\bar{v}^2 + f\bar{v} + g$ Freeway	Expressway	Main road	Secondary road
Indicator	R^2	Clear	0.511	0.548	0.573	0.595
		Rainy	0.556	0.568	0.577	0.604
	MSE	Clear	0.532	2.715	6.648	8.084
		Rainy	1.428	2.758	7.002	8.313
Coefficient	a	Clear	5.07×10^{-11}	5.34×10^{-10}	1.45×10^{-9}	4.43×10^{-9}
		Rainy	6.92×10^{-11}	5.70×10^{-10}	1.44×10^{-9}	4.80×10^{-9}
	b	Clear	-2.13×10^{-8}	-1.64×10^{-7}	-3.96×10^{-7}	-1.02×10^{-6}
		Rainy	-2.87×10^{-8}	-1.72×10^{-7}	-3.94×10^{-7}	-1.09×10^{-6}
	c	Clear	3.55×10^{-6}	1.95×10^{-5}	4.12×10^{-5}	8.88×10^{-5}
		Rainy	4.67×10^{-6}	2.01×10^{-5}	4.12×10^{-5}	9.39×10^{-5}
	d	Clear	-2.96×10^{-4}	-1.13×10^{-3}	-2.06×10^{-3}	-3.67×10^{-3}
		Rainy	-3.78×10^{-4}	-1.15×10^{-3}	-2.06×10^{-3}	-3.83×10^{-3}
	e	Clear	1.29×10^{-2}	3.25×10^{-2}	5.05×10^{-2}	7.40×10^{-2}
		Rainy	1.58×10^{-2}	3.29×10^{-2}	5.07×10^{-2}	7.62×10^{-2}
	f	Clear	-0.290	-0.453	-0.592	-0.711
		Rainy	-0.333	-0.460	-0.598	-0.725
	g	Clear	1.686	1.982	2.202	2.360
		Rainy	1.909	2.017	2.227	2.380

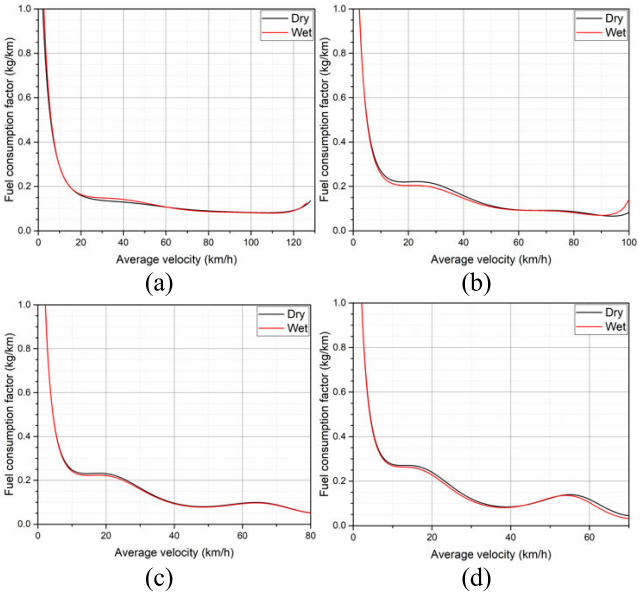


FIGURE 22. NR model predicted fuel consumption factors under two weather conditions for (a) freeway, (b) expressway, (c) main road, (d) secondary road.

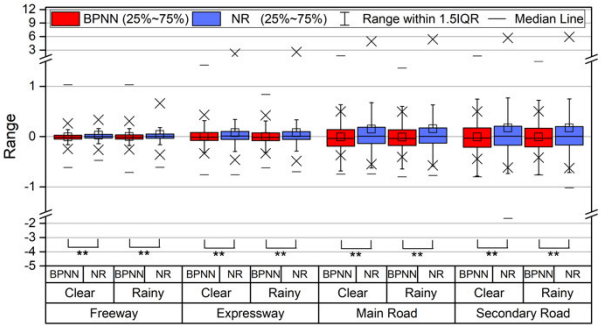


FIGURE 23. Box-plot of BPNN and NR model predicted fuel consumption factor error under two weather conditions of four road types (means the population mean is significantly different at 0.01 level).**

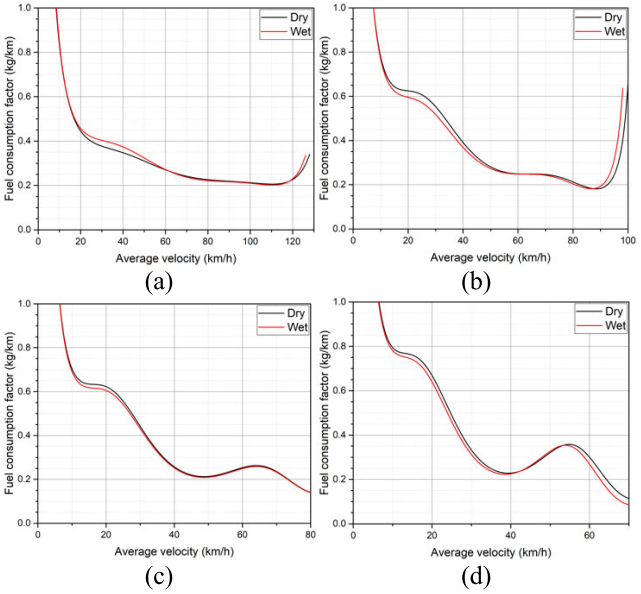


FIGURE 24. NR model predicted emission factors under two weather conditions for (a) freeway, (b) expressway, (c) main road, (d) secondary road.

instantaneous data collected from 10,944 taxis in operation for 2 consecutive months in Shenzhen. The fuel consumption and emission factor of four road types and two weather conditions are analyzed. The main conclusions reached in this study can be outlined as follows:

(1) The fuel consumption factor and emission factor of urban roads in Shenzhen were calculated and predicted with both the NR and BPNN methods. The BPNN model outperformed the NR method with the data in this research.

(2) Different road types have varied fuel consumption and emission factors, while freeway and expressway are less distinguishing, main road and secondary road are similar. The difference ratio among four road types is up to 17% for fuel consumption factor, and 15% for emission factor according to BPNN predictions.

(3) In low-speed range, freeway and expressway have lower fuel consumption and emission factors, while main road and secondary road have lower ones in middle-speed range. With clear traffic condition that vehicles running around the speed limit, four road types have similarly low fuel consumption and emission factors.

(4) In middle-speed range, rainy weather condition makes driver behaviors less aggressive, and has fuel consumption factor lower than clear weather condition by 3.9% to 6.7% for different road types, and 3.8% to 5.2% in terms of emission factor.

(5) The main reason for different features of the four road types in low- and middle-speed range are that the traffic conditions and driving patterns tend to vary much at the same average speed for freeway, expressway, main road, and secondary road.

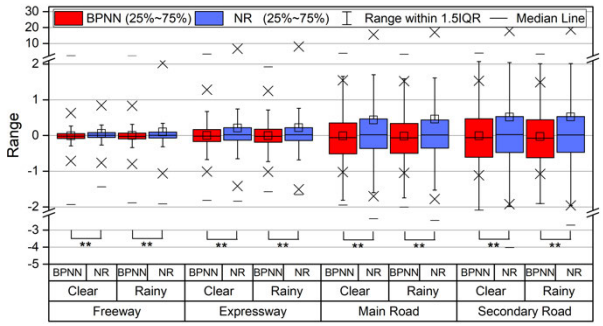


FIGURE 25. Box-plot of BPNN and NR model predicted emission factor error under two weather conditions of four road types (** means the population mean is significantly different at 0.01 level).

However, because of the limited data availability, there are still potential improvements that can be made in the future work. 1) More detailed data should be collected to generate denser CMEM simulation results. 2) If the data could include information from various types of vehicles, the model would be more suitable for different kinds of on-road vehicles, and would be more helpful in practice. Therefore, in a future study, it is suggested that finer data from different vehicles be added to the database to make the model more general and applicable.

APPENDIX

See Tables 7 and 8. See Figures 22–25.

REFERENCES

[1] *Beijing Transportation Development Annual Report 2017*, BT Inst., Accra, Ghana, 2017.
[2] *China Vehicle Environmental Management Annual Report*, Ministry Ecol. Environ. People's Republic China (MoEaE), Beijing, China, 2018.
[3] *China Statistical Yearbook 2018*. National Bureau of Statistics of China (NBoS), Beijing, China, 2018.
[4] D. L. Greene, H. H. Baker, Jr., and S. E. Plotkin, "Reducing greenhouse gas emissions from U.S. Transportation," Pew Center Global Climate Change, Arlington, VA, USA, Jan. 2011. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/23588
[5] H. Wang, C. Chen, C. Huang, and L. Fu, "On-road vehicle emission inventory and its uncertainty analysis for Shanghai, China," *Sci. Total Environ.*, vol. 398, nos. 1–3, pp. 60–67, Jul. 2008.
[6] G. Lenaers. (2009). *Real Life CO_2 Emission and Consumption of Four Car Powertrain Technologies Related to Driving Behaviour and Road Type*. [Online]. Available: https://doi.org/10.4271/2009-24-0127
[7] M. V. Faria, G. O. Duarte, R. A. Varella, T. L. Farias, and P. C. Baptista, "How do road grade, road type and driving aggressiveness impact vehicle fuel consumption? Assessing potential fuel savings in Lisbon, Portugal," *Transp. Res. D, Transp. Environ.*, vol. 72, pp. 148–161, Jul. 2019.
[8] H. Y. Tong, W. T. Hung, and C. S. Cheung, "On-road motor vehicle emissions and fuel consumption in urban driving conditions," *J. Air Waste Manage. Assoc.*, vol. 50, no. 4, pp. 543–554, Apr. 2000.
[9] Y. Li, G. Tang, J. Du, N. Zhou, Y. Zhao, and T. Wu, "Multilayer perceptron method to estimate real-world fuel consumption rate of light duty vehicles," *IEEE Access*, vol. 7, pp. 63395–63402, 2019.
[10] P. Ping, W. Qin, Y. Xu, C. Miyajima, and K. Takeda, "Impact of driver behavior on fuel consumption: Classification, evaluation and prediction using machine learning," *IEEE Access*, vol. 7, pp. 78515–78532, 2019.
[11] Y. Yao, X. Zhao, Y. Zhang, C. Chen, and J. Rong, "Modeling of individual vehicle safety and fuel consumption under comprehensive external conditions," *Transp. Res. D, Transp. Environ.*, vol. 79, Feb. 2020, Art. no. 102224.
[12] S. Zhang, D. Peng, Y. Li, L. Zu, M. Fu, H. Yin, and Y. Ding, "Study on the real-world emissions of rural vehicles on different road types," *Environ. Pollut.*, vol. 273, Mar. 2021, Art. no. 116453.
[13] E. Demir, T. Bektaş, and G. Laporte, "A review of recent research on green road freight transportation," *Eur. J. Oper. Res.*, vol. 237, no. 3, pp. 775–793, Sep. 2014.
[14] T. J. Barlow and P. G. Boulter, "Emission factors 2009: Report 2—A review of the average-speed approach for estimating hot exhaust emissions," Berkshire, Wokingham, U.K., Trl Published Project Rep., 2009. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/4248/report-2.pdf
[15] 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.
[16] I. Ben Dhaou, "Fuel estimation model for ECO-driving and ECO-routing," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 37–42.
[17] J. Sun, D. Niu, S. Chen, and K. Li, "Development and investigation of a dynamic eco-driving speed guidance strategy for signalized highway traffic," in *Proc. Transp. Res. Board 92nd Annu. Meeting*, Washington, DC, USA, Jan. 2013, Paper 13-2272.
[18] M. Fu, J. A. Kelly, and J. P. Clinch, "Estimating annual average daily traffic and transport emissions for a national road network: A bottom-up methodology for both nationally-aggregated and spatially-disaggregated results," *J. Transp. Geography*, vol. 58, pp. 186–195, Jan. 2017.
[19] B. Jing, L. Wu, H. Mao, S. Gong, J. He, C. Zou, G. Song, X. Li, and Z. Wu, "Development of a vehicle emission inventory with high temporal–spatial resolution based on NRT traffic data and its impact on air pollution in Beijing—Part 1: Development and evaluation of vehicle emission inventory," *Atmos. Chem. Phys.*, vol. 16, no. 5, pp. 3161–3170, Mar. 2016.
[20] M. Nyhan, S. Sobolevsky, C. Kang, P. Robinson, A. Corti, M. Szell, D. Streets, Z. Lu, R. Britter, S. R. H. Barrett, and C. Ratti, "Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model," *Atmos. Environ.*, vol. 140, pp. 352–363, Sep. 2016.
[21] D. Tuia, M. O. de Eicker, R. Zah, M. Osses, E. Zarate, and A. Clappier, "Evaluation of a simplified top-down model for the spatial assessment of hot traffic emissions in mid-sized cities," *Atmos. Environ.*, vol. 41, no. 17, pp. 3658–3671, Jun. 2007.
[22] X. Ma, "Towards intelligent fleet management: Local optimal speeds for fuel and emissions," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 2201–2206.
[23] B. T. Morris, C. Tran, G. Scora, M. M. Trivedi, and M. J. Barth, "Real-time video-based traffic measurement and visualization system for energy/emissions," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1667–1678, Dec. 2012.
[24] W. Li, G. Wu, Y. Zhang, and M. J. Barth, "A comparative study on data segregation for mesoscopic energy modeling," *Transp. Res. D, Transp. Environ.*, vol. 50, pp. 70–82, Jan. 2017.
[25] X. Huang and H. Peng, "Eco-routing based on a data driven fuel consumption model," 2018, *arXiv:1801.08602*. [Online]. Available: http://arxiv.org/abs/1801.08602
[26] X. Che, Q. Yu, G. Ren, and R. Liang, "Average fuel consumption calculation model of touring coach based on OBD data—A case study in city X," in *Proc. Int. Conf. Civil, Transp. Environ.*, 2016, pp. 1370–1374.
[27] H. Liu, X. Chen, Y. Wang, and S. Han, "Vehicle emission and near-road air quality modeling for Shanghai, China: Based on global positioning system data from taxis and revised MOVES emission inventory," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2340, no. 1, pp. 38–48, Jan. 2013.
[28] W. Huang, Y. Guo, and X. Xu, "Evaluation of real-time vehicle energy consumption and related emissions in China: A case study of the Guangdong–Hong Kong–Macao greater bay area," *J. Cleaner Prod.*, vol. 263, Aug. 2020, Art. no. 121583.
[29] T. Li, J. Wu, A. Dang, L. Liao, and M. Xu, "Emission pattern mining based on taxi trajectory data in Beijing," *J. Cleaner Prod.*, vol. 206, pp. 688–700, Jan. 2019.
[30] Z. He, W. Zhang, and N. Jia, "Estimating carbon dioxide emissions of freeway traffic: A spatiotemporal cell-based model," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 5, pp. 1976–1986, May 2019.
[31] D. Karbowski and K. Zhang, "Evaluation of energy consumption of vehicles along a stretch of congested freeway," in *Proc. Transp. Res. Board 91st Annu. Meeting*, Washington, DC, USA, Jan. 2012, Paper 12-4736.
[32] C. Pasquale, I. Papamichail, C. Roncoli, S. Sacone, S. Siri, and M. Papageorgiou, "Two-class freeway traffic regulation to reduce congestion and emissions via nonlinear optimal control," *Transp. Res. C, Emerg. Technol.*, vol. 55, pp. 85–99, Jun. 2015.
[33] Y. Arfat, S. Usman, R. Mehmood, and I. Katib, "Big data tools, technologies, and applications: A survey," in *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*, R. Mehmood, S. See, I. Katib, and I. Chlamtac, Eds. Cham, Switzerland: Springer, 2020, pp. 453–490.

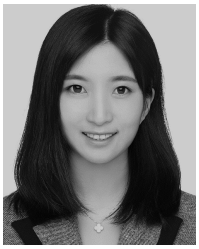
[34] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of artificial intelligence in transport: An overview," *Sustainability*, vol. 11, no. 1, p. 189, Jan. 2019.
[35] O. Azeez, B. Pradhan, H. Shafri, N. Shukla, C.-W. Lee, and H. Rizeei, "Modeling of CO emissions from traffic vehicles using artificial neural networks," *Appl. Sci.*, vol. 9, no. 2, p. 313, Jan. 2019.
[36] D. Kim and J. Lee, "Application of neural network model to vehicle emissions," *Int. J. Urban Sci.*, vol. 14, no. 3, pp. 264–275, Dec. 2010.
[37] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017.
[38] S. Kaffash, A. T. Nguyen, and J. Zhu, "Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis," *Int. J. Prod. Econ.*, vol. 231, Jan. 2021, Art. no. 107868.
[39] S. Kanarachos, J. Mathew, and M. E. Fitzpatrick, "Instantaneous vehicle fuel consumption estimation using smartphones and recurrent neural networks," *Expert Syst. Appl.*, vol. 120, pp. 436–447, Apr. 2019.
[40] E. Alomari, I. Katib, and R. Mehmood, "Iktishaf: A big data road-traffic event detection tool using Twitter and spark machine learning," *Mobile Netw. Appl.*, pp. 1–16, Aug. 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11036-020-01635-y
[41] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, "Smarter traffic prediction using big data, in-memory computing, deep learning and GPUs," *Sensors*, vol. 19, no. 9, p. 2206, May 2019.
[42] Z. Feng, M. Yang, C. Ma, K. Jiang, Y. Lei, W. Huang, Z. Huang, J. Zhan, and M. Zhou, "Driving anger and its relationships with type a behavior patterns and trait anger: Differences between professional and non-professional drivers," *PLoS ONE*, vol. 12, no. 12, Dec. 2017, Art. no. e0189793.
[43] T. Chen, N. N. Sze, S. Newnam, and L. Bai, "Effectiveness of the compensatory strategy adopted by older drivers: Difference between professional and non-professional drivers," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 77, pp. 168–180, Feb. 2021.
[44] T. Chen, N. N. Sze, and L. Bai, "Safety of professional drivers in an ageing society—A driving simulator study," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 67, pp. 101–112, Nov. 2019.
[45] M. Barth, "The comprehensive modal emission model (CMEM) for predicting light-duty vehicle emissions," in *Proc. 4th Transp. Planning Air Qual., Persistent Problems Promising Solutions*. Reston, VA, USA: ASCE, 2010, pp. 126–137.
[46] M. Barth, F. An, J. Norbeck, and M. Ross, "Modal emissions modeling: A physical approach," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1520, no. 1, pp. 81–88, Jan. 1996.
[47] G. Scora and M. Barth, "Comprehensive modal emissions model (CMEM), version 3.01, user's guide," Center Environ. Res. Technol., Univ. California, Riverside, CA, USA, Tech. Rep., 2006.
[48] 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, Jan. 2013.
[49] H. Wang and L. Fu, "Developing a high-resolution vehicular emission inventory by integrating an emission model and a traffic model: Part 1—Modeling fuel consumption and emissions based on speed and vehicle-specific power," *J. Air Waste Manage. Assoc.*, vol. 60, no. 12, pp. 1463–1470, Dec. 2010.
[50] S. Bai, Y.-C.-E. Chiu, and D. A. Niemeier, "A comparative analysis of using trip-based versus link-based traffic data for regional mobile source emissions estimation," *Atmos. Environ.*, vol. 41, no. 35, pp. 7512–7523, Nov. 2007.
[51] M. H. Askariyeh and M. Arhami, "Projecting emission reductions from prospective mobile sources policies by road link-based modelling," *Int. J. Environ. Pollut.*, vol. 53, nos. 1–2, pp. 87–106, 2013.
[52] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *J. Transp. Eng.*, vol. 128, no. 2, pp. 182–190, Mar. 2002.
[53] S. K. Pathak, V. Sood, Y. Singh, and S. A. Channiwala, "Real world vehicle emissions: Their correlation with driving parameters," *Transp. Res. D, Transp. Environ.*, vol. 44, pp. 157–176, May 2016.
[54] L. Jing, J. H. Cheng, J. Y. Shi, and H. Fei, "Brief introduction of back propagation (BP) neural network algorithm and its improvement," in *Proc. Adv. Comput. Sci. Inf. Eng.*, 2012, pp. 553–558.
[55] W. W. Hsieh and B. Tang, "Applying neural network models to prediction and data analysis in meteorology and oceanography," *Bull. Amer. Meteorol. Soc.*, vol. 79, no. 9, pp. 1855–1870, Sep. 1998.
[56] K. Hornik, *Approximation Capabilities of Multilayer Feedforward Networks*. Amsterdam, The Netherlands: Elsevier, 1991.
[57] *Annual Report on Road Network Density in Major Chinese Cities*, China Acad. Urban Planning Des. (CAoUP D), Fujian, province, 2018.
[58] R. Raymond, T. Morimura, T. Osogami, and N. Hirosue, "Map matching with hidden Markov model on sampled road network," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, 2012, pp. 2242–2245.
[59] X. Wang, H. Liu, R. Yu, B. Deng, X. Chen, and B. Wu, "Exploring operating speeds on urban arterials using floating car data: Case study in Shanghai," *J. Transp. Eng.*, vol. 140, no. 9, Sep. 2014, Art. no. 04014044.
[60] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," presented at the 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, Aug. 2014, doi: 10.1145/2623330.2623653.
[61] Y. Huang, D. Sun, and L.-H. Zhang, "Effects of congestion on drivers' speed choice: Assessing the mediating role of state aggressiveness based on taxi floating car data," *Accident Anal. Prevention*, vol. 117, pp. 318–327, Aug. 2018.

RUI SHANG received the B.S. degree from Tongji University, China, in 2013, and the M.S. degree from Tsinghua University, China, in 2016, where he is currently pursuing the Ph.D. degree with Tsinghua-Berkeley Shenzhen Institute. His research interests include intelligent transportation systems, cooperative driving, and sustainable transportation systems.

YI ZHANG (Member, IEEE) received the B.S. and M.S. degrees from Tsinghua University, China, in 1986 and 1988, respectively, and the Ph.D. degree from the University of Strathclyde, U.K., in 1995. He is currently a Professor of control science and engineering with Tsinghua University. His active research areas include intelligent vehicle-infrastructure cooperative systems, analysis of urban transportation systems, urban road network management, traffic data fusion and dissemination, and urban traffic control and management. His current research interests include intelligent transportation systems, advanced control theory and applications, advanced detection and measurement, and systems engineering.

ZUO-JUN MAX SHEN (Member, IEEE) received the Ph.D. degree from Northwestern University, Evanston, IL, USA, in 2000.

He is the Chancellor's Professor with the Department of Industrial Engineering and Operations Research and the Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley, CA, USA. He has been active in integrated supply chain design and management, applied optimization, decision making with limited information, public transportation systems, and transportation systems sustainability.



YI ZHANG received the B.S. degree from Tsinghua University, the M.S. degree from Imperial College London, and the Ph.D. degree from the University of Cambridge. She is currently an Assistant Professor with the Future Human Habitat Division, Tsinghua Shenzhen International Graduate School (SIGS), Tsinghua University. Her research interests include application planning of renewable energy at city scale with city information modeling (CIM), intelligent building energy systems, big data analysis on energy consumption behavior, and electrical vehicles operation and management.