Emission pattern mining based on taxi trajectory data in Beijing

Tingting Li a, *, Jianping Wu a, Anrong Dang b, Lyuchao Liao a, Ming Xu a

a Department of Civil Engineering, Tsinghua University, Beijing, 100084, China
b School of Architecture, Tsinghua University, 100084, China

Abstract
Traffic-related air pollution has been one of the major environmental problems in China. It is urgent to explore the urban traffic emission patterns for the low-carbon urban planning and traffic management. With this purpose, a new urban traffic emission analysis model is proposed in this paper. The traffic analysis zones (TAZs) are treated as the analysis unit. Then the spatial and temporal dynamic emission patterns are studied based on taxi GPS data in Beijing. The whole urban area of Beijing is divided into 33 TAZs depending on the feature of road networks. And the trip patterns of TAZs are extracted. The instantaneous emissions of CO2, NOx, VOC and PM within and between TAZs are estimated. The relationships between emissions and road densities are studied. The results demonstrated that (1) the highest taxi trips during the day occur at 10:00, 16:00 and 20:00. (2) The variations of the 4 pollutants within and between TAZs are similar. The emissions of TAZs with business centers, entertainment centers and transportation hubs are obviously higher than others. For TAZs within the 5th Ring Road, the northern emissions are stronger than the southern. Emission patterns can be divided into 3 types, corresponding to the time periods of 0:00–3:00, 3:00–6:00 and 6:00–24:00. (3) There is a positive relationship between emissions within and between TAZs and road densities. And when the road densities of TAZs are bigger than 0.6, emissions within TAZs will rise up obviously with the increasing road densities.

1. Introduction

Urban travel demand is growing quickly with the urbanization process in China. And it causes many traffic problems. The rapidly increasing greenhouse gas (GHG) emissions and air pollution caused by fossil fuel from road traffic have attracted much attention. It has become an imperative task to conduct energy-saving and emission reduction work. The central and local municipal governments in China have centered on emission mitigation (Wu et al., 2017). In the aspect of vehicles, there are efforts about the development of China 1–6 emission standards (MEE and SAMR, 2001, 2005b, c, 2013, 2016), enhanced I/M programs for in-use vehicles (Wu et al., 2011), yellow-labeled vehicles classification and regulations (State Council P. R. China, 2013), improvement on vehicle engines (Huo et al., 2012a) and motorcycles and heavy-duty truck restrictions (MEE and SAMR, 2005a). In the aspect of fuel, the main efforts contain the fuel quality improvement (Yue et al., 2015) and new energy vehicle promotions (State Council P. R. China, 2012). In the aspect of traffic management, there are policies about public traffic infrastructure enhancement, public transport subsidy (MOHURD. P. R. China, 2016), bus lanes (Yu et al., 2015), bus rapid transit system (Liu and Teng, 2014), new energy buses (Lin and Tan, 2017) and taxis, odd and even number restrictions (Wu et al., 2017) and old vehicle regulations (Zhang, 2014). As for economic measures, there are car purchase restrictions (Diao et al., 2016), increased parking fees, new energy vehicle subsidies (Helveston et al., 2015; Ma et al., 2017; Xu et al., 2017) and tax reduction. Besides, the policy about the congestion fee is being considered.

The advanced information and communication technologies promote the development of smart city. At the same time, various traffic detection data is applied into the broader research field. The diversified traffic detection data can be divided into 3 classes. The first one is the individual travel data, containing the smart card records of buses and subway (Huang et al., 2017b), bike-sharing...
data (Zhang et al., 2015), car-sharing data (He et al., 2014; Yu et al., 2017), cellphone signal data (Mao et al., 2016), cellphone navigation data and other application data of cellphone. The second one is the regulation data of enterprises, such as logistic data of commercial vehicles (Joubert and Meintjes, 2015). The third one is the road detection data, including GPS trajectory data (Hsueh and Chen, 2018; Li et al., 2015), electronic toll data based on RFID (Tseng et al., 2014), vehicle detection data of inductive loop detectors, microwave detectors, infrared detectors, laser radar detectors, ultrasonic detectors and closed-circuit television cameras (Seo et al., 2017). The appearance of big data mining techniques makes it possible to explore the patterns of large-scale urban traffic and air pollution. Compared with the social media data, road detection data and mobile communication network data, taxi GPS trajectory data has advantages of wide coverage, high spatial resolution and low installation and maintenance cost. There have been related studies based on GPS records in urban traffic management and control such as traffic congestion and incident detection (D’Andrea and Marcelloni, 2017; Munoz-Organero et al., 2018), traffic state estimation (Hsueh and Chen, 2018; Zhan et al., 2017) and route choice (Ciscal-Terry et al., 2016).

In addition, energy-saving and emission-reduction work has become a global issue, leading to some research on traffic emissions (Huang et al., 2017a, 2017b; Luo et al., 2017). There are two types of approaches in constructing high-resolution traffic emission inventories, that is, top-down and bottom-up methods (Liu et al., 2018). Top-down approaches are applied into the estimation of annual vehicle emission inventories, in which the emission is calculated relying on the average trips and traffic flow, average velocity, average mileage and fuel type and then allocated to each cell (Gómez et al., 2018; Puliafito et al., 2015). But this allocation method may result in low accuracies and uncertainties in the spatial emission estimation (Gómez et al., 2018). What’s more, the restricted access of high-emitting vehicles in some urban area may cause estimation errors. Thus, many works are relied on the bottom-up rather than top-down approaches. Some scholars demonstrated that the data collected in finer spatial scale in bottom-up method could be helpful for more accurate emission estimation (López-Aparicio et al., 2017; Pallavidino et al., 2014). Besides, the traffic simulation models are used to improve emission inventories (Hofer et al., 2018; Osorio and Nanduri, 2015; Xu et al., 2018; Zhu and Zhang, 2017). However, the emission factor varies according to road conditions, vehicle mileage and locations. The real-time driving states are neglected in these aforementioned works, which are necessary for more accurate emission estimation (Nyhan et al., 2016). Even in the MOBILE model (USEPA, 2012), there are only 14 types of driving states. Early in 2003, Frey et al. (2003) demonstrated that the emissions of hydrocarbons and carbon monoxide in the acceleration period were 5 times greater than the idling period. To solve the above problems, the vehicle driving states with speed and acceleration data should be extracted for more precise emission inventory estimation.

The taxi GPS trace has advantages of monitoring the whole trip and reflecting fine-gained individual space-time trajectories. The individual traces contain the instantaneous speed, accelerations, positions and directions (Nyhan et al., 2016), which are precise to estimate the traffic emissions. Relying on GPS data, some researchers allocated the accumulated emissions calculated by microscopic emission model into the road network and observed the spatial distributions of emissions (Luo et al., 2017; Nyhan et al., 2016). Other scholars studied emission inventories in the resolution of grid cells which split the geographic area into several parts, such as 1 x 1 km resolution (Gómez et al., 2018; Gois et al., 2004; Wang et al., 2010). However, the trips are not related to a single location or a single road, they are regional behaviors. And a trip covers several roads and regions. Thus it is necessary for scholars to analyze traffic emissions caused by trips from the viewpoint of regions. Particularly, traffic analysis zones (TAZs) are defined based on regional attributes such as geographic features. The TAZ is the basic analysis unit of the traditional four-step urban travel demand forecasting model (Agrawal et al., 2018; Sun et al., 2016). The trip demand between the TAZs is estimated, which is useful for the heavily loaded area identification. Then the planners make decisions by taking the feature of TAZs into account. Poku-Boansi and Cobbinah (2017) studied the land use mode and urban travel based on TAZs. Jang et al. (2017) explored characteristics of urban built environment by observing traffic crash occurrences of TAZs. Lee et al. (2018) analyzed the spatial patterns of vehicle crash proportion at TAZs level. TAZs are important analysis units in the related research. But to authors’ knowledge, there is a lack of the research about traffic emissions of traffic analysis zones.

A new urban traffic emission analysis model is proposed in this paper. Previous studies are related to emissions on roads. But there is a lack of research related to emission patterns of TAZs. A TAZ is a basic unit in traffic planning and management. Thus the TAZ is regarded as the unit of real-time urban traffic emission patterns. The purpose of this study is to explore the pattern of traffic emissions of TAZs in the urban area. Hence, the whole urban area of Beijing is divided into 33 TAZs based on the feature of road networks. The GPS data of 12409 taxis in Beijing are analyzed. With the help of a microscopic emission model, emissions of CO, NOx, VOC and PM are estimated. The emissions within and between the TAZs thereby are identified. What’s more, the relationships between road features and emissions of TAZs are explored.

2. Data and methodology

2.1. Data source and data processing

Emissions within and between the TAZs not only represent spatial distributions of traffic pollution, but also reflect the trend of urban trips. As a general case in this study, Beijing is one of the largest metropolises all over the world, which has experienced explosive growth in recent years. At the end of the year 2015, the population of Beijing is up to 21.71 million with the GDP per capita RMB 106.28 thousand. The municipal administrative area is 16410.54 km². And there are 68284 taxis in Beijing according to Wen et al. (2016). The data in this paper originates from taxi GPS trajectory records in Beijing on 1st November, 2012 (workday), which contains 12409 taxis. Each taxi is sampled average each 32 s and there are 68284 taxis in Beijing according to Wen et al. (2016). The data in this paper originates from taxi GPS trajectory records in Beijing on 1st November, 2012 (workday), which contains 12409 taxis. Each taxi is sampled average each 32 s on a 24-h basis as Table 1 shows. The major area with taxi data intensively available is within the 6th Ring Road. And the longitude and latitude of the research area is from 116.164 to 116.575 and 39.743 to 40.098. As Table 2 displays, each data record consists of 32 s on a 24-h basis as Table 1 shows. The major area with taxi data intensively available is within the 6th Ring Road. And the longitude and latitude of the research area is from 116.164 to 116.575 and 39.743 to 40.098. As Table 2 displays, each data record consists of
eliminated. Besides, the records with continuous time, the same taxi ID and operation state are extracted as representative trips.

2.2. Emission model

A complete taxi trajectory can be partitioned into several subdivisions which correspond to GPS records with speed and acceleration information. Pollutant emissions of each subdivision can be estimated referring to the microscopic emission model (Int Panis et al., 2006). Then they are summed up for working out emissions of each trip. The instantaneous traffic emission model is applied to analyze emissions (Nyhan et al., 2016; Osorio and Nanduri, 2015). And Harrington et al. (1998) believed that the emission rates can be calculated according to the base emission rate and the correction factors. In this paper, real-time emissions of pollutants CO2, NOx, VOC and PM are estimated. For a taxi whose ID is signed with letter n, the emission rate of pollutant p in time interval $\Delta t$ can be calculated as follows.

$$E_{n,t}^p = E_{n,\text{basic}}(t)\cdot C_{n,\text{p}}^i \cdot C_{n,\text{f}}$$

$E_{n,t}^p$ is the instantaneous emission rate of pollutant p (g/s) for the vehicle n. $E_{n,\text{basic}}(t)$ is the basic emission rate proposed by Int Panis et al. (2006). And the variable p can represent CO2, NOx, VOC and PM. $C_{n,\text{p}}^i$, $C_{n,\text{f}}$ and $C_{n,\text{p}}^5$ are correction coefficients of vehicle emission control technology, taxi mileage and driving environment.

(1) Basic emission rate

$$E_{n,\text{basic}}(t) = \max\left[ E_{0}^p f_1^p + f_2^p v_n(t) + f_3^p v_n(t)^2 + f_4^p a_n(t) + f_5^p a_n(t)^2 \right]$$

Variables $v_n(t)$ and $a_n(t)$ represent the speed (m/s) and acceleration (m/s²) of car n at time t respectively. $E_{0}^p$ is the low bound emission rate of pollutant p (g/s). Variables $f_1^p$, $f_2^p$, $f_3^p$, $f_4^p$, $f_5^p$ and $f_6^p$ serve as the emission constants differing with taxi ID and pollutants. And their values also vary depending on the instantaneous moving state of the car. For pollutants NOx and VOC, the acceleration $a_n(t)$ should be compared with $-0.5$ m/s² firstly. The specific values are determined according to emission functions of petrol cars proposed by Int Panis et al. (2006).

(2) Correction coefficient of vehicle emission control technology

The coefficient $C_{n,\text{f}}^i$ corrects the emission difference caused by vehicle emission control technologies of various emission standards, which can be indirectly reflected by the average emission factors. The China 4 emission standard has been conducted since 2008 for light-duty gasoline vehicles (Wu et al., 2011). And the share of gasoline taxis in Beijing was 100% in 2010 (Yang, 2018; Zhang et al., 2014). The oldest taxis are 5 years old (Lin et al., 2009), which are in compliance with China 4 emission standard. Then the taxis in this paper are believed to be gasoline cars with China 4 emission standard. Besides, the vehicle emission standards China 1–5 are considered to be equivalent to Euro 1–5 standards (Yue et al., 2015).

$$C_{n,\text{f}}^i = \frac{E_{n,t}^p}{\sum_{i=0}^{3} P_i \cdot E_{n,t}^p}$$

As Eq. (3) shows, $E_{n,t}^p$ means the emission factor of pollutant p in China 4 standard. The variable i means the emission standard, for instance, 0 for Euro 0, 1 for Euro 1. $P_i$ is the ratio of vehicles in Euro i standard in Panis’ measurement (Int Panis et al., 2006). And the variable $E_{n,t}^p$ can represent the emission factor of various pollutants which are listed in Table 3.

(3) Correction coefficient of taxi mileage

The basic emission rates are not adequate for emission estimation of taxis in Beijing. Therefore, the taxi correction coefficient is

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>CO2 (g/km)</th>
<th>NOx (g/km)</th>
<th>VOC (mg/km)</th>
<th>PM (mg/km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euro 0</td>
<td>228</td>
<td>1.9</td>
<td>469.3</td>
<td>7.00</td>
</tr>
<tr>
<td>Euro 1</td>
<td>212</td>
<td>1.0</td>
<td>80.7</td>
<td>1.90</td>
</tr>
<tr>
<td>Euro 2</td>
<td>204</td>
<td>0.47</td>
<td>56.8</td>
<td>0.56</td>
</tr>
<tr>
<td>Euro 3</td>
<td>193</td>
<td>0.23</td>
<td>25.6</td>
<td>0.53</td>
</tr>
<tr>
<td>Euro 4</td>
<td>178</td>
<td>0.05</td>
<td>14.9</td>
<td>0.49</td>
</tr>
</tbody>
</table>
put forward in this paper. The age of the vehicle is very important to emission estimation. The corrected emission rate reflects the degradation rate considering the vehicle mileage. As Eq. (4) shows, \( R_i \) can represent the age ratio of taxis in Beijing, which is listed in Table 4 (Lin et al., 2009). \( M_{\text{MEAN}} \) is the average mileage (Zhang et al., 2014). \( D_i^p \) is the mileage degradation rate which is function of the pollutant and mileage. Lin et al. (2009) also researched the relationship between the vehicle age and mileage in China, which was helpful in calculating the value of \( D_i^p \). And the parameters are listed in Table 5. For VOC and PM, there is little degradation research, which will be supplemented in the further field study.

\[
C_p = \sum_{i=1}^{5} R_i \cdot D_i^p \tag{4}
\]

\[
D_i^p = g^p \times M_{\text{MEAN}} + b^p \tag{5}
\]

(4) Correction coefficient of environment

The environment correction coefficient \( C_p \) reflects the influence of the geographical location on emissions. And it is the integration of altitude, temperature and humidity. In this paper, the temperature factor is considered. Joumard et al. (2009) researched the relationships between temperature and emission as Eq. (6) shows. \( T \) is the value of temperature with unit ‘C. The average temperature of 1st November, 2012 is defined as 9°C (php.weather.sina.com, 2012). Variables \( g^p \) and \( h^p \) are the correction parameters of pollutant \( p \).

\[
C_e = g^p \times T + h^p \tag{6}
\]

2.3. Map segmentation

Differing from previous studies related to traffic emissions of urban roads, the spatial emission patterns within and between TAZs are explored in this paper. The research area is divided into several TAZs with the following 3 rules.

(1) Each region is rational and meaningful in real Beijing map.
(2) The features of various regions are comparable.
(3) There is no obvious difference in road densities for the same TAZ.

The urban area is split by arterial roads into various parts (Yuan et al., 2012). And the roads have strong influence on socio-economic activities. For the first rule, the city map can be segmented based on original road networks. Citizens live in these regions and take daily trips. Infrastructures with different functions are located in these regions. Moreover, the area of each TAZ should be approximate, leading to the meaningful comparison. After the segmentation operation of arterial roads, these obtained small regions are clustered with the third rule.

2.3.1. Morphological operation

Mathematical morphology is a kind of theory in image and signal processing, of which the basic idea is to identify and analyze the image by utilizing structural elements with particular shape to measure and extract the corresponding shape in an image. In this paper, urban roads are extracted as line elements. And the intersections, roundabouts, lanes and other elements of real road networks are simplified by the morphological operation, leading to the lower-complexity road networks. For instance, there are several closed inter-cells at roundabout in Fig. 2(a) where the road network binaryzation operation has been completed. The morphological operation in this paper contains 2 steps.

(1) The first step is road dilation. Dilation is the basic operation of morphological process. The excess network information such as the small connected parts caused by lanes, overpass, bridges and other elements can be eliminated with dilation. Letter A and B are defined as the sets in two-dimensional integer space \( \mathbb{Z}^2 \), which can be denoted as the expression \( A, B \subseteq \mathbb{Z}^2 \). Dilation operations of set A with set B is signified in Eq. (7). After parameter verification, proper structural element B is chosen. And the result proves that the gap between roads and bridges can be eliminated.

\[
A \oplus B = \{a + b | a \in A, b \in B\} = \bigcup_{b \in B} A_b \tag{7}
\]

(2) The second step is network thinning. In the image thinning process, the foreground pixel \( p \) is deleted depending on the connectivity of the 8 pixels, defined as \( a_1, a_2, a_3, \ldots, a_8 \) (Lam et al., 1992). They besiege the pixel in the neighborhood structure as Fig. 1 shows. The connectivity is defined as \( X_H(p) \) (Hilditch, 1969).

Fig. 1. Neighborhood structure of pixel p.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CO₂</th>
<th>NOx (Ntziachristos and Samaras, 2010)</th>
<th>VOC</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g^p )</td>
<td>0</td>
<td>3.986 e-06</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( b^p )</td>
<td>1</td>
<td>0.932</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( M_{\text{MEAN}} \geq 160,000 \text{ km} )</td>
<td>1</td>
<td>1.57</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 4
Fleet composition.

<table>
<thead>
<tr>
<th>Taxi age (year)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{\text{MEAN}} (\times 10^4 \text{ km}) ) (Zhang et al., 2014)</td>
<td>12.6</td>
<td>25.2</td>
<td>37.8</td>
<td>50.4</td>
<td>63.0</td>
</tr>
<tr>
<td>( R_i ) (%)</td>
<td>14.5</td>
<td>35.4</td>
<td>25.2</td>
<td>16.1</td>
<td>8.8</td>
</tr>
<tr>
<td>(Lin et al., 2009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Emission degradation rate due to mileage for Euro 4 petrol cars.
Complete road connection information.

Morphological operation, the network in Fig. 2(b) is generated with the spatial cluster model based on TAZ features. The related definitions in spatial cluster model are listed in Table 6.

The small regions with analogous road features are grouped into big TAZs. And the emission patterns in diverse TAZs are explored. The TAZ set \( U \) is defined with two rules: (1) road density similarity. The road features of small regions in the same TAZ are most similar. This goal can be depicted as \( S_{k_i,k_j} \leq S_{k_i,k_0} \), where \( k_i, k_j \in U \), \( k_0 \in U \), \( k_0 \neq k_i, k_j \), and \( N_{k_i,k_j} = 1 \) and \( N_{k_j,k_0} = 1 \) are satisfied. (2) Area restriction, which can be expressed as \( \text{Area}_{k_i} < \lambda \). The solution procedure of spatial cluster model is shown in Table 7.

The elbow method (Hardy, 1994) is applied to choose the proper number of clusters and evaluate the result by standard deviation, which can be expressed in Eq. (13).

\[
\text{Std} = \frac{1}{M} \sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} \left( D_{k_i} - D_{U_j} \right)^2
\]

\( M \) denotes the number of cluster regions in the graph, and \( N \) denotes the number of small regions for each big region \( U \). Based on the variation trend of \( \text{Std} \), number 33 is determined as the cluster number as Fig. 4 displays. And these 33 TAZs do not destroy characteristics of the region formed by arterial roads such as the 2nd Ring Road, the 3rd Ring Road, the 4th Ring Road and the 5th Ring Road.

3. Result

3.1. Trip patterns of TAZs

The origin and destination points of one trip are identified by the operation state of GPS data. So do the time, longitude, latitude and speed information. Trip patterns of the 33 TAZs are analyzed with trip trajectories of taxis. The trip number of each TAZ in \( j \)th

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e_k )</td>
<td>( { c_k, 1, c_k, 2, c_k, 3, \ldots, c_k, n } ), the edge set of the region ( k )</td>
</tr>
<tr>
<td>( U )</td>
<td>( { U_1, U_2, \ldots, U_n } ), the set of regions</td>
</tr>
<tr>
<td>( \text{Area}_{k_i} )</td>
<td>the area of region ( k )</td>
</tr>
<tr>
<td>( \text{Area}_{U_i} )</td>
<td>the area of region ( U_i )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>a predefined maximum expected area of the traffic analysis zone ( U_i )</td>
</tr>
<tr>
<td>( l_{k_i} )</td>
<td>the length of road ( i ) in region ( k )</td>
</tr>
<tr>
<td>( D_k )</td>
<td>( D_k = \sum_{i=1}^{n} l_{k_i}/\text{Area}_{k_i} ), road density of region ( k )</td>
</tr>
<tr>
<td>( N_{k_i,k_j} )</td>
<td>the adjacent relation of region ( k_i ) and ( k_j ), if ( k_i \neq k_j ) and ( F \neq \Phi ), ( N_{k_i,k_j} = 1 )</td>
</tr>
<tr>
<td>( F )</td>
<td>( F = e_k/r_{k_i} )</td>
</tr>
<tr>
<td>( D_{k_i} )</td>
<td>( D_{k_i} = (D_{k_i} - D_{k_i_{\text{min}}})/(D_{k_{i_{\text{max}}}} - D_{k_{i_{\text{min}}}}) )</td>
</tr>
<tr>
<td>( S_{k_i,k_j} )</td>
<td>( S_{k_i,k_j} =</td>
</tr>
</tbody>
</table>
Gout
Gin
Ga
hour is defined as follows.

\[
G_a(j) = G_a^{in}(j) + G_a^{out}(j)
\]  

(14)

\[
G_a^{in}(j) = \sum_{T} \{ n \epsilon \{1, N\} | D_n \epsilon C_a \}
\]  

(15)

\[
G_a^{out}(j) = \sum_{T} \{ n \epsilon \{1, N\} | O_n \epsilon C_a \}
\]  

(16)

\[G_a^{in}(j)\] means the number of inflow trips in TAZ \(a\), and \[G_a^{out}(j)\] is the number of outflow trips. Variable \(T\) represents the set of trips, that is, all trips of \(N\) taxis in \(j\)th hour. In this paper, the trips represent the driving process with and without customers. \(C_a\) is the coordinate set of TAZ \(a\). Variable \(O_n\) indicates the origin point of the trip, and \(D_n\) indicates the destination point. For the hourly trip patterns, the time interval is defined as one hour. The hourly trip number \(G_a(j)\) represents the trip number of TAZ \(a\) in \(j\)th hour on 1st November, 2012. The relative trip number is defined as \(G_a(j)'\) in Eq. (17). And the hourly trip number of all TAZs is defined as \(G(j)\) in Eq. (18).

\[
G_a(j)' = \frac{G_a(j) - G_{min}}{G_{max} - G_{min}}
\]  

(17)

\[
G(j) = \sum_{a=1}^{33} G_a(j)
\]  

(18)

\(G_{min}\) and \(G_{max}\) are the minimum and maximum of \(G_a(j)\). And the value of \(G_a(j)'\) is shown in Fig. 5. It can be observed from Fig. 5 that the relative trip number of most TAZs changes greatly during the day, especially for the TAZ identified with ID number 10, 11, 12, 14, 16–23. In Beijing map, these TAZs are located within the 4th Ring Road or in the northern area between the 4th Ring Road and the 5th Ring Road, confirming that the lively area is urban center. Evidently, the high relative trip numbers are in the daytime and early night (6:00–24:00).

The numbers of real-time trips \(G(j)\) and running taxis are shown in Fig. 6. The trip number changes greatly during the day. It begins to rise up at 5:00 in the morning. And the first peak appears around 9:00–10:00. Then it drops at noon. And the trip number reaches a low level around 13:00. In the time period of 13:00–16:00, trip number goes up steadily. And the highest value appears at around 15:00–16:00 rather than the peak hours (7:00–9:00 and
17:00–19:00) of urban traffic demand. From 17:00 to 18:00, the trip number keeps getting smaller owing to the heavier and heavier traffic. After the turning point at around 19:00, arrangements of dining out and other leisure activities cause another trip peak. And after 23:00, taxi trips decrease quickly, implying that people tend to go to sleep. As for the running taxi number, it has the same variation trend with the trip number before 16:00. The taxi number at around 16:00 is more than the morning and evening peak hours, which is similar with the trip number. It can be explained that some taxi drivers in Beijing tend to not work in peak hours because of the traffic congestion (http://news.sina.com.cn, 2013). The average trip time in rush hours is longer than other time, leading to higher operating cost. It is investigated that the operating revenue is less than cost in peak hours. Thus some taxi drivers usually choose to change shifts, take a rest and park taxis in morning and evening rush hours. Compared with rush hours, there are more running taxis in the afternoon, causing the highest trip number in the time period of 15:00–16:00.

3.2. Emission patterns

3.2.1. Emission patterns between TAZs

The spatial emission patterns of TAZs are explored. The emission quantities of pollutant \( p \) from TAZ \( a \) to TAZ \( b \) in the time period of \( t_1 - t_2 \) are defined in Eq. (19).

\[
E^{ex}_{a, b}(t_1, t_2) = \sum_{l=0}^{k} E^p_l(t) \Delta t
\]  

(19)

\( E^p_l(t) \) is the emission rate of taxi trips from TAZ \( a \) to TAZ \( b \). Variable \( k \) is the trip number from TAZ \( a \) to \( b \). And \( l \) means the travel time. The accumulated emissions of pollutant CO\(_2\), NO\(_x\), VOC and PM are estimated according to Eq. (19). As Fig. 7 depicts, the CO\(_2\) emissions are accumulated for each 3 hours, leading to 8 time intervals for the 24-h data. The straight lines connect geometrical centers of TAZs. The color and width of lines stand for emission quantities. For instance, \( E^{ex}_{CO_2}_{a, b}(0, 3) \) means the CO\(_2\) emission quantity caused by trips between TAZ \( a \) and TAZ \( b \) in the time period of 0:00–3:00. The relative emission quantity \( \frac{E^{ex}_{a, b}(t_1, t_2)}{E^{ex}_{a, b}(t_1, t_2)_{\text{max}} - E^{ex}_{a, b}(t_1, t_2)_{\text{min}}} \) is defined in Eq. (20). \( E^{ex}_{a, b}(t_1, t_2)_{\text{max}} \) and \( E^{ex}_{a, b}(t_1, t_2)_{\text{min}} \) are the minimum and maximum of \( E^{ex}_{a, b}(t_1, t_2) \).

\[
E^{ex}_{a, b}(t_1, t_2)_{\text{max}} = \frac{E^{ex}_{a, b}(t_1, t_2)_{\text{max}} - E^{ex}_{a, b}(t_1, t_2)_{\text{min}}}{E^{ex}_{a, b}(t_1, t_2)_{\text{max}} - E^{ex}_{a, b}(t_1, t_2)_{\text{min}}}
\]  

(20)

Fig. 7 shows the value of \( E^{ex}_{CO_2}_{a, b}(t_1, t_2) \) between each two TAZs. It can be observed that emissions center on the daily time and the early night. Besides, emissions are chiefly located within the 5th Ring Road, the downtown area. From 0:00 to 3:00, the major emissions are distributed in the eastern area within the 5th Ring Road. And the high-emission area covers Houhai, Nanluoguxiang, Sanlitun and Workers’ Stadium, the lively recreational areas in Beijing. The lowest emissions occur in the time period of 3:00–6:00. With the beginning of human activities, emissions within the 5th Ring Road rise up during 6:00–9:00. Apparently, in the time period of 9:00–12:00, the emissions between TAZs reach a high level. For the business areas such as Wudaokou and Xidan, and transportation junctions like Beijing South Railway Station and Beijing West Railway Station, there tend to be strong emissions.
between them and other TAZs. It reveals that the major daily human activities are related to business areas and traffic hubs. Similar with the former time intervals, spatial relationships in the time period of 12:00—15:00 keep stable. After a drop between 15:00 and 18:00, emission quantities in the period of 18:00—24:00 remain unchanged. By comparing emissions in different time intervals, it can be identified that there exists temporal and spatial trip disequilibrium between TAZs. And the hot spots practically are located within the 5th Ring Road. The instantaneous distribution of emissions indicates that activities in northern area are livelier than southern area.

In order to explore the emission of various pollutants, emission quantities during the whole day are accumulated, i.e., \( E_{a,b}^{\text{CO}_2}(0,24) \), \( E_{a,b}^{\text{NO}_x}(0,24) \), \( E_{a,b}^{\text{VOC}}(0,24) \), \( E_{a,b}^{\text{PM}}(0,24) \). Then the value of \( E_{a,b}^{\text{CO}_2}(0,24) \), \( E_{a,b}^{\text{NO}_x}(0,24) \), \( E_{a,b}^{\text{VOC}}(0,24) \) and \( E_{a,b}^{\text{PM}}(0,24) \) can be obtained according to Eq. (20). As Fig. 8 shows, the emission distributions of the 4 pollutants are similar. The emission quantities are pretty strong between TAZs in urban center, with a downtrend from urban center to suburbs. In order to acquire the main spatial emission patterns, the first 12 strongest emission quantities between 2 TAZs of the 4 pollutants are extracted as a new network as Fig. 9 presents. It is highlighted that the main emission network covers all TAZs within the 4th Ring Road, with 10 points and 12 edges. On the one hand, in the perspective of TAZ type, core areas embody large business centers and important transportation hubs. On the other hand, the emission quantities in eastern TAZs are stronger than other areas, especially the northern area. And the emission quantities between the southern TAZs are stronger than the northern TAZs. During the whole day, the pollutant emissions show a polycentric distribution.

### 3.2.2. Emission patterns within TAZs

The accumulated emission quantity within TAZ \( a \) in the time period of \( t_1 - t_2 \) is defined as \( E_{a}^{\text{in}}(t_1, t_2) \). As Eq. (21) shows, \( E_{a}^{\text{in}}(t) \) is the emission rate of taxi trips within TAZ \( a \). Variable \( k \) is the trip number within TAZ \( a \). And \( l \) means the travel time. The relative emission quantities are defined as \( E_{a}^{\text{in}}(t_1, t_2) \).

\[
E_{a}^{\text{in}}(t_1, t_2) = \sum_{t=0}^{k} \sum_{n=1}^{k} E_{a}^{\text{in}}(t) \Delta t
\]

(21)

\[
E_{a}^{\text{in}}(t_1, t_2) = \left[ E_{a}^{\text{in}}(t_1, t_2) - E_{a}^{\text{in}}(t_1) \right] / \left( E_{a}^{\text{in}}(t_1, t_2) \right)
\]

(22)

\( E_{a}^{\text{in}}(t_1, t_2) \) and \( E_{a}^{\text{in}}(t_1, t_2) \) are maximum and minimum of \( E_{a}^{\text{in}}(t_1, t_2) \). The value of \( E_{a}^{\text{in}}(t_1, t_2) \) is shown in Fig. 10. The CO2 emission variations within TAZs and between TAZs are similar. It hints that for a certain TAZ, the CO2 emissions within this TAZ have a positive relationship with the emissions between this TAZ and other TAZs. It is clear that a small amount of CO2 emission is caused in the northern area between the 4th Ring Road and the 5th Ring Road. For instance, the value of \( E_{a}^{\text{CO}_2}(t_1, t_2) \) in the northern area within the 5th Ring Road can be up to 0.75—1, which is 10 times of the southern area within the 5th Ring Road from 9:00 to 12:00. And a slightly downtrend occurs at around 21:00. It can be highlighted that CO2 emission quantity generated by the short-distance taxi trips in central area is distinctly higher than the fringe area.

### 3.3. Correlations between emission quantities and TAZ features

Region segmentation based on spatial cluster model guarantees
the maximum similarity of clustered areas. Road characteristics have an influence on TAZ trips (Crane and Crepeau, 1998), further affecting emissions between TAZs. TAZs with high road densities may attract more trips, resulting in an influence on the speed and emissions. The road density of TAZ $a$ is defined as variable $D_a$. The relative road density $D_a'$ is defined as follows.

$$D_a' = (D_a - D_{min})/(D_{max} - D_{min})$$

(23)

$D_{max}$ and $D_{min}$ are the maximum and minimum of $D_a$. Fig. 11 shows the value of $D_a$. It can be observed that the road densities within the 5th Ring Road are higher than the fringe area, similar with traffic emissions.

3.3.1. Correlations between emission quantities between TAZs and TAZ features

Accumulated emissions between TAZ $a$ and other TAZs during the whole day are defined as $E_{a,b}^{ex}$ $(0, 24)$. Variable $U$ represents the set of TAZs. And the value of $E_{a,b}^{ex} (0, 24)$ is normalized to $E_{a,b}^{ex} (0, 24)'$ according to Eq. (20). The relationships between relative emissions $E_{a,b}^{ex} (0, 24)'$ of pollutant CO$_2$, NO$_x$, VOC and PM and relative road densities $D_a'$ are indicated in Fig. 12. It implies that the relationships between road densities and the emissions of 4 pollutants are similar. For TAZs with road densities higher than 0.6, the emissions rise up with the increasing road densities. The coordinates are divided into 4 parts by the median values of road density and emission. And the 4 parts correspond to high road density-high emission, high road density-low emission, low road density-low emission and low road density-high emission. Most TAZs are located in high road density-high emission and low road density-low emission area. Thus it can be denoted that there is a positive relationship between emission quantities and road densities.

3.3.2. Correlations between emission quantities within TAZs and TAZ features

The accumulated emissions within TAZ $a$ during the whole day can be presented as $E_{a}^{in} (0, 24)$. And the relative emissions $E_{a}^{in} (0, 24)'$ can be obtained according to Eq. (20). The relationships of $E_{a}^{in} (0, 24)'$ and $D_a'$ are exhibited in Fig. 13. When road densities of TAZs are in the range of 0–0.6, the emission quantities within TAZs will stay a low level. And for TAZs whose densities are higher than 0.6, the positive relationships between emission quantities and road densities are obvious. It can be observed from Figs. 12 and 13 that TAZs with road densities higher than 0.6 are located within the 4th Ring Road and middle-west and northeast area between the 4th Ring Road and the 4th Ring Road. And high road densities correspond to more business activities, recreation activities and trip demand.

4. Discussion

The emission patterns of TAZs are studied in this paper, which are helpful for urban planning and management.

(1) The temporal and spatial emission variations of TAZs are identified with GPS data. And the significant disparity in
emissions across TAZs is verified. The government is suggested to reduce the emission disparity and improve the supply of infrastructures with different functions in TAZs. That is, for different regions, the functions of business and entertainment should be satisfied as much as possible so that citizens don’t need to take long-distance trips with various purposes. What’s more, the urban planners should also consider the emission bearing capacity of regions.

2. As the basic unit of traffic planning, the features of TAZs such as land use, population composition and infrastructures should be investigated during the first stage of planning. However, the current traffic emissions of TAZs are not considered in the previous study. As the major environmental problem, the air pollution hinders the economic development in China. Traffic planners can get policy support from this study.

3. The public transport priority policy has been applied in many cities in China, which is expected to relieve air pollution. Our results lay a foundation for better public transit assignment. With the dynamic emission inventories in this paper, the government officials are advised to increase bus lines between high-emission TAZs in the high-emission time period. Then the bus accessibility can be improved, resulting in more convenient trips and less traffic emissions. Furthermore, the taxi services also can be improved by designing more taxi reception stations in high-emission TAZs.
5. Conclusions

The 33 TAZs of the research area are defined as the analysis units. The urban trip and emission patterns are explored based on real-time taxi trajectories of 12409 taxis in Beijing. A new traffic emission patterns analysis model is proposed. Analysis is performed on emissions of CO₂, NOₓ, VOC and PM, temporal and spatial variations and the relationship with the TAZ features. The conclusions are as follows.

(1) There exist 3 peak time periods for TAZ trips, that is 10:00, 16:00 and 20:00. And trips in TAZs within the 4th Ring Road or in northern area between the 4th Ring Road and the 5th Ring Road are higher than others and fluctuate greatly.

(2) Generally speaking, the variations of the four pollutants are similar between TAZs. Spatially, emission quantities are related to the number and type of infrastructures and locations of TAZs. The emissions of TAZs with business centers, entertainment centers and transportation hubs are obviously higher than others. Emissions of TAZs within the 5th Ring Road are higher than the fringe TAZs in Beijing. And for TAZs within the 5th Ring Road, the northern emissions are stronger than the southern. Temporally, emissions can be divided into 3 types, corresponding to time periods of 0:00–3:00, 3:00–6:00 and 6:00–24:00. In the time period of 0:00–3:00, the emissions between TAZs with entertainment functions are highest. In the time period of 3:00–6:00, emissions keep a low level. In the time period of 6:00–24:00, emissions of different TAZs varies obviously in time and space. For emissions within TAZs, the variations of emissions are analogous with emissions between TAZs.

(3) There is a positive relationship between road densities and emissions within and between TAZs. Furthermore, the patterns between road densities and emission quantities resemble across various pollutants. When the road densities of TAZs are bigger than 0.6, emissions within TAZs will rise up obviously with the increasing road densities. It can be concluded that TAZ road densities have an influence on traffic emissions within and between TAZs.

(4) The policy implications of results are discussed, including the region function planning, traffic planning, the public transport improvement and optimal design of charging stations and new energy vehicles promotion.

(5) The result of this paper is meaningful for urban taxi emission analysis. It should be noted that this study can be extended. For emission analysis in other traffic scenes, the emission estimation model should be localized. In the further study, the experiment on emission estimations and comparison will be conducted.

Acknowledgements

This research is supported in part by National Natural Science Foundation of China (U1509205) and Tsinghua University Initiative Scientific Research Program (2015THZ01). The authors are grateful to anonymous reviewers for their constructive comments.

References


Carslaw, D.C., Bever, S.D., Tate, J.E., Westmoreland, E.J., Williams, M.L., 2011. Recent evidence concerning higher NOx emissions from passenger cars and light duty vehicles. Atmos. Environ. 45 (39), 7053–7063.


Fig. 13. Correlations between the relative emission quantities within TAZs and road densities (the relationship between $E^{th}_{i,j}$ and $D_{i,j}$).
Cities of China Environment Science and Engineering, Tsinghua University.