Article

Estimating Potential Demand of Bicycle Trips from Mobile Phone Data—An Anchor-Point Based Approach

Yang Xu [1,2], Shih-Lung Shaw [1,3,4,*], Zhixiang Fang [3,4] and Ling Yin [5]

[1] Department of Geography, University of Tennessee, Knoxville, TN 37996, USA; yxu30@utk.edu
[2] Senseable City Laboratory, SMART Centre, Singapore 138602, Singapore
[3] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,
    Wuhan University, Wuhan 430079, China; zxfang@whu.edu.cn
[4] Collaborative Innovation Center of Geospatial Technology, 129 Luoyu Road, Wuhan 430079, China
[5] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China;
    yinling@siat.ac.cn
[*] Correspondence: sshaw@utk.edu; Tel.: +1-865-974-6036

Abstract: This study uses a large-scale mobile phone dataset to estimate potential demand of bicycle trips in a city. By identifying two important anchor points (night-time anchor point and day-time anchor point) from individual cellphone trajectories, this study proposes an anchor-point based trajectory segmentation method to partition cellphone trajectories into trip chain segments. By selecting trip chain segments that can potentially be served by bicycles, two indicators (*inflow* and *outflow*) are generated at the cellphone tower level to estimate the potential demand of incoming and outgoing bicycle trips at different places in the city and different times of a day. A maximum coverage location-allocation model is used to suggest locations of bike sharing stations based on the total demand generated at each cellphone tower. Two measures are introduced to further understand characteristics of the suggested bike station locations: (1) accessibility; and (2) dynamic relationships between incoming and outgoing trips. The accessibility measure quantifies how well the stations could serve bicycle users to reach other potential activity destinations. The dynamic relationships reflect the asymmetry of human travel patterns at different times of a day. The study indicates the value of mobile phone data to intelligent spatial decision support in public transportation planning.

1. Introduction

Bike sharing systems have received increasing attention in the past few decades. Many cities around the world are promoting bicycle use to mitigate urban problems related to public health, traffic congestion, energy consumption, and air pollution. Bike sharing systems offer short-term bike rental services to individuals for point-to-point trips. A successful bike sharing system could encourage people's use of bikes for short distance trips and alleviate traffic pressure in congested urban areas. Unfortunately, it is not easy to determine where investments and resources should be allocated when implementing these bike sharing systems. Among the various factors that could be considered, knowing where the demands are and when they occur is of primary importance.

Travel surveys and census data have been widely used in past studies [1–3] to estimate the demand of bicycle usage, and provide decision support for locating new cycling facilities such as bike sharing stations. However, collecting such data can be costly and time-consuming. Moreover, the

amount of information that can be collected by the conventional methods is largely constrained by available resources. Recent advancements of location-aware technologies have provided many new data sources (e.g., smart card data and mobile phone data) for understanding how people move around in their daily lives. These new datasets enable us to obtain timely and spatially detailed information on human travel patterns. However, few studies have leveraged these data sources to estimate potential demand of bicycle trips, which serves as valuable information for planning a bike sharing system.

In recent years, researchers have used mobile phone data to study human mobility patterns and people's use of urban space. Among these studies, considerable efforts have been devoted to uncover people's activity anchor points (e.g., home and workplace) as well as movement patterns among these locations [4–9]. Such information reflects how people organize their trips among important activity destinations, and sheds light on people's daily trip chains [10–12]. These activity anchor points and trip chains can be used to estimate travel demand related to various transportation modes (e.g., cycling) in a city. Hence, this study uses an actively tracked mobile phone dataset collected in Shenzhen, China to estimate potential demand of bicycle trips in the city. The main contributions of this study are as follows:

- By identifying two important anchor points (night-time anchor point [*NTA*], and day-time anchor point [*DTA*]) from individual cellphone trajectories, we introduce an anchor-point based trajectory segmentation method to partition cellphone trajectories into meaningful trip chain segments. By selecting trip chain segments that fall within particular ranges of travel distance along the road network, two indicators (*inflow* and *outflow*) are generated at the cellphone tower level to estimate potential demand of incoming and outgoing bicycle trips at different places in the city and different times of a day. The two indicators reflect the intensity and daily rhythms of people's short distance trips at a relatively fine spatial resolution, and can be further used to suggest locations of bike sharing stations.
- Based on the total demand (i.e., sum of *inflow* and *outflow*) generated at each cellphone tower, a *maximum coverage* location-allocation model is used to suggest locations of bike sharing stations under four different scenarios (e.g., 300, 600, 900, and 1200 bike stations). Two measures are introduced to further understand the characteristics of the suggested bike station locations: (1) accessibility; and (2) dynamic relationships between incoming and outgoing trips. The accessibility measure quantifies how well the stations could serve bicycle users to reach other potential activity destinations. The dynamic relationships between incoming and outgoing trips reflect the asymmetry of human travel patterns at each bike station over time, which serve as useful information for the operation of a bike sharing system (e.g., distribution and redistribution of bicycles among the bike stations).

2. Literature Review

2.1. Bike Sharing Systems

Bike sharing systems have received growing attention in recent years. According to a report [13] provided by the Institute for Transportation & Development Policy (ITDP) in 2013, more than 600 cities (examples of these bike sharing systems include Vélib in Paris, France (http://www.velib.paris/), Bicing in Barcelona, Spain (https://www.bicing.cat/), Call-a-Bike in Germany (http://www.callabike.de/), Cycle Hire in London, United Kingdom, and Ecobici in Mexico City, Mexico (https://www.ecobici.df.gob.mx/)) around the world have established their own bike sharing systems and more are starting every year. The evolution of bike sharing systems over the past 50 years can be categorized into three generations [14,15]. The first generation of bike sharing systems, also known as the Free Bike System, was implemented in Amsterdam, Netherlands in 1965. The system was provided for public use at no charge, and the bicycles were unlocked so that users could drop them off at any place they wanted. However, the bike sharing system suffered from problems such as theft and vandalism, and collapsed within a short period of time. The second generation of bike

sharing systems, known as the coin-deposit system, was first established in Nakskov, Denmark in 1993 (which was followed by a larger bike sharing program launched in Copenhagen in 1995). Users could pick up and return the bicycles at specific locations using a coin deposit. The third generation of bike sharing systems, known as the information technology based system, was first introduced in England in 1996. The third generation incorporated many new technologies such as smartcards, mag-stripe cards, and mobile phone access [15]. Some researchers also provided an outlook to the fourth generation of bike sharing systems [16,17], which will incorporate more advanced technologies such as improved distribution, ease of installation, tracking, pedal assistance, and anti-theft mechanism.

2.2. Forecasting Bicycle Travel Demand

To establish a successful bike sharing system, planners need to obtain a good understanding of potential travel demand in relation to factors such as land topography, connectivity of transportation networks, land-use diversity, weather, and safety [1,18,19]. According to Porter, Suhrbier and Schwartz [20], previous studies usually adopted four broad categories of methods to estimate bicycle trip demands, which are aggregate-level methods, attitudinal surveys, discrete choice models, and regional travel models (e.g., four-step travel demand models). Most of these methods rely on detailed information of human activity patterns (e.g., surveys), or many assumptions about human travel behavior (e.g., discrete choice models). For example, Landis [21] proposed a Latent Score (LDS) model based on a probabilistic gravity model to estimate the amount of bicycle trips that would occur on each road segment. Clark [22] used a four-step travel demand model to estimate the length and travel time of trips in Bend, Oregon to identify travels that could be made by bicycles. Rybarzcyk and Wu [23] introduced the bicycle level of service index and demand potential index to analyze the spatial relationships between bicycle supply and demand. The demand of bicycle trips was estimated based on population distribution and locations of parks, recreation areas, schools and businesses. Wardam, Tight and Page [24] developed a mode choice model that combined revealed preference data (with individuals' actual mode choices) and stated preference data (with hypotheses on individual choices among different alternatives) to predict future trends in commuter cycling in Great Britain. Although travel surveys and regional travel demand models are valuable for estimating potential demand of bicycle trips, they usually involve tremendous efforts and financial resources to collect the data. Moreover, many travel demand models used "zone structures that are too large to be of much use in deciding on the size and location of bike sharing stations" (p. 56) [13]. New data and analytical methods are needed to gain an improved understanding of human travel demand and to better assist decision making in transportation planning.

2.3. Mobile Phone Data for Travel Behavioral Analysis

Recent advancements of location-aware technologies have produced many new data sources for understanding the whereabouts of people in space and time. These new datasets enable studies of human activities "at low cost and on an unprecedented scale" [7]. For example, many studies have used mobile phone data to characterize and predict human mobility patterns [25–30], and to better understand various aspects of urban dynamics [31–34]. Among these studies, considerable efforts have been devoted to uncover people's use of urban space and daily rhythms of urban flows. However, there has been limited research on estimating potential demand of bicycle trips from mobile phone data.

In the past few years, there were some studies that used mobile phone data to better understand human travel behavior, especially the movement patterns that were tied to people's major activity locations (e.g., home and workplace). For example, Iqbal et al. [35] used call detail records (CDRs) in Dhaka, Bangladesh to generate tower-to-tower transient origin-destination (OD) matrices. Similarly, Alexander et al. [36] used CDRs collected in the Boston metropolitan area over a period of two months to estimate OD trips by purposes (e.g., home-based work trips, home-based other trips, and non-home-based trips). Dong et al. [37] used CDRs to suggest traffic zone division in urban areas to

assist travel demand forecast. Wang et al. [38] used mobile phone data collected in San Francisco and Boston area to evaluate urban road usage patterns. It is clear that mobile phone data can be leveraged to uncover human travel demand associated with different transportation modes and activity types in various urban contexts.

2.4. Bike Stations and Location-Allocation Models

One of the most important tasks of planning a bike sharing system is to determine the location of bike stations. Well placed bike sharing stations would ensure that the system meets the current demand and stimulate people's use of bicycles in the future. Many studies have given their thoughts to where bike stations should be located under particular scenarios. For example, Larsen, Patterson and El-Geneidy [3] proposed a prioritization index calculated at the grid-cell level to demonstrate how to prioritize cycling infrastructure investments. The prioritization index was aggregated from several indicators including OD of actual bicycle trips, OD of short car trips, cyclists' route preferences, and concentration of bicycle crashes. Martinez et al. [39] proposed a heuristic algorithm, which encompassed a mixed integer linear program (MILP) and a p-median location-allocation problem to optimize the location of bike sharing stations in Lisbon, Portugal. The locations of bike stations were determined based on a list of factors related to user demand, the required investment, and operational costs. García-Palomares, Gutiérrez and Latorre [40] used the population and number of jobs at the building level to estimate potential demand of bicycle trips in central Madrid. The authors adopted two location-allocation models with different objective functions (i.e., minimize impedance and maximize coverage) to suggest facility locations of bike sharing stations.

Some studies adopted location-allocation models to suggest the optimal locations of bike stations in relation to the distribution of potential demand. These location-allocation models aim at determining the number and/or locations of facilities to meet some predefined objectives while satisfying the requirements at the demand points [41]. The location-allocation models could vary depending on the specific objectives. For example, the p-median problem and the p-center problem are two typical forms of location-allocation models [42,43]. The objective of the p-median problem is to locate p facilities to minimize the total weighted travel cost from the demand points to the facilities. The p-center problem aims at providing p facilities to minimize the maximum distance from a demand point to its closest facility. Toregas et al. [44] introduced the Location Set Covering Problem with the objective of determining the minimal number of facilities such that all demand points fall within a specified maximal service distance from a facility. Based on this model, Church and Revelle [45] formulated the Maximal Covering Location Problem (MCLP), which maximizes the population (or demand) within the service distance of the facilities by locating a fixed number of facilities.

3. Study Area and Dataset

Shenzhen is a major financial and technology center in southern China (see Figure 1A). The city situates north of Hong Kong and covers a total area of 2050 km^2. It has an estimated population of 15 million as of 2014 [46]. As shown in Figure 1B, the city has six administrative districts and four management new districts (Guangming and Longhua are two management new districts subordinate to Bao'an district; and Pingshan and Dapeng are two management new districts subordinate to Longgang district). Shenzhen was a small finishing village when it became China's first Special Economic Zone (SEZ) in 1979. The SEZ comprised only Nanshan, Futian, Luohu and Yantian districts until 1 July 2010, and was then expanded to include all other districts. The southern and northern parts of Shenzhen have very different socioeconomic and demographic characteristics. The four districts in the southern part of Shenzhen (i.e., Nanshan, Futian, Luohu and Yantian) are commonly known as *Guan Nei*, which are highly developed areas in terms of finance, technology, education, and tourism. The other six districts are usually known as *Guan Wai*, with manufacturing as its major industry. According to a recent travel survey [47], non-motorized trips accounted for a large percentage of total trips in Shenzhen (walk: 50.0%; and bicycle/moped: 6.2%). The city government

considers cycling as an effective transportation mode and plans to improve the corresponding facilities in the next few years. It is thus important to study where such facilities (e.g., bike sharing stations) should be built to best accommodate people's travel needs.
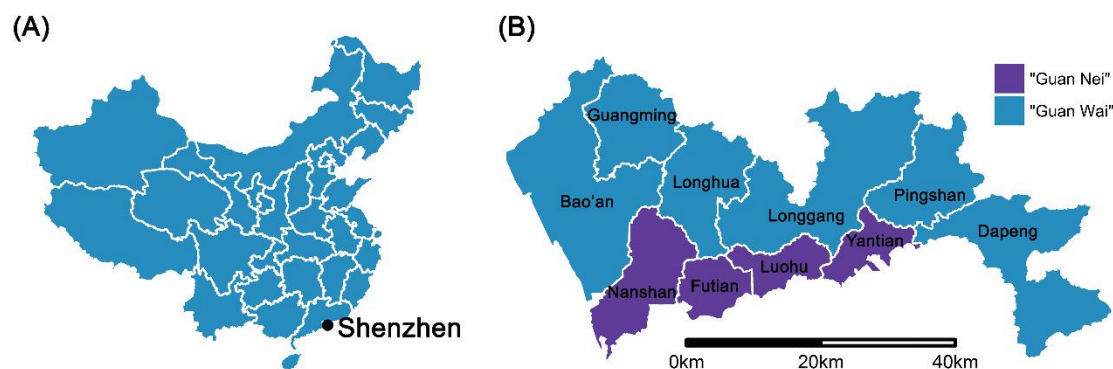

Figure 1. (**A**) Shenzhen's location in China; and (**B**) Shenzhen's administrative districts.

This study uses an actively tracked mobile phone dataset (the mobile phone dataset used in this study was acquired through research collaboration with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, and the research was approved by the Institutional Review Board (IRB)) collected on a weekday (23 March 2012) in Shenzhen, China. The dataset had been anonymized by the mobile phone carrier before it was made available to this research. Hence, the dataset only includes arbitrary unique IDs for mobile subscribers that do not reveal their identity (e.g., phone number). The number of mobile subscribers sampled in each administrative district is in agreement with the population distribution recorded by the census data [47], with a Pearson correlation coefficient of 0.99 [9]. Note that we have removed the mobile subscribers with power on or power off event during the study period, since it is difficult to infer their locations when cellphones are disconnected from the cellular network. The remaining dataset after filtering these individuals covers 5.8 million cellphones, with their locations reported approximately once every hour as the x, y coordinates of the serving cellphone tower. The dataset does not include location records for the 23:00–24:00 time window. Each individual cellphone, therefore, has 23 observations in the study day (Table 1). The spatial configuration of cellphone towers could vary in different parts of the study area. The densities of cellphone towers are generally higher in populated urban areas. The average nearest distance among the cellphone towers in this dataset is 0.19 km.

Table 1. Example of an individual's cellphone location records.

User ID	Record ID	Time Window in Which Location Was Reported (t)	Longitude of Cellphone Tower (x)	Latitude of Cellphone Tower (y)
932 *****	1	(00:00–01:00)	113.*****	22.*****
932 *****	2	(01:00–02:00)	113.*****	22.*****
932 *****	3	(02:00–03:00)	113.*****	22.*****
...	...	...	113.*****	22.*****
932 *****	23	(22:00–23:00)	113.*****	22.*****

4. Methodology

This section first introduces how we generate important activity anchor points from individual cellphone trajectories. Then, an anchor-point based trajectory segmentation method is proposed to partition the cellphone trajectories into trip chain segments. These trip chain segments are then analyzed to derive potential demand of bicycle trips. We use a maximum coverage location-allocation model to suggest locations of bike sharing stations. As this model aims to locate a fixed number of

facilities such that the total demand within a specified impedance cutoff (i.e., service radius) of the facilities is maximized, it can be used to provide reasonable suggestions on where to place bike sharing stations to best accommodate people's travel needs. Finally, we characterize the accessibility as well as the dynamic relationships between the incoming and outgoing trips at these bike station locations.

4.1. Anchor Point Extracion and Trajectory Segmentation

As shown in Table 1, an individual's cellphone trajectory *T* can be represented as follows:

$$T = \{P_1\left(x_1, y_1, t_1\right), P_2\left(x_2, y_2, t_2\right), \ldots, P_i\left(x_i, y_i, t_i\right)\} \tag{1}$$

where P_i denotes the *i*th ($i = 1, 2, \ldots, 23$) cellphone location record; x_i and y_i denote the coordinates of the serving cellphone tower; and t_i represents the one-hour time window in which the location was recorded.

Activity anchor points have been frequently used in past studies [48,49] to denote a person's major activity locations such as home, workplace, favorite restaurants, etc. These activity anchor points serve as important activity origins and destinations of people's daily travels. One challenge of using mobile phone data to determine an individual's activity anchor points is that an individual's cellphone location record could switch among adjacent cellphone towers due to either cellphone load balancing [50] or signal strength variation [51]. Hence, it is necessary to consider these issues when estimating individual activity anchor points.

In this paper, we introduce *activity anchor point (AAP)* as a set of cellphone towers that are geographically concentrated and where an individual spent a certain amount of time. To derive *AAPs* for a cellphone trajectory *T*, we first calculate the frequency (i.e., number of time windows) of each unique cellphone tower traversed by *T*. We then select the most visited cellphone tower, and group all the cellphone towers that are within 0.5 km of the selected tower into a cluster. We then select the next most visited tower and perform the same grouping process. The process is iterated until all cellphone towers in *T* are processed. Finally, we calculate the number of cellphone location records (i.e., observations) assigned to each cluster. Any cluster with two or more cellphone locations is identified as an *AAP*. The remaining clusters (i.e., isolated cellphone towers) are defined as *random cellphone towers*.

Note that we choose a constant threshold of 0.5 km for two reasons. First, although we are aware that cellphone tower density could vary within a city, choosing a constant threshold enables us to consistently evaluate individual cellphone trajectories in a city. Second, as the average nearest distance among cellphone towers is 0.19 km, choosing 0.5 km addresses the problem of signal switches among nearby cellphone towers, and keeps individual movements which occurred between different activity clusters (i.e., *AAPs*).

Figure 2 shows an example of an individual's cellphone trajectory in a three-dimensional space-time system proposed by Hägerstrand [52]. The cellphone tower locations of this individual are grouped into four clusters, which include three *AAPs* (clusters A, B and C) and one *random cellphone tower* (cluster D). The red lines represent movements occurred within clusters (i.e., intra-cluster movements), and the green lines denote inter-cluster movements.

Note that intra-cluster movements could be caused by issues of cellphone signal switches or individual movements that are very short in distance (i.e., within walkable distance). These intra-cluster movements are not used to generate potential demand of bicycle trips. Thus, we merge cellphone tower in each cluster of *T* to derive a generalized cellphone trajectory *T'*. We choose the cellphone tower with the highest frequency in each cluster as the representative cellphone tower. As illustrated in Figure 3, given a cellphone trajectory *T*, four representative cellphone towers (A, B, C and D) that correspond to the four clusters are used to derive the generalized cellphone trajectory *T'*. The generalized cellphone trajectories in the mobile phone dataset are then used to derive individual trip chain segments.
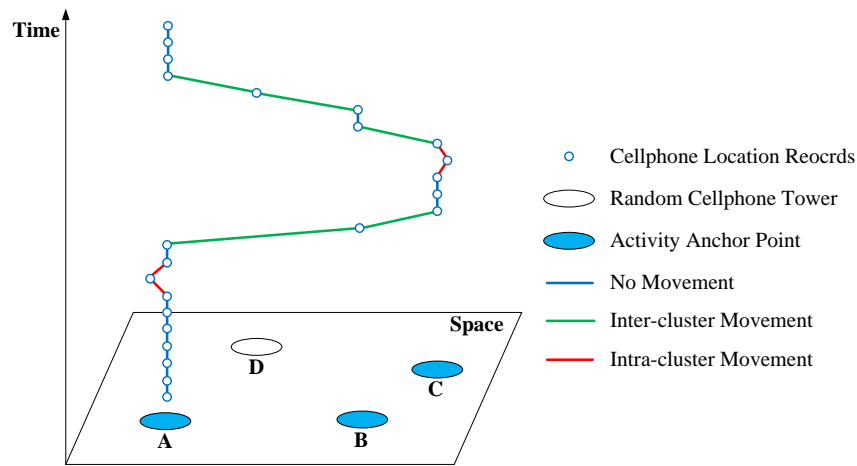
Figure 2. An individual's cellphone trajectory *T* and concepts of: (1) activity anchor point (*AAP*); (2) random cellphone tower; (3) inter-cluster movements; and (4) intra-cluster movements.
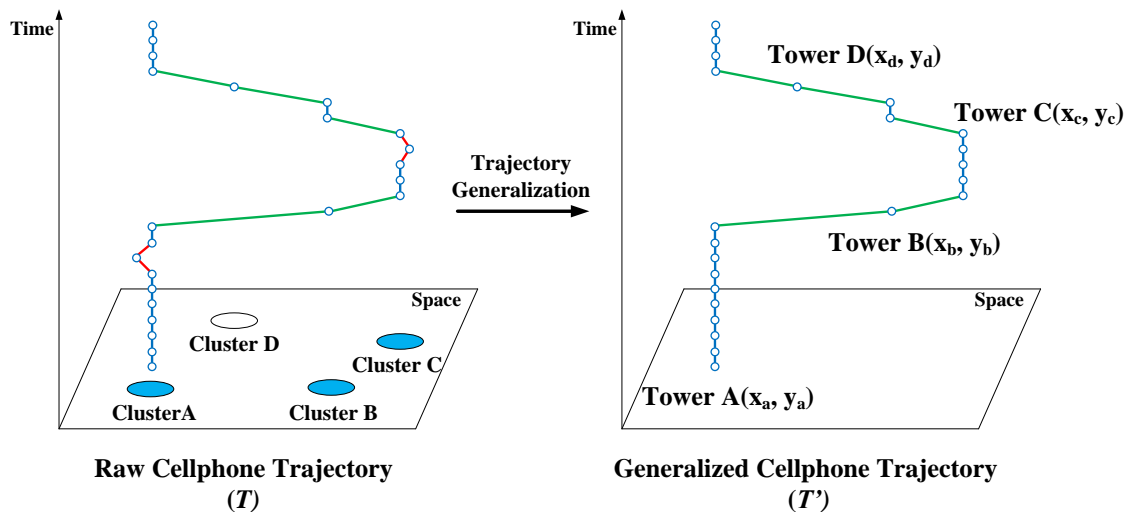

Figure 3. Derive generalized cellphone trajectory (*T'*) from an individual's raw cellphone trajectory (*T*) using the representative cellphone tower of each cluster.

4.2. Trajectory Segmentation Based on Trip Chain Analysis

Trip chaining often describes a travel, with possible intermediate stops, between an individual's activity anchor points (e.g., home and workplace). The trip chaining behavior reflects the complexity of human travel patterns and is an important factor that drives individual mode choice [53]. In this study, we estimate two important activity anchor points—the night-time anchor point (*NTA*) and day-time anchor point (*DTA*)—as approximate individual home location and workplace. These two anchor points are used to partition individual cellphone trajectories into trip chain segments.

According to [54], the normal hours of sleep and work for people in Shenzhen are 00:00 to 07:00 and 09:00–18:00, respectively. For each cellphone trajectory *T'*, the duration of stay at different representative cellphone towers during these two time periods are used to identify individual *NTA* and *DTA*. Considering people's daily routines in most big cities in China, we adopt the approach proposed in [55] to derive the two activity anchor points. In particular, we define *NTA* as the representative cellphone tower with a minimum of four hours of stay between 00:00 and 07:00, and *DTA* as the tower with a minimum of six hours of stay between 09:00 and 18:00. Based on this rule, we are able to estimate *NTA* and *DTA* for 99% and 85% of all individuals in the dataset, respectively. According to

our analysis, 55% of the individuals have both *NTA* and *DTA* extracted that correspond to different representative cellphone towers; 30% have both *NTA* and *DTA* extracted that correspond to the same representative cellphone tower; 14% have only *NTA* extracted; and the remaining 1% have neither *NTA* nor *DTA* extracted. In this study, individuals with neither of the two anchor points extracted are not considered when generating potential demand of bicycle trips.

　　We then use *NTA* and *DTA* to partition cellphone trajectories into trip chain segments. For a trajectory T', each trip chain segment after partition refers to a list of consecutive cellphone records which originated and ended at either *NTA* or *DTA*. Table 2 shows the four types of trip chain segments derived in this study. *ND* refers to the trip chain segments that started at *NTA* and ended at *DTA*. The *InTransit* locations refer to other cellphone towers traversed by the trip chain segment. These *InTransit* locations could refer to intermediate stops of the trip, or random cellphone towers captured by the mobile phone dataset. Similarly, *NN* refers to the trip chain segments that both started and ended at *NTA*. *DD* denotes the segments that both started and ended at *DTA*. Note that *InTransit* locations do not always exist in *ND* or *DN* trip chain segments. For example, an individual could be located at *NTA* during a certain one-hour time window and at *DTA* during the next time window.

Table 2. Four types of trip chain segments derived from individual cellphone trajectories.

Type	Representation of Trip Chain Segment	Typical Activity Patterns
ND	*NTA–InTransit–DTA*	Home–(Drop mail to FedEx)–Workplace
NN	*NTA–InTransit–NTA*	Home–(Shop at grocery store)–Home
DN	*DTA–InTransit–NTA*	Workplace–(Dine at restaurant)–Home
DD	*DTA–InTransit–DTA*	Workplace–(Meet with others at Starbucks)–Workplace

4.3. Generate Potential Demand of Bicycle Trips

　　According to a recent travel survey in Shenzhen [47], the average trip distances for walking and cycling in this city are 1.6 km and 4.8 km, respectively. However, it is pointed out in the survey that the average walking trip distance in Shenzhen is generally higher than that of other domestic and foreign cities (usually 1 km) due to an underdevelopment of cycling facilities. Hence, we consider 1 km as a reasonable walk distance, and use 1 km and 5 km as the spatial thresholds to filter the trip chain segments.

　　For each trip chain segment, we first calculate its *range*, which is defined as the maximum distance (i.e., shortest path distance along road network) between all pairs of cellphone towers traversed by the segment. The trip chain segments with *range* between 1 km and 5 km are used to generate potential demand. We use this filtering strategy to exclude those trip chain segments that are either within a reasonable walk distance, or beyond normal travel distance for bicycles. The reason of using *range* to filter each segment is that an individual might have intermediate stops during a trip chain segment. If the distance: (1) between the origin and destination of this trip chain segment; or (2) between an intermediate stop (i.e., *InTransit* location) and the origin (or destination) is beyond normal travel distance for cycling, the individual is unlikely to use bicycle for this particular trip.

　　As individual cellphone trajectories were recorded at the cellphone tower level, the potential demand is thus aggregated by individual cellphone towers. In this study, two basic types of demand, $inflow_p$ and $outflow_p$, are extracted at each cellphone tower p during different time periods of the study day:

$$inflow_p = \left(I_1^p, I_2^p, I_3^p, \ldots, I_{22}^p \right) \tag{2}$$

$$outflow_p = \left(O_1^p, O_2^p, O_3^p, \ldots, O_{22}^p \right) \tag{3}$$

$$total\_inflow_p = \sum_{i=1}^{22} I_i^p \tag{4}$$

$$total\_outflow_p = \sum_{i=1}^{22} O_i^p \tag{5}$$

As shown in Equations (2) and (3), I_i^p and O_i^p refer to the volume of incoming/outgoing trips at cellphone tower p during a particular time interval i, respectively (e.g., $i = 1$ denotes the time interval between time windows t_1 (00:00–01:00) and t_2 (01:00–02:00)). As illustrated in Table 1, each cellphone tower trajectory covers 23 time windows in the study day. Hence, each of the $inflow_p$ and $outflow_p$ has 22 observations. As shown in Equations (4) and (5), $total\_inflow_p$ and $total\_outflow_p$ refer to the total amount of incoming/outgoing trips at cellphone tower p for the entire day, respectively. The two measures are used as the input for a maximum coverage location-allocation model to suggest locations of bike sharing stations.

We next introduce how $inflow_p$ and $outflow_p$ are extracted from the trip chain segments. Note that a trip chain segment *TS* can be represented as a series of cellphone tower locations:

$$TS = \{P_1(x_1, y_1, t_1), P_2(x_2, y_2, t_2), \ldots, P_i(x_i, y_i, t_i)\} \tag{6}$$

where P_i denotes an individual cellphone tower at time interval i, x_i and y_i denote the (x, y) coordinates of P_i, and t_i represents the ith one-hour time window during which the cellphone location was recorded. By comparing each pair of consecutive cellphone towers (P_i and P_{i+1}) in *TS*, we assign a unit of demand to $O_i^{P_i}$ (i.e., a unit of *outflow* to cellphone tower P_i at time interval i), and a unit of demand to $I_i^{P_{i+1}}$ (i.e., a unit of *inflow* to cellphone tower P_{i+1} at time interval i) if P_i and P_{i+1} refer to different representative cellphone towers:

$$x_i \neq x_{i+1} \ or \ y_i \neq y_{i+1} \tag{7}$$

We repeat this procedure until all trip chain segments (*ND, NN, DN, DD*) are processed.

4.4. Suggest Facility Locations of Bike Stations

This study uses the *maximum coverage* location-allocation model in ArcGIS 10.1 to suggest locations of bike sharing stations. The objective of this model is to locate a fixed number of facilities (i.e., bike stations) such that the total demand within a specified *impedance cutoff* (i.e., service radius) of the facilities is maximized. When configuring the *maximum coverage* module, the individual cellphone towers in the actively tracked mobile phone dataset (5928 in total) are used as both *demand points* and the *candidate locations* of the facilities. The *weight* at each demand point (i.e., cellphone tower) p is calculated as the sum of $total\_inflow_p$ and $total\_outflow_p$, since they correspond to the number of drop-off and pick-up activities of potential bicycle trips, respectively. These two types of activities are both considered as travel demand when planning bike sharing stations in a city [40]. The *impedance cutoff* is chosen at 500 meters (road network distance) to approximate the service radius of bike sharing stations, which serves as a reasonable walking distance from activity origins/destinations to the closest bike sharing stations for bicycle pick-up/drop-off activities. For the *number of facilities* (*N*) to be located, we define four different scenarios (*N = 300, N = 600, N = 900, N = 1200*) and compare the outcomes (e.g., percentage of potential demand that can be covered) among the four scenarios. Once the facility locations are determined (in each of the four scenarios), the location-allocation model will allocate the demand points to the facilities. A demand point that is inside the *impedance cutoff* of one facility is allocated to that facility, while a demand point that falls within the *impedance cutoff* of two or more facilities is allocated to its nearest facility. Any demand point that falls outside of all facilities' *impedance cutoff* is not served by any facility.

4.5. Characterization of Bike Stations

In this study, two measures are introduced to assess the bike stations once their locations are determined. First, we introduce an accessibility measure to evaluate how well the stations could

serve bicycle users to reach other potential activity destinations. We then investigate the dynamic relationships between the incoming and outgoing trips that are allocated to each bike station.

In order to measure these two characteristics of bike stations, we first retrieve the demand points that are allocated to each bike station, and calculate the total demand allocated to each station. Specifically, for each bike station q, we introduce $inflow\_C_q$ and $outflow\_C_q$ to represent the total amount of incoming and outgoing trips that are allocated to the station, respectively:

$$inflow\_C_q = \left(J_1^q, J_2^q, J_3^q, \ldots, J_{22}^q \right) \tag{8}$$

$$outflow\_C_q = \left(K_1^q, K_2^q, K_3^q, \ldots, K_{22}^q \right) \tag{9}$$

J_i^q and K_i^q refer to the number of incoming and outgoing trips that are allocated to q during time interval i (e.g., 1, 2, 3, $\ldots$, 22), respectively:

$$J_i^q = \sum_{m=1}^{n} I_i^m * C_{qm} \tag{10}$$

$$K_i^q = \sum_{m=1}^{n} O_i^m * C_{qm} \tag{11}$$

where n denotes the total number of demand points (i.e., cellphone towers) in the study area. C_{qm} takes the value of 1 if demand point m is allocated to the bike station q, and is 0 otherwise. Note that:

$$total\_inflow\_C_q = \sum_{i=1}^{22} J_i^q \tag{12}$$

$$total\_outflow\_C_q = \sum_{i=1}^{22} K_i^q \tag{13}$$

By doing so, we are able to aggregate the incoming and outgoing trips from the demand points to each bike sharing station over different time intervals of a day.

The concept of accessibility has been widely used in transportation studies to describe how well a location could reach other potential activity destinations [56]. In order to represent the accessibility of each bike station, we adopt a gravity-based measure that has been used in previous studies [19,40] to quantify bicycle accessibility. For each bike station q, the accessibility A_q is calculated as follows:

$$A_q = \sum_{k=1}^{n-1} \frac{total\_inflow\_C_k * M_{qk}}{\left(D_{qk} \right)^{\alpha}} \tag{14}$$

where n denotes the total number of bike stations (e.g., 300, 600, 900, and 1200). M_{qk} takes the values of 1 if the road network distance between station q and station k is less than $5km$, and is 0 otherwise. D_{qk} is the road network distance between station q and station k, and α takes the value of 2 (which is the default value of the gravity based measure) to reflect the distance decay effect. Note that we use $total\_inflow\_C_k$ (i.e., number of incoming trips allocated to each station) in order to approximate the total amount of opportunities (i.e., activities) at station k.

We next introduce $netflow_q$ to reflect the dynamic relationships between the incoming and outgoing trips allocated to each bike station q:

$$netflow_q = \left(Net_1^q, Net_2^q, Net_3^q, \ldots, Net_{22}^q \right) \tag{15}$$

For each particular time interval i, Net_i^q is calculated as the net volume of trips (i.e., outgoing − incoming) normalized by the total number of trips:

$$Net_i^q = \frac{K_i^q - J_i^q}{K_i^q + J_i^q} \tag{16}$$

The value of Net_i^q ranges from −1.0 to 1.0. A positive value indicates that station q serves as a *trip producer* at time interval i, while a negative value indicates that the station serves as a *trip attractor* during that time interval. The temporal characteristics of $net\,flow_q$ indicate the asymmetry of human travel patterns at different times of a day. In order to assess the temporal characteristics of $net\,flow_q$, this study uses the k-means clustering method to group the bike stations. The clustering results can help us examine the temporal characteristics of $net\,flow_q$ associated with different bike stations and their geographic distribution.

5. Analysis Results

5.1. General Statistics

By analyzing the generalized cellphone trajectories of 5.8 million individuals in the dataset, we are able to derive a total of 7,086,241 trip chain segments of which the *range* falls between 1 km and 5 km. As shown in Table 3, we have 1,636,494 *ND* segments (24.3%) and 1,480,342 *DN* segments (22.0%). The percentages of *ND* and *DN* segments are close to each other, which reflects the regularity of human travel patterns between *NTA* and *DTA* during the day. The number of *NN* segments is 3,159,753 (47.0%), which suggests a large proportion of trips around individual NTA. We also identify 449,652 *DD* segments, which account for only 6.7% of the total number of trip chain segments.

Table 3. Number and percentage of extracted trip chain segment by type.

Type of Trip Chain Segment	Amount	Percentage of Total
ND	1,636,494	24.3%
NN	3,159,753	47.0%
DN	1,480,342	22.0%
DD	449,652	6.7%

Figure 4 illustrates the temporal distribution of trip chain segments by type. As illustrated in Figure 4A, the majority of *ND* segments occurred during morning rush hours since *ND* segments mainly correspond to commuting activities during this time period (i.e., time windows 7, 8 and 9). We also observe a local peak at time window 13, which is presumably explained by people who went back home from their workplace during the lunch break. Similar temporal patterns are observed for *DN* segments (see Figure 4C). The number of *DN* segments reached its peak around afternoon rush hours but decayed slowly during night time. The identified patterns can be potentially explained by two reasons. First, people chose different times to get off work in order to avoid traffic congestion. Second, some people might need to work overtime and leave their workplaces late in the evening. The concentration of *DN* segments during night time suggests that the operation hours of bike sharing stations should include these time periods to meet people's travel needs. As illustrated in Figure 4B, the volume of *NN* segments remains relatively consistent over time. The *DD* segments mainly concentrate during normal work hours, with its peak around time interval 12 (see Figure 4D).
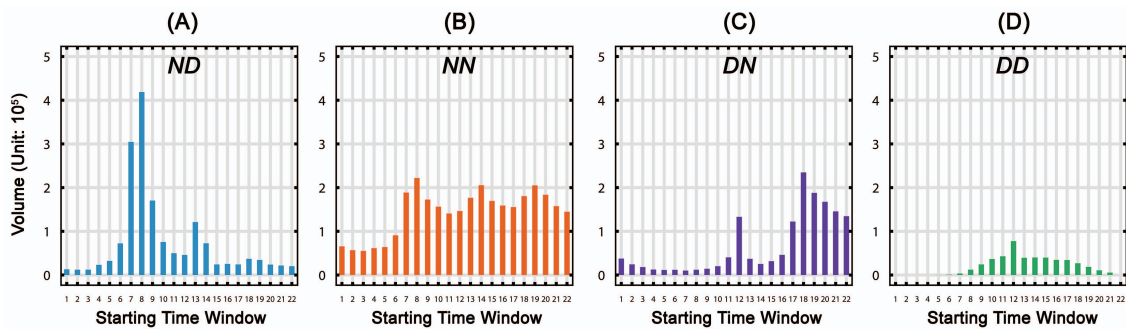
Figure 4. Temporal distribution of trip chain segments by type: (**A**) *ND* trip chain segments; (**B**) *NN* trip chain segments; (**C**) *DN* trip chain segments; and (**D**) *DD* trip chain segments.

5.2. Spatiotemporal Distributions of Potential Demand

The spatial and temporal dynamics of potential demands serve as critical information for planning and operation of bike sharing stations. As $outflow_p$ and $inflow_p$ are generated at the cellphone tower level, we use kernel density maps to illustrate the geographic distribution of potential demand at different times of the day. As a bike sharing station only serves nearby demand points, a small search radius should be used to fit a density surface to reflect the geographic patterns of demand. In this study, we choose 1 km as the search radius to produce the density maps.

In this section, several key time intervals are chosen to illustrate geographic distributions of $outflow_p$ and $inflow_p$. For example, Figure 5A shows the density pattern of $outflow_p$ at time interval 8 (i.e., 07:00–08:00 to 08:00–09:00). Areas with a high density of demand mainly locate at south Futian, southwest Bao'an, southwest Nanshan, and central Longhua. These areas generated a large number of potential bicycle trips in the early morning. By further overlaying the density map with land use map, we find that these areas are mainly residential neighborhoods in Shenzhen. For example, areas a, b, c, f and g are places with many residential apartments. Areas d and e cover several "urban villages" (e.g., Shangsha Village and Huanggang Village) in Shenzhen. These "urban villages" usually refer to densely populated areas with a large migrant population [57]. Figure 5B shows the density pattern of $outflow_p$ at time interval 9 (i.e., 08:00–09:00 to 09:00–10:00). Certain areas in south Nanshan and south Futian still generated a large number of trips, while the intensity of $outflow_p$ became lower in the northern part of Shenzhen as compared to the previous time interval. As discussed above, the northern districts (i.e., Guan Wai) in Shenzhen are mainly industry-oriented areas with a large number of migrant workers, while the districts in the south (i.e., Guan Nei) offer more employment opportunities related to education, technology, and commerce. The identified patterns are likely to be caused by the differences of work schedules between Guan Nei and Guan Wai.

Figure 5C illustrates the density pattern of $inflow_p$ at time interval 8 (i.e., 07:00–08:00 to 08:00–09:00). Several areas with a high density of demand are highlighted on the map. We notice that certain industrial parks (e.g., Foxconn Technology Park, Yantian Industrial Park) in the northern districts attracted a large number of trips in the early morning. In southern Shenzhen, however, the areas with a high density of demand mainly cover commercial districts and business centers. Figure 5D shows the density pattern of $inflow_p$ at time interval 9 (i.e., 08:00–09:00 to 09:00–10:00). The industrial parks in the north attracted much fewer trips during time interval 9 as compared to time interval 8. However, the commercial areas and business centers in Futian and Luohu continued to attract a large number of trips. The area with the highest density is Huaqiang North, which is the largest commercial center in Shenzhen and is known for its business of computer hardware and electronic products.
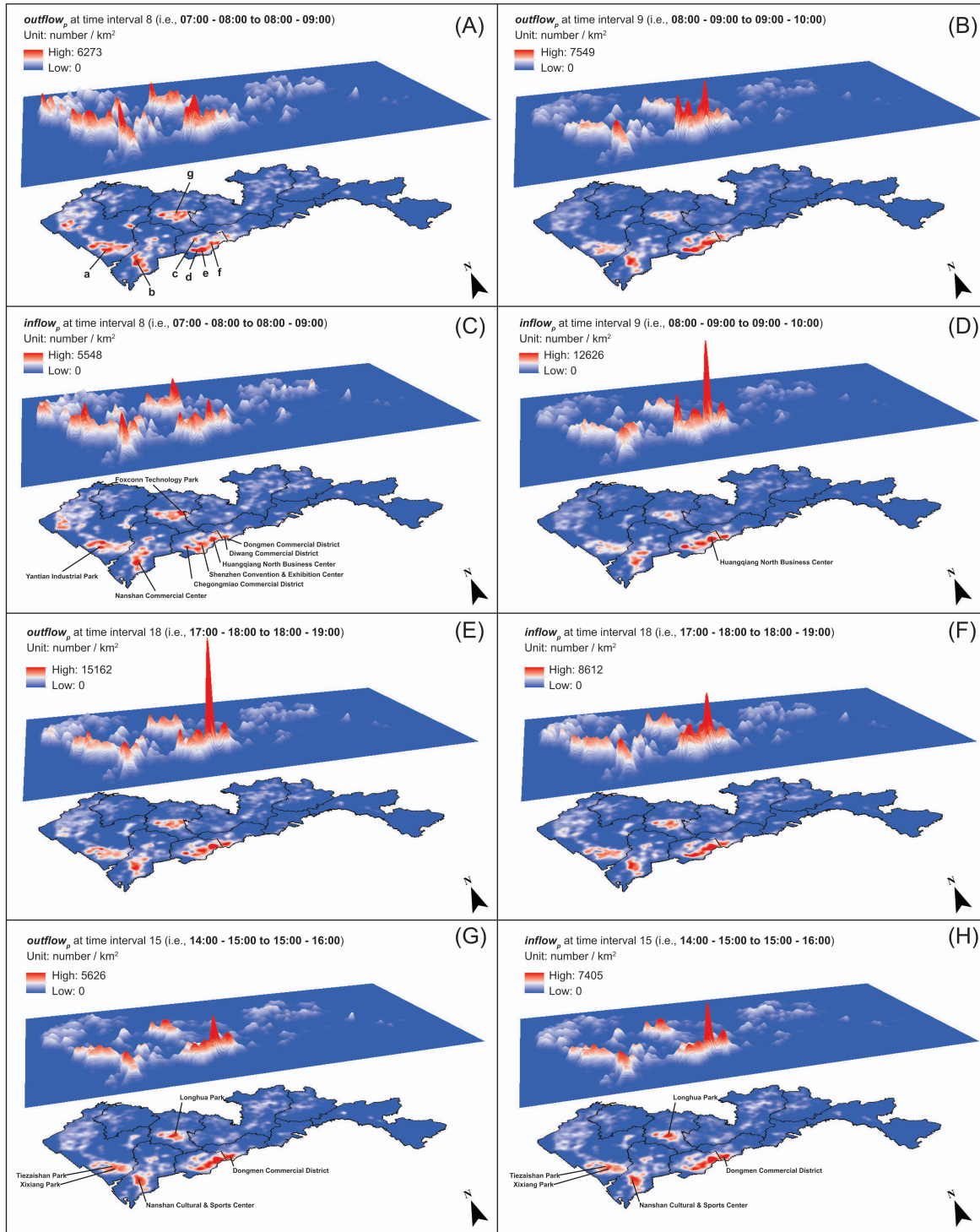
Figure 5. Spatial distribution patterns of: (**A**) $outflow_p$ during time interval 8; (**B**) $outflow_p$ during time interval 9; (**C**) $inflow_p$ during time interval 8; (**D**) $inflow_p$ during time interval 9; (**E**) $outflow_p$ during time interval 18; (**F**) $inflow_p$ during time interval 18; (**G**) $outflow_p$ during time interval 15; and (**H**) $inflow_p$ during time interval 15.

Figure 5E,F illustrate the geographic patterns of $outflow_p$ and $inflow_p$ at time interval 18 (i.e., 17:00–18:00 to 18:00–19:00), respectively. We find that areas which generated a large amount of trips during this time interval (see Figure 5E) also attracted a notable amount of trips in the morning (see Figure 5C,D). Similarly, areas which attracted many trips in the late afternoon (see Figure 5F) also

generated a large number of trips during morning rush hours (see Figure 5A,B). The analysis results reflect the regularity and rhythms of human travel patterns in Shenzhen.

We next examine the density patterns of $outflow_p$ and $inflow_p$ at time interval 15 (i.e., 14:00–15:00 to 15:00–16:00). As morning and afternoon rush hours refer to the time periods when a large number of *ND* and *DN* segments occurred, we choose this particular time interval to better understand the dynamics of travel demand related to other types of trip chain segments (e.g., *NN*). By comparing Figure 5G and 5H, we notice that the density patterns of $outflow_p$ and $inflow_p$ are very similar to each other at time interval 15. Areas that generated more trips tended to also attract more trips at the same time. Note that a considerable proportion of potential demand at time interval 15 was extracted from *NN* segments (see Figure 4). Many areas with a high density of $outflow_p$ and $inflow_p$ during this time interval are associated with recreational and shopping activities. For example, we find many parks (e.g., Longhua Park, Xixiang Park, and Tiezaishan Park) with a high density of potential demand during this time interval. These parks are open and free to the public and offer various sports and recreational facilities. The Nanshan Cultural & Sports Center, funded by the local government, has several art schools, amateur sports schools, cultural centers and theaters, which offer different types of recreational activities. The Dongmen commercial district in Luohu integrates commerce, tourism, shopping, and recreation as its core functions. It seems that the potential demand during this time period is strongly tied to people's leisure activities.

5.3. Suggested Locations of Bike Sharing Stations

In this study, 5928 unique cellphone towers in the dataset are used as both demand points and candidate facility locations. As described in Section 4.4, the total demand (i.e., weight) at each cellphone tower *p* is calculated as the sum of the incoming (i.e., $total\_inflow_p$) and outgoing (i.e., $total\_outflow_p$) trips. Figure 6 illustrates the geographic distribution of these cellphone towers and the density of total demand (using 1 km search radius). Areas with a high density of total demand mainly locate in central Longhua, southwest Bao'an, south Nanshan, southwest Luohu, and Futian.
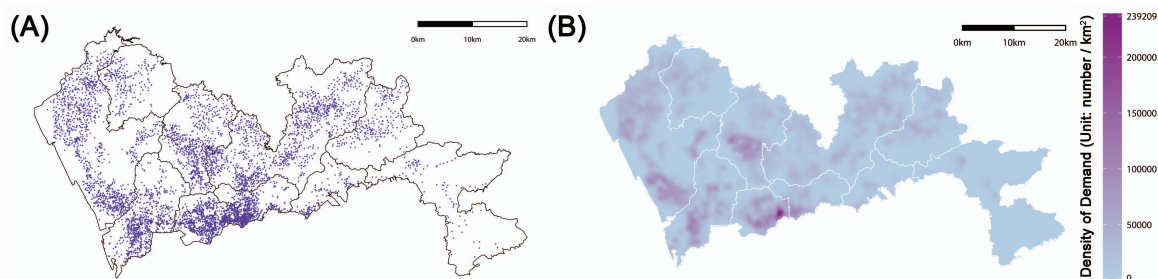

Figure 6. (**A**) Spatial distribution of cellphone towers (5928 in total); and (**B**) density of total demand (Unit: $number/km^2$).

Figure 7 shows the locations of bike sharing stations derived from the location-allocation model. When the number of facilities (*N*) equals 300 (see Figure 7A), the majority of bike stations are located around the areas with a high density of total demand (e.g., central Longhua, southwest Bao'an, southwest Nanshan, southwest Luohu, and Futian). When *N* is set to 600 (see Figure 7B), the density of bike stations at those areas starts to increase. As *N* increases to 900 and 1200 (see Figure 7C,D), the bike stations gradually cover certain areas in the northern part of Shenzhen (e.g., Guangming, Longgang, and Pingshan Districts). Note that the location-allocation model derives a few bike stations that are isolated from the majority of other bike sharing stations under the four scenarios (*N = 300, N = 600, N = 900,* and *N = 1200*). These bike station locations should not be considered during the real planning stage.
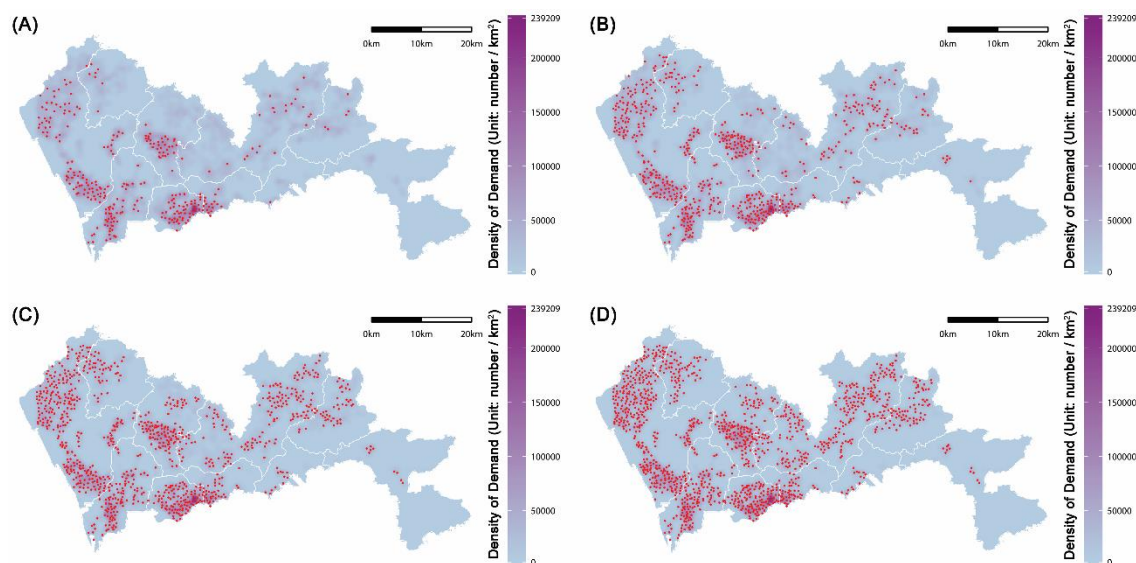
Figure 7. Locations of bike sharing stations derived from the maximum coverage location-allocation model: (**A**) 300 facilities; (**B**) 600 facilities; (**C**) 900 facilities; and (**D**) 1200 facilities.

Table 4 summarizes the percentage of total demand that can be covered by the bike sharing stations under the four different scenarios. The solution of $N = 300$ covers a considerable percentage of total demand (40.2%) since most stations are located in areas with a very high density of demand. As N increases from 300 to 1200, the percentage of demand covered gradually increases from 40.2% to 84.6%, which shows a diminishing return by adding more bike stations.

Table 4. Percentage of total demand covered by the bike sharing stations.

Number of Stations (N)	Amount of Total Demand Covered	Percentage of Total Demand Covered	Increment	Increment Percentage
300	9,888,085	40.2%	—	—
600	14,860,322	60.4%	4,972,237	20.2%
900	18,325,887	74.5%	3,456,565	14.1%
1200	20,829,777	84.6%	2,503,890	10.1%

5.4. Accessibility of the Bike Stations

Figure 8 shows the accessibility of bike stations under the four different scenarios. When $N = 300$, bike stations with high accessibility are mainly located in areas (e.g., central Longhua, southwest Bao'an, southwest Nanshan, southwest Luohu, and Futian) where the density of total demand is high (see Figure 6B). As N changes to 600, there is an increase of overall accessibility for bike stations in those areas. However, the majority of bike stations in northern Shenzhen still experience low accessibility. As N changes to 900 and 1200, we observe a slight increase of accessibility for bike stations in northern Shenzhen but the trend is not obvious.

Figure 9 shows the average accessibility of bike stations by administrative districts (Dapeng and Yantian are not included in this particular analysis due to a very small number of bike stations). In general, bike stations in Futian have the highest average accessibility under all four scenarios, followed by Bao'an, Longhua, Luohu, and Nanshan. Bike stations in Guangming, Longgang, and Pingshan have relatively low accessibility. As N increases from 300 to 1200, we observe an overall increase of the average accessibility for bike stations in most districts. However, as N changes from 900 to 1200, the average accessibility in particular districts (e.g., Futian, Longhua, and Nanshan) remains stable or even decreases. This is because when N becomes very large, the new bike stations

added to these districts tend to be located in peripheral areas where the density of demand is relatively low. On the one hand, there are fewer potential activity destinations (i.e., opportunities) around these bike stations, which causes their low accessibility. On the other hand, these newly added bike stations do not noticeably improve the accessibility of nearby bike stations due to their low level of available opportunities (i.e., $total\_inflow\_C_q$). The analysis results indicate that, in districts where potential demand is concentrated in particular areas, adding more bike stations can lead to noticeable improvement (of average accessibility) at the beginning but will experience a diminishing return as N becomes larger. However, for districts where potential demand is more uniform over space (e.g., Longgang and Pingshan), adding more bike stations will enhance the overall accessibility of the stations in a more consistent manner.
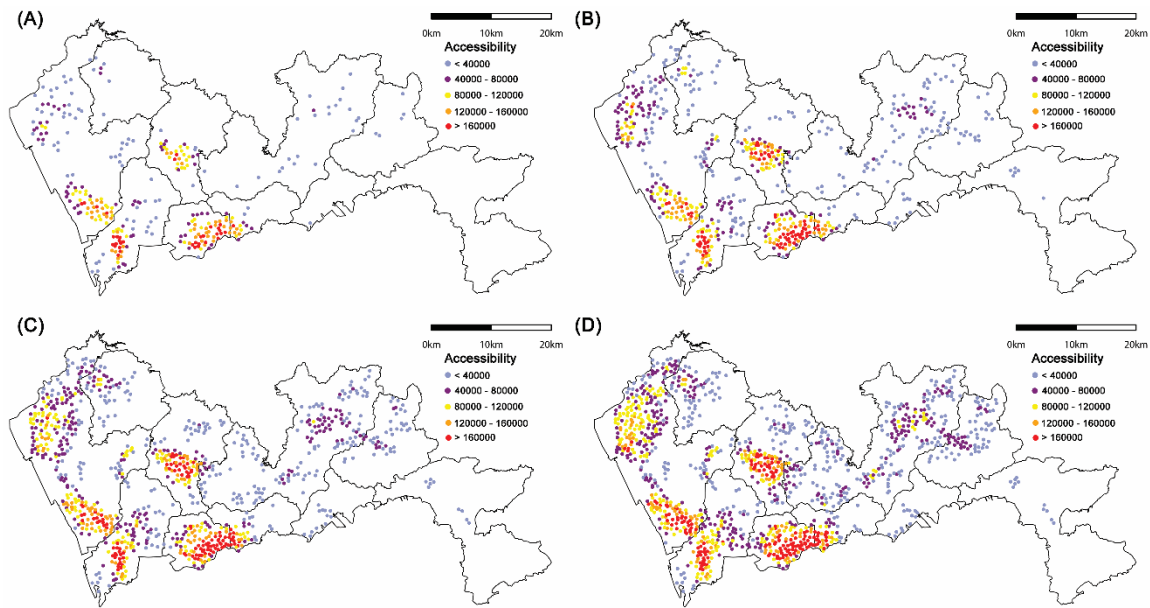
Figure 8. Accessibility of the bike stations: (**A**) 300 facilities; (**B**) 600 facilities; (**C**) 900 facilities; and (**D**) 1200 facilities.
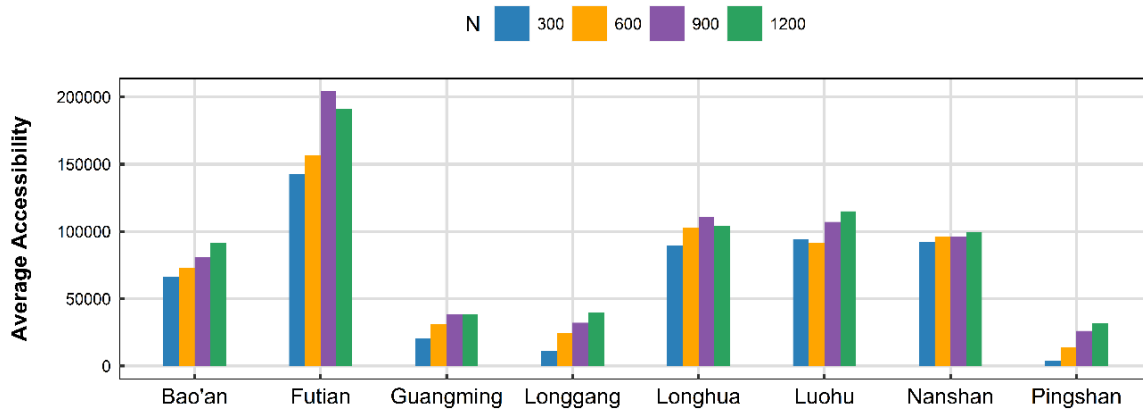

Figure 9. Average accessibility of bike stations by districts under the four different scenarios.

5.5. Dynamic Relationships Between Incoming and Outgoing Trips at the Bike Stations

In this section, we use *N = 1200* as an example to illustrate how $netflow_q$ can be used to better understand the relationship between the incoming and outgoing trips allocated to the bike sharing stations. The k-means clustering method is used to group the bike stations into different clusters based on the temporal patterns of $netflow_q$. In order to determine a proper number of clusters, we evaluate how the total within-cluster variance changes as we increase the number of clusters. As shown in Figure 10, when the number of cluster changes from 1 to 40, the total within-cluster variance drops notably at the beginning of the curve, and then decays slowly as the number of clusters becomes larger. In our analysis, we choose seven as the cluster size to perform the k-means since further increasing the number of clusters does not improve the result much.
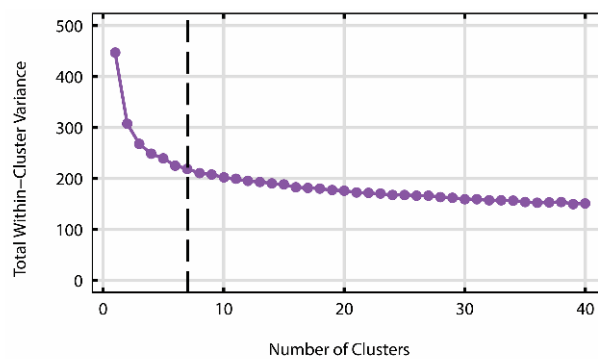
Figure 10. Relationship between total within-cluster variance and number of clusters derived from the k-means clustering method.

Figure 11 shows the average values (i.e., mean center) of $netflow_q$ of the seven clusters (C1 to C7). The incoming and outgoing trips allocated to the bike stations in C1 tend to be in balance throughout the entire day. The characteristics of these bike stations can be described as *mixed* usage patterns. The bike stations in C2 serve as trip attractors in the morning, and trip producers in the late afternoon and evening. However, the overall difference between the incoming and outgoing trips is smaller as compared to that of clusters C3 and C4. Thus, the bike stations in C2 can be described as *weak morning attractor—late afternoon and evening producer*. Similarly, bike stations in C6 can be described as *weak morning producer—late afternoon and evening attractor*. For C3 and C4, the average values of $netflow_q$ reach almost 0.4 in the morning, which indicates a relatively large difference between the incoming and outgoing trips. Hence, the bike stations in C3 and C4 can be described as *strong morning producer—late afternoon and evening attractor*. The difference between C3 and C4 is that the morning peak of C3 occurred at time interval 7 (i.e., 06:00–07:00 to 07:00–08:00) and only lasted for two hours. However, the morning peak of C4 occurred at time interval 8, and the bike stations in this cluster serve as trip producers during the entire morning. Likewise, stations in C5 and C7 can be described as *strong morning attractor—later afternoon and evening producer*. Similar to the difference between C3 and C4, the bike stations in C5 serve as trip attractor during the entire morning. We also notice that the bike stations in certain clusters (e.g., C2, C3, C6, and C7) have opposite directions of $netflow_q$ at time interval 12 and 13. According to Figure 4, there is a considerable amount of *ND*, *DN*, and *DD* trip chain segments around noon time. It is likely that certain individuals left their workplaces for particular activities (e.g., went to restaurants or returned to home), and then went back to their workplace.

We next examine the spatial distributions of the seven clusters. As shown in Figure 12, the bike stations in C1 are widely spread across different districts in Shenzhen. The bike stations are likely to be located at places with mixed land use patterns. The incoming and outgoing trips allocated to these stations are balanced with each other during the entire day. C3, C4 and C6 correspond to *morning producer—late afternoon and evening attractor*. Similar to C1, the bike stations in C6 are widely distributed across Shenzhen. However, we notice that the bike stations in C3 and C4 have a general north–south divide. The stations in C3 are mainly located in Guan Wai, and the ones in C4 are mainly distributed in Guan Nei. As described previously, Guan Wai covers mainly industrial-oriented areas with many migrant workers, while Guan Nei offers more diverse employment opportunities such as technology, commerce, education. The spatial and temporal patterns of C3 and C4 suggest that people in Guan Wai have more rigid work hours than people in Guan Nei. Hence, planners should expect different bicycle usage patterns between the bike stations in C3 and in C4. For example, the bike stations in C3 would need more free bicycles than open docks in the early morning, while bike stations in C4 need to have sufficient number of bicycles during the entire morning to satisfy the outgoing trips.
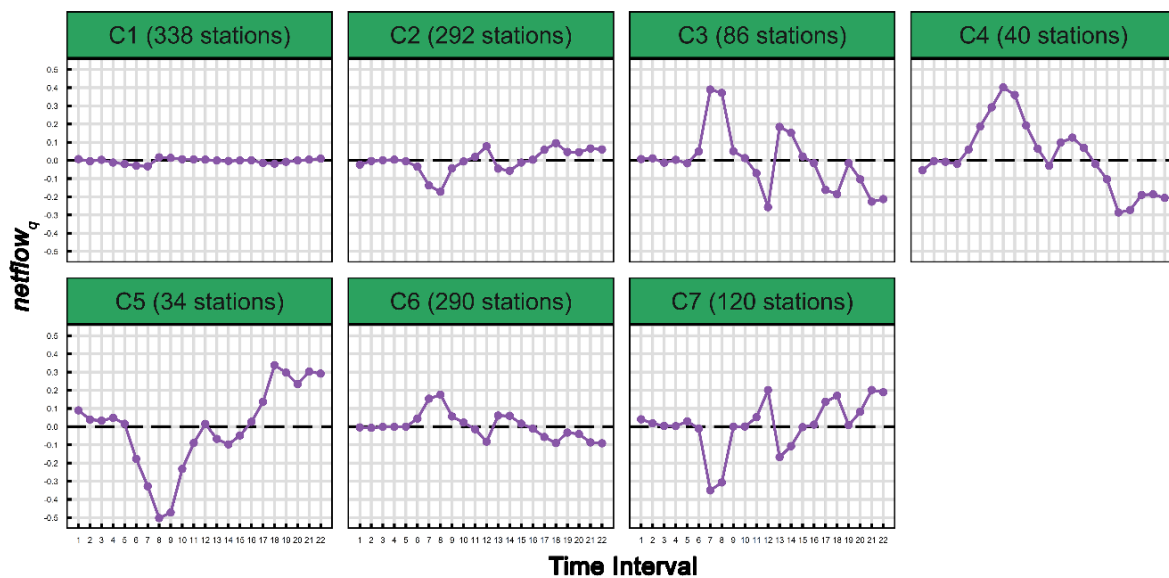

Figure 11. Temporal patterns of $netflow_q$ of the seven clusters derived from k-means clustering algorithm.

C2, C5 and C7 correspond to *morning attractor—late afternoon and evening producer*. The bike stations in these clusters (especially C5 and C7) should have enough free bike docks in the morning and an adequate number bicycles in the evening. Similarly, there is a north–south divide of the distribution patterns of C5 and C7. The bike stations in C5 cover the major employment and commercial centers in Guan Nei (see Figure 5C). The ones in C7 are mainly located in Guan Wai. In general, the difference of people's travel patterns between Guan Nei and Guan Wai should be regarded as an important factor for the planning and operation of bike stations in Shenzhen.

The temporal patterns of *netflow* of the seven clusters and their geographic distributions reveal an asymmetry of human travel patterns in Shenzhen. Such information can be valuable to decision making. For example, the suggested locations with "mixed" patterns could be good candidates for placing bike sharing stations since the incoming and outgoing trips tend to balance with each other during a day. For other suggested locations where incoming and outgoing trips have an imbalance, potential costs and strategies of allocating bicycles can be evaluated before the bike sharing stations are selected. Furthermore, the seven clusters can be overlaid with other data sources (e.g., land use data) to gain a deeper understanding of how urban flows are shaped by various characteristics of the

built environment such that the findings can be generalized to other cities in support of urban and transportation planning.
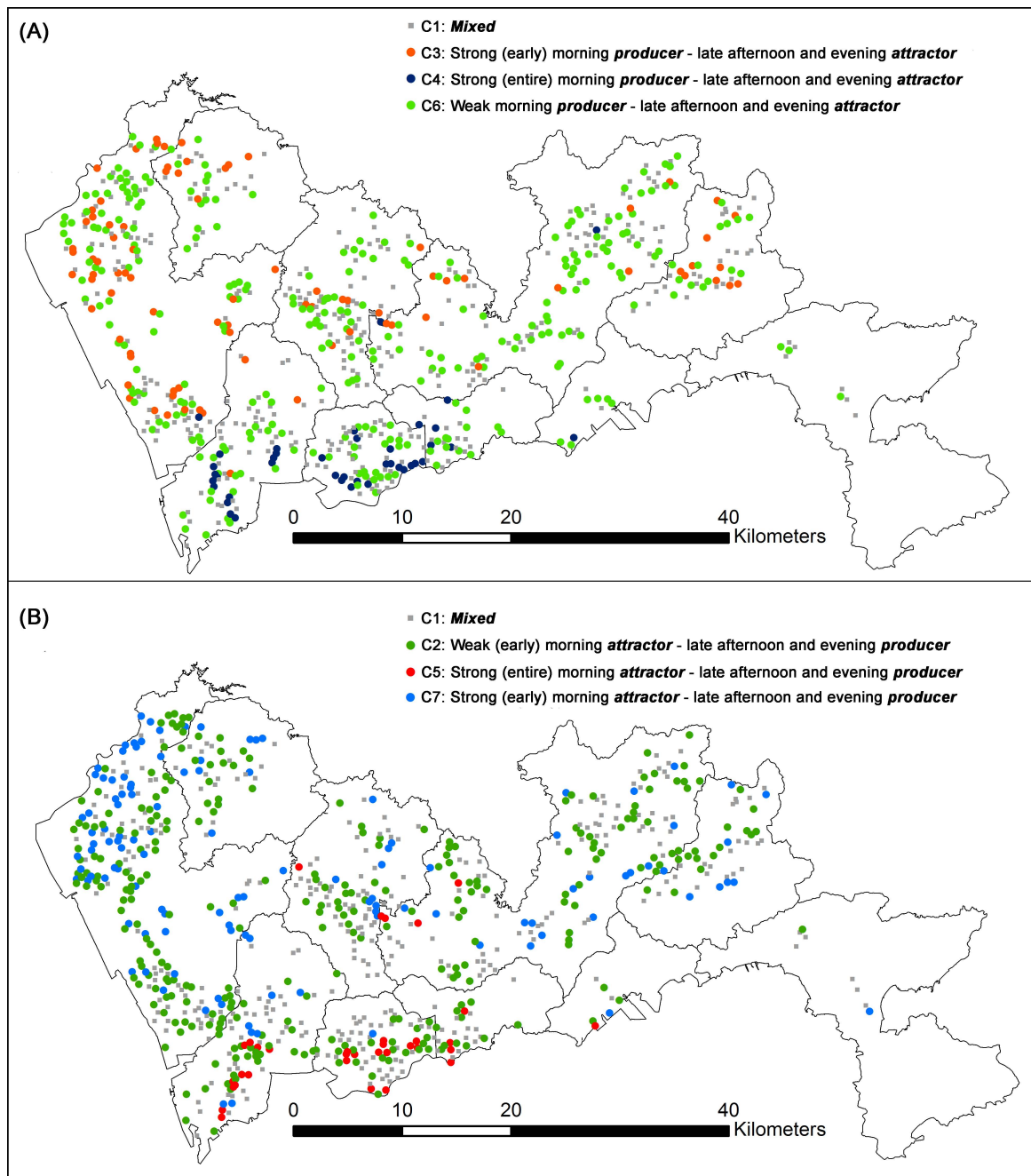

Figure 12. (**A**) Spatial distributions of C1, C3, C4 and C6; and (**B**) spatial distributions of C1, C2, C5, and C7.

6. Conclusions

Using Shenzhen, China as a case study, this research demonstrates how large scale mobile phone data can be used to uncover potential demand of bicycle trips in a city, and to provide suggestions to the locations of bike sharing stations. By identifying two important anchor points (night-time anchor point [*NTA*] and day-time anchor point [*DTA*]) from individual cellphone trajectories, we propose an anchor-point based trajectory segmentation method to partition the cellphone trajectories into trip chain segments. These trip chain segments refer to the tours that start and end at individuals' major

activity locations (e.g., home and workplace), and serve as the basic elements for estimating potential bicycle trips. Two indicators, *inflow* and *outflow*, are generated at the cellphone tower level to estimate potential demands of incoming and outgoing trips at different places in the city and different times in a day. The two indicators reflect the intensity and daily rhythms of people's short distance trips at a relatively fine spatial resolution.

By applying a maximum coverage location-allocation model, we offer suggestions to the locations of bike sharing stations under four different scenarios. The solution with 300 bike stations (*N = 300*) covers a considerable proportion (40.2%) of the total demand in the city. As *N* increases from 300 to 1200, the percentage of demand covered increases from 40.2% to 84.6%. However, the average accessibility of bike stations in districts where potential demands concentrate in a few areas (e.g., Futian, Longhua, and Nanshan) has a diminishing return as *N* becomes larger. Bike stations in districts where potential demands are more uniform over space (e.g., Longgang and Pingshan) have steady improvements (of accessibility) as *N* gets larger.

A k-means algorithm is performed to distinguish the dynamic relationships between the incoming and outgoing trips allocated to the bike stations. Seven clusters (C1 to C7) are derived to illustrate the unique characteristics of these bike stations (using *N = 1200* as an example). C1 refers to the bike station locations with *mixed* travel patterns. These locations could be good candidates for placing bike stations due to the balance of incoming and outgoing trips throughout the entire day. C3, C4 and C6 are identified as *morning producer—late afternoon and evening attractor*, which means at these stations, more bicycles should be available in the morning to satisfy the outgoing trips. C2, C5 and C7 refer to *morning attractor—late afternoon and evening producer*. These stations should have more open docks in the morning to absorb the incoming trips. Note that stations in C3, which are mainly located in the northern part of Shenzhen (Guan Wai), serve as a trip producer only in the early morning. While the ones in C4, which are mainly located in the south (Guan Nei), serve as a trip producer during the entire morning. A similar difference is observed between C5 and C7. The temporal difference of human travel patterns between the north and south, which is potentially related to the local industry and employment structures, should be regarded as an important factor for the planning of bike sharing stations in Shenzhen.

There are several aspects of this study that could be further enhanced in future studies. First, the current research is conducted using mobile phone data collected on a weekday. It would be beneficial to analyze mobile phone data collected on both weekdays and weekends to gain a more comprehensive view of potential bicycle trip demands in a city. Second, as the sampling rate of this mobile phone dataset is one hour, it is possible that the InTransit locations defined in our analysis do not capture all intermediate stops of people's daily trip chains. It means that the current analysis may underestimate the potential demand of bicycle trips at some intermediate stops that are not captured by the mobile phone data. Incorporating mobile phone data covering a longer time period (e.g., several months) can improve the identification of these intermediate stops based on the repetitive patterns of individual travel behavior. Third, the suggestions for placing bike sharing stations are provided based on the potential demand derived from mobile phone data. Other factors such as land topography, safety, current infrastructures of bike lanes, and connectivity to nearby transit stations [3,58] should be considered in future studies to further evaluate the suitability of specific bike station locations. In sum, this study enhances our understanding of the spatial-temporal dynamics of potential bicycle trips in Shenzhen. The proposed methods can be applied to mobile phone data and similar data sources collected in other cities to support intelligent spatial decisions in public transportation planning.

References

1. Cervero, R.; Duncan, M. Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area. *Am. J. Public Health* **2003**, *93*, 1478–1483. [CrossRef] [PubMed]

2. Barnes, G.; Krizek, K. Estimating bicycling demand. *Transp. Res. Rec. J. Transp. Res. Board* **2005**. [CrossRef]

3. Larsen, J.; Patterson, Z.; El-Geneidy, A. Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. International. *J. Sustain. Transp.* **2013**, *7*, 299–317. [CrossRef]

4. Candia, J.; González, M.C.; Wang, P.; Schoenharl, T.; Madey, G.; Barabási, A.-L. Uncovering individual and collective human dynamics from mobile phone records. *J. Phys. A Math. Theor.* **2008**, *41*, 224015. [CrossRef]

5. Ahas, R.; Silm, S.; Järv, O.; Saluveer, E.; Tiru, M. Using mobile positioning data to model locations meaningful to users of mobile phones. *J. Urban Technol.* **2010**, *17*, 3–27. [CrossRef]

6. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.

7. Becker, R.; Cáceres, R.; Hanson, K.; Isaacman, S.; Loh, J.M.; Martonosi, M.; Rowland, J.; Urbanek, S.; Varshavsky, A.; Volinsky, C. Human mobility characterization from cellular network data. *Commun. ACM* **2013**, *56*, 74–82. [CrossRef]

8. Xu, Y.; Shaw, S.L.; Zhao, Z.; Yin, L.; Fang, Z.; Li, Q. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation* **2015**, *42*, 625–646. [CrossRef]

9. Xu, Y.; Shaw, S.L.; Zhao, Z.; Yin, L.; Lu, F.; Chen, J.; Fang, Z.; Li, Q. Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Ann. Am. Assoc. Geogr.* **2016**, *106*, 489–502.

10. Strathman, J.G.; Dueker, K.J.; Davis, J.S. Effects of household structure and selected travel characteristics on trip chaining. *Transportation* **1994**, *21*, 23–45. [CrossRef]

11. McGuckin, N.; Zmud, J.; Nakamoto, Y. Trip-chaining trends in the United States: Understanding travel behavior for policy making. *Transp. Res. Rec. J. Transp. Res. Board* **2005**, *1917*, 199–204. [CrossRef]

12. Golob, T.F.; Hensher, D.A. The trip chaining activity of Sydney residents: A cross-section assessment by age group with a focus on seniors. *J. Transp. Geogr.* **2007**, *15*, 298–312. [CrossRef]

13. Institute for Transportation & Development Policy. The Bike-Share Planning Guide. Available online: https://www.itdp.org/wp-content/uploads/2014/07/ITDP_Bike_Share_Planning_Guide.pdf (accessed on 25 October 2015).

14. DeMaio, P.J. Smart bikes: Public transportation for the 21st century. *Transp. Q.* **2003**, *57*, 9–11.

15. Wang, Z.-G.; Kong, Z.; Xie, J.-H.; Yin, L.-E. The 3rd generation of bike sharing systems in Europe: Programs and implications. *Urban Transp. China* **2009**, *4*, 7–12.

16. Shaheen, S.; Guzman, S.; Zhang, H. Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transp. Res. Rec. J. Transp. Res. Board* **2010**. [CrossRef]

17. DeMaio, P. Bike-sharing: History, impacts, models of provision, and future. *J. Public Transp.* **2009**, *12*, 41–56. [CrossRef]

18. Harkey, D.; Reinfurt, D.; Knuiman, M. Development of the bicycle compatibility index. *Transp. Res. Rec. J. Transp. Res. Board* **1998**, *1636*, 13–20. [CrossRef]

19. Iacono, M.; Krizek, K.J.; El-Geneidy, A. Measuring non-motorized accessibility: Issues, alternatives, and execution. *J. Transp. Geogr.* **2010**, *18*, 133–140. [CrossRef]

20. Porter, C.; Suhrbier, J.; Schwartz, W. Forecasting bicycle and pedestrian travel: State of the practice and research needs. *Transp. Res. Rec. J. Transp. Res. Board* **1999**, 94–101. [CrossRef]

21. Landis, B.W. Bicycle system performance measures. *ITE J.* **1996**, *66*, 18–26.

22. Clark, D. Estimating future bicycle and pedestrian trips from a travel demand forecasting model. In Proceedings of the 67th Annual Meeting of the Institute of Transportation Engineers, Washington, DC, USA, 14 July 1997.

23. Rybarczyk, G.; Wu, C. Bicycle facility planning using GIS and multi-criteria decision analysis. *Appl. Geogr.* **2010**, *30*, 282–293. [CrossRef]

24. Wardman, M.; Tight, M.; Page, M. Factors influencing the propensity to cycle to work. *Transp. Res. A Policy Pract.* **2007**, *41*, 339–350. [CrossRef]

25. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [CrossRef] [PubMed]

26. Bayir, M.A.; Demirbas, M.; Eagle, N. Mobility profiler: A framework for discovering mobility profiles of cell phone users. *Pervasive Mobile Comput.* **2010**, *6*, 435–454. [CrossRef]

27. Phithakkitnukoon, S.; Horanont, T.; di Lorenzo, G.; Shibasaki, R.; Ratti, C. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Human Behavior Understanding*; Springer: Berlin, Germany, 2010; pp. 14–25.

28. Song, C.; Qu, Z.; Blumm, N.; Barabási, A.-L. Limits of predictability in human mobility. *Science* **2010**, *327*, 1018–1021. [CrossRef] [PubMed]

29. Calabrese, F.; Diao, M.; di Lorenzo, G.; Ferreira, J.; Ratti, C. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transp. Res. C Emerg. Technol.* **2013**, *26*, 301–313. [CrossRef]

30. De Montjoye, Y.-A.; Hidalgo, C.A.; Verleysen, M.; Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. *Sci. Rep.* **2013**, *3*. [CrossRef] [PubMed]

31. Ahas, R.; Aasa, A.; Mark, Ü.; Pae, T.; Kull, A. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tour. Manag.* **2007**, *28*, 898–910. [CrossRef]

32. Ratti, C.; Sevtsuk, A.; Huang, S.; Pailer, R. *Mobile Landscapes: Graz in Real Time*; Springer: Berlin, Germany, 2007.

33. Reades, J.; Calabrese, F.; Ratti, C. Eigenplaces: Analysing cities using the space-time structure of the mobile phone network. *Environ. Plan. B Plan. Des.* **2009**, *36*, 824–836. [CrossRef]

34. Vieira, M.R.; Frias-Martinez, V.; Oliver, N.; Frias-Martinez, E. Characterizing dense urban areas from mobile phone-call data: Discovery and social dynamics. In Proceedings of the Social 2010 IEEE Second International Conference on Computing (SocialCom), Minneapolis, MN, USA, 20–22 August 2010.

35. Iqbal, M.S.; Choudhury, C.F.; Wang, P.; González, M.C. Development of origin-destination matrices using mobile phone call data. *Transp. Res. C Emerg. Technol.* **2014**, *40*, 63–74. [CrossRef]

36. Alexander, L.; Jiang, S.; Murga, M.; González, M.C. Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. C Emerg. Technol.* **2015**, *58*, 240–250. [CrossRef]

37. Dong, H.; Wu, M.; Ding, X.; Chu, L.; Jia, L.; Qin, Y.; Zhou, X. Traffic zone division based on big data from mobile phone base stations. *Transp. Res. C Emerg. Technol.* **2015**, *58*, 278–291. [CrossRef]

38. Wang, P.; Hunter, T.; Bayen, A.M.; Schechtner, K.; González, M.C. Understanding road usage patterns in urban areas. *Sci. Rep.* **2012**. [CrossRef] [PubMed]

39. Martinez, L.M.; Caetano, L.; Eiró, T.; Cruz, F. An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon. *Procedia Soc. Behav. Sci.* **2012**, *54*, 513–524. [CrossRef]

40. García-Palomares, J.C.; Gutiérrez, J.; Latorre, M. Optimizing the location of stations in bike-sharing programs: a GIS approach. *Appl. Geogr.* **2012**, *35*, 235–246. [CrossRef]

41. Rushton, G. *Optimal Location of Facilities*; COM Press: Wentworth, NH, USA, 1979.

42. Hakimi, S.L. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Oper. Res.* **1965**, *13*, 462–475. [CrossRef]

43. Suzuki, A.; Drezner, Z. The p-center location problem in an area. *Locat. Sci.* **1996**, *4*, 69–82. [CrossRef]

44. Toregas, C.; Swain, R.; ReVelle, C.; Bergman, L. The location of emergency service facilities. *Oper. Res.* **1971**, *19*, 1363–1373. [CrossRef]

45. Church, R.; Velle, C.R. The maximal covering location problem. *Pap. Reg. Sci.* **1974**, *32*, 101–118. [CrossRef]

46. Shenzhen Daily. "Shenzhen: Most Crowded in China". 2012. Available online: http://szdaily.sznews.com/html/2012-05/30/content_2063502.htm (accessed on 3 October 2015).

47. Transport Commission of Shenzhen Municipality. Guidelines of Transport Planning and Design for Pedestrian and Bicycle Systems in Shenzhen. 2011. Available online: http://www.szpl.gov.cn/xxgk/ztzl/zxcgh/jtghcgg.pdf (accessed on 31 August 2015).

48. Dijst, M. Two-earner families and their action spaces: A case study of two Dutch communities. *GeoJ* **1999**, *48*, 195–206. [CrossRef]

49. Schönfelder, S.; Axhausen, K.W. Activity spaces: Measures of social exclusion? *Transp. Policy* **2003**, *10*, 273–286. [CrossRef]

50. Csáji, B.C.; Browet, A.; Traag, V.A.; Delvenne, J.-C.; Huens, E.; Van Dooren, P.; Smoreda, Z.; Blondel, V.D. Exploring the mobility of mobile phone users. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 1459–1473. [CrossRef]

51. Isaacman, S.; Becker, R.; Cáceres, R.; Martonosi, M.; Rowland, J.; Varshavsky, A.; Willinger, W. Human mobility modeling at metropolitan scales. In Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, Ambleside, UK, 25–29 June 2012.

52. Hägerstrand, T. What about people in regional science? *Pap. Reg. Sci.* **1970**, *24*, 7–24. [CrossRef]

53. Ye, X.; Pendyala, R.M.; Gottardi, G. An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transp. Res. B Methodol.* **2007**, *41*, 96–113. [CrossRef]

54. Xu, N.; Ling, Y.; Jinxing, H. Identifying Home-work locations from short-term, large-scale, and regularly sampled mobile phone tracking data. *Geomat. Inf. Sci. Wuhan Univ.* **2014**, *39*, 750–756.

55. Long, Y.; Zhang, Y.; Cui, C. Identifying commuting pattern of Beijing using bus smart card data. *Acta Geogr. Sin.* **2012**, *67*, 1339–1352.

56. Hansen, W.G. How accessibility shapes land use. *J. Am. Inst. Plan.* **1959**, *25*, 73–76. [CrossRef]

57. Wei, L.; Yan, X. Transformation of "urban village" and feasible mode. *City Plan. Rev.* **2005**, *7*, 9–14.

58. Dill, J. Bicycling for transportation and health: The role of infrastructure. *J. Public Health Policy* **2009**, *30*, S95–S110. [CrossRef] [PubMed]