Mapping spatio-temporal patterns and detecting the factors of traffic congestion with multi-source data fusion and mining techniques

Jinchao Song\textsuperscript{a,⁎}, Chunli Zhao\textsuperscript{b,c,⁎⁎}, Shaopeng Zhong\textsuperscript{d}, Thomas Alexander Sick Nielsen\textsuperscript{e}, Alexander V. Prishchepov\textsuperscript{a}

\textsuperscript{a} Department of Geosciences and Natural Resource Management, University of Copenhagen, 1350 Copenhagen, Denmark
\textsuperscript{b} Faculty of Engineering, Department of Technology and Society, Transport & Roads, Lund University, 221 00 Lund, Sweden
\textsuperscript{c} K2, the Swedish Knowledge Centre for Public Transport, Lund, Sweden
\textsuperscript{d} School of Transportation and Logistics, Dalian University of Technology, 116024 Dalian, China
\textsuperscript{e} Danish Road Directorate, Section for Planning and Analysis, Niels Juel's Gade 13, 1022 København K, Denmark

ARTICLE INFO

Keywords:
Traffic congestion
Land use
Spatiotemporal pattern
Multi-source data

ABSTRACT

The study focuses on mapping spatiotemporal patterns and detecting the potential drivers of traffic congestion with multi-source data. First, based on real-time traffic data retrieved from an online map, the k-means clustering algorithm was applied to classify the spatiotemporal distribution of congested roads. Then, we applied a geographical detector (Geo-detector) to mine the potential factors for each spatiotemporal pattern. The results showed six congestion patterns for intra-regional roads and inter-regional roads on weekdays. On both intra-regional and inter-regional roads, congestion density reflected by building height was the strongest indicator during the morning peak period. Public facilities such as hospitals, tourist sites and green spaces located near areas of employment or residential areas contributed to congestion during and off-peak hours. On intra-regional roads, the sparse road network and greater distance from the city center contribute to congestion during peak hours. On inter-regional roads, the number of bus stops contributed most to the early evening peak congestion, while the design of the entrances to large buildings in mixed business areas and public service areas increased the level of congestion. The results suggest that land use should be more mixed in high-density areas as this would reduce the number of trips made to the city center. However, mixed land-use planning should also be combined with a detailed design of the microenvironment to improve accessibility for different travel modes in order to increase the efficiency of traffic and reduce congestion. The innovative approach can be potentially applied in traffic congestion and land use planning studies elsewhere based on real-time multi-source data.

1. Introduction

Urbanization in developing countries has been rapid during the past 20 years (Cohen, 2006). An increasing number of vehicles and population density have created excessive demand for urban infrastructure. These trends pose a challenge to urban planners and policymakers as they need to plan land use in order to decrease traffic congestion with limited space in inner cities (Alberti, 2008; Lauf et al., 2016). Understanding the complex interplay between spatiotemporal patterns of traffic congestion and their drivers has been a particular challenge in urban management. Road networks provide accessibility for citizens, but congestion leads to increased gas consumption, traffic-related pollution (Bruneckreef & Holgate, 2002; Clougherty, Wright, Baxter, & Levy, 2008; Dons, Poppel, Kochan, Wets, & Int, 2013; Wang, Henderson, Shibi, Allen, & Brauer, 2013) and time delay (Lv, Duan, Kang, Li, & Wang, 2015). Due to congestion, people in Beijing spend twice as much time travelling as they would otherwise (Raincent, 2016). Despite the abundance of studies on route optimization and planning, there is still insufficient knowledge of the factors governing congestion and its spatio-temporal pattern. This highlights a need to study the potential drivers of traffic congestion and systematical solutions how to reduce congestions.

In their search for ways to alleviate traffic congestion, previous studies have focused on road design and construction (Cervero, 2010; Fulton, Noland, Thomas, & Carolina, 2019), congestion prediction (Kong, Xu, Shen, Wang, & Yang, 2016), traffic light realignment (Wen, 2017), etc.

⁎ Corresponding author.
⁎⁎ Corresponding author at: K2, the Swedish Knowledge Centre for Public Transport, Lund, Sweden.

E-mail address: songjinchao08@163.com (J. Song).

https://doi.org/10.1016/j.compenvurbsys.2019.101364
Received 24 February 2019; Received in revised form 13 June 2019; Accepted 3 July 2019
0198-9715/ © 2019 Elsevier Ltd. All rights reserved.
and fourth ring roads, 3.61 million between the fourth and fifth ring roads (China Statistics, 2017). Specifically, transportation by bus, private car and taxi accounted for 25%, 31.9% and 3.6%, respectively (Beijing Municipal Commission of Transport, 2016). The main flows of commuter traffic were from the residential areas (Tiantongyuan, Wangjing and Tongzhou) to the central business district (CBD) and from the residential areas (Huilongguan, Mentougou and Fangshan) to an area with high employment density (Zhongguancun) (Long et al., 2015) (Fig. 2).

For historical reasons, all former imperial palaces lie on the axis of the old city of Beijing. Groups of buildings were laid out on both sides to achieve perfect symmetry. The subsequent urban planning of Beijing inherited these symmetrical characteristics. The city has expanded layer by layer over the past few decades. Hence, the study area was divided into four quadrants (Fig. 1c) using a vertical axis along the central axis and a horizontal axis along Chang'an Road. The national standard of functional zones in urban planning (People's Republic of China, Ministry of Construction, 1998) classifies land parcels by their (designated) use. The classification was included in the analysis as the following seven categories (Fig. 1d): residential, commercial, public services, work (employment), transportation, non-built-up and waterbodies as an indicator of land use composition that may influence congestion patterns.

2.2. Data collection and pre-processing

In research on urban form and travel behavior, urban form is considered to be a contributory cause with a probabilistic influence on travel behavior (Naess, 2006; Naess, 2013). Ewing and Cervero (Ewing & Cervero, 2010) suggested the five Ds (density, diversity, design, destination accessibility, and distance to the transit) model to define the key characteristics of the urban form including functional zones/land use. The four Ds have been studied to determine their on travel behavior. The model has been broadly applied in recent studies on urban form and behavior (Zhang, Li, Ding, Zhao, & Huang, 2016; Zhao, Nielsen, Olafsson, Carstensen, & Fertner, 2018; Zhao, Nielsen, Olafsson, Carstensen, & Meng, 2018). In line with the 'Ds' model suggested by Ewing and Cervero (2010), congestion is considered to be a consequence of travel behavior that is highly dependent on cars as well as the organization of land use in a city. Thus, this paper takes Ewing and Cerveros 'Ds' as a starting point for exploring how urban form factors are related to congestion. The factors considered in the study are: Density, diversity, design and distance to the transit. The factors represent the four Ds were controlled as independent variables and listed in Table 1.

2.3. Dependent variable

The dependent variables were defined as congested road lengths based on real-time traffic data. Real-time traffic data were acquired from AutoNavi™, a map service provider in China, which is widely used for navigation. Users of service upload real-time location-based information to the computing platform. Data retrieved by the spliced HTTP included traffic status (free-flowing, partially congested and congested), direction and speed of vehicles by road segment and time-of-day. For expressways and major or minor arterials, the online traffic map shows red when the average speed is below 20 km/h or 15 km/h, respectively, in a given direction. We parsed weekly traffic data every five minutes from November 6 (Monday) to November 12 (Sunday), 2017. For each layer, we converted the lines to points with a step size of 10 m. Therefore, there were 288 layers in the daily congestion dataset and 123,194 points for each layer.

3. Method

The study applied a hierarchical clustering approach, k-means...
Fig. 1. a) Road network classification. Blue line represents inter-regional road and black line represents intra-regional road; the red line represents congested road; c) very high resolution satellite imagery; d) functional zone mapped by satellite remote sensing imagery and POIs; e) four regions divided by Chang'an Street and the central axis: quadrant 1 (Q1), quadrant 2 (Q2), quadrant 3 (Q3), quadrant 4 (Q4); The coding guideline for vector polygonal zones can be read as follows: the first number represents the quadrant, the second number represents the ring road, and the third number represents the location in a clockwise sequence; f) building data including height; g) population density mapped by satellite remote sensing imagery; h) locations of bus stops. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
clustering, to classify the spatiotemporal congestion patterns over a day and a week. Using the classification of congested roads, we analyzed the change in congested road length over 24 h and compared the congestion patterns between weekdays and weekends. Geo-detector was conducted to explore the relationships between spatiotemporal congestion patterns and correlated factors (Fig. 3).

3.1. Classification with k-means clustering

To understand the clustering characteristics in congestion over time (for example, morning and evening peaks) and space, we clustered congestion patterns by using time as the independent variable. Given points X with cases labelled by point ID, and variables labelled by traffic status, the grouping procedure used partitioning to classify points through k-means clustering. This process is often formulated as follows by a mathematical machine learning program (Arai & Barakbah, 2007; Huang, 1998; Jianliang, Haikun, & Ling, 2009):

\[ J = \sum_{i=1}^{k} \sum_{j=1}^{n} d_i(x_j, c_i) \]

where \( x_i \) represents a y-dimensional dataset with 288 columns (00:00, 00:05, 00:10...) and 123,194 rows (point 1, point 2...), while \( d_i(x_j, c_i) \)

Table 1
The definition of the four “Ds” built-environment variables and the description of the relevant independent variables.a

<table>
<thead>
<tr>
<th>Four Ds</th>
<th>Attributes</th>
<th>Representative independent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity</td>
<td>The proportion of functional zones. Functional zones were mapped with very high-resolution imagery and points of interests (Song, Lin, Li, &amp; Prishchepov, 2018), including residential areas, commercial zone, public service areas, work areas, and transportation areas, water bodies and non-built-up area.</td>
<td>Proportion of functional zones was calculated: Area (functional zone)/Area (statistic unit)</td>
</tr>
<tr>
<td>Density</td>
<td>Population density, building density, building height</td>
<td>The population density (30 m) was retrieved by location-based service data and VIIRS night-time lights (Song et al., 2019). Population density = (PD1 + PD2 + PD3 + ...PDn)/n, PDn represents the population density in a pixel with spatial resolution of 30 m. Building density = Area (building area)/Area (statistic unit). Building height in each statistical unit = BH1 + BH2 + ... + BHn, BHn represents building height of each building polygon. Building data with area and building height were parsed from the AutoNavi™ online map.</td>
</tr>
<tr>
<td>Design</td>
<td>Characteristics of street network, building area and building height. Number of entrances and exits for highway and ring road (for inter-regional roads); Distance to city centre</td>
<td>Road length = road length / Area (statistic unit). Points of entrance and exit for highways and ring roads (numbers and density) were parsed from the AutoNavi™ online map. Number of entrances and exits for highway and ring road = numbers (entrance and exits) / Area (statistic unit). Distance to city centre is measured by the distance between the congested road and Tian’anmen.</td>
</tr>
<tr>
<td>Distance to transit</td>
<td>Number of bus stops and metro stations per unit area</td>
<td>Number of bus stops and metro stations = numbers (bus stops and metro stations) / Area (statistic unit)</td>
</tr>
</tbody>
</table>

* Statistic unit, zone split by the axes and ring roads for intra-regional congestion analysis and buffer (700 m) of road section for inter-regional congestion analysis.
**Table 2**

Types of interaction between covariates.

<table>
<thead>
<tr>
<th>Type of interaction</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear-weakening</td>
<td>( q(X_1 \cap X_2) &lt; \min(q(X_1), q(X_2)) )</td>
</tr>
<tr>
<td>Independent</td>
<td>( q(X_1 \cap X_2) = q(X_1) + q(X_2) )</td>
</tr>
<tr>
<td>Nonlinear-enhancing</td>
<td>( q(X_1 \cap X_2) &gt; \min(q(X_1), q(X_2)) )</td>
</tr>
</tbody>
</table>

represents the distance between point \( x_i \) and the cluster centroid \( c_i \).

1. Define \( k \) centroids \( C = c_1, c_2, \ldots, c_k \) using \( X \) as the initial cluster center of the clustering space and set \( k = 6 \) based on expert knowledge and our repeated experiments;
2. Assign each instance to its closest center: if \( d_j(x_i, c_i) < d_m(x_i, c_m) \), then assign \( x_i \) to cluster \( c_i \);
3. Recalculate the center of the cluster \( c_i \^* \);
4. If \( c_i \^* = c_i \), we end the program; then \( c_1^{*}, c_2^{*}, \ldots, c_k^{*} \) represents the final cluster. In this study, we set the number of iterations to 100.

**3.2. Detection of driving factors and interactions**

Geo-detector can quantitatively measure the degree of stratified heterogeneity (Ju et al., 2016; Wang et al., 2016). The analyses can be carried out with a single-factor detector and an interaction detector. In order to explore relationship between the potential factors and congestion, it is necessary to examine the interactions between the factors. The urban form factors are highly correlated, and their effects are difficult to separate. Furthermore, the traffic conditions are likely to be under the influence of several factors at a time. Therefore, both single-factor detector and interaction detector are applied in this study. The length of congested roads and roads with slow traffic (Fig. 4) were synchronous on both weekdays and weekends. The total length of congested roads was longer during the morning peak than during the morning peak on both weekdays and weekends. The temporal patterns in congested lengths on weekdays were different from those on weekends. On weekdays, the congested length during the morning peak started to increase at 6 a.m., with the maximum value (10 km) appearing at 8 a.m. On Saturday, the congested length during the morning peak started to increase at 8 a.m., with the maximum (5 km) appearing at 11 a.m. On Sunday, the congested length (2 km) during the morning peak was shorter than those on weekdays and Saturday. The maximum congested lengths in the evening peak appeared at approximately 6 p.m. on both weekdays and weekends. Here we focused on analyzing spatiotemporal patterns of congestion on weekdays.

**4. Results**

**4.1. Changes in congested length over time**

The k-means clustering results (Figs. 5 and 6) showed six clusters for intra-regional roads and inter-regional roads on weekdays, including morning peak hours (A) at 06:30–09:00 (A1) and 06:30–11:00 (A2),...
morning and evening peak hours (B) at 06:30–10:00 and 17:00–19:30 (B1), daytime hours at 06:30–11:00 and 12:30–15:00 (B2) and evening peak hours (C) at 17:00–19:00 (C1) and 17:00–21:00 (C2).

4.2.1. Intra-regional road analysis

The result revealed that the length of congested roads in the most congested zones were 6–9 km, 2 km and 3–6 km during the A, B and C periods, respectively (Fig. 5). During the morning peak hours (Fig. 5. A1, A2), the congested zone was mainly located outside the western (A1) and northwestern fourth ring road (A2), with congested road lengths greater than 6 km. The congested road lengths ranged from 2 km to 3 km during the morning and evening peaks (Fig. 5. B1) and in the daytime (B2). During the evening peak hour (Fig. 5. C1, C2), the congested zone was mainly located in the eastern part of Beijing, with congested road lengths more than 5 km.

The q-values represent the most influential factors based on k-means clustering (Table 3). For A1 (06:30–09:00) and B1 (06:30–10:00, 17:00–19:30), building height was the most influential single factor. The interactions enhanced the level of (q = 1) congestion when ‘building height’ coexisted with ‘work’, ‘population density’ and ‘water body’ during A1 (06:30–09:00), as well as ‘residence’ and ‘public service’ during B1 (06:30–10:00 17:00–19:30). During the B2 period (06:30–11:00 and 12:30–15:00), the interactions between ‘work area’ and ‘building density’, ‘public service’ and ‘road density’ were enhanced nonlinearly (q = 1). During the A2 (06:30–11:00), the interaction between ‘distance to city center’ and ‘green space’ area, ‘road density’ and ‘building height’ nonlinearly enhanced the effect on congestion. During the C1 period (17:00–19:00), ‘road density’ contributed most to

---

Fig. 4. Temporal patterns of congested length (the vertical axis represents the length of the congested road, and the horizontal axis represents the time of data acquisition, which occurred every five minutes).
**Fig. 5.** Congested clusters with an independent variable of data acquisition time on secondary roads and branch roads on weekdays (A represents the morning peak and includes A1 and A2, B represents congested roads in the daytime and includes B1 and B2, and C represents the evening peak and includes C1 and C2). Red zones represent the maximum congested length during that period. The arrow represents the direction of congested road. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**
The most influential factors and interactions for each spatiotemporal congestion pattern on intra-regional roads.

<table>
<thead>
<tr>
<th>Time</th>
<th>Ds</th>
<th>Single factor</th>
<th>q</th>
<th>Positive (+)/negative(−)</th>
<th>Interacting S</th>
<th>q_inter</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:30–09:00</td>
<td>Density</td>
<td>Building height</td>
<td>0.28</td>
<td>+</td>
<td>Water body</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population density</td>
<td></td>
<td></td>
<td>Population density</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Work/employment area</td>
<td></td>
<td></td>
<td>Work/employment area</td>
<td>1</td>
</tr>
<tr>
<td>06:30–11:00</td>
<td>Design</td>
<td>Distance to city center</td>
<td>0.31</td>
<td>+</td>
<td>Green space</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Road density</td>
<td></td>
<td></td>
<td>Road density</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Building height</td>
<td></td>
<td></td>
<td>Residence</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Residence</td>
<td></td>
<td></td>
<td>Public service</td>
<td>1</td>
</tr>
<tr>
<td>06:30–10:00 17:00–19:30</td>
<td>Density</td>
<td>Building height</td>
<td>0.38</td>
<td>+</td>
<td>Residence</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public service</td>
<td></td>
<td></td>
<td>Distance to center</td>
<td>1</td>
</tr>
<tr>
<td>06:30–11:00 12:30–15:00</td>
<td>Design</td>
<td>Work/employment area</td>
<td>0.27</td>
<td>+</td>
<td>Road density</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Building density</td>
<td></td>
<td></td>
<td>Public service</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Residential area</td>
<td></td>
<td></td>
<td>Commercial zone</td>
<td>0.78</td>
</tr>
<tr>
<td>17:00–19:00</td>
<td>Design</td>
<td>Road density</td>
<td>0.18</td>
<td>−</td>
<td>Residential area</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commercial zone</td>
<td></td>
<td></td>
<td>Metro station</td>
<td>0.69</td>
</tr>
<tr>
<td>17:00–21:00</td>
<td>Diversity</td>
<td>Commercial zone</td>
<td>0.19</td>
<td>+</td>
<td>Work/employment area</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Water body</td>
<td></td>
<td></td>
<td>Water body</td>
<td>0.73</td>
</tr>
</tbody>
</table>
congestion, and its interaction with ‘residence area’, ‘commercial zone’ and number of ‘metro stations’ enhanced their effects on the length of congested roads. During the C2 period (17:00–21:00), the interaction between ‘commercial zone’ and ‘work’ area and ‘water body’ area significantly enhanced their effects on traffic congestion. Group statistics for Ds influence indicated that the ‘Design’ affected congestion during three periods (A2, B2 and C1), while ‘Density’ affected congestion during two periods (A1, B1) (Table 3).

4.2.2. Inter-regional road analysis

The length and direction of inter-regional road congestion exhibited spatiotemporal heterogeneity (Fig. 6). The most congested quadrants were Q4 (20 km of road congestion) and Q3 (23 km of road congestion) in A1 and A2, respectively. Given temporal heterogeneity, the congested road lengths of more than 45 km (A1), 48 km (A2) and 44 km (C1) during the evening peak were longer than in other time segments. During the morning peak hours (A1, A2), the mean direction of congested roads was towards the city center in all quadrants. However, the difference between A1 and A2 was that the mean direction of Q1 in A1 was southwest, while the mean direction was southeast in A2. Also, the direction of congested roads of Q2 in A1 was west, while it was north in A2. The difference between B1 and B2 was that the congested roads in B1 were clustered within the third ring road in Q1 and Q4, while they were dispersed in B2. During the evening peak, congested roads were mainly located in Q1 and Q4 with an approximate length of 20 km in C1, while the mean direction in all quadrants was approximately south. In C2, congested roads were mainly located in Q1 with a length of

![Fig. 6. Spatio-temporal distribution of congested roads on expressways and arterial roads on weekdays. Blue arrows represent the linear directional mean in each quadrant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image)

![Fig. 7. Congested length statistics in each quadrant on inter-regional roads during the different period on weekdays (the quadrant partition in Fig. 1c is Q1, Q2, Q3, and Q4).](image)
10 km and a mean direction of northeast (Fig. 7).

During the A1 period (06:30–09:00), 'building height' was the most influential factor, and its interaction with 'work' area, 'public service' area and 'population density' nonlinearly enhanced congestion. During the A2 period (06:30–11:00), the number of 'entrances and exits' on highways and ring roads was the most influential factor, while 'population density' was the main factor during the B1 period (06:30–10:00 and 17:00–19:30) and B2 period (06:30–11:00 and 12:30–15:00). The number of 'bus stops' contributed most to the early evening peak (C1, 17:00–19:00), and its interaction with 'business area', 'public service' area and 'building density' reached 0.8, which indicated non-linear enhancement for traffic congestion. 'Public service' area, as an auxiliary factor, contributed to congestion during all periods. Group statistics for the influence of the Ds indicated that 'Density' played an important role in congestion during the A1 period (06:30–09:00), B1 period (06:30–10:00, 17:00–19:30), and B2 period (06:30–11:00 and

<table>
<thead>
<tr>
<th>Time</th>
<th>Ds</th>
<th>Single factor</th>
<th>(q)</th>
<th>Positive(+)/negative(−)</th>
<th>Interacting S</th>
<th>(q_{\text{inter}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:30–09:00</td>
<td>Density</td>
<td>Building height</td>
<td>0.34</td>
<td>+</td>
<td>Work/ employment area</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Public service</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Building density</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Population density</td>
<td>0.68</td>
</tr>
<tr>
<td>06:30–11:00</td>
<td>Design</td>
<td>Entrance and exist on highway</td>
<td>0.3</td>
<td>+</td>
<td>Metro station</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Public service</td>
<td>0.53</td>
</tr>
<tr>
<td>06:30–10:00</td>
<td>Density</td>
<td>Population density</td>
<td>0.42</td>
<td>+</td>
<td>Metro station</td>
<td>0.77</td>
</tr>
<tr>
<td>06:30–11:00</td>
<td>Density</td>
<td>Population density</td>
<td>0.37</td>
<td>+</td>
<td>Metro station</td>
<td>0.68</td>
</tr>
<tr>
<td>17:00–19:00</td>
<td>Distance to transit</td>
<td>Bus stops</td>
<td>0.51</td>
<td>+</td>
<td>Metro station</td>
<td>0.74</td>
</tr>
<tr>
<td>17:00–21:00</td>
<td>Distance to transit</td>
<td>Building height</td>
<td>0.25</td>
<td>+</td>
<td>Metro station</td>
<td>0.87</td>
</tr>
</tbody>
</table>

![Fig. 8. Spatial distribution of the functional zone, population density, metro station and bus stop and congestion status: a) Traffic congestion around the park during morning peak; b) Traffic congestion around Xiehe Hospital in the daytime; c) Traffic congestion around the shopping mall during the evening rush hour.](image-url)
12:30–15:00), while ‘distance to transit’ played an important role during the C1 period (17:00–19:00) and C2 period (17:00–21:00) (Table 4).

5. Discussion

5.1. Intra-regional road analysis

‘Distance to the city center’ was the strongest factor for the congestion during the A2 period (06:30–11:00) (Table 3), the further away from city center, the more congestion likely to occur. Since the middle of the 1990s, Beijing has been undergoing transformation from a monocentric to a polycentric urban structure. High employment density sub-centers have been established (Huang et al., 2017). The area outside the fourth ring road was intensively developed in this transformation process, especially in the northern part (Wang et al., 2009; Zhao, Nielsen, Olafsson, Carstensen, & Meng, 2018; Zhao, 2011). The congestion during the A2 period occurred at one of the high density employment sub-centers, where many high-tech jobs are located, i.e., in IT and at universities. However, the sub-center was developed often without affiliating sufficient public services, which may result in employees having to travel to the public facilities in the city center, thereby contributing to congestion (Huang et al., 2017). The result revealed that a long distance from the city center, the high density of jobs and population increased the congestion in the area (e.g., Fig. 5. A1, 461 and B1, 353), where has a long distance to the city center. Lower total travel distance might reduce traffic congestion under reasonable urban function distribution in a compacted zone. This result supports previous findings that building characteristics and distance to the city center have an impact on the mode choices, car dependence, and hence, the traffic status (Downs, 2005).

Regarding ‘diversity’, it is reflected by the land uses (functional zone) that displayed impact on the congestion. First, water bodies and urban green space (Fig. 8a) including the restored green belt, parks usually interrupt the connectivity of the roads, which reduces the density of the road network, resulting in traffic congestion. The result supported the previous findings that a sparse road network reduces accessibility and connectivity, as it creates long distance between the entrance roads, which encourages the use of cars rather than alternative modes (ITF, 2018). Consequently, vehicles have to go through the same road to pass the area, resulting in traffic congestion. In the newly developed Chinese urban areas, the length of blocks ranges between 400 and 800 m, which is much longer than the average block length of 50 m in Tokyo and 120 m in Paris, London and Manhattan (ITF, 2018). The large size of the blocks in many Chinese cities has serious consequences in terms of accessibility (World Bank, 2014), which calls for a radical redesign of such areas to improve accessibility. Plans to open up closed areas in order to improve road connectivity should be considered in future planning policy. Dense road networks not only enhance the capacity of the roads, but may also encourage passengers to use alternatives to cars. Consequently, it reduces high-risk congestion spots and eases pressure on traffic (Yao et al., 2018).

The impact of ‘public service’ (Diversity) exemplified in B1 (06:30–10:00 and 17:00–19:30) and B2 (06:30–11:00 and 12:30–15:00). The distribution of functional zones contributed to congestion. For instance, during the B1 period (Fig. 5, 451), the occurrence of congestion is attributed to the presence of a primary school, which implies that the biggest source of vehicles is the parents dropping off and picking up their children. This intensified traffic congestion on the roads during the rush hour. This supported the previous finding that driving-to-school trips account for about 15% of all trips in the morning rush-hour (Lu, Sun, & Zheng, 2017) with the average speed on school days being 29.6% slower than during the school summer holidays in Beijing (Yang, Liu, & He, 2016). Another reason is that the population may exceed the designed capacity because of an increase in the school-aged population and car ownership over the past decade (Beijing Government, 2010). During the B2 period (06:30–11:00 and 12:30–15:00), the congested roads were located in a mixed land-use area with business, public service facilities, including tourist areas and hospitals. The tourist area and hospital receive many visitors on a daily basis. For instance, the Xiehe hospital attract twelve thousand visitors (Fig. 8b) per day (Xiehe, 2017), who arrive by private cars or taxi, which creates heavy congestion in the area. More importantly, the traffic congestion was worsened and continued during the off-peak hours while it located next to the tourism resources and the commercial area with high job density. The result indicates that large public facilities such as the Xiehe hospital, which provide services to the population not only in Beijing, but also patients from other places in China, can cause severe congestion when they are located in dense residential, tourism or employment centers. Such large public facilities receive many visitors every day, which means that dedicated traffic management strategies that take the timing of demand in the area into consideration as well as detailed street design should be developed to mediate demand and capacity. Cities with many tourist sites may need to classify the roads regarding usage by tourists in order to determine the levels of traffic intensity, especially during peak periods (Saenz-de-miera & Rossello, 2012). Furthermore, the shopping center (Fig. 5, C1 and C2) located near the intersections increased traffic congestion. The shopping center attracted large flows of traffic and reduced vehicle speed. This may be due to the entrance design of the shopping mall. The negative impact may be addressed by adjusting the location of the entrances in order to relocate the traffic flows onto the branch roads instead of only being concentrated on the main road crossing the intersections.

5.2. Inter-regional road analysis

Commuting between the residential zone and the work area contributed to traffic congestion on the inter-functional roads. For instance, according to Figs. 2, 5 and 6, the most congested time in Q1 was C1 (17:00–19:00) because it was the rush hour when people commute home to the residential areas (Wangjijng and Tiantongyuan) from the CBD. The most congested time in Q2 was A2 (06:30–11:00) mostly due to people commuting from the residential area (Yizhuang) to the CBD. The most congested time in both Q3 and Q4 was A1 (06:30–09:00), which was likely due to people commuting from the residential areas (Fangshan and Daxing, Mentougou and Huiyongguan) to the work area (Zhongguancun). For example, Baizhifang (Fig. 2) (120,000 people/km²) is home to many migrants who live in old, compact buildings. Xizhimen is one of the biggest traffic hubs and is surrounded by railway stations, commercial zone and public service facilities (zoo and exhibition center), and metro stations (Fig. 2) just inside the second ring road. These cases were correlated to the studies that commuting between place of employment and residence contributes to congestion on inter-functional roads, resulting in one-way traffic flow and clear road directions of congestion during peak periods (Fig. 2) (Long et al., 2015).

Based on both intra and inter-regional road analysis, we conclude that the separation between a residential area and other functional areas results in long commuting distance, which contributes to severe traffic congestion. Therefore, increasing mixed land use would improve the situation, which is in line with previous literature that finds that mixed land use reduces travel distances and, hence, car dependency (Land, Factors, Travel, & Litman, 2018). However, increasing the extent of mixed land use at the macro level is insufficient. In the case of both intra and inter-regional roads, congestion is connected to the spatial distribution of large public facilities, such as shopping malls, hospitals, tourist sites, green spaces, which demonstrates that the mixed level of land use in a certain degree. However, to achieve an efficient traffic environment would require a design that allowed citizens to access an area with different transport modes. To improve access to hospitals and schools, perhaps more entrances can be open for citizens to access with different transport modes. Redesigning the entrances may decentralize
the car flows, hence reducing congestion. Opening more access to crossing the green belt area, and connecting the employment center and shopping malls might be a solution for relieving the congestion. Therefore, at the same time as increasing the mixed level of land use at the macro level, improving the detail design at the micro level is suggested in order to achieve an efficient traffic environment, which would result in social benefits in the form of improved liveability and a healthier city.

5.3. Multi-source data fusion and data analytical tool

The potential factors system were constructed based on multi-source data. Multi-source data improved the spatiotemporal resolution of real-time traffic information. For instance, GPS systems on vehicles uploaded high spatiotemporal traffic information. Points of interests and high-resolution remote sensing imagery enabled us to understand patterns of land use and jobs-housing imbalance (Song et al., 2018). Location-based service data provides detailed population density and mobility, etc. Multi-source data assisted in locating the congested road sections and identifying the factors.

While multi-source data offered detailed information, new techniques should be considered to extract useful knowledge from the mass of data for enhanced decision-making. The Geo-detector identified the most influential factors and quantitatively characterized the interactions between pairs of factors that contributed to traffic congestion, thus exploring urban complexity and heterogeneity. Open discussions in this subject include, but is not limited to, application of travel origins and destinations, the balance between residential area and employment area. Diverse location-based service data, video and image recognition techniques could help to predict traffic flow and reduce commuting times, as well as control traffic lights and evacuate vehicles and intervene the land use planning.

6. Conclusion

In this study, we used multi-source data to mine spatiotemporal patterns of traffic congestion and identify the most influential factors affecting traffic congestion. Our findings suggested six clusters for intra-regional and inter-regional roads on weekdays, including the most congested periods: the morning and evening peaks. On both intra-regional and inter-regional roads, building height, which implies a high concentration of jobs, was the most significant single indicator during the morning peak. On intra-regional roads during the late evening peak, the interactions among business area, work area, and water body area significantly enhanced their effects on traffic congestion. On inter-regional roads, the number of bus stops contributed most to the early evening peak; the interaction of this factor with business area, public service area and building density enhanced the effects on congestion. A combined strategy is suggested for future planning with regard to reducing congestion. Planning for mixed land use would be more effective if it was combined with detailed design of the micro-environment to enhance accessibility for different travel modes. The proposed approach can be used to identify traffic congestion hotspots and the potential urban form factors that impact congestion, thereby providing planners with essential information for the production of strategies to alleviate congestion.

Acknowledgment

This research has been supported by the China Scholarship Council Foundation, the National Science Foundation of China (Project No. 71701030), the Humanities and Social Sciences Youth Foundation of the Ministry of Education of China (Project No. 17YJCZH265), China Postdoctoral Science Foundation (Project No. 2018T110223).

Code availability

The code used for mapping functional zone in this study has been deposited in GitHub:

Traffic congestion data: https://github.com/soho1990/Jinchao-Song/tree/master/traffic%20congestion

References

Lauf, S., Haase, D., & Kleinschmit, B. (2016). The effects of growth, shrinkage, population aging and preference shifts on urban development—A spatial scenario analysis of