

## Research Article

# A Novel Trip Coverage Index for Transit Accessibility Assessment Using Mobile Phone Data

**Zhengyi Cai, Dianhai Wang, and Xiqun (Michael) Chen**

*College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China*

Correspondence should be addressed to Xiqun (Michael) Chen; [chenxiqun@zju.edu.cn](mailto:chenxiqun@zju.edu.cn)

Received 30 November 2016; Accepted 18 January 2017; Published 13 February 2017

Academic Editor: Xiaoyue Liu

Copyright © 2017 Zhengyi Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Transit accessibility is an important measure on the service performance of transit systems. To assess whether the public transit service is well accessible for trips of specific origins, destinations, and origin-destination (OD) pairs, a novel measure, the Trip Coverage Index (TCI), is proposed in this paper. TCI considers both the transit trip coverage and spatial distribution of individual travel demands. Massive trips between cellular base stations are estimated by using over four-million mobile phone users. An easy-to-implement method is also developed to extract the transit information and driving routes for millions of requests. Then the trip coverage of each OD pair is calculated. For demonstrative purposes, TCI is applied to the transit network of Hangzhou, China. The results show that TCI represents the better transit trip coverage and provides a more powerful assessment tool of transit quality of service. Since the calculation is based on trips of all modes, but not only the transit trips, TCI offers an overall accessibility for the transit system performance. It enables decision makers to assess transit accessibility in a finer-grained manner on the individual trip level and can be well transformed to measure transit services of other cities.

## 1. Introduction

Public transportation plays an important role in solving traffic problems in urban cities. It is well recognized among transportation planners that transit accessibility is an important measure of the service performance. The Transit Capacity and Quality of Service Manual summarized spatial, temporal, information, and capacity availability factors of public transit systems [1]. A major concern in the public transit sector has been the adequate assessment of access to transit services. Measures of transit accessibility are important in assessing existing transit services, allocating investments, and making decisions on the land development [2]. Transit accessibility has been one of the key indicators of transit planning, performance evaluation, and quantification of the level of service. Transit system planners design the layout of transit lines and stops to improve accessibility and enhance the transit attractiveness. One of the recent research concerns is the extent to which public transit systems enable the less privileged population of privately nonmotorized travelers to access the systems more conveniently, efficiently, and comfortably [3]. To achieve this objective, the first question

we need to answer is: *How to assess accessibility of public transit for trips in a network with spatially and temporally nonuniform travel demands?*

As shown in Table 1, modeling public transit accessibility has attracted numerous research efforts in the past decades, primarily including the evaluation of the spatial coverage [4–6], temporal coverage [7–10], and trip coverage [11–13]. Since passengers' departure time varied according to their own need, Owen and Levinson calculated continuously in time for the evaluation of transit systems rather than at a single of a few departure times [14]. On the other hand, more and more research combined the spatial coverage and temporal coverage to assess transit accessibility. For instance, Mavoa et al. combined public transit, walking accessibility index, and transit frequency to calculate accessibility at the parcel level [15]. Mamun et al. combined the spatial coverage, temporal coverage, and trip coverage to measure public transit performance [16]. El-Geneidy et al. incorporated both travel time and transit fares to determine whether people residing in socially disadvantaged neighborhoods [17].

Based on the literature review it is evident that (1) most assessment models belong to the physical location-based

TABLE 1: Summary of transit accessibility measures.

Category	Measure description	Application	Reference
Physical access to transit	Proximity to transit stops in time or distance	Measuring accessibility for local transit operators in London	Hillman and Pool [5]
	Quarter-mile buffers around transit routes	Transit coverage in the Queen Anne Community of Seattle	Nyerges [6]
	Pedestrian average and maximum walking distance to transit stops	Three neighborhood plans for a 23.3 ha site	Aultman-Hall et al. [4]
Accessibility to destination	Travel-impedance measurements (e.g., travel distance, time, or cost)	Mass/light rapid transit systems in Singapore	Liu and Zhu [11]
	Public transport relative accessibility percentage (transit catchment area and population by transit within 60 min)	90 sites in the south east of England	Gent and Symonds [12]
	Transit accessibility index (TAI) and transit dependence index (TDI)	TransCAD-based transit accessibility measure (TAM) software tool	Bhat et al. [13]
Temporal accessibility	Span and headway of transit service and time-of-day distribution of travel demand	Numerical illustration	Polzin et al. [7]
	Space-time accessibility measures with opportunities and human activity-travel behavior	Commercial and industrial land parcels of Portland Metropolitan Region, Oregon	Kim and Kwan [8]
	Dynamic activity opportunities that can be reached within a prespecified time limit with known transit schedules	Southern California Association of Governments megaregion	Lei et al. [9]
	Rate of access poverty among population	Regional transportation plan scenarios from the San Francisco Bay Area	Golub and Martens [10]

proximity analysis, while few studies take into account the transit coverage of individual travels in a real-world large-scale network; (2) most of the existing transit accessibility measures account for the spatial and temporal coverage, while very few studies consider the trip coverage; (3) the accessibility metrics produced by most existing tools are therefore static in the sense that they describe the transit system but consider less the temporal variability of individual travel demands. To assess whether the public transit service is well accessible for trips of specific origins, destinations, and origin-destination (OD) pairs, a public transit accessibility measure coping with the trip coverage is needed to provide a more reasonable assessment of transit quality of service.

There are some reasons for the gap of the previous studies. In the past, multiple sources of data required to evaluate transit accessibility considering individual travel demands are difficult to collect and consequently extensive efforts are required in order to obtain the useful data. In particular,

it is difficult to measure real-world travel demands due to the small amount of household survey data in the past. In addition, many surveys are zone based and unable to describe individual travel behavior. In some cases, public transit operational data (e.g., stops, routes, schedules, frequencies, and hours of operation) may be hard to access and data fusion could become uneasy due to their inconsistent formats in the time scale and data particle size [18].

Fortunately, the recent advent of data collection technologies, for example, mobile phone signaling data and automated vehicle location, has shifted a data-poor environment to a data-rich environment and offered opportunities to conduct comprehensive transit system performance evaluation. For example, cell phone signaling data have emerged to be a widely used resource to measure both individual travel behavior and network demand, for example, individual human mobility patterns [19, 20], estimation of OD matrices [21], and OD trip purposes [22]. On the other hand, more

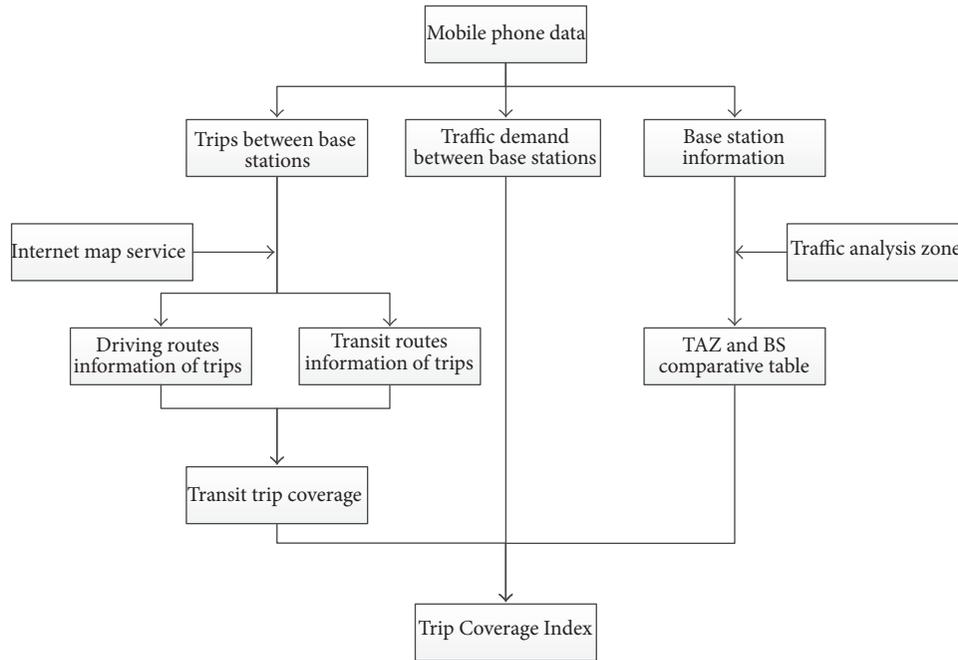


FIGURE 1: Estimation framework of TCI.

and more web map service data become readily available for public use, for example, Google Maps APIs [23], Baidu Map API [24], and AMAP Open Platform [25], which can provide massive on-demand transit trip planning services in real time. Ma and Wang developed a data-driven platform for online transit performance monitoring using automated fare collection and automated vehicle location [26]. Ma et al. developed a series of data mining methods to identify the spatiotemporal commuting patterns of public transit riders using one-month transit smart card data [27]. Therefore, new accessibility indicators taking into account individual trips will definitely provide a more powerful tool.

This paper is aimed at presenting a new public transit quality of service measure, the Trip Coverage Index (TCI), which takes into account both the trip coverage for transit systems and the spatial distribution of heterogeneous and dynamic individual travel demands. The TCI provides a quantitative measure of transit accessibility on the basis of massive trips collected from mobile phone data. The transit accessibility information is extracted from the Baidu Map with the Python code implementation, for example, the access to transit facilities, transit routes (shortest in time/length and alternatives), transit on-vehicle time, and OD connectivity. The novel measure of transit service performance fills the research gap that the conventional spatial coverage index does not consider the coverage to individual trips or the percentage of travel demands that can be served by the transit systems.

The rest of the paper is organized as follows: Section 2 presents the methodological approach to the trip coverage analysis, which is different from the conventional spatial coverage of transit services. In Section 3, an illustrative and

tractable numerical example is employed to present how to calculate TCI and compare it with the conventional measure of spatial coverage. Section 4 shows the field data utilized in this paper and presents results of a real-world city-wide case study that applies TCI to the transit network of Hangzhou, China. Finally, Section 5 concludes the paper and outlooks the future research.

## 2. Methodology

In this part, we first propose a new method to acquire the transit route information for millions of trips determined from the mobile phone data automatically based on online map and programming; then a new public transit quality of service measure (TCI) is proposed considering the access to transit facilities, transit routes information, driving routes information, and OD connectivity. The development of the proposed TCI requires several steps and the framework is shown in Figure 1. The first step is to acquire travel flows between cellular base stations using mobile phone data. Second, the information of transit services and driving routes between each OD pair is then extracted by accessing the online map service for millions of times. Third, the transit trip coverage from base station  $m$  to base station  $n$  is estimated based on the data retrieved from the online map. Fourth, TCI from Zone  $i$  to Zone  $j$  is estimated using the transit trip coverage and travel demand between the base stations. Each of the key procedures of the transit accessibility assessment will be presented in the following sections.

**2.1. Trip Estimation.** In this section, we introduce the mobile phone data and present the methods used to determine trips from the mobile phone data.

**2.1.1. Mobile Phone Signaling Data.** The dataset used in this study consists of two tables in the database: one is the base station table and the other is the anonymous table of mobile phone records. The mobile phone record is generated when a device connects to the cellular network in any of the following cases:

- (i) when the phone makes or receives a call;
- (ii) when the phone sends or receives a message;
- (iii) when the phone is switched on or off;
- (iv) when the user moves from one base station to another; or
- (v) when the system sends the periodic location update request on the phone, for example, 2 h.

The mobile phone signaling data contain Call Details Records (CDR), which were previously utilized to estimate OD demands in numerous related studies [19–22]. Each record of the mobile phone signaling data contains an anonymous user ID, base station ID, and timestamp at the instance of the phone communication with the base station. The base station table contains the base station ID, longitude, and latitude. There are more than 52 thousand of base stations in the urban area of Hangzhou, China. The average covering radius of each base station is less than 100 m.

**2.1.2. Determining Trips.** In order to infer trips from the mobile phone signaling data, the first step is to filter out noise resulting from one base station to another. The call balancing is conducted by the mobile service provider, which creates the appearance of false movements, and distinguishes users' stay locations. Once the stay locations are determined, we evaluate the trips as paths between a user's consecutive locations. To achieve this, we estimate the trips by employing the method of using mobile phone traces data [20]. The estimation is carried out as follows:

- (i) Each mobile phone signaling record  $R_i(k)$  is characterized by a position  $p_i(k)$  expressed by latitude, longitude, and a timestamp  $t_i(k)$  for the  $k$ th observation of a given anonymous user  $i$ .
- (ii) Then the signaling records are connected into a sequence of records  $\{R_i(1), R_i(2), \dots, R_i(n)\}$  according to their time series.
- (iii) If the signaling record series  $\{R_i(q), R_i(q+1), \dots, R_i(z)\}$  satisfy the criteria, (I)  $t_i(z) - t_i(q) > \Delta T$ ; (II)  $\max\{p_i(j), p_i(k)\} < \Delta S, \forall q \leq j, k \leq z$ , which mean a user should stay in an area with the radius  $\Delta S$  (set as 200 m) over a certain time interval  $\Delta T$  (set as 30 min). Then the points  $\{R_i(q), R_i(q+1), \dots, R_i(z)\}$  are fused together by selecting the point with the longest stay time as the stay location.
- (iv) We evaluate paths between a user's stay locations at consecutive points, and the stay locations are assumed to be trip origins or destinations.

## 2.2. Extracting Transit Routes from an Online Map

**2.2.1. Online Map Service.** Calculating the trip coverage indicators requires a database with transit data such as the transit network, road network, operational transit information, and bus stops. Based on those data, we know how many transit lines serve the trips from base station  $m$  to  $n$  and how large the distance is from base station  $m$  to  $n$  by transit line  $l$  and other associated information. Some literatures collected data based on Google Transit or GTFS (General Transit Feed Specification), a supplemental service to Google maps [3, 9, 18]. In those studies, the public transit network in a GIS format and the road network data were required, however, which were difficult to acquire as the data might be from different sources and difficult to use since the data must share the same coordinate, scale, context, and so forth [18].

More and more online map services provide path navigation in China, for example, Baidu Map and AMAP. If the user selects the transportation mode, enters the origin and destination, and chooses a departure time on the map website, it will return the route planning information including the trip distance, trip time, and suggested routes from the origin to destination. Some online map services provide open resources to developers, which are mostly in the form of the Application Programming Interface (API). The API is a set of predefined user applications and the operating system's function, by means of which programmers can easily achieve the underlying operating system feature development or packages. Launched in April, 2010, Baidu Map API [24] not only includes the basic interface to build maps, but also provides information such as local search, route planning, and other data services, through which we can acquire the route information of trips. However, since it needs to search the transit route information for millions of trips completed by over 4 million mobile phone users for a week in Hangzhou, it is impossible to manually acquire such huge information.

As shown in Figure 2, we propose a new method to acquire the transit route information for millions of trips automatically. The trip database stores millions of trips on a local computer server (each trip includes origin/destination geolocations and departure and arrival time). The Python code extracts one piece of trip data and makes an API request to the Baidu Map server via HTTP for the transit route data and the server will return data in the form of XML or JSON after it queries the back-end database. After that we call JSON parsing functions [28] and store the result to the database.

**2.2.2. Transit Routes Data.** The response of the transit route information from the Baidu Map API contains fruitful information and we just extract the useful information for assessing transit accessibility, for example, the taxi route information and bus route information. An example of response is shown in Table 2. The status code indicates whether the online service returns valid results, 0 means a correct record, and 1 means invalid information. The taxi route information includes the taxi distance of the trip, taxi travel time, and monetary costs. The bus route information is much more complicated. There may be several suggested bus schemes per trip and several segments per bus scheme.

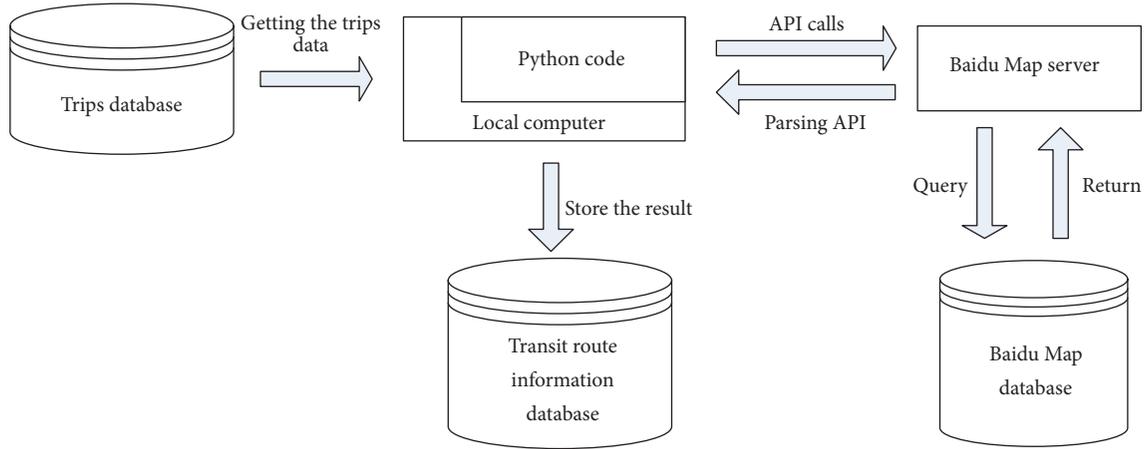


FIGURE 2: Web scraping of transit information.

TABLE 2: A sample of the transit route information from the Baidu Map API.

Status	0	0: correct record 1: error record
Taxi	Taxi_distance (m)	4,540
	Taxi_travel time (s)	503
	Total distance (m)	4,584
	Total travel time (s)	2,276
Bus scheme 1	Scheme_type	1
	Segment1_type	5
	Segment1_distance (m)	311
	Segment2_type	3
	Segment2_distance (m)	3523
	Segment3_type	5
Bus scheme 2	Segment3_distance (m)	750
	...	...
		0: correct record 1: error record

In addition, the `segment_type` code means the travel model (e.g., 5 for walking, 3 for bus).

Generally, there are three segments per scheme in a direct transit route without transfer, which means that the trip distance  $d_{m,n,l}$  from origin base station  $m$  to destination base station  $n$  by taking transit line  $l$  consists of the access distance by walking, in-vehicle distance by bus, and egress distance by walking, given by

$$d_{m,n,l} = d_{m,n,l}^a + d_{\text{in-vehicle}} + d_{m,n,l}^e. \quad (1)$$

**2.3. Trip Coverage Index.** The conventional evaluation criterion for the transit service includes the transit spatial

coverage area, which is usually estimated using the buffer area covered by the transit route or by the area within a walking distance threshold of a transit stop or transit route [1]. The walking distance threshold is modified for various features, for example, the percent elderly in the population and street connectivity [29]. It is commonly accepted by transit planners and researchers that bus transit users are willing to walk up to 1/4 mile (400 m) to reach their nearest transit stop [30–33]. The government agencies and researchers of China use 500 m as the buffer radius to evaluate the transit serving area [34, 35]. The sensitivity of the walking distance threshold will be analyzed in Section 4. In the context of transit, a traveler may transfer from one bus route to another and continue to reach

his/her destination. According to Modesti and Sciomachen [36], more than two times of transfers in a transit trip are generally intolerable for transit users, such that two transfers can be chosen as the maximum value allowed per trip.

Based on the aforementioned idea, this paper presents the binary connectivity parameter ( $\gamma_{m,n,l}$ ) to indicate whether two regions are connected by transit services. Here we only consider walking to reach the transit station and other options such as bike, park, and ride have not been considered. For any trip from base station  $m$  to  $n$ , if (1) there exists a transit line  $l$  that connects the two regions; (2) both the access and egress distances are smaller than the preselected walking distance threshold; and (3) the transfer count  $k_{m,n,l}$  is less than the transfer tolerance threshold  $N$ , then the binary connectivity parameter  $\gamma_{m,n,l}$  is 1; and 0, otherwise. The binary connectivity parameter is given by

$$\gamma_{m,n,l} = \begin{cases} 1, & d_{m,n,l}^a \leq d_0, d_{m,n,l}^e \leq d_0, k_{m,n,l} \leq N, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

$\forall l, m, n.$

The research concern is the extent of the efficiency and attractiveness of the public transport system compared to private cars. As is known to all, there are many factors that may influence the travel mode choice such as travel time, transit distance, transit fare car parking fare, and weather. However, this paper is not concerned with individual travel mode choice behavior but provides the trip-level assessment of transit accessibility, so we just take into account the travel time and travel distance. On the other hand, the developed index (i.e., TCI) is applied to assess accessibility of public transit for all trips, not only those currently or likely to be using transit and no matter whether he/she owns cars or not. Instead, they can rent a car or take a taxi to reach the destination for those have no access to personal car. Therefore, the trip coverage from base station  $m$  to  $n$  served by the transit line  $l$ , that is,  $TC_{m,n,l}$ , is measured by the ratio of the driving distance to the transit distance from base station  $m$  to  $n$  by transit line  $l$  and the ratio of the total travel time by driving to the total travel time by transit line  $l$ , given by

$$TC_{m,n,l} = \alpha \frac{d_{m,n}}{d_{m,n,l}} + (1 - \alpha) \frac{T_{m,n}}{T_{m,n,l}}, \quad 0 \leq \alpha \leq 1, \quad (3)$$

where the weighting factor  $\alpha$  can be determined according to the preference on travel distance or travel time by decision makers. The default value is 0.5.

Considering there may be several transit lines or no transit lines serving the trip, we select the maximum value between 0 and  $TC_{m,n,l}$  multiplied by the binary connectivity parameter  $\gamma_{m,n,l}$  as the trip coverage score  $TC_{m,n}$  for the OD pair in the network, given by

$$TC_{m,n} = \max \{ \gamma_{m,n,l} \cdot TC_{m,n,l}, 0 \}, \quad (4)$$

where the trip coverage score  $TC_{m,n}$  takes into account the OD pairwise transit accessibility which is rarely considered by previous studies.

There are some short trips, of which the distance is shorter than the walking distance threshold. In other words, it is unnecessary to take bus for this trip. So when calculating TCI, we only consider the trips with a distance longer than twice of the service access distance threshold, that is, 1,000 m.

The spatial relationship between TAZs and base stations (BSs) can be obtained by the ArcGIS spatial analysis toolbox. The TAZ-BS membership table can be obtained, which is the foundation of calculating TCI from TAZ  $i$  to  $j$ . TCI from TAZ  $i$  to  $j$  is defined as the weighted average trip coverage score ( $TC_{m,n}$ ) by the travel demand, given by

$$TCI_{i,j} = \frac{\sum_{P_{m,n} \in \{d_{m,n} > 2d_0\}} TC_{m,n} \cdot P_{m,n}}{\sum_{P_{m,n} \in \{d_{m,n} > 2d_0\}} P_{m,n}}. \quad (5)$$

TCI can be used to quantify the coverages of origin TAZ  $i$  and destination TAZ  $j$ , given by

$$TCI_i = \frac{\sum_j TCI_{i,j} \cdot P_{i,j}}{\sum_j P_{i,j}},$$

$$TCI_j = \frac{\sum_i TCI_{i,j} \cdot P_{i,j}}{\sum_i P_{i,j}}, \quad (6)$$

$$TCI = \frac{\sum_i \sum_j TCI_{i,j} \cdot P_{i,j}}{\sum_i \sum_j P_{i,j}}.$$

### 3. An Illustrative Numerical Example

This section provides a tractable numerical example to illustrate the application of TCI to the assessment of transit accessibility. As shown in Figure 3, we consider a road network of four zones served by three transit lines, and each line has four stops. The dashed circles represent 500-meter buffers around each transit stop.

Table 3 shows the travel demands between base stations and the trip coverage results for the trips. Columns 4 and 5 of Table 3 show which TAZs base stations  $m$  and  $n$  belong to, respectively. Column 6 provides the number of transit lines serving the OD pair. Columns 7–13 present the transit route information which can be obtained from the online map for a real-world network. The binary connectivity parameter for each transit line estimated by (2) is shown in Column 14.

The illustrative numerical example helps understand the difference between the proposed measure and the conventional spatial coverage measure. There are two lines and two bus stops serving BS1, while there is only one transit line and one bus stop serving BS2. At the same time, there are two bus lines serving trips of BS1–5 and only one bus line serving trips of BS2–5. It is reasonable to expect a higher level of transit coverage for BS1–5 than that of BS2–5. However, for trips of BS1–5,  $d_{m,n,l}^e$  of both line 1 and line 2 (see Column 11 in bold in Table 3) exceed the preselected distance threshold (500 m), and the binary connectivity parameters for both lines are zero, which means no buses can offer services to trips of BS1–5 given the stop buffer distance threshold.

Column 15 shows the trip coverage of different OD pairs, that is,  $TC_{m,n}$  estimated by (3)–(4). For BS1–6,  $TC_{1,6}$  is

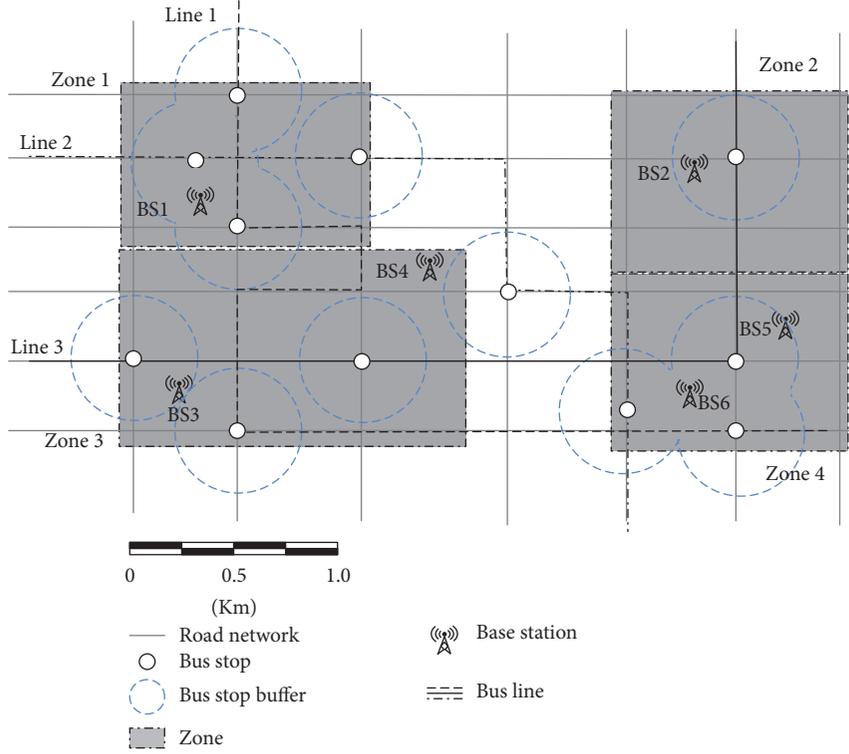


FIGURE 3: An illustrative road and transit networks.

TABLE 3: Trip coverage calculation.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$BS_m$	$BS_n$	$P_{m,n}$	Zone $i$	Zone $j$	Bus line	$d_{m,n}$	$T_{m,n}$	$d_{m,n,l}^a$	$T_{m,n,l}$	$d_{m,n,l}^e$	$d_{in.veh}$	$d_{m,n,l}$	$\gamma_{m,n,l}$	$TC_{m,n}$
1	2	0	1	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
<b>1</b>	<b>3</b>	4	1	3	1	<b>1400</b>	613	300	1225	400	2100	<b>2800</b>	1	<b>0.500</b>
<b>1*</b>	<b>5*</b>	<b>8*</b>	<b>1*</b>	<b>4*</b>	1/2*	3700	1245/1353	300/250	2125/2178	<b>700/1100</b>	4500/3300	5500/4650	0/0	0.000
<b>1*</b>	<b>6*</b>	<b>3*</b>	<b>1*</b>	<b>4*</b>	1/2*	3300	1566/1227	300/200	1875/1470	450/450	4500/3300	5250/3950	1	0.835
2	1	1	2	1	NA	2600	NA	NA	NA	NA	NA	NA	0	0.000
2	3	12	2	3	3	3400	1198	300	1550	300	3800	4400	1	0.773
2	5	8	2	4	3	<b>1400</b>	827	300	975	450	900	<b>1650</b>	1	<b>0.848</b>
3	1	7	3	1	1	1400	613	400	1225	300	2100	2800	1	0.500
3	2	15	3	2	3	3400	1198	300	1550	300	3800	4400	1	0.773
3	5	2	3	4	3	3200	1293	300	1475	450	2900	3650	1	0.877
<b>5*</b>	<b>1*</b>	<b>3*</b>	<b>4*</b>	<b>1*</b>	1/2*	3700	1873/1695	1100/700	2279/1800	250/300	4500/3300	5850/4300	0/0	0.000
<b>6*</b>	<b>1*</b>	<b>8*</b>	<b>4*</b>	<b>1*</b>	1/2*	3300	1580/1311	450/450	1893/1571	300/200	4500/3300	5250/3950	1	0.835
6	2	8	4	2	3	1600	945	450	975	300	900	1650	1	0.970
6	3	3	4	3	3	2600	1236	450	1475	300	2900	3650	1	0.712

Note: NA = not applicable.

calculated as follows and other OD pairs can be calculated in the same way:

$$\begin{aligned}
 TC_{1,6} &= \max \{ \gamma_{1,6,1} \cdot TC_{1,6,1}, \gamma_{1,6,2} \cdot TC_{1,6,2}, 0 \} \\
 &= \max \left\{ 1 \times \left( 0.5 \times \frac{3300}{5250} + 0.5 \times \frac{1566}{1875} \right), 1 \right. \\
 &\quad \left. \times \left( 0.5 \times \frac{3300}{3950} + 0.5 \times \frac{1227}{1470} \right), 0 \right\} = 0.835.
 \end{aligned} \tag{7}$$

Table 3 shows that driving distances of BS1–3 and BS2–5 have the same value of 1400 m, but the transit route distance of BS1–3 is longer than that of BS2–5, which means the transit route of BS1–3 makes a detour and the connectivity level is lower than that of BS2–5 (see Columns 7, 13, and 15 in bold in Table 3). This situation is embodied in the calculation of  $TC_{m,n}$ .

Finally, TCI from TAZ  $i$  to  $j$  should incorporate the trip coverage with the travel demand of the trip. TCI for TAZ 1 to

TABLE 4: TCI and travel demand (in brackets).

Origin	Destination				TCI <sub>i</sub>
	1	2	3	4	
1	NA	0.000 (0)	0.500 (4)	<b>0.228</b> (11)	0.300 (15)
2	0.000 (1)	NA	0.773 (12)	0.848 (8)	0.765 (21)
3	0.500 (7)	0.773 (15)	NA	0.877 (2)	0.702 (24)
4	<b>0.607</b> (11)	0.970 (8)	0.712 (3)	NA	0.753 (22)
TCI <sub>j</sub>	0.536 (19)	0.841 (23)	0.706 (19)	0.526 (21)	0.658 (82)

Note: NA = not applicable.

4 can be calculated according to (5) and the results are shown in Table 4.

$$\begin{aligned} TCI_{1,4} &= \frac{TC_{1,5} \cdot p_{1,5} + TC_{1,6} \cdot p_{1,6}}{p_{1,5} + p_{1,6}} \\ &= \frac{0 \times 8 + 0.835 \times 3}{8 + 3} = 0.228. \end{aligned} \quad (8)$$

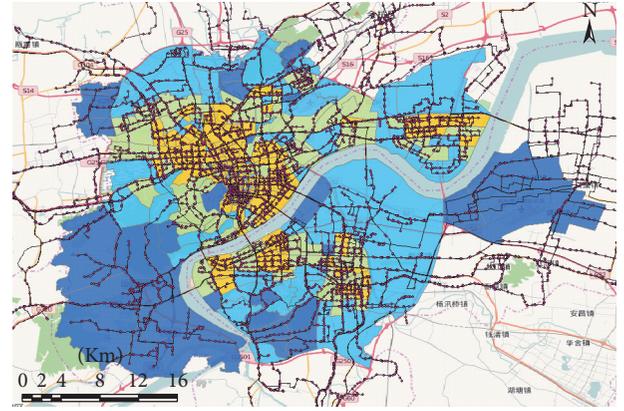
The TCI<sub>i</sub> for TAZ 1 as the origin can be calculated by

$$\begin{aligned} TCI_1 &= \frac{\sum_j TCI_{1,j} \cdot p_{1,j}}{\sum_j p_{1,j}} = \frac{0.5 \times 4 + 0.228 \times 11}{4 + 11} \\ &= 0.300. \end{aligned} \quad (9)$$

Similarly, other TCI<sub>i</sub> and TCI<sub>j</sub> can be obtained. Results of the trip coverage as well as the travel demand of each OD pair are shown in Table 4. It has been realized that the trips in the opposite direction have the same trip coverage scores as Table 3, for example, trips of BS1–6 and BS6–1 (see italic rows in Table 3) and trips of BS1–3 and BS3–1. This is because the bus lines are set two ways in this numerical example. However, the zone-to-zone TCIs in the opposite directions show different scores, for example, TCI<sub>1,4</sub> and TCI<sub>4,1</sub> highlighted in Table 4. Recalling the contents indicated by asterisk of Table 3, we find that the demands from Zone 1 to Zone 4 and from Zone 4 to Zone 1 are different in opposite directions, which means the transit system covering more travel demands has a higher value of TCI.

TCI also offers a way to quantify the transit service level of OD pairs that require a transfer between transit lines. Equation (3) can be improved by considering the transfer distance  $d_{transfer}$  and transfer travel time  $T_{transfer}$ , given by

$$\begin{aligned} TC_{m,n,l} &= \alpha \frac{d_{m,n}}{\sum_l d_{m,n,l} + d_{transfer}} \\ &+ (1 - \alpha) \frac{T_{m,n}}{\sum_l T_{m,n,l} + T_{transfer}}, \quad 0 \leq \alpha \leq 1. \end{aligned} \quad (10)$$



Coverage radius meter



FIGURE 4: Layout of transit and the average base station coverage radius in terms of TAZs in Hangzhou, China.

The spatial coverage is the proportion of the area served by transit stops, which can be calculated by the Transit Capacity and Quality of Service Manual [1]. This method uses a buffer (set as 500 m) around each stop to define the spatial coverage of bus services. Table 5 shows the zonal data of the spatial coverage calculations and the corresponding TCI results. The buffer area for each stop is calculated using the ArcGIS toolbox and the overlapped buffers are calculated only once. The results show that the spatial coverage of Zone 1 is much higher than that of Zone 2, while the TCI of Zone 1 is lower than that of Zone 2, which are highlighted in Table 5.

## 4. Case Study

4.1. Study Area and Data. In this section, TCI is applied to a case study in Hangzhou, China, to assess the transit

TABLE 5: Zonal data of spatial coverage and TCI.

Zone, $i$	Zone area (km <sup>2</sup> )	Bus stop buffer (km <sup>2</sup> )	Spatial coverage	TCI <sub><math>i</math></sub>	TCI <sub><math>j</math></sub>
1	0.880	0.680	<b>0.773</b>	<b>0.300</b>	<b>0.536</b>
2	1.026	0.283	<b>0.275</b>	<b>0.765</b>	<b>0.841</b>
3	1.650	0.807	0.489	0.702	0.706
4	1.026	0.503	0.490	0.753	0.526

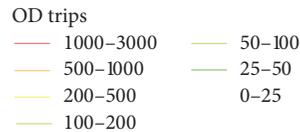
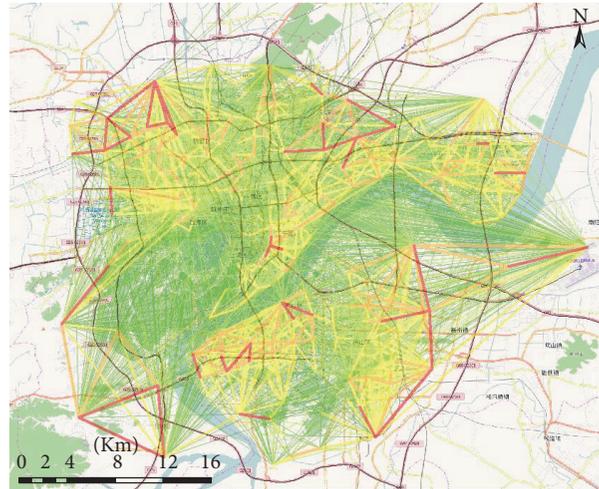


FIGURE 5: Desire lines of OD trips during the AM peak hours.

accessibility. Hangzhou is the capital and most populous city of Zhejiang Province, China. As shown in Figure 4, the study area contains the Shangcheng District, Xiacheng District, Jianggan District, Binjiang District, Xihu District, Gongshu District, and part of Xiaoshan District. The study area is 955 km<sup>2</sup>, and it contains 540 TAZs with 4.43 million residents. As shown in Figure 4, there are 912 transit lines and 18,508 transit stops in the transit network of Hangzhou.

The mobile phone signaling data used in this study consist of two tables, that is, the base station table and the anonymous mobile phone records table collected from 4.17 million mobile phone users in Hangzhou over one month between August and September, 2015. The position accuracy of a trip is determined by the coverage radius of base stations. There are 41,823 base stations in the 540 TAZs, and the average BS coverage radius for each TAZ is shown in Figure 4. Results show that the average coverage radius of 90% base stations is less than 100 m. The remaining 10% are distributed in less populated areas such as the mountainous and wetland areas shown in Figure 4.

The study time periods are AM peak hours (7:00–9:00) and PM peak hours (17:00–19:00). After processing the mobile phone signaling data using the method proposed in

Section 2.1, we obtain 2,816,910 trips in AM peak hours and 2,756,187 trips in PM peak hours on September 8, 2015, a regular working day. The desire lines of trips are shown in Figure 5. The average trip distance is 5.68 km and more than 50% of the trips are less than 3 km.

We also obtain spatial and temporal distributions of the population density using the trip information, for example, origin, destination, and timestamp. As shown in Figure 6, the population distribution is dispersed before 7:00 and after 20:00 and is aggregating during working hours. These observations are consistent with the daily experience.

**4.2. Results.** Combining both the mobile phone data and transit information extracted from the online map service, we are able to calculate the TCI for different time of day. The analytical results are as follows: the distribution of TCI <sub>$i$</sub>  in the AM peak hours and PM peak hours are shown in Figures 7(a)-7(b). We can see a totally different picture of TCI <sub>$i$</sub>  in the AM peak hours as compared to that in PM peak hours, which indicates that some of the TCI <sub>$i$</sub>  vary during different periods while the transit routes and departure interval are the same, which is similar to the findings of [37]. Based on those pictures, we should have different principles and

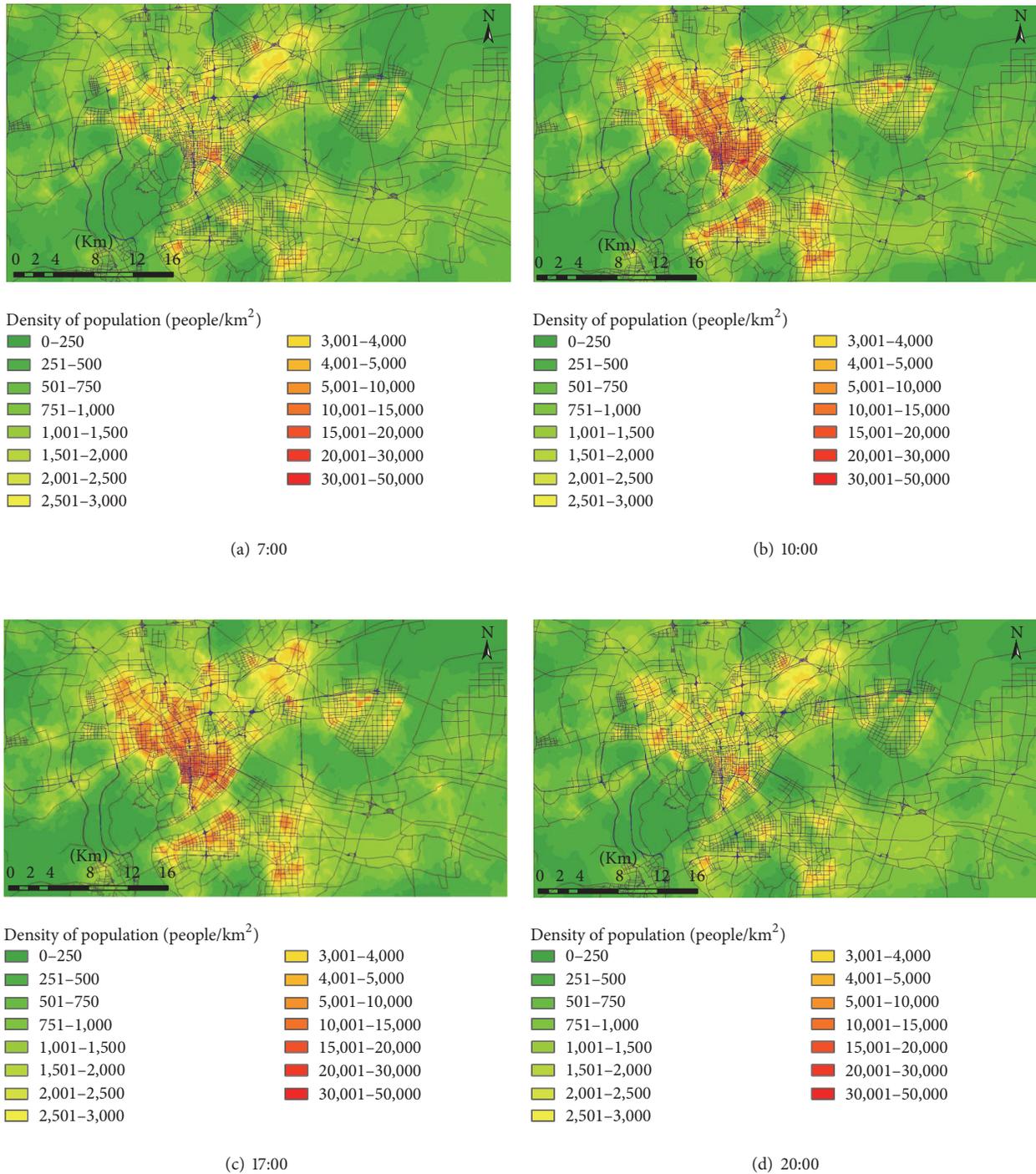
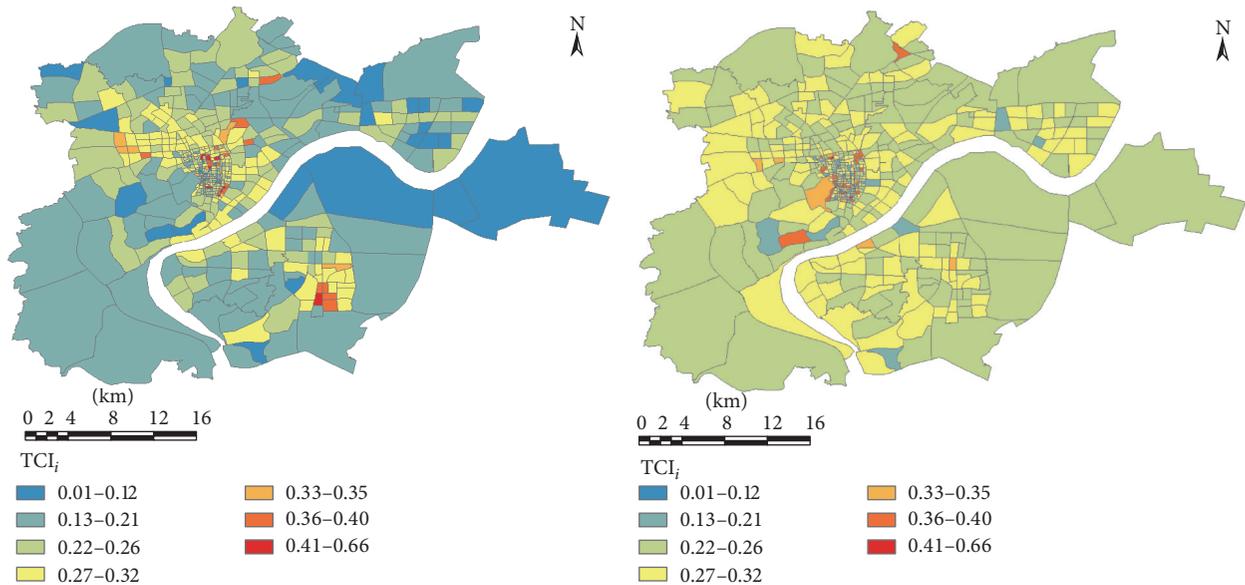


FIGURE 6: Spatiotemporal distributions of population density using mobile phone signaling data (September 8, 2015).

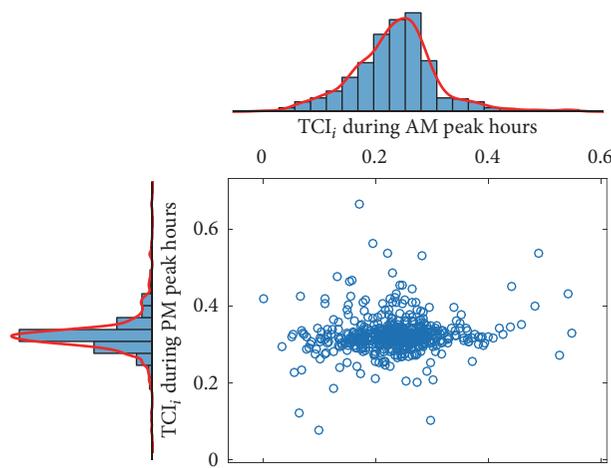
strategies in terms of deploying our bus-related resources and services. Most of the  $TCI_i$  in the AM peak hours are higher than those in the PM hours according to Figure 7(c) and the outliers show that  $TCI_i$  of some TAZs significantly vary with time periods and different travel demands. Transit operators should reschedule transit routes in a dynamic manner to be more consistent with travel demands.

As shown in Figure 8, results are compared with the spatial transit coverage during the AM and PM peak hours, respectively.  $TCI_i$  of TAZs is a skewed distribution, and most of the  $TCI_i$  is lower than 0.5, which means the transit system provides a poor coverage for the travel demand, while the spatial transit coverage calculated by the buffer method shows higher scores in both peak hours, which means most of



(a) Distribution of  $TCI_i$  of the TAZs in the AM peak hours

(b) Distribution of  $TCI_i$  of the TAZs in the PM peak hours



(c) Comparison of  $TCI_i$  of TAZs between AM and PM peaks

FIGURE 7: Distributions of TCI of TAZs during peak hours ( $d = 1000$  m,  $\alpha = 0.5$ ).

the population can access to the bus stops in the walking threshold. The comparison between  $TCI_i$  and the spatial transit coverage shows the bus stops which can be accessed in the walking threshold may not lead travelers to their destinations by transit services.

In order to further explore the sensitivity of the walking distance threshold, acceptable times of transfer, and the weighting factor  $\alpha$ , we summarize the statistics of TCI in terms of these factors in Figure 9. As shown in Figure 9(a), as the walking distance threshold rises from 500 m to 900 m, the TCI of whole network rises significantly; then it slows down after the walking distance threshold is longer than 900 m,

which means that transit operators could gain more trip coverage by improving bus services for those travel demands that the walking distance is under 900 m.

As the acceptable transfer times increase from 0 to 2, the TCI increases both during the AM peak hours and PM peak hours, which is comparable with experience. The results suggest that increasing the transit route crossings would provide a better transit service.

We also explore the interaction between the weighting factor  $\alpha$  and TCI. The larger value of  $\alpha$  is, the more weight of travel distance is considered in the TCI. Figure 9(c) shows that TCI increases with  $\alpha$  in all transfer scenarios, which

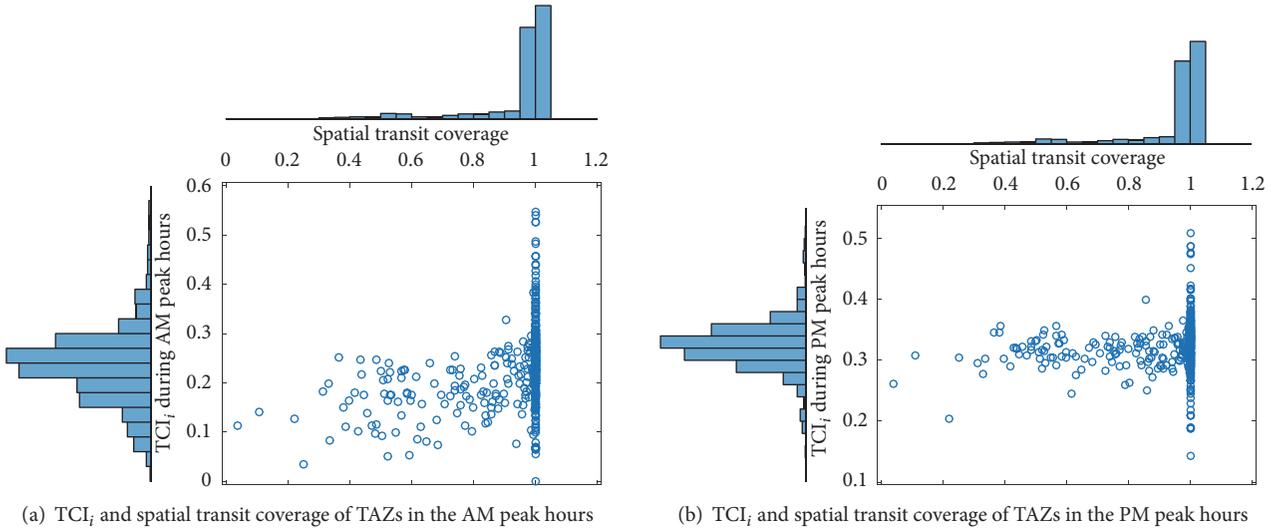


FIGURE 8: Comparison of  $TCI_i$  with the spatial transit coverage during AM and PM peak hours ( $d = 1000$  m,  $\alpha = 0.5$ ).

means TCI is sensitive to ratio of the bus travel time and the driving time. So we should promote the travel time reliability of buses to provide a better transit service.

## 5. Conclusions

In this paper, the novel TCI is proposed for measuring transit connectivity and accessibility. It is built on the existing transit service measures and allows us to analyze the transit connectivity and accessibility for massive trips between the origin and destination, as well as the transit coverage from or to a TAZ. This paper is among the first attempts considering the connectivity of trips from point to point and real-world complicated travel demand in a large-scale urban area. The TCI developed in this paper provides the capability to quantify the level of accessibility of the transit system and vary the assessment of transit accessibility with the temporal and spatial change of travel demands.

This paper also presents an easy-to-implement method to acquire the transit route information for millions of trips based on the online map. Since the data is acquired automatically using computer programming, it is possible to easily construct the data repository and analyze large public transit networks.

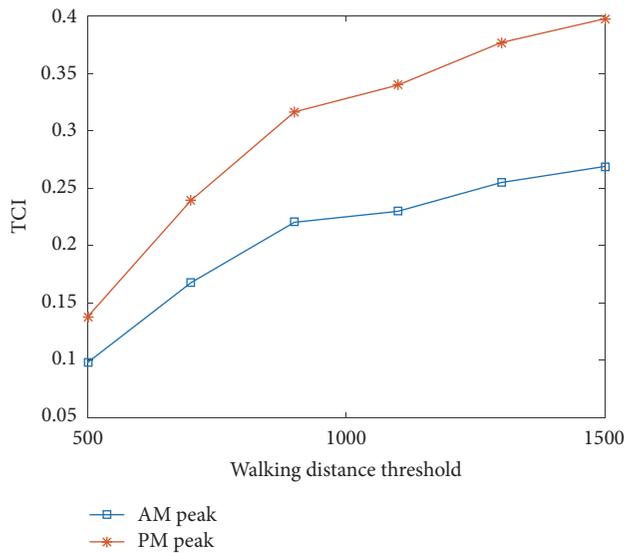
TCI can be applied to all trips, not only those currently or likely to be using transit, such that TCI is demonstrated as an overall measure of transit accessibility and can be used to measure how the transit system reaches its target, which is to provide services for more potential users.

Through the case study of Hangzhou, we find that fluctuations in the travel demand in different time periods make TCI distributing diversely across the city, which means transit operators should reschedule transit routes in a dynamic way to be consistent with travel demands. The sensitivity analysis is performed to determine how the walking distance threshold, times of transfer, and the weighting factor would impact

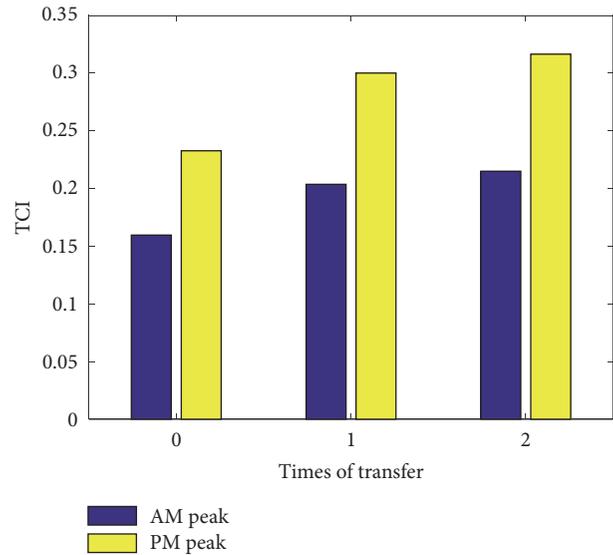
the network-wide TCI. The results can provide operators targeted measures to improve transit services.

## Notations

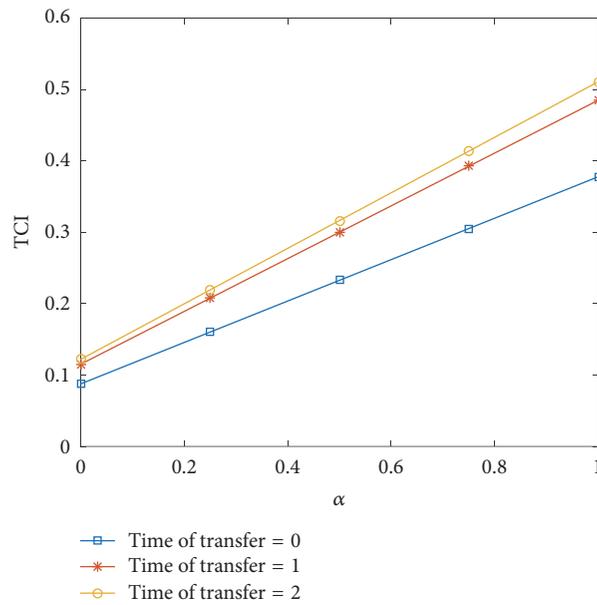
$d_0$ :	Service access distance threshold
$d_{m,n}$ :	Distance from base station $m$ to $n$ by car
$T_{m,n}$ :	Travel time from base station $m$ to $n$ by car
$d_{m,n,l}$ :	Total distance from base station $m$ to $n$ by transit $l$
$k_{m,n,l}$ :	Transfer count from base station $m$ to $n$ by transit $l$
$T_{m,n,l}$ :	Total travel time from base station $m$ to $n$ by transit $l$
$d_{m,n,l}^a$ :	Access distance from base station $m$ to $n$ by transit $l$
$d_{m,n,l}^e$ :	Egress distance from base station $m$ to $n$ by transit $l$
$i$ :	Origin traffic analysis zone (TAZ)
$j$ :	Destination TAZ
$l$ :	Transit line
$m$ :	Origin base station
$n$ :	Destination base station
$p_{m,n}$ :	Travel demand from base station $m$ to $n$
$p_{i,j}$ :	Travel demand from TAZ $i$ to $j$
$TC_{m,n,l}$ :	Trip coverage from base station $m$ to $n$ by transit $l$
$TC_{m,n}$ :	Trip coverage from base station $m$ to $n$
$TCI_{i,j}$ :	Trip coverage index from TAZ $i$ to $j$
$TCI_i$ :	Trip coverage index of origin TAZ $i$
$TCI_j$ :	Trip coverage index of destination TAZ $j$
TCI:	Trip coverage index of a transit network
$\gamma_{m,n,l}$ :	Binary connectivity parameter, 1 if a transit line $l$ connects the trip from $m$ to $n$ with $d_{m,n,l}^a$ and $d_{m,n,l}^e$ smaller than $d_0$ ; 0 otherwise.



(a) Transit network TCI with respect to the walking distance threshold ( $\alpha = 0.5$ )



(b) Transit network TCI with respect to transfers ( $d = 1000$  m,  $\alpha = 0.5$ )



(c) Transit network TCI with respect to  $\alpha$  ( $d = 1000$  m)

FIGURE 9: Sensitivity analysis for transit network.

### Competing Interests

The authors declare that they have no competing interests.

### Acknowledgments

This research is financially supported by Zhejiang Provincial Natural Science Foundation of China under Grant no. LR17E080002, National Natural Science Foundation of China under Grant nos. 51508505, 51338008, and 51278454, and Hangzhou Municipal Science and Technology Commission under Grant no. 20142013A57. Mr. Yanlei Cui helped in

processing some of the data used in this paper, and his assistance is gratefully acknowledged.

### References

- [1] KFH Group, *Transit Capacity and Quality of Service Manual*, Report 100, Transit Cooperative Research Program, Washington, DC, USA, 2nd edition, 2003.
- [2] S. Al Mamun and N. E. Lowmes, "Measuring service gaps: accessibility-based transit need index," *Transportation Research Record*, vol. 2217, pp. 153–161, 2011.

- [3] K. Fransen, T. Neutens, S. Farber, P. De Maeyer, G. Deruyter, and F. Witlox, "Identifying public transport gaps using time-dependent accessibility levels," *Journal of Transport Geography*, vol. 48, pp. 176–187, 2015.
- [4] L. Aultman-Hall, M. Roorda, and B. W. Baetz, "Using GIS for evaluation of neighborhood pedestrian accessibility," *Journal of Urban Planning and Development*, vol. 123, no. 1, pp. 10–17, 1997.
- [5] R. Hillman and G. Pool, "GIS-based innovations for modelling public transport accessibility," *Traffic Engineering and Control*, vol. 38, no. 10, pp. 554–559, 1997.
- [6] T. Nyerges, "Geographic information system support for urban/regional transportation analysis," in *Geography of Urban Transportation*, S. Hanson, Ed., pp. 240–265, The Guilford Press, New York, NY, USA, 1995.
- [7] S. E. Polzin, R. M. Pendyala, and S. Navari, "Development of time-of-day-based transit accessibility analysis tool," *Transportation Research Record*, no. 1799, pp. 35–41, 2002.
- [8] H.-M. Kim and M.-P. Kwan, "Space-time accessibility measures: a geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration," *Journal of Geographical Systems*, vol. 5, no. 1, pp. 71–91, 2003.
- [9] T. Lei, Y. Chen, and K. Goulias, "Opportunity-based dynamic transit accessibility in Southern California: measurement, findings, and comparison with automobile accessibility," *Transportation Research Record*, no. 2276, pp. 26–37, 2012.
- [10] A. Golub and K. Martens, "Using principles of justice to assess the modal equity of regional transportation plans," *Journal of Transport Geography*, vol. 41, pp. 10–20, 2014.
- [11] S. Liu and X. Zhu, "Accessibility analyst: an integrated GIS tool for accessibility analysis in urban transportation planning," *Environment and Planning B: Planning and Design*, vol. 31, no. 1, pp. 105–124, 2004.
- [12] C. Gent and G. Symonds, "Advances in public transport accessibility assessments for development control—a proposed methodology," in *Planning and Transport, Research and Computation (PTRC) Annual Transport Practitioners' Meeting*, 2005.
- [13] C. R. Bhat, S. Bricka, J. La Mondia, A. Kapur, J. Y. Guo, and S. Sen, "Metropolitan area transit accessibility analysis tool," Tech. Rep. 0-5178-P3, Texas Department of Transportation, 2006.
- [14] A. Owen and D. M. Levinson, "Modeling the commute mode share of transit using continuous accessibility to jobs," *Transportation Research Part A: Policy and Practice*, vol. 74, pp. 110–122, 2015.
- [15] S. Mavoia, K. Witten, T. McCreanor, and D. O'Sullivan, "GIS based destination accessibility via public transit and walking in Auckland, New Zealand," *Journal of Transport Geography*, vol. 20, no. 1, pp. 15–22, 2012.
- [16] S. A. Mamun, N. E. Lownes, J. P. Osleeb, and K. Bertolaccini, "A method to define public transit opportunity space," *Journal of Transport Geography*, vol. 28, pp. 144–154, 2013.
- [17] A. El-Geneidy, D. Levinson, E. Diab et al., "The cost of equity: assessing transit accessibility and social disparity using total travel cost," in *Proceedings of the Transportation Research Board 95th Annual Meeting*, 16-3715, Washington, DC, USA, 2016.
- [18] Y. Hadas, "Assessing public transport systems connectivity based on Google Transit data," *Journal of Transport Geography*, vol. 33, pp. 105–116, 2013.
- [19] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [20] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: a mobile phone trace example," *Transportation Research Part C*, vol. 26, pp. 301–313, 2013.
- [21] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin-destination matrices using mobile phone call data," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 63–74, 2014.
- [22] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin-destination trips by purpose and time of day inferred from mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 240–250, 2015.
- [23] "Google Maps APIs," 2016, <https://developers.google.com/maps>.
- [24] "Baidu Map Open Platform," 2016, <http://lbsyun.baidu.com>.
- [25] AMAP, <http://lbs.amap.com>.
- [26] X. Ma and Y. Wang, "Development of a data-driven platform for transit performance measures using smart card and GPS data," *Journal of Transportation Engineering*, vol. 140, no. 12, Article ID 04014063, 2014.
- [27] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.
- [28] R. Lawson, *Web Scraping with Python*, Packt Publishing Ltd, 2015.
- [29] S. Biba, K. M. Curtin, and G. Manca, "A new method for determining the population with walking access to transit," *International Journal of Geographical Information Science*, vol. 24, no. 3, pp. 347–364, 2010.
- [30] A. T. Murray, "Strategic analysis of public transport coverage," *Socio-Economic Planning Sciences*, vol. 35, no. 3, pp. 175–188, 2001.
- [31] F. Zhao, L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan, "Forecasting transit walk accessibility: regression model alternative to buffer method," *Transportation Research Record*, no. 1835, pp. 34–41, 2003.
- [32] J. Gutiérrez, O. D. Cardozo, and J. C. García-Palomares, "Transit ridership forecasting at station level: an approach based on distance-decay weighted regression," *Journal of Transport Geography*, vol. 19, no. 6, pp. 1081–1092, 2011.
- [33] Q. Chen and J. Pan, "Coverage rate about the bus stop serving area," *Urban Public Transport*, vol. 2, pp. 17–18, 2002.
- [34] Y. Wang and Q. Li, "Evaluation of the public bus network in Beijing," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 5, pp. 135–141, 2007.
- [35] Ministry of Construction of the People's Republic of China, *Code for Transport Planning on Urban Road*, Ministry of Construction of the People's Republic of China, Beijing, China, 1995.
- [36] P. Modesti and A. Sciomachen, "A utility measure for finding multiobjective shortest paths in urban multimodal transportation networks," *European Journal of Operational Research*, vol. 111, no. 3, pp. 495–508, 1998.
- [37] W. A. Xu, Y. Ding, J. Zhou, and Y. Li, "Transit accessibility measures incorporating the temporal dimension," *Cities*, vol. 46, pp. 55–66, 2015.

