

Article

Unfolding Spatial-Temporal Patterns of Taxi Trip based on an Improved Network Kernel Density Estimation

Boxi Shen ¹, Xiang Xu ^{2,*} , Jun Li ¹, Antonio Plaza ³ and Qunying Huang ⁴ 

¹ Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; shenbx@mail2.sysu.edu.cn (B.S.); lijun48@mail.sysu.edu.cn (J.L.)

² Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528402, China

³ Hyperspectral Computing Laboratory Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain; aplaza@unex.es

⁴ Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA; qhuang46@wisc.edu

* Correspondence: xuxiang@zsc.edu.cn

Received: 9 October 2020; Accepted: 13 November 2020; Published: 15 November 2020



Abstract: Taxi mobility data plays an important role in understanding urban mobility in the context of urban traffic. Specifically, the taxi is an important part of urban transportation, and taxi trips reflect human behaviors and mobility patterns, allowing us to identify the spatial variety of such patterns. Although taxi trips are generated in the form of network flows, previous works have rarely considered network flow patterns in the analysis of taxi mobility data; Instead, most works focused on point patterns or trip patterns, which may provide an incomplete snapshot. In this work, we propose a novel approach to explore the spatial-temporal patterns of taxi travel by considering point, trip and network flow patterns in a simultaneous fashion. Within this approach, an improved network kernel density estimation (imNKDE) method is first developed to estimate the density of taxi trip pick-up and drop-off points (ODs). Next, the correlation between taxi service activities (i.e., ODs) and land-use is examined. Then, the trip patterns of taxi trips and its corresponding routes are analyzed to reveal the correlation between trips and road structure. Finally, network flow analysis for taxi trip among areas of varying land-use types at different times are performed to discover spatial and temporal taxi trip ODs from a new perspective. A case study in the city of Shenzhen, China, is thoroughly presented and discussed for illustrative purposes.

Keywords: pick-up and drop-off points (ODs); network kernel density estimation (NKDE); land-use data; map-matching

1. Introduction

Urban mobility data are of significant importance to urban development and play an important role in understanding urban traffic system [1,2], typically include taxi trajectories [3,4], bus trajectories [5], smart card records for transportation [6], bike sharing trajectories [7,8] and other public transport systems (underground, tramway, railway, etc). Among them, the taxi is an important part of urban transportation, and the taxi trips reflect human behaviors and mobility patterns, allowing us to identify the spatial variety of mobility patterns. Specifically, point patterns, trip patterns, and network flow patterns offer a valuable opportunity to get valuable insights in taxi mobility, which is one of the most important parts of urban mobility [9]. It is important for the urban traffic application nowadays.

It is essential to understand the insights that mobility data imply. Taxi mobility research has been primarily focusing on the following three aspects. First, most studies mainly considered point

patterns, such as taxi trajectories, which can give relevant insight into the passengers' behavior [10,11], trip purposes [12,13], and spatial patterns [14,15]. Second, some studies further considered trip patterns, such as leveraged taxi trips, whose pick-up and drop-off points (ODs) [16] are used for spatial pattern analysis and identification of trip purpose and spatial distribution. Note, in this paper, we define the locations of pick-up events as Os, the locations of drop-off events as Ds, the pairs of O and D are ODs. Third, few studies considered network flow patterns, and many methodologies generally extracted one single pattern for exploitation [17,18].

While examining point patterns, it is important to explore the relationship between land-use types and human mobility data [19], as human activities are closely related to land-use types [20]. Many studies have traced the human's travel behaviors based on land-use types [9,10]. The existing inverse relationship between human activities and land-use types (i.e., extracting land-use categories from residential activities) has triggered many studies [16,17,21,22]. Multiple features extracted from taxi trajectory data, such as the outflow, inflow, net flow (inflow–outflow) and net flow ratio ($(\text{inflow} - \text{outflow}) / (\text{inflow} + \text{outflow})$) can support land-use classification [23,24]. For example, Ge et al. proposed an integration framework to fuse multiple features and increase the accuracy of land-use classification [25]. These taxi trajectory-based features were computed with a grid cell-based method, and the taxi trajectory data were not converted into a road network space. The impact of different land-use and districts on taxi trip ODs may vary in time series. Understanding the relationship between land-use and taxi service activities can provide relevant insights about how to optimize transport planning.

Previously, the network kernel density estimation (NKDE) algorithm, which considers a kernel density function based on road network distance instead of the Euclidean distance, was used to analyze the point events along road networks [26]. The analysis based on NKDE can provide more precise patterns in network-related scenarios, offering information about the most densely occupied road network segments around point events. While NKDE has been widely used to examine traffic accidents [27], economic activities [28], central business districts [29], and the accessibility of points of interest (POIs) [30], few studies have leveraged it for exploring the spatial patterns of ODs along road networks.

In trip patterns, many studies leveraged taxi trip ODs to analyze the characteristics [16]. An important research area is to discover people's behavioral patterns by analyzing the taxi trip ODs. Such ODs, computed from the taxi track data, have been explored for trip pattern analysis, trip purpose and spatial distribution by means of integrated spatiotemporal geographic information system (GIS) toolkits. Meanwhile, taxi services are one of the most important driving activities, and have been widely used to quantify transport characteristics [31]. For example, transportation theory demonstrates that drivers minimize travel time for route choice behavior [32], and some studies found that taxi trajectories can give relevant insights into passengers' behavior [10,11], trip purposes [12,13], and spatial patterns [14–16], henceforth drawing plenty of attention for research.

In fact, taxi trips are generated in a network flow, which is more intuitive to analyze the spatial travel under the network flow patterns. Network flow considers locations, links, and the interaction between locations by the number of links [33]. Shen et al. aim to uncover spatial and temporal patterns such as people's location characteristics and space-time movements (whose trends vary over time) through the analysis of a large volume of taxi data. Liu et al. proposed a spatially-embedded network model to discover intra-city spatial interactions [21].

To sum up, existing works only consider one kind of spatial pattern. Since the taxi mobility data itself is available in point frame format, and the associated human behavior is in network frame format, it is natural and essential to analyze spatial patterns via the three sources of information. In this work, we pursue the integration of point patterns, trip patterns, and network flow patterns to provide a better identification and understanding of the spatial variety of taxi mobility data. Specifically, we propose to identify the spatial variety of travel patterns from taxi mobility data by considering the point, trip and

network aspects, simultaneously. In order to achieve this goal, there are three challenging issues that need to be addressed, as detailed below:

- (1) First of all, although much progress has been made in the literature regarding the relationship between land-use and transportation (good performance has been reported from spatial distributions, spatial statistics, and spatial analytics), there are few studies focused on the relationship between taxi spatial patterns and land-use. With this consideration in mind, we propose a new method, called improved network kernel density estimation (imNKDE), which is able to estimate the OD density efficiently from a large amount of taxi trajectory data, and to further identify the spatial patterns from OD density and land-use data via Poisson regression.
- (2) The relationships and spatial characteristics of a taxi trip and its corresponding routes have not been sufficiently investigated, and these items are helpful to find the distribution of the hottest road segments. While a taxi trip with passengers likely follows the shortest route, i.e., the shortest path [32], the trip eventually selected by a taxi driver may be impacted by various factors of the road network and structure, such as travel time, travel speed, the number of road lanes and left turns, and the proportion of highways. As such, it is still unclear to what extent a taxi trip follows the shortest path. Here, we propose the use of trip patterns to measure the relationship between taxi trips and taxi routes based on their similarities. A regression model is introduced to further investigate the potential factors that may affect the taxi driver's choices for taxi trips.
- (3) Another important issue that we consider is that the current network flow is designed for integrating point analytics, which is not suitable for trip analytics. As mentioned, trip patterns are essential; however, previous studies did not show the spatial relationship between long and short taxi trips. Therefore, further network flow analytics should be considered to reveal the distance-based effects among taxi trips. In this work, we propose to use network flow pattern analytics for modeling the differences between taxi trip ODs and land-use data. More specifically, we use inter-zonal based and inner-zonal based spatial interaction analysis to capture these differences.

A case study in the city of Shenzhen, China, has been selected for validation purposes. The taxi mobility data, along with the land-use data and road network data are considered. The main contributions of our work can be summarized as follows:

- (1) An improved NKDE (imNKDE) is proposed to process a large amount of taxi trajectories for the estimation of OD density.
- (2) By jointly considering the taxi trip ODs and road network data, it is observed that taxi drivers prefer roads with more lanes or highways.
- (3) We identify network flow patterns which are used to discover spatial interactions between different districts and land-use.

The remainder of the paper is organized as follows. Section 2 describes some related works. Section 3 introduces the research data. Section 4 describes the methodology for identifying spatial patterns for urban mobility data. Detailed results and analyses are given in Section 5. Section 6 concludes the paper with some remarks and hints at plausible future research lines.

2. Related Work

2.1. Point Patterns for Taxi Trajectory Analytics

Mobility data fostered one of the most emerging and active innovation areas for sustainable urban transport [34–37]. Taxi trajectories represent an important road-based mobility data source. Zhou et al. analyzed the urban functional structures and people's activities by using functionally critical network locations based on taxi trajectories [38]. Functional regions can be defined by taxi

trajectories, using clustering methods [39–42]. Nevertheless, focusing on the taxi trip ODs may lead to a loss of transport information on the whole trip, thus overlooking shortest path properties along the trip and routes. Correspondingly, it is essential to conduct spatial analytics for taxi trajectory events along road networks. NKDE has been widely used to detect urban hotspots along road networks based on taxi trajectories [43–45]. Delso et al. proposed an integrated model to measure the pedestrian-habitat suitability of streets [46]. Land-use type also has strong relationships with taxi trajectories. Pan et al. proposed an improved clustering algorithm to perform land-use classification using taxi data [24], however, this work did not consider taxi trip ODs along road networks and did not identify the spatial patterns between ODs and land-use types. Traffic volume or trip length have been used to explain traffic volume or trip length (from a land-use perspective) by analyzing the dependence between human's behaviors and land-use types [47,48].

2.2. Taxi Trip Patterns Analytics

Understanding route choice behavior of taxi trajectories is essential to sustainable urban transport. Wardrop's transportation theory demonstrates that drivers minimize travel time for route choice behavior [33]. Following the shortest path is one of the major features of human route choice behavior, because the shortest path potentially exhibits the minimum travel time [34]. However, Yao et al. found that taxi drivers tend to choose the route with faster travel speed, less left turns and more proportion of highways (only considering 221 taxi trips in Beijing) [49]. Sun et al. showed that travel distance, travel time and road preference have high influence on taxi drivers' route choices (e.g., Shenzhen taxi trajectory data [50]). The total number of taxi trips in Sun et al.'s work was around 4000, which is much smaller than the total trips in real-world taxi trajectory scenarios. In turn, the multinomial logistic regression model has been used to model route choice, which is relatively influenced by actual travel time [51]. As mentioned before, taxi drivers tend to choose the path with the shortest distance or fastest time [51,52]. As a result, it is critical to discover which taxi trips (with passengers) follow the shortest path in the presence of different factors, using larger-scale taxi trajectory data sets.

2.3. Network Flow Pattern for Taxi Trajectory Analytics

The taxi network flow considers locations, links, and the interaction between locations by the number of links [53]. Visualization of the network flow for taxi trips (using grid-cell based counting) has the potential to reflect the spatial interactions among different parts [53,54], but this method only considers the properties of taxi trajectories. Yang et al. proposed a flow map method called MapLinks for analyzing ODs [55]. The OD-Wheel method revealed some details of local patterns in taxi trajectories [56]. Further research is needed towards the inclusion of network flows in taxi mobility patterns.

3. Study Area and Dataset

3.1. Study Area

Shenzhen city in Guangdong province, China, is selected as our study area. Shenzhen is located in the southeast of China, adjacent to Hong Kong (see Figure 1). As a special economic zone and a modern metropolitan city in China, Shenzhen, is one of the first-tier cities in China. Meanwhile, Shenzhen is one of the most economically efficient cities in mainland China. The Luohu district, Nanshan district and Futian district are the central districts as the administration, finance, culture, and information centers of the Shenzhen city. In the end of 2019, there are nine administrative districts and one new district. The total area of Shenzhen is 1997.47 square kilometers, with a built-up area of 927.96 square kilometers. Its permanent population is 13,438,800. As one of the most densely populated cities in China, there is a strong traffic fluidity in Shenzhen. The 2019 digital representation of Shenzhen road network is composed of 72,357 road segments and 51,074 network nodes, as available from an OpenStreetMap (<http://www.openstreetmap.org/>) using OSMnx [57,58] (see Figure 2).

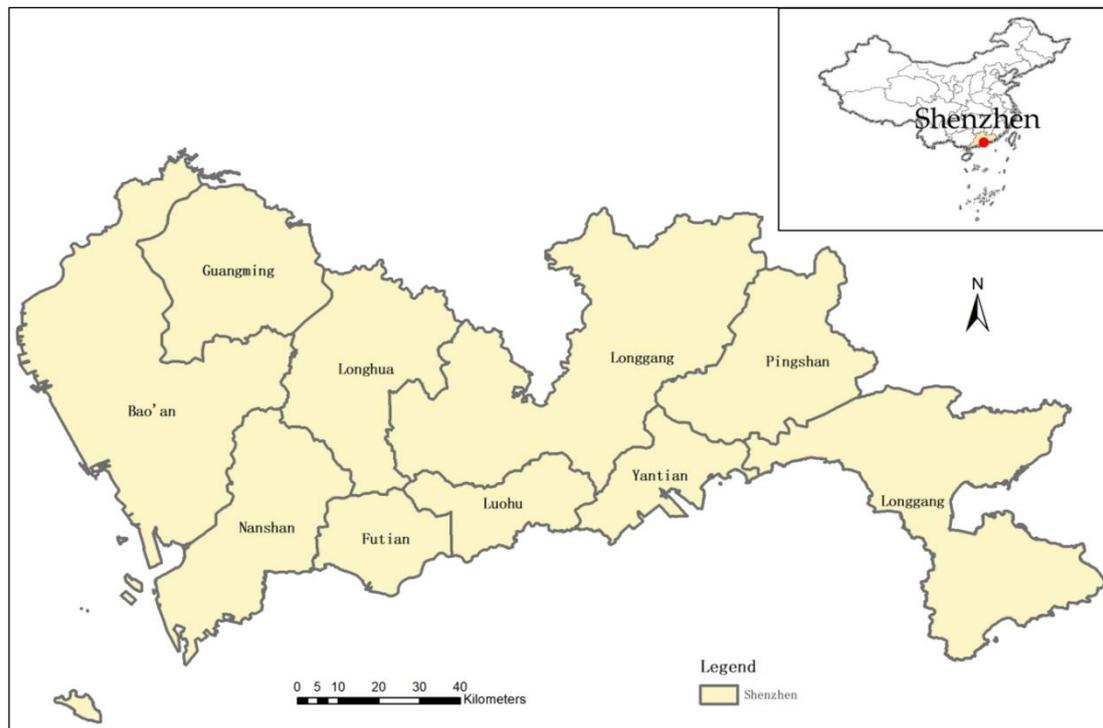


Figure 1. Location map of Shenzhen.

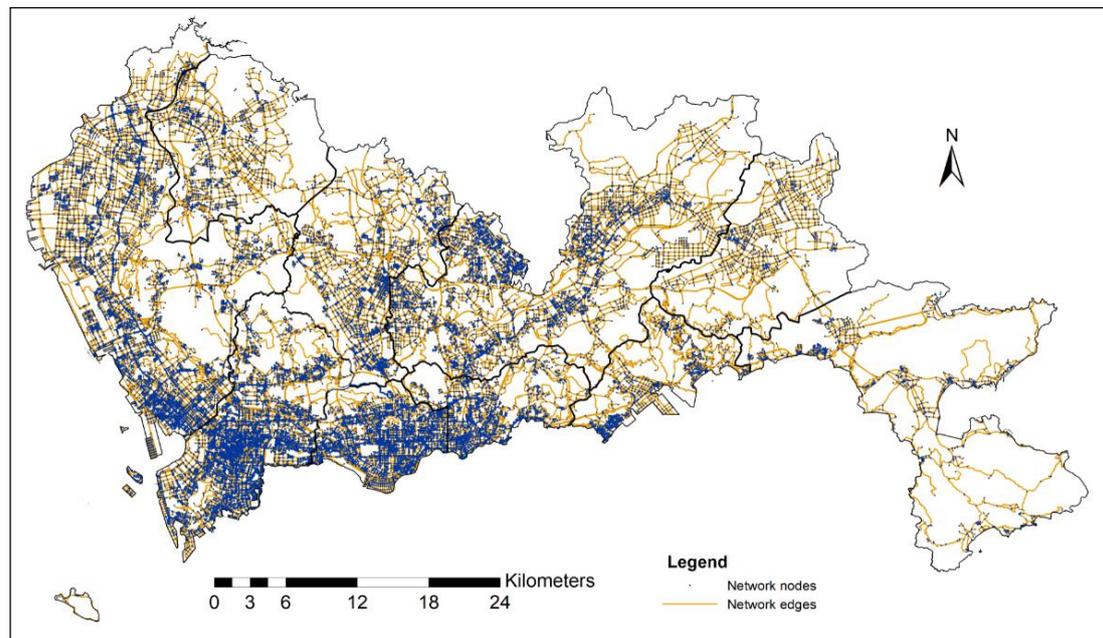


Figure 2. Digital representation of road network in Shenzhen.

3.2. Taxi Mobility Data

The taxi mobility data in Shenzhen is generated in this work from the STL dataset (<https://github.com/cbdog94/STL>) in [59]. Table 1 describes the metadata for this dataset. This taxi mobility dataset was collected in September 2009. The taxi mobility data used in this study mainly focused on the downtown area.

Table 1. Data description for Shenzhen taxi mobility data.

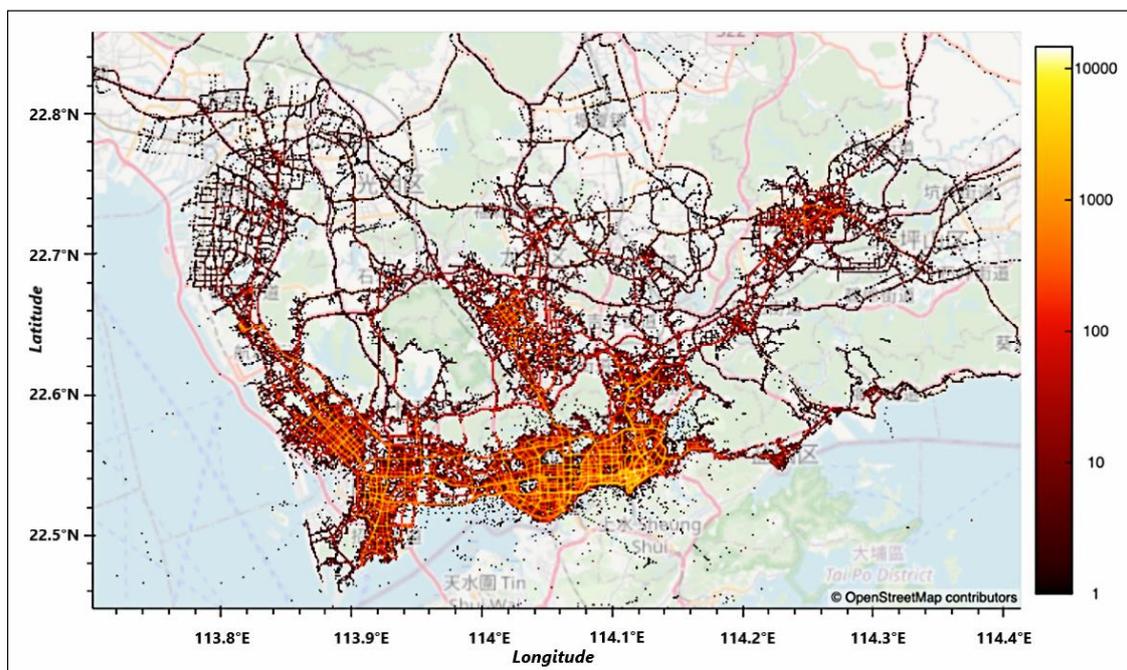
Attribute of the Dataset	Description of the Dataset	Information
Data size	35.16 GB	Total size of the taxi mobility data
Number of taxis	7475	Total number of taxis in the whole dataset
Number of Taxi Trajectories	6,068,516	Total number of taxi trajectories
Sampling rate	10–30 s	Interval time between two adjacent global positioning system (GPS) records

Each record for taxi movement contains the taxi ID, time, longitude, latitude, speed, direction, occupied status, and other information. Table 2 shows a description of each record for the Shenzhen taxi dataset.

Table 2. Description for one record in the Shenzhen Taxi dataset.

Field Name	Format	Sample Value
Taxi id	string	B001B1
Time	string	2009-09-23 21:39:45
Longitude	float number	114.06316
Latitude	float number	22.52787
Speed	integer (km/hour)	26
Direction	integer	0
Occupied	integer (1-with passenger, 0-without)	1
Other	integer number	0

Figure 3 shows the heat map of the taxi trajectories for the Shenzhen taxi dataset with passengers occupied. In this figure, the road network can be clearly observed. The taxi trajectories are aligned to road segments, and the study area has spatially varying color distributions.

**Figure 3.** Heat map of the taxi trajectories (with passengers) in Shenzhen.

3.3. Land-Use Data

Thirteen types are considered for the land-use data, including apartments, business, culture, education, facility, grassland, health, parking, sightseeing, sports, subway, transport, water and others (Figure 4). Shenzhen is one of the “sponge cities” (on a pilot list) in China. An important aspect is that the “sponge city” pilots contain plenty of grassland and water. The land-use types with grassland and water are also considered in this study. The apartment data were generated from the map-overlay result from the building dataset and facility layer. The land-use data were collected in 2014. It was obtained from Shenzhen municipal planning and land resources commission and Shenzhen municipal statistics bureau. According to [60], the construction land in the downtown area, especially the transportation land, has developed slowly since the year of 2009. In this area, the transportation land is mainly expanded to the periphery and surrounding area. Besides, as aforementioned, the taxi mobility data mainly focused on the downtown area, which has minor changes from 2009 to 2014 from both aspects of land-use types and transportation structures. Therefore, even the land-use data used in this study were collected in 2014, its transportation land, especially in the downtown area has a small change compared with that in 2009. It is therefore appropriate to use land-use data from 2014 and taxi trip data from 2009 for the taxi trip patterns analysis.

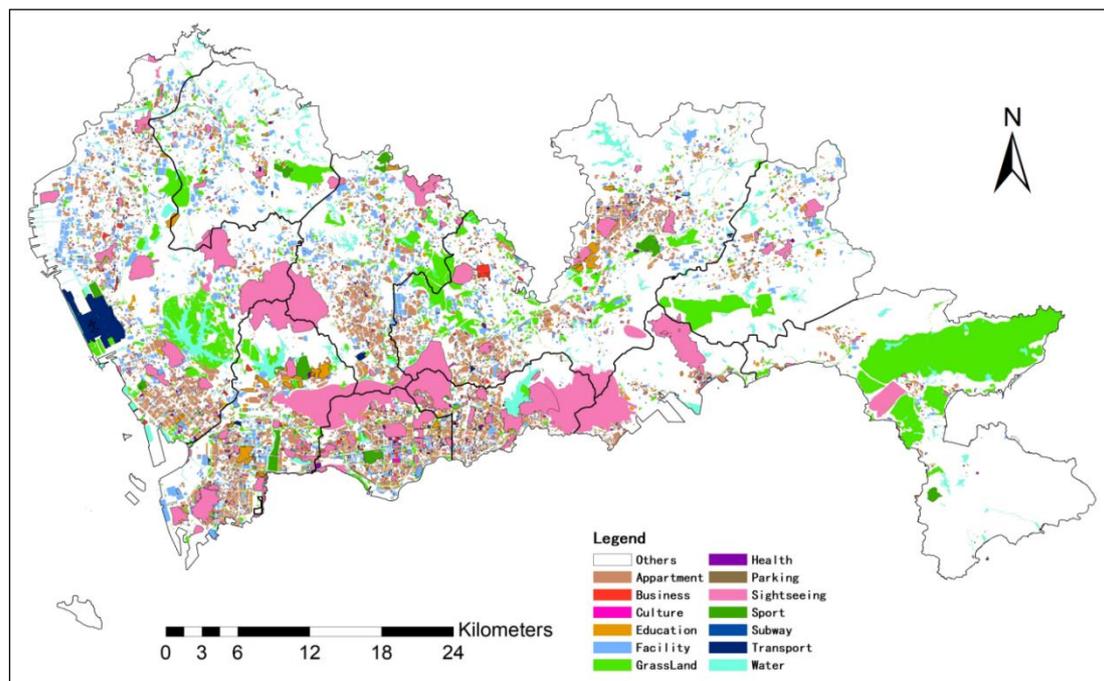


Figure 4. Land-use map in Shenzhen.

4. Methodology

In this section, we present the proposed methodology for identifying multiple mobility patterns in Shenzhen. For illustrative purpose, the workflow adopted for discovering spatial-temporal patterns is demonstrated in Figure 5. First, pick-up points (Os), drop-off points (Ds), OD-trips and time series data are obtained from taxi datasets. In this study, the spatial-temporal patterns of taxi trip are unfolded from three aspects: point patterns, trip patterns and network flow patterns. For point patterns, we propose an imNDKE method to estimate OD densities based on ODs and road network. In order to reveal the point patterns of taxi trip, OLS method is used to simulate the relationship between land-use and O density and D density, respectively. For trip patterns, related with the road network, we calculate the shortest-paths for OD trips. Next, to analyze the relationship between taxi trip and shortest-path, the coincidence rate (CR) of an OD trip and the corresponding shortest-path is calculated. Meanwhile,

the spatial distribution of “hot road” is also demonstrated. For network flow patterns, the network flow among different districts is obtained from the spatial and temporal analysis of Os, Ds, OD-trips and time series. We design a chord diagram to visualize flows and quantify flow data. Based on the methods mentioned above, the spatial and temporal patterns of taxi trip are analyzed from the point, trip and network flow patterns.

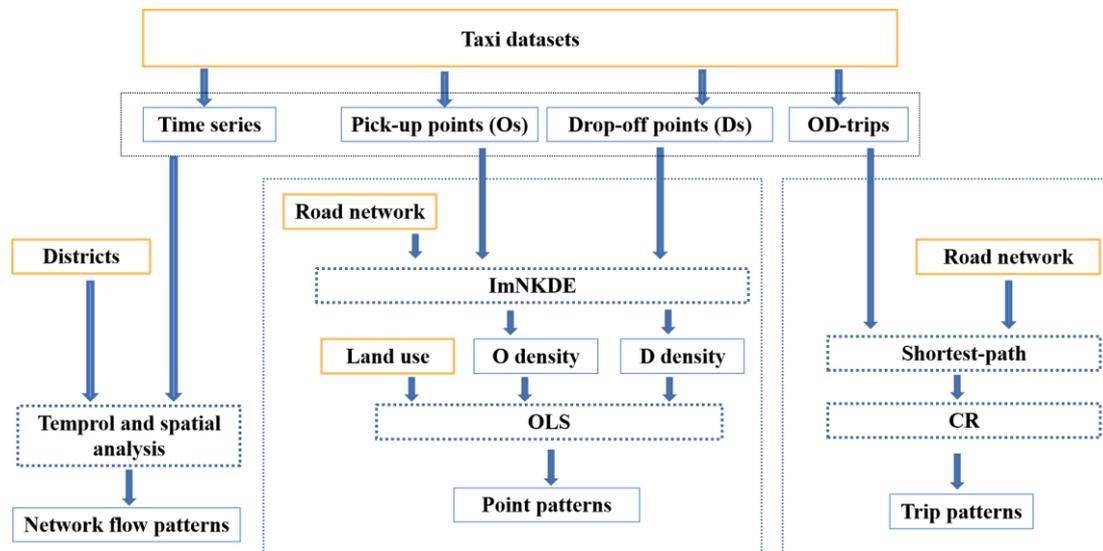


Figure 5. The workflow for discovering spatial-temporal patterns.

4.1. Preprocessing: Computing ODs and Map Matching

The taxi trip ODs are generated in the geo-processing step. The taxi data are cleaned and pre-processed by using spatial operations, in which when the occupied feature has an abnormal status (i.e., the value is neither 0 nor 1), or when the taxi stopped by the same location for more than 10 min, the data record is removed from consideration. Then, we perform the extraction of the taxi trip OD from the occupied status of each trajectory as follows:

- If the previous state for the taxi trajectory is not occupied (without passengers) and the current status of the taxi trajectory is occupied (with passengers), then the current taxi trajectory is the potential pick-up point (O).
- If the passenger status of the previous taxi trajectory is occupied and the current trajectory is not occupied at the next tracking point, then the previous track point is an alternative drop-off point (D).
- If the time interval between O and D is smaller than a minimum time t (5 min), this OD pair is ignored due to the reason that the time taken by the passenger is too short, and we assume that a single passenger occupied task cannot be completed within this period. With this in mind, the obtained taxi trip ODs are illustrated in Figure 6.

Then, we perform map matching operators for the taxi trajectories onto the road network, using a hidden Markov model (HMM) [61,62]. The efficient shortest path algorithm is frequently used in map-matching and taxi trip analytics along the road network. As the A* algorithm, one of the popular shortest path algorithms, performs very well in comparison with other methods in real-world road networks [63,64], we use a bidirectional A* Dijkstra algorithm with binary heap. Note that, if one record of the taxi trajectory does not have one map matching result, this record will be ignored.

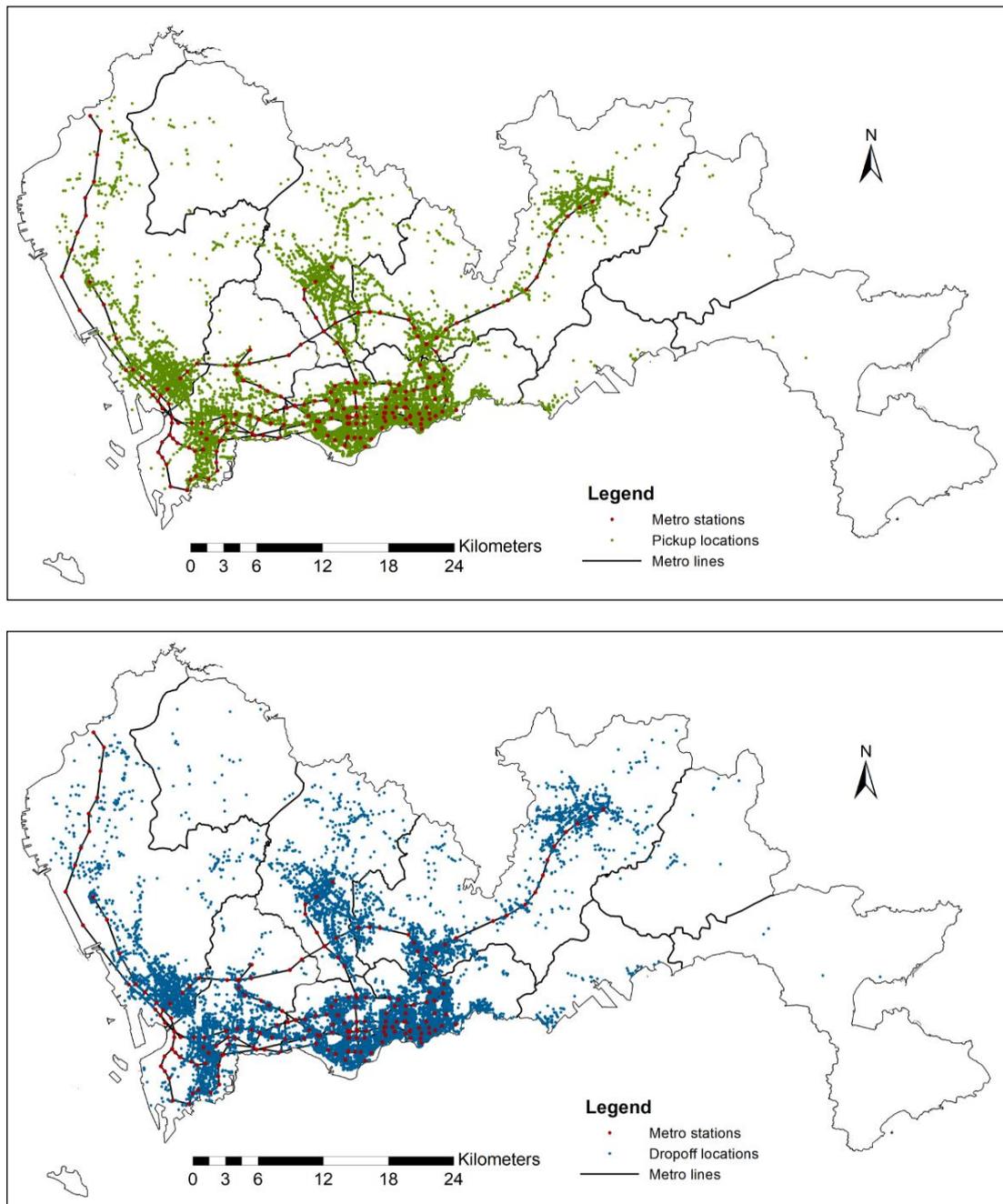


Figure 6. Taxi trip pick-up (Top) and taxi drop-off locations (Bottom).

4.2. An Improved Network Kernel Density Estimation (imNKDE)

In this section, we first briefly introduce the traditional network kernel density estimation (NKDE) method. Then, we detailed the estimation of the ODS' density by using our proposed imNKDE method, and the detection of the spatial taxi mobility patterns between the taxi data and land-use data along the road networks.

4.2.1. NKDE

The KDE (a nonparametric method) is used for data surface estimation [65,66]. The kernel function analysis is essentially a weighted distance method. It is based on the idea that the influence of distant point on the target grid is less than the influence of relatively close point. In the kernel function

estimation function, the factors affecting the kernel function estimation are the type of the kernel function and the radius of kernel function search. There are many types of kernel functions, such as Gaussian, uniform kernel, triangular kernel, and gamma kernel. The Gaussian function is the most commonly used in the literature. Xie and Yan argued that using different kernel functions has little effect on the result of the density estimation [67]. From the influence factors of the kernel estimation function, the search radius has a more significant influence on the kernel estimation result than the use of different kernel functions.

Based on KDE, NKDE is widely used to identify the hotspots and evaluate ODs along with the road network. The bandwidth for NKDE has important influence in the detection of the smoothness of the spatial patterns, in which it is observed that the narrow bandwidths (between 20 and 250 m) are more appropriate for identifying local effects at smaller scales [67]. Let $NKDE(x)$ be the network kernel density estimation in a point x . Then we can obtain $NKDE(x)$ as the sum of the different densities of the individual kernels for each point [68], as shown in Equation (1):

$$NKDE(x) = \sum_{i=1}^n K_{y_i}(x), \quad (1)$$

where y_i is the O/D point event, with $i = 1, 2, \dots, n$, and n is the total number of all O/D points. K_{y_i} is the kernel function of point y_i . $K_{y_i}(x)$ is the value at a point x of different kernels for each point, and $NKDE(x)$ is the value of NKDE at a point x . In this study, following previous research, we choose the Gaussian kernel. Notice that, the NKDE value at an O/D point of the road network presents the OD density.

4.2.2. ImNKDE

Our proposed improved NKDE (imNKDE) method aims at tackling the problem of large amounts of taxi datasets, so as to efficiently achieve the OD density. When the number of trajectories is huge, the procedure of querying each road edge and nearby taxi events is time-consuming and challenging for the NKDE method. Actually for the OD density estimation, in practice we compute the densities for the pick-up (O) events and drop-off (D) events individually which means that, for each location, its density estimation is independently estimated with regards to the other locations. Quite opposite, in our work we propose to estimate the OD densities of the subareas via NKDE, as it is observed that NKDE is very efficient when the number of trajectories is relatively small. More specifically, we divide the whole road network into different sub-networks and introduce a shared-memory parallel computing approach to estimate the OD densities for each subarea. Suppose we have n road network edges, we divide the road network edges into k parts (k equals to the number of threads). All threads simultaneously execute the density estimation via NKDE, where a shared memory parallel computing method (in which different threads share the spatial indices of taxi events) is proposed for efficient processing. After running all the computing tasks for each road edge and querying nearby taxi events, we aggregate the results as the final OD density. The parameters for the imNKDE method are set as follows. The search radius for the Gaussian kernel function is set to 200 m. The equal split method is chosen with ten segments over road network [26].

Figure 7 demonstrates the characteristic of NKDE and imNKDE methods, and the difference between them. As shown in Figure 7, for NKDE, the O and D events are calculated individually for the whole area. Due to the large amount of data, and the complex network structure, it is time-consuming for the estimation. To improve the calculation efficiency, we divide the road network into different segments, and estimate the O and D density at the same time for different segments in our proposed imNKDE. Specifically, in order to improve the efficiency and avoid the loss of spatial information while decomposing the road network, OD densities for each segments are calculated with a shared memory parallel computing method. Next, the OD densities of segments have been aggregated to obtain the final OD densities. In general, compared with the original NKDE, our proposed imNKDE method has

significantly improved the efficiency while achieving the same results for OD densities estimation with the traditional NKDE method.

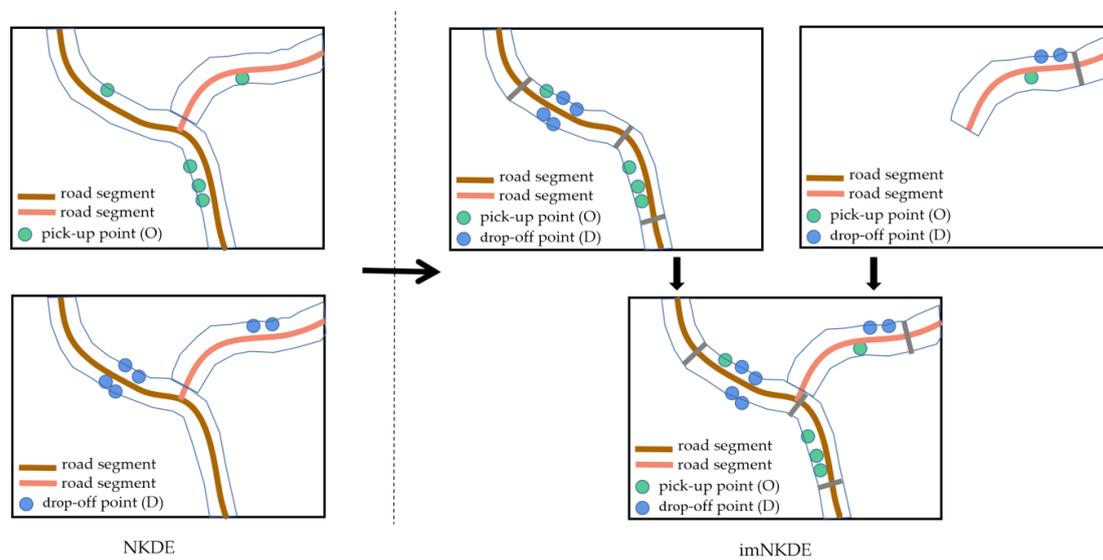


Figure 7. Schematic for imNKDE.

Finally, we use the proposed imNKDE to explore the relationship between taxi trip and land-use. Herein we consider the taxi trip OD density, and land-use data. Poisson regression is used to explore the relationship between the imNKDE and the land-use. For the relationship analysis, the imNKDE value is snapped into the 500 m * 500 m grid-based dataset. Then, the land-use type value is aggregated to the grid-based dataset. The Ordinary Least Squares (OLS) regression is a simple and effective multivariate regression method to estimate unknown factors for Poisson regression. The formulation for the OLS model is defined by Equation (2):

$$r = \beta_0 + \sum_{k=1}^m \beta_k x_k + \varepsilon, \quad (2)$$

where r is the dependent variable, i.e., the land-use type, β_0 is the intercept, m is the number of independent variables, x_k is the independent variable, β_k is the corresponding estimated coefficient, and ε is the residual. Here, we calculate the relationship between taxi trip OD density and different land-use types respectively. Therefore, for different land-use types, r is imNKDE value, i.e., O density or D density. As we only consider the land-use type as independent variable, here, $m = 1$, and x_1 is the land-use type, thus β_1 is used as the coefficient between land-use and imNKDE in the following parts.

If the p -value for the OLS result between the imNKDE and the land-use type is less than 0.05, then the variables are significant in affecting imNKDE in the regression. If the coefficient for the regression result is positive, the grids with high percentage of land-use values have high imNKDE.

4.3. The Taxi Trip Patterns in Shenzhen (Metropolis)

The taxi trips and routes are used to describe the trip patterns. In this study, we analyze the difference between actual taxi trip and the corresponding shortest path, which is calculated considering the distance. The coincidence rate (CR) is considered to describe the relationship between a taxi trip and a shortest path. Let $T = \{t_1, \dots, t_k, \dots, t_m\}$ be the taxi trip that is map-matched from the taxi trajectory, where t_k is the road segment for the taxi trip, m is the number of road segments for the taxi trip; let $S = \{s_1, \dots, s_i, \dots, s_n\}$ be the corresponding shortest path, where s_i is the road segment for the shortest path, n is the number of road segments for the shortest path; $P = S \cap T = \{p_1, \dots, p_j, \dots, p_{n_c}\}$

be the intersection of set S and T , with p_j being the road segment where passed by both the shortest path and the actual taxi trip, and n_c the number of these road segments. Then we have CR as follows:

$$CR(T, S) = \frac{\sum_{p_j \in P} L_{p_j}}{\sum_{s_i \in S} L_{s_i}}, \quad (3)$$

where L_{s_i} denotes the length of road segment s_i , and L_{p_j} denotes the length of road segment p_j . The value of CR ranges from 0 to 1. When the CR value is 1, it means that the taxi trip is the same as the shortest path. When the CR value is 0, it indicates that there is no overlapping between a taxi trip and the shortest path. Here we also use the OLS for analysis purposes. If the p-value for the OLS result between the value on road segments and the frequency of the taxi trips is less than 0.05, which means the route choice is strongly dependent on the properties of the road network, instead of the shortest path.

4.4. Network Flow Patterns

In this work, we use the proposed imNKDE with the network flow method, to detect the temporal patterns for taxi OD-trips among different regions. Taxi data are essential for understanding interaction patterns among different regions. The network flow among different districts can reveal the flow of taxi movement patterns. A chord diagram, representing network flows or connections between network nodes, has been designed for visualizing flows, using circular plots to quantify flow data [69,70]. Each component is represented by a fragment on the outer part of the circular layout. Then, network flow arcs are drawn between each pair of the network nodes. The width of the arc is proportional to the size of the network flow.

As the chord diagram can reveal the taxi trips' interactions among different regions, a network flow matrix is built among different districts and different land-use regions. The point-in-polygon operator is performed for each pair of taxis in the OD flow. Notice that, if a pick-up (O) point is located in region A and the drop-off (D) point is located in region B, then the item value of network flow matrix from region A to region B is increased by 1 unit.

5. Experiments and Results

This section presents experiments on the spatial pattern analysis from taxi trajectories in Shenzhen. We first investigate the relationship between the taxi data and land-use data. Then, the trip patterns of taxi trips and the shortest paths are analyzed, and the correlation between trips and road structure is revealed. Finally, network flow analytics for taxi data are studied for discovering spatial and temporal taxi trip ODs.

5.1. Point Pattern Analytics

5.1.1. Relationship between Metro Stations and Taxi ODs

The relationship between ODs and the metro stations reveals the trip purposes. The drop-off location of a trip (related to work, shopping, school, leisure and business) greatly affects the choices of trip mode and the trip travel behavior. Taking metro stations as an example, we discuss the spatial distance of the taxi pick-up events and taxi drop-off events (Table 3). The metro stations and lines in Shenzhen are shown in Figure 6. Table 4 shows the statistical results for only taxi pick-up events, only taxi drop-off events, the pair of pick-up events and drop-off events near the metro stations. The percentage for the pair of pick-up events and drop-off events is smaller than that of only pick-up events and only drop-off events. This is mainly because that when the origin and destination are both near the metro stations, people may choose the subway instead of taxi for travel. Besides, the numbers of the pick-up events are more than those of the drop-off events. The time of taxi data we used for

analysis is 21:39:45. In this situation, people are more likely to use taxi services after taking subways more than before taking subways.

Table 3. Statistics table of land-use types.

Types	Area (km ²)	Percentage
Apartment	289.0484	14.47%
Business	10.7002	0.54%
Culture	0.8329	0.04%
Education	29.6237	1.48%
Facility	213.3491	10.68%
Grassland	357.9826	17.92%
Health	2.4545	0.12%
Parking	0.0974	0.00%
Sightseeing	200.5232	10.04%
Sports	13.8547	0.69%
Subway	0.8973	0.04%
Transport	12.3457	0.62%
Water	0.1379	0.01%
others	865.6224	43.34%
total	1997.47	100.00%

Table 4. Percentage of taxi pick-up and drop-off events of ODs near metro stations.

Radius (Meters)	Only Pick-Up Events	Only Drop-Off Events	Pick-Up Events and Drop-Off Events
100	5.10%	2.24%	0.18%
200	10.29%	6.40%	1.58%
300	16.64%	11.82%	5.4%
400	22.77%	17.84%	11.13%
500	27.98%	23.03%	17.31%

5.1.2. Relationship between Taxi Trip OD Density and Land-Use

Figure 8 (top) shows the pick-up (O) locations and Figure 8 (bottom) demonstrates the drop-off (D) locations obtained by the proposed imNKDE method over the road network. The green color denotes the lower imNKDE value, while the red color denotes the higher imNKDE value. In this figure, we can observe that the results for the Os and Ds over the road network share similar patterns in the city-scale. The highest two regions are Futian and Luohu Districts.

Table 5 shows the correlation results of OLS between the OD density and land-use. There is significant correlation between taxi trip ODs and land-use types of business, facility, grassland, health, subway, and transport. This indicates that there are more OD events occurring in these regions. Furthermore, a significant negative correlation can be observed between the ODs and the land-use with apartment, parking, sightseeing, sports and water. Finally, it can be observed that, for the class of education, there is significant correlation between the drop-off events and the education class, while the correlations between the pick-up events and the education class is the other way around. This means there are more drop-off events than pick-up events in the education area. For the sightseeing region, there is a significant negative correlation between the drop-off events and sightseeing class, while for the pick-up events, the correlation is still negative however not significant. This means that there are few drop-off events in sightseeing regions. This is because that the time of the taxi data was collected at 21:39:45, when people may not go sightseeing at this time.

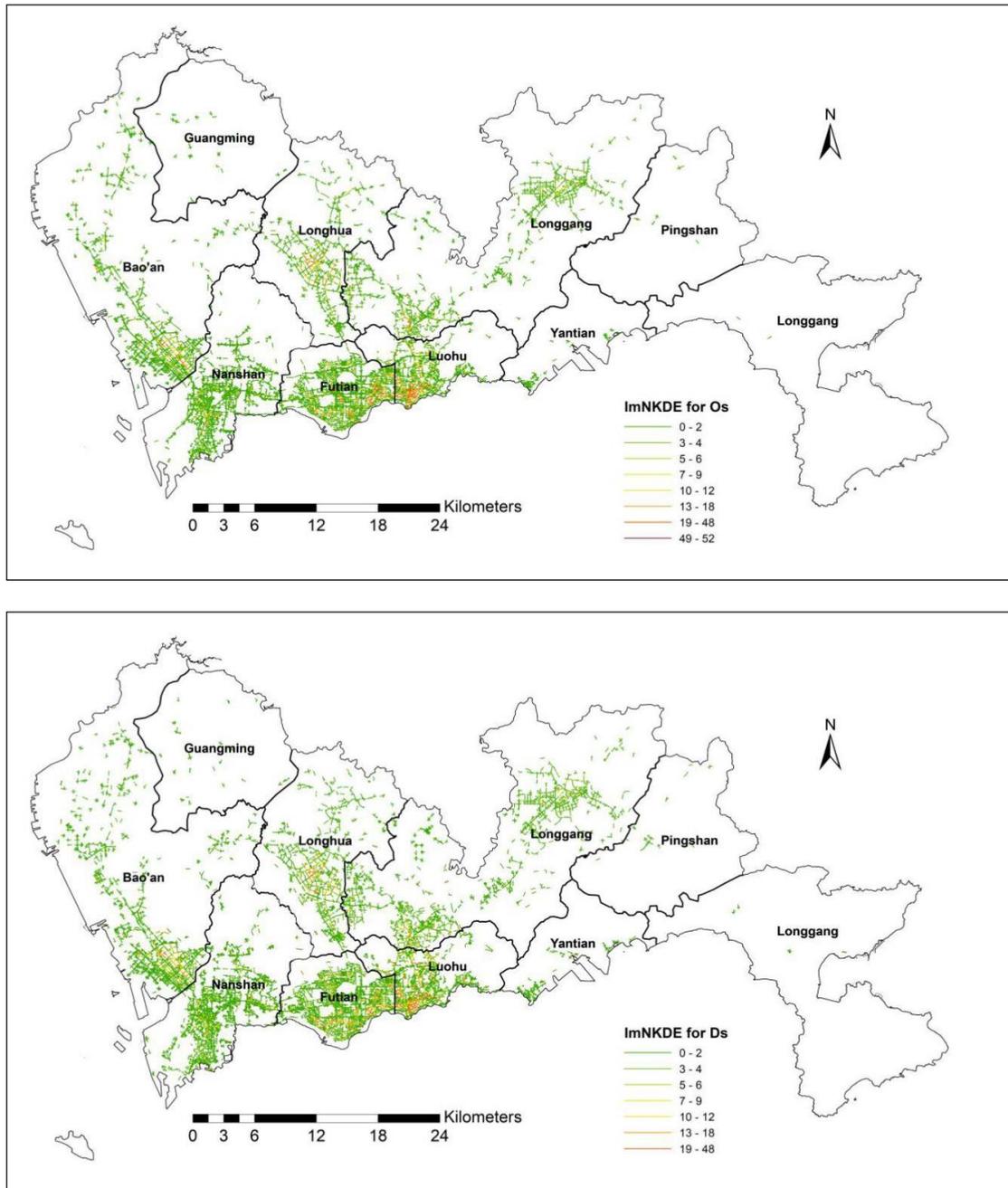


Figure 8. ImNKDE for Os (Top) and Ds (Bottom).

Table 5. Correlation results between ODs and land-use.

Types	Pick-Up Events		Drop-Off Events	
	Coefficient	<i>p</i> -Value	Coefficient	<i>p</i> -Value
Apartment	-9.527×10^{-11}	<0.05	-6.556×10^{-11}	<0.05
Business	1.958×10^{-5}	<0.05	1.471×10^{-5}	<0.05
Culture	7.877×10^{-6}	0.105	8.882×10^{-7}	0.852
Education	5.212×10^{-7}	0.619	3.745×10^{-6}	<0.05
Facility	9.672×10^{-6}	<0.05	1.114×10^{-5}	<0.05
Grassland	2.838×10^{-13}	<0.05	3.543×10^{-13}	<0.05
Health	2.235×10^{-5}	<0.05	2.844×10^{-5}	<0.05
Parking	-0.0001	0.107	-7.783×10^{-5}	0.528
Sightseeing	-7.093×10^{-9}	0.991	-2.071×10^{-6}	<0.05
Sports	-3.282×10^{-6}	<0.05	-3.333×10^{-6}	<0.05
Subway	0.0002	<0.05	0.0002	<0.05
Transport	1.327×10^{-5}	<0.05	1.079×10^{-5}	<0.05
Water	-5.584×10^{-6}	<0.05	-3.374×10^{-6}	<0.05

5.2. Trip Patterns for Taxi Trajectory Analytics

Figure 9 shows the hot road segments in taxi trip and shortest paths. Here, “hot road” means that the road is used by more taxi trajectories. In Figure 9, hot roads index represents the number of taxi trajectories or shortest paths using this road. It can be observed that the maximum value of hot roads in taxi trips is larger than that in the shortest paths. This indicates that more diverse patterns exist in taxi trips than that in the shortest paths. This is expected, as (in reality) drivers are likely to choose a route with easy access, instead of the shortest path. Furthermore, the hot road segments for taxi trips and shortest paths exhibit similar distributions in the city-scale. Notice that there are two regions with a high number of hot roads in both maps: Futian District and Luohu District.

In order to further investigate the road structure that affect the taxi driver’s choices for taxi trips, we perform an OLS analysis on the road structure and hot roads. Concerning the road structure, in this work we consider two main factors, i.e., the number of road lanes and the road functional classes. The obtained results are shown in Table 6. It can be observed that there is significant correlation between the number of road lanes and hot roads (the coefficient is 0.0439) with *p*-value smaller than 0.05. This means that a road with more lanes is more likely to be a hot road, i.e., a road with more taxi trips. This is reasonable, as (in reality) drivers favor roads with more lanes in any case. Furthermore, a negative correlation is observed between the road functional classes and the hot roads (the coefficient is -0.0405), which indicates that hot roads are more likely to happen in low road functional classes, i.e., highways. In summary, we can conclude that taxi drivers prefer roads with more lanes or highways. Normally, roads with more lanes and highways mean that there are less traffic jams, thus reducing the time spent for a trip. Drivers are looking for the shortest path with less time.

Table 6. Correlations results between the road structure and hot roads.

Factor Variable	Coefficient	<i>p</i> -Value
Number of road lanes	0.0439	<0.05
Road functional classes	-0.0405	<0.05

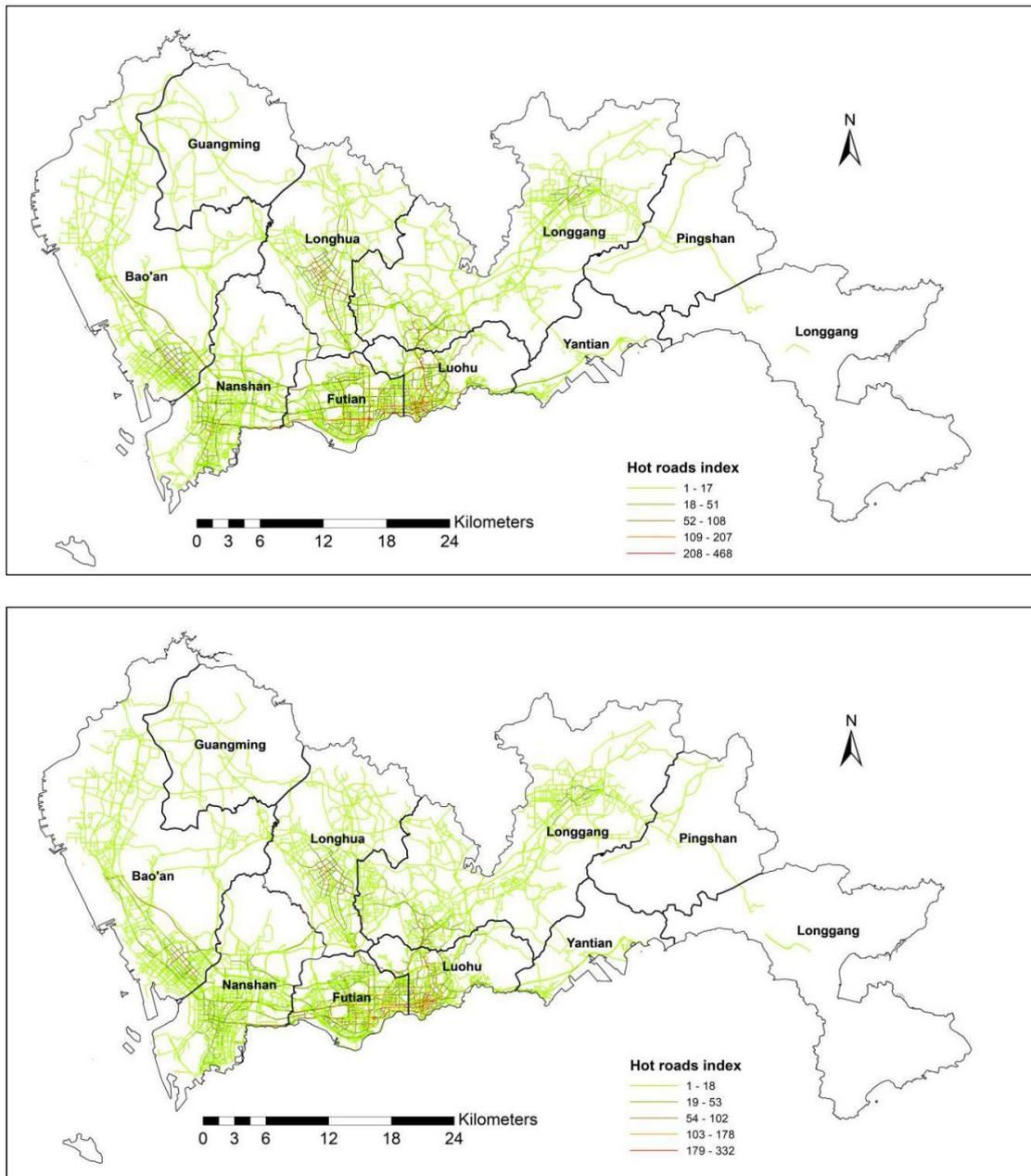
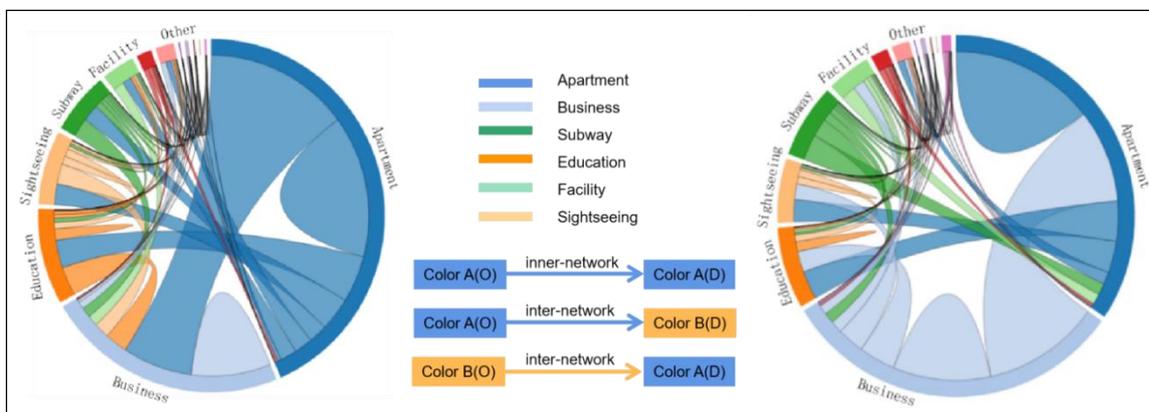


Figure 9. Hot roads in taxi trips (top) and shortest paths (bottom).

5.3. Network Flow Patterns between Taxi Trip ODs and Land-Use Types

The network flow patterns, from a spatial viewpoint (i.e., land-use types, along with the temporal characteristics) are investigated in this study. For the taxi trip ODs, and according to time series, they are divided into morning rush hours (07:00–09:00) and evening rush hours (17:00–19:00) on Weekdays (Monday, Tuesday, Wednesday, Thursday and Friday) and weekends (Saturday and Sunday). In Figures 10 and 11, different color of the circle in the outside means different land-use type. The arc means the inter-network trip, which means the O and D are in different land-use type regions, or the inner-network trip, which means the O and D are in regions with the same land-use type. For example, in Figures 10 and 11, the orange part of the circle represents the education region. First, the orange arc in the orange part means that the taxi trip is from education region to education region. Second, the orange arc between orange part and other parts means that the taxi trip is from education region to other regions. For example, the orange arc between orange part and light blue part means that people

travel from education region to business region. Third, the arc in other colors between orange part and other parts means that the taxi trip is from other regions to education region. Such as the blue arc between orange part and blue part means people travel from apartment region to education part. In Figure 10, the network flow of taxi trip ODs on weekdays, and Figure 11 illustrates the network flow of taxi trip ODs on weekends, respectively. The results indicate that the top six land-use types OD activities are apartment, business, education, sightseeing, subway, and facility. In the morning rush hours on weekdays, people tend to go out of the apartment and enter in the business area by taxi. It is interesting to derive the same observation for weekends. This is somehow understandable, as Shenzhen is a very busy city and many people work on weekends. In the evening rush hours, the most abundant network flow is from the business region to the apartment region on both weekdays and weekends. This is expected, since people are back to home after work. It should be noted that the evening rush hour on weekends shares very similar patterns as the evening rush hour on weekdays. However, in the morning rush hours, there are some differences among different regions in the network flows. First of all, in the subway region (on weekdays), the network flow is only from apartment to subway, while (on weekends) it is the other way around, as there are only outgoing network flows on weekends. This might be due to the fact that, on weekends, people at apartments may receive visitors who are likely choose metro as part of their trip. Furthermore, in the education region, on weekdays, there is a network flow from education to business. This most likely happens because parents send kids to the school first and then go to their business. As it can be observed, there is a network flow from apartment to education. However, on weekends, there is still a network flow from apartment to education. This is because parents very often take kids to training schools on weekends. However, there is no network flow from education to business. Instead, there is a network flow from business to education. This is most likely because parents go to pick up kids after training from work.



(a) Morning rush hours (07:00-09:00)

(b) Evening rush hours (17:00-19:00)

Figure 10. Network flow of taxi trips in Shenzhen City on weekdays.

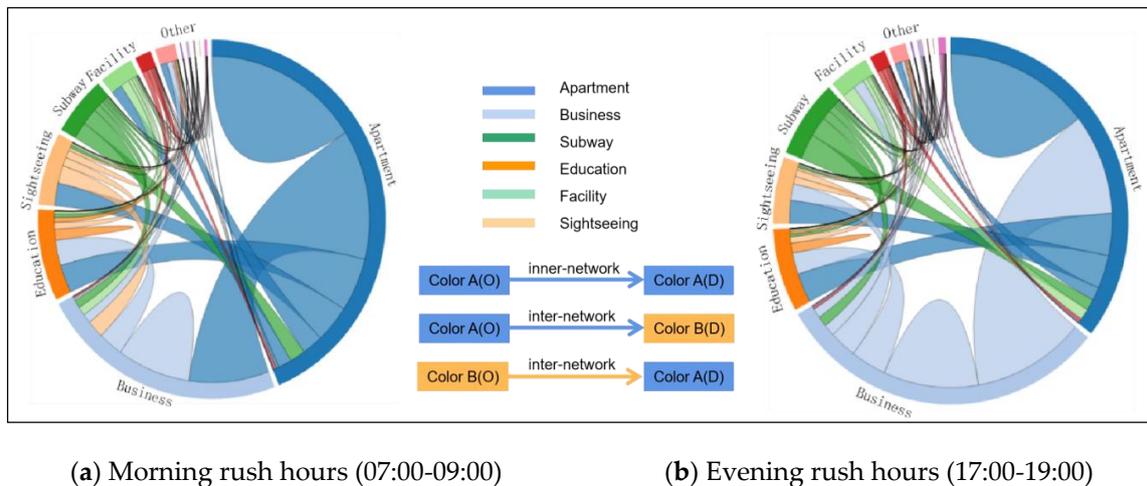


Figure 11. Network flow of taxi OD trips in Shenzhen City on weekends.

6. Conclusions and Future Work

The better understanding of taxi trip patterns is important for transport planning optimization. First, with the country's vigorous promotion of new energy fields and a series of policy support, there are more and more new energy vehicles. It is very important to choose the appropriate sites for charging stations. The spatial distribution of OD densities based on imNKDE could help us know the main area of taxi trip, which could provide reference for the site selection. Meanwhile, the knowledge of ODs densities could guide the setting of stops for taxis. The proper setting of taxi stops can effectively reduce traffic jams and enhance the city image. Second, due to some reasons, such as the traffic jams, road maintenance and others, there are many taxi trips do not follow the shortest path rule. Therefore, with the relationship analysis between taxi trip and shortest path, the decision makers could know the possible deficiencies in the current transportation network, and design new strategies to improve the traffic network and promote the development of traffic. Last but not least, the network flow pattern of taxi trip could reflect the relationship between taxi trip and land-use, it can provide a guidance for traffic planning in other cities.

In this paper, we have developed a new method to identify the spatial variety in travel patterns from taxi mobility data and land-use. A main innovation of our method is that it considers the point, trip and network aspects simultaneously, as opposed to other existing methods. Another important contribution of our work is the development of a new improved network kernel density estimation (imNKDE) algorithm, with the capability of efficiently estimating the density of OD pairs from massive taxi trajectory data, to identify the spatial patterns from such density and land-use data. We also introduce the use of trip patterns to measure the relationship between taxi trips and taxi routes based on their similarities. Last but not least, another important contribution of this work is the incorporation of network flow pattern analytics to model the differences between taxi trip ODs and land-use data.

Our experimental results, conducted using a case study in the city of Shenzhen, China (including taxi mobility data, land-use data and road network data), demonstrate that our newly developed method can process large amounts of taxi trajectories and accurately identify network flow patterns, which are further exploited to discover the spatial interactions among different districts and land-use areas. Our spatial-temporal analytics is multi-source data-driven since there are more cities can be applicable using our methods. Although our method uses an efficient shared memory parallel implementation to estimate the OD densities for each subarea, in the future we will develop a graphics processing unit (GPU)-based implementation that will accelerate the processing of large amounts of data even more, aiming at real-time processing of trajectory data, which may greatly assist in traffic

monitoring and control. Meanwhile, we will further analyze the relationship between taxi trip patterns and land-use and provide a reference for traffic planning.

Author Contributions: Conceptualization, Jun Li and Boxi Shen; methodology, Boxi Shen and Qunying Huang; software, Boxi Shen; validation, Boxi Shen, Xiang Xu and Qunying Huang; formal analysis, Boxi Shen and Antonio Plaza; investigation, Boxi Shen and Qunying Huang; resources, Boxi Shen; data curation, Boxi Shen; writing—original draft preparation, Boxi Shen; writing—review and editing, Boxi Shen, Antonio Plaza, Xiang Xu and Qunying Huang; visualization, Boxi Shen; supervision, Jun Li and Xiang Xu; project administration, Boxi Shen; funding acquisition, Jun Li. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Jiangxi Provincial Natural Science Foundation under Grant 20192BAB217003, by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19090104) and the National Key Research and Development Program of China under Grant No.2017YFB0502900.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zhang, D.; Zhao, J.; Zhang, F.; He, T. Urbancps: A cyber-physical system based on multi-source big infrastructure data for heterogeneous model integration. In Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems, Seattle, WA, USA, 14–16 April 2015; ACM: New York, NY, USA, 2015; pp. 238–247.
- Zhang, D.; Zhao, J.; Zhang, F.; He, T.; Lee, H.; Son, S.H. Heterogeneous model integration for multi-source urban infrastructure data. *ACM Trans. Cyber-Phys. Syst.* **2017**, *1*, 4. [[CrossRef](#)]
- Yuan, J.; Zheng, Y.; Zhang, C.; Xie, W.; Xie, X.; Sun, G.; Huang, Y. T-drive: Driving directions based on taxi trajectories. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; ACM: New York, NY, USA, 2010; pp. 99–108.
- Liu, D.; Weng, D.; Li, Y.; Bao, J.; Zheng, Y.; Qu, H.; Wu, Y. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Trans. Vis. Comput. Graph.* **2016**, *23*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Garg, N.; Ramadurai, G.; Ranu, S. Mining bus stops from raw gps data of bus trajectories. In Proceedings of the 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, India, 3–7 January 2018; pp. 583–588.
- Seaborn, C.; Attanucci, J.; Wilson, N.H. Analyzing multimodal public transport journeys in london with smart card fare payment data. *Transp. Res. Rec.* **2009**, *2121*, 55–62. [[CrossRef](#)]
- Lin, D.; Zhang, Y.; Zhu, R.; Meng, L. The analysis of catchment areas of metro stations using trajectory data generated by dockless shared bikes. *Sustain. Cities Soc.* **2019**, *49*, 101598. [[CrossRef](#)]
- He, T.; Bao, J.; Li, R.; Ruan, S.; Li, Y.; Tian, C.; Zheng, Y. Detecting vehicle illegal parking events using sharing bikes' trajectories. In Proceedings of the Sigkdd International Conference, London, UK, 19–23 August 2018; ACM: New York, NY, USA, 2018; pp. 340–349.
- Zhou, Z.; Yu, J.; Guo, Z.; Liu, Y. Visual exploration of urban functions via spatio-temporal taxi OD data. *J. Vis. Lang. Comput.* **2018**, *48*, 169–177. [[CrossRef](#)]
- Shi, J.; Tao, L.; Li, X.; Xiao, Y.; Atchley, P. A survey of taxi drivers' aberrant driving behavior in Beijing. *J. Transp. Saf. Secur.* **2014**, *6*, 34–43. [[CrossRef](#)]
- Chen, X.M.; Zahiri, M.; Zhang, S. Understanding ride splitting behavior of on-demand ride services: An ensemble learning approach. *Transp. Res. Part C Emerg. Technol.* **2017**, *76*, 51–70. [[CrossRef](#)]
- Wang, P.; Fu, Y.; Liu, G.; Hu, W.; Aggarwal, C. Human mobility synchronization and trip purpose detection with mixture of hawkes processes. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; ACM: New York, NY, USA, 2017; pp. 495–503.
- Rose, J.M.; Hensher, D.A. Demand for taxi services: New elasticity evidence. *Transportation* **2014**, *41*, 717–743. [[CrossRef](#)]
- Ferreira, N.; Poco, J.; Vo, H.T.; Freire, J.; Silva, C.T. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2149–2158. [[CrossRef](#)]

15. Shen, J.; Liu, X.; Chen, M. Discovering spatial and temporal patterns from taxi-based floating car data: A case study from nanjing. *Gisci. Remote Sens.* **2017**, *54*, 617–638. [[CrossRef](#)]
16. Chen, C.; Zhang, D.; Ma, X.; Guo, B.; Wang, L.; Wang, Y.; Sha, E. Crowddeliver: Planning city-wide package delivery paths leveraging the crowd of taxis. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1478–1496. [[CrossRef](#)]
17. Cai, H.; Jia, X.; Chiu, A.S.; Hu, X.; Xu, M. Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet. *Transp. Res. Part D Transp. Environ.* **2014**, *33*, 39–46. [[CrossRef](#)]
18. Guo, D.; Zhu, X.; Jin, H.; Gao, P.; Andris, C. Discovering spatial patterns in origin-destination mobility data. *Trans. GIS* **2012**, *16*, 411–429. [[CrossRef](#)]
19. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.-L.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [[CrossRef](#)]
20. Boarnet, M.; Crane, R. The influence of land use on travel behavior: Specification and estimation strategies. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 823–845. [[CrossRef](#)]
21. Liu, X.; Gong, L.; Gong, Y.; Liu, Y. Revealing travel patterns and city structure with taxi trip data. *J. Transp. Geogr.* **2015**, *43*, 78–90. [[CrossRef](#)]
22. Liu, Y.; Kang, C.; Gao, S.; Xiao, Y.; Tian, Y. Understanding intra-urban trip patterns from taxi trajectory data. *J. Geogr. Syst.* **2012**, *14*, 463–483. [[CrossRef](#)]
23. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-use classification using taxi gps traces. *IEEE Trans. Intell. Transp. Syst.* **2012**, *14*, 113–123. [[CrossRef](#)]
24. Liu, X.; Kang, C.; Gong, L.; Liu, Y. Incorporating spatial interaction patterns in classifying and understanding urban land use. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 334–350. [[CrossRef](#)]
25. Ge, P.; He, J.; Zhang, S.; Zhang, L.; She, J. An integrated framework combining multiple human activity features for land use classification. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 90. [[CrossRef](#)]
26. Okabe, A.; Satoh, T.; Sugihara, K. A kernel density estimation method for networks, its computational method and a gis-based tool. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 7–32. [[CrossRef](#)]
27. Xie, Z.; Yan, J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach. *J. Transp. Geogr.* **2013**, *31*, 64–71. [[CrossRef](#)]
28. Timothée, P.; Nicolas, L.-B.; Emanuele, S.; Sergio, P.; Stéphane, J. A network based kernel density estimator applied to barcelona economic activities. In *International Conference on Computational Science and its Applications*; Springer: Berlin, Germany, 2010; pp. 32–45.
29. Yu, W.; Ai, T.; Shao, S. The analysis and delimitation of central business district using network kernel density estimation. *J. Transp. Geogr.* **2015**, *45*, 32–47. [[CrossRef](#)]
30. Li, Q.; Zhang, T.; Wang, H.; Zeng, Z. Dynamic accessibility mapping using floating car data: A network-constrained density estimation approach. *J. Transp. Geogr.* **2011**, *19*, 379–393. [[CrossRef](#)]
31. Tang, J.; Liu, F.; Wang, Y.; Wang, H. Uncovering urban human mobility from large scale taxi gps data. *Phys. A Stat. Mech. Its Appl.* **2015**, *438*, 140–153. [[CrossRef](#)]
32. Wardrop, J.G. Road paper. Some theoretical aspects of road traffic research. *Proc. Inst. Civ. Eng.* **1952**, *1*, 325–362. [[CrossRef](#)]
33. Vrotsou, K.; Fuchs, G.; Andrienko, N.; Andrienko, G. An interactive approach for exploration of flows through direction-based filtering. *J. Geovis. Spat. Anal.* **2017**, *1*, 1. [[CrossRef](#)]
34. Golledge, R.G. Path selection and route preference in human navigation: A progress report. In *International Conference on Spatial Information Theory*; Springer: Berlin, Germany, 1995; pp. 207–222.
35. Goldman, T.; Gorham, R. Sustainable urban transport: Four innovative directions. *Technol. Soc.* **2006**, *28*, 261–273. [[CrossRef](#)]
36. Mallus, M.; Colistra, G.; Atzori, L.; Murrioni, M.; Pilloni, V. Dynamic carpooling in urban areas: Design and experimentation with a multi-objective route matching algorithm. *Sustainability* **2017**, *9*, 254. [[CrossRef](#)]
37. Hodson, M.; Geels, F.W.; McMeekin, A. Reconfiguring urban sustainability transitions, analysing multiplicity. *Sustainability* **2017**, *9*, 299. [[CrossRef](#)]
38. Zhou, Y.; Fang, Z.; Thill, J.-C.; Li, Q.; Li, Y. Functionally critical locations in an urban transportation network: Identification and space-time analysis using taxi trajectories. *Comput. Environ. Urban Syst.* **2015**, *52*, 34–47. [[CrossRef](#)]

39. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and pois. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; ACM: New York, NY, USA, 2012; pp. 186–194.
40. Castro, P.S.; Zhang, D.; Chen, C.; Li, S.; Pan, G. From taxi gps traces to social and community dynamics: A survey. *ACM Comput. Surv.* **2013**, *46*, 17. [[CrossRef](#)]
41. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering urban functional zones using latent activity trajectories. *IEEE Trans. Knowl. Data Eng.* **2014**, *27*, 712–725. [[CrossRef](#)]
42. Qi, G.; Li, X.; Li, S.; Pan, G.; Wang, Z.; Zhang, D. Measuring social functions of city regions from large-scale taxi behaviors. In Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, 21–25 March 2011; pp. 384–388.
43. Zhao, P.; Liu, X.; Shen, J.; Chen, M. A network distance and graph-partitioning-based clustering method for improving the accuracy of urban hotspot detection. *Geocarto Int.* **2019**, *34*, 293–315. [[CrossRef](#)]
44. Tang, L.; Kan, Z.; Zhang, X.; Sun, F.; Yang, X.; Li, Q. A network kernel density estimation for linear features in space–time analysis of big trace data. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1717–1737. [[CrossRef](#)]
45. Xia, Z.; Li, H.; Chen, Y.; Liao, W. Identify and delimitate urban hotspot areas using a network-based spatiotemporal field clustering method. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 344. [[CrossRef](#)]
46. Delso, J.; Martín, B.; Ortega, E.; Van De Weghe, N. Integrating pedestrian-habitat models and network kernel density estimations to measure street pedestrian suitability. *Sustain. Cities Soc.* **2019**, *51*, 101736. [[CrossRef](#)]
47. Antipova, A.; Wang, F.; Wilmot, C. Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Appl. Geogr.* **2011**, *31*, 1010–1018. [[CrossRef](#)]
48. Liu, Y.; Wang, F.; Xiao, Y.; Gao, S. Urban land uses and traffic ‘source-sink areas’: Evidence from gps-enabled taxi data in shanghai. *Landsc. Urban Plan.* **2012**, *106*, 73–87. [[CrossRef](#)]
49. Yao, E.J.; Pan, L.; Yang, Y.; Zhang, Y.S. Taxi driver’s route choice behavior analysis based on floating car data. In *Applied Mechanics and Materials*; Trans Tech Publications: Stafa-Zurich, Switzerland, 2013; pp. 2036–2039.
50. Sun, D.; Zhang, C.; Zhang, L.; Chen, F.; Peng, Z.-R. Urban travel behavior analyses and route prediction based on floating car data. *Transp. Lett.* **2014**, *6*, 118–125. [[CrossRef](#)]
51. Li, J. *Do Taxi Drivers Choose the Shortest Routes?* Technische Universität München: Munich, Germany, 2017.
52. Tang, J.; Jiang, H.; Li, Z.; Li, M.; Liu, F.; Wang, Y. A two-layer model for taxi customer searching behaviors using gps trajectory data. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3318–3324. [[CrossRef](#)]
53. Veloso, M.; Phithakkitnukoon, S.; Bento, C. Sensing urban mobility with taxi flow. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Chicago, IL, USA, 1 November 2011; ACM: New York, NY, USA, 2011; pp. 41–44.
54. Veloso, M.; Phithakkitnukoon, S.; Bento, C. Urban mobility study using taxi traces. In Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis, Beijing, China, 18 September 2011; ACM: New York, NY, USA, 2011; pp. 23–30.
55. Yang, Y.; Dwyer, T.; Jenny, B.; Marriott, K.; Cordeil, M.; Chen, H. Origin-destination flow maps in immersive environments. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 693–703. [[CrossRef](#)] [[PubMed](#)]
56. Lu, M.; Liang, J.; Wang, Z.; Yuan, X. Exploring od patterns of interested region based on taxi trajectories. *J. Vis.* **2016**, *19*, 811–821. [[CrossRef](#)]
57. Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **2017**, *65*, 126–139. [[CrossRef](#)]
58. Boeing, G. OSMnx: A python package to work with graph-theoretic OpenStreetMap street networks. *J. Open Source Softw.* **2017**, *2*, 1–4. [[CrossRef](#)]
59. Cheng, B.; Qian, S.; Cao, J.; Xue, G.; Yu, J.; Zhu, Y.; Li, M.; Zhang, T. STL: Online detection of taxi trajectory anomaly based on spatial-temporal laws. In *International Conference on Database Systems for Advanced Applications*; Springer: Berlin, Germany, 2019; pp. 764–779.
60. Li, G.; Zhang, J.; Zhang, J.; Jin, X. Temporal and spatial analysis of land-use structure in Shenzhen. *Geomat. Spat. Inf. Technol.* **2017**, *11*, 26–30.
61. Newson, P.; Krumm, J. Hidden markov map matching through noise and sparseness. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; ACM: New York, NY, USA, 2009; pp. 336–343.
62. Yang, C.; Gidofalvi, G. Fast map matching, an algorithm integrating hidden markov model with precomputation. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 547–570. [[CrossRef](#)]

63. Zeng, W.; Church, R.L. Finding shortest paths on real road networks: The case for A*. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 531–543. [[CrossRef](#)]
64. Wang, S.; Gao, S.; Feng, X.; Murray, A.T.; Zeng, Y. A context-based geoprocessing framework for optimizing meetup location of multiple moving objects along road networks. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1368–1390. [[CrossRef](#)]
65. Anderson, T.K. Kernel density estimation and k-means clustering to profile road accident hotspots. *Accid. Anal. Prev.* **2009**, *41*, 359–364. [[CrossRef](#)]
66. Borruso, G. Network density estimation: A GIS approach for analysing point patterns in a network space. *Trans. GIS* **2008**, *12*, 377–402. [[CrossRef](#)]
67. Xie, Z.; Yan, J. Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst.* **2008**, *32*, 396–406. [[CrossRef](#)]
68. Delso, J.; Martín, B.; Ortega, E. A new procedure using network analysis and kernel density estimations to evaluate the effect of urban configurations on pedestrian mobility. The case study of vitoria–gasteiz. *J. Transp. Geogr.* **2018**, *67*, 61–72. [[CrossRef](#)]
69. Abel, G.J.; Sander, N. Quantifying global international migration flows. *Science* **2014**, *343*, 1520–1522. [[CrossRef](#)] [[PubMed](#)]
70. Sorichetta, A.; Bird, T.J.; Ruktanonchai, N.W.; Zu Erbach-Schoenberg, E.; Pezzulo, C.; Tejedor, N.; Waldock, I.C.; Sadler, J.D.; Garcia, A.J.; Sedda, L. Mapping internal connectivity through human migration in malaria endemic countries. *Sci. Data* **2016**, *3*, 160066. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).