Transportation Research Part A 105 (2017) 154–166

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

An optimization approach for the placement of bicycle-sharing stations to reduce short car trips: An application to the city of Seoul

Chung Park, So Young Sohn*

Department of Information & Industrial Engineering, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Republic of Korea

A R T I C L E   I N F O

Keywords:
Bicycle-sharing stations
Short trip of cars
Location-allocation model
Spatial pattern analysis
Spatial big data

A B S T R A C T

Substantial motor vehicle exhaust, a primary cause of air pollution, is emitted on short car trips of three miles or less. Bicycles have been considered an optimum means of completing these short trips because the bicycle is an environmentally friendly, economical, and convenient vehicle. Accordingly, many countries have adopted public bicycle-sharing systems to reduce the use of private vehicles for short trips in central downtown areas. In this paper, we propose a new framework, based on taxi trajectory data, for locating bicycle-sharing stations most efficiently to replace short automobile trips. The proposed framework is applied to Gangnam-*gu*, a district within the city of Seoul, Korea. Results using two different location-allocation models are demonstrated. As expected, when the *p-median* model was implemented, the selected stations were more scattered over the whole district, whereas when the MCLP model was implemented, the stations were more concentrated on central areas. Our approach is applicable to any city considering a bicycle-sharing system and can contribute to the system's efficiency in improving environmental conditions in a central downtown area.

1. Introduction

The number of private vehicles has been increasing globally due to rapid economic development and population growth, producing considerable environmental problems such as noise, traffic congestion, and air pollution (Dunlap and Jorgenson, 2012). According to the US auto magazine *Ward's Auto*, the number of cars owned worldwide has doubled every 15 years since the 1970s. This growth rate has not abated, as the World Trade Organization reported that global auto registrations have increased by 16% from 2010 to 2013. According to South Korea's National Statistical Office, the number of automobile registrations increased from about 1 million in 1985 to 25 million in 2016.

Substantial motor vehicle exhaust, a primary cause of air pollution, is emitted on short trips of three miles or less. Public bicycle-sharing systems have been adopted to reduce the use of private vehicles for short trips in central downtown areas (Braun et al., 2016; Lin and Yang, 2011). Bicycling is an efficient means of mobility over short distances that essentially produce no externalities such as air pollution and traffic congestions (Zahabi et al., 2016).

For several reasons, a bicycle-sharing system is particularly well suited to replace short-distance automobile trips. First, bicycles can serve a similar purpose to private cars for short trips such as shopping and commuting and are more efficient in downtown areas. Second, there is only a negligible speed difference between bicycles and automobiles in downtown settings. Jensen et al. (2010) analyzed 11 million bicycle trajectories in downtown Lyon, France, demonstrating that bicycle trips were actually faster than car travel. Whereas the average speed of cars in European cities varies from 10 to 15 km/h, Jensen et al. (2010) suggested that cyclists

* Corresponding author.
E-mail address: sohns@yonsei.ac.kr (S.Y. Sohn).

http://dx.doi.org/10.1016/j.tra.2017.08.019
Received 23 December 2016; Received in revised form 1 June 2017; Accepted 15 August 2017
0965-8564/

average 13.5–15 km/h. According to the Seoul Transport Operation and Information Service (TOPIS), the average speed of all automotive vehicles in downtown Seoul during 2015 was 17.9 km/h. Third, Martinez et al. (2012) recommended that bicycle-sharing systems be introduced as an alternative to public transportation or private cars because they have greatly progressed in the recent years and have become integrated with other modes of transport in urban areas. Moreover, riding bicycles reduces carbon dioxide emissions and helps to ease urban traffic congestion, since bicycles need less road space than cars. In sum, bicycles have been considered an optimum means of short to distance travel because they are environmentally friendly, economical, and convenient (Liu et al., 2012).

Based on these advantages, many countries have sought to expand their bicycling infrastructures and particularly to establish bicycle-sharing systems (Lin and Yang, 2011). Among the many examples, Paris started "Velib," the world's largest bicycle-sharing program with more than 20,000 short to term rental bikes, in 2007. In the same year, Barcelona, Spain introduced "Bicing," with about 6000 bikes distributed at 400 stations across the entire city (Kaltenbrunner et al., 2010). Various U.S. cities, such as New York and Washington, have implemented similar programs. Seoul has expanded its public bicycle-sharing system, "Seoul to bike," to about 5600 bikes in 460 stations.

Although policies related to bicycle-sharing systems have mainly aimed to reduce private vehicle trips (Braun et al., 2016), primarily in downtown areas, no study has examined the systems' effectiveness. Therefore, we propose a new framework for locating bicycle-sharing stations most efficiently to replace short automobile trips. Moreover, potential demand for bicycles has normally been considered static and deterministic in previous studies (Lin and Yang, 2011; Lin et al., 2013). In the present study, we use real-scale floating population data recorded on an hourly basis, to estimate potential demand more accurately. These data come from Geo-vision, a subsidiary of SK Telecom, which has the majority of subscribers in Korea. Geo-vision can report the number of people who are present in each 50 m × 50 m cell of area by sensing the location of mobile phone signals every hour. Our case study looks at Gangnam-*gu*, an administrative division of the city of Seoul, Korea that has no bicycle-sharing stations and suffers from severe traffic congestion.

In Section 2, we review previous studies on bicycle sharing. In Section 3, we introduce the data and method used in this study, followed by the results and empirical analysis in Section 4. Section 5 presents conclusions, suggestions for future studies and limitations of this study.

2. Literature review

Facility location is one of the most important factors in strategic decision making by the public sector (Frade and Ribeiro, 2015). Previous studies on where to locate bicycle-sharing stations have used various approaches. Models have been developed in the field of operations research (Frade and Ribeiro, 2015; Lin et al., 2013; Shu et al., 2013; García-Palomares et al., 2012; Lin and Yang, 2011). Frade and Ribeiro (2015) applied a maximum covering location approach to design the bicycle-sharing system in such a way as to maximize the demand covered by selected bicycle stations and satisfy the available budget as a constraint. García-Palomares et al. (2012) proposed a GIS to based methodology to determine the spatial distribution of potential trip demand, and they selected bicycle station sites by using two location to allocation models, the minimum impedance model (*p-median*) and maximum coverage location problem (MCLP). Their methods of estimating travel demand were quite remarkable, in that they conducted GIS analysis to detect the distribution of potential user demand.

Lin et al. (2013) suggested a mathematical model of bicycle-sharing stations with stipulated service levels and bicycle stocks. Given data on origins, destinations, bicycle candidate stations, and travel demands from origins to destinations, they determined the location of bicycle stations and lanes to minimize the sum of travel costs on networks from origins to destinations, fixed costs, bicycle inventory costs, and penalty costs for uncovered demands. Lin and Yang (2011) devised a mathematical model that considered users' travel costs, the setup costs for bicycle lanes, the facility costs of bike stations, and the availability rate for bicycle pick-ups at stations. In addition, Shu et al. (2013) introduced a network flow model for effective allocation of bicycles to sharing stations. Given the locations of bicycle stations, they calculated the flow of bicycles within the network and the number of trips and demonstrated the efficiency of periodic redistribution of bicycles.

Approaches based on connecting bicycle users to public transportations or private cars have also been introduced. Liu et al. (2012) suggested a new scheme for a Beijing public bicycle system by referring to worldwide examples of system management and analyzed why the first attempt to establish such a system in Beijing failed. Romero et al. (2012) proposed a method to consider private automobiles and public bicycles simultaneously using the genetic algorithm. They indicated that greater travel distance to the stations makes public bicycles uncompetitive in comparison to the private cars.

How to balance the bicycle-sharing stations to prevent them from becoming too full or too empty has also been debated. Kloimüllner et al. (2014) aimed to reach specified levels at the end of the process, so that the selected stations would fulfill demand and keep any stations from being empty or completely full of bicycles during the rebalancing process, which would result in user dissatisfaction. Erdoğan et al. (2015) proposed the Static Bicycle Rebalancing Problem, seeking not only to decide the amount of bicycles to be assigned to each station but also to minimize the cost of stations to be stopped by a few vehicles. Papazek et al. (2014) developed a methodology to minimize the absolute difference between target and final utilization levels at all bicycle-sharing stations. Rainer-Harbach et al. (2013) introduced a redistribution method between bicycle stations to keep them from becoming empty by using Variable Neighborhood Search (VNS) with an embedded Variable Neighborhood Descent (VND).

However, none of these research studies considered the problem of reducing short trips in private vehicles. In view of this gap in the literature, the present paper proposes a framework for locating bicycle-sharing stations most efficiently so as to reduce such car trips. In addition, we conduct an empirical test of this framework in Gangnam-*gu*, an administrative division of Seoul, Korea.

Table 1
Framework of this study.

Step 1. Select candidate sites	Step 2. Extract potential demand sites	Step 3. Solve the models
√ Derive the tax trajectory data in which driving patterns are similar to bicycle travel behavior	√ Select the demand sites by considering the number of floating population	√ Select optimum sites by solving the location-allocation models. √ Determine the desired capacity of selected bicycle stations

3. Methodology

Our research suggests the following framework for locating bicycle-sharing stations. In step 1, we select candidate sites for bicycle stations; in step 2, we extract potential demand sites; and in step 3, we solve the location to allocation problem to determine the location of bicycle stations. Table 1 further describes the tasks involved at each step.

3.1. Select candidate sites for bicycle-sharing stations

As explained above, we aim to reduce the use of private cars for short trips by establishing bicycle-sharing stations. Therefore, the trajectory data for private vehicles with driving patterns similar to those of bicycles should be extracted.

There are several key characteristics of bicycle riding in any setting, such as time, weather, and distance traveled. In the case of Barcelona, Kaltenbrunner et al. (2010) calculated the average number of bicycles available at the bicycle-sharing stations, showing that the bicycle use rate was highest between 8:00 and 10:00 a.m., between 2:00 and 4:00 p.m., and after 7:00 p.m. on weekdays. Nankervis (1999) found that the greater the likelihood of rain, the fewer commuters ride bicycles. Flynn et al. (2012) verified the relationship between commuting to work by bicycle and weather. They showed that rain, snow, and extreme temperatures had strong negative influences on commuting to work by bicycle. Since the characteristics of bicycle use patterns vary across countries, policymakers should extract trajectory data for private vehicles that are similar to bicycle riding patterns in their country in terms of time, weather condition, and driving distance. Accordingly, we extracted pick-up locations from the taxi trajectory to reflect typical bicycle riding patterns. Finally, after calculating the frequency of taxi riding at each pick-up location, we would be able to identify the top to value sites for bicycle-sharing stations. Policymakers can adjust the number of sites.

We used trajectory data from taxis instead of from private vehicles for three reasons (Cai et al., 2016). First, due to privacy issues related to car travel, most studies of vehicle travel patterns have been based on taxi trajectory data. Although taxi trajectories differ from private vehicle travel patterns, they exhibit similar traits to the travel patterns of individuals. Second, taxis can be regarded as pervasive sensors that provide valuable information about mobility dynamics and traffic conditions. Third, taxi trajectories typically offer more detailed information with more sophisticated spatial to temporal solutions and have larger sample sizes than other datasets used in previous studies.

3.2. Extract potential demand sites

We selected the potential demand sites for bicycles from among the following places, consistent with prior literature: (1) metro stations, (2) shopping centers, (3) parks and (4) residences.

García-Palomares et al. (2012) proposed that the required bicycle stations should be located at train and metro stations to connect users to public transportation networks. Krykewycz et al. (2010) tested a preference score approach to test potential users' preferences with regard to a bicycle-sharing system. They used three categories of input factors to calculate the preference scores: network/facility factors, trip attraction factors, and trip origin factors. They contended that bicycle demand was positively associated with those factors. The network/facility factors include proximity to rail stations, bicycle to friendly streets, and bus stops. Based on these previous studies, we chose to consider the population near metro stations as the source of potential demand for a bicycle-sharing system. Krykewycz et al. (2010) also suggested that preference for riding a bicycle is determined by trip attraction factors, which indicate proximity to shopping centers, parks, and recreation areas. This reinforced the justification for using shopping centers and parks as potential demand sites.

We also considered residences such as apartment complexes as potential demand sites for public bicycle use because many people use bicycles to commute or exercise from home (Flynn et al., 2012).

García-Palomares et al. (2012) created a layer of points containing the number of people related to each building to assess potential bicycle demand in their study. Li et al. (2015) stated that since the spatial accessibility of facilities varies by type, different radii should be chosen when drawing the buffer zone for different infrastructure types. They suggested that the radius of the buffer zone should be 1 km for metro stations and 3 km for parks, using planning criteria from the Shanghai Planning Bureau as their reference. In this research, we constructed potential demand sites by creating buffer zones with radii of 100 m around metro stations and shopping centers, 200 m around residences, and 300 m around parks. Policymakers depending on their own city's characteristics can determine the radii for a specific system. As shown in Fig. 1, we summed up the total floating population in the buffer zone, creating the demand sites (cells). At each demand site, the meaning of the floating population is somewhat different. In cases of metro stations, shopping centers, and parks, floating population means the average number of people who are moving in real time,

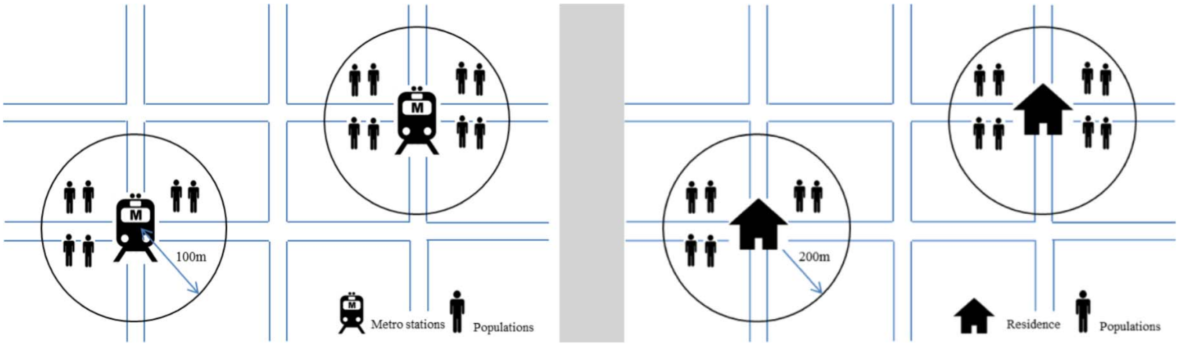

Fig. 1. Illustrative example of creating demand sites (cells).

excluding the population residing in a specific space, on an hourly basis. In cases of residential areas, the data on floating population include both the population residing in residential areas as well as the people moving within the cell.

3.3. Solve the models

The location to allocation model is a widely used method in planning service for regional development (Rahman and Smith, 2000). This model is used to find where facilities should be located to meet some predefined objective. Several methods are commonly used to solve the location to allocation model, such as minimum impedance (*p*- median) and the maximum coverage location problem (MCLP). This study used *p-median* and MCLP models, separately for the following reasons.

First, previous studies using bi-level optimization processes or optimization of multi-attribute functions are mostly based on simulations using a small size of data (Lin et al., 2013; Lin and Yang, 2011; Romero et al., 2012), because those functions have so many objective functions, variables and constraints. However, in this study, we deal with spatial big data such as floating population and taxi trajectory data. Therefore, we use the most representative and relatively simple models to make computing more efficient. Indeed, the *p-median* and MCLP models used in this study were utilized for empirical analysis in other studies (Bandyopadhyay and Singh, 2012; García-Palomares et al., 2012).

Second, this study has purposes to present a framework that policy makers can refer to selecting public bicycle locations to reduce the driving distance of cars. Therefore, we used *p-median* and MCLP which are relatively convenient and intuitive models. Among these models, policy makers can select one model upon their discretion. If they want citizens to use the facility at a lower cost, then they can choose *p-median*. If they want more people to use the facility, then they can choose MCLP (Gwak et al., 2017). García-Palomares et al. (2012) concluded that the *p-median* model is more advantageous with respect to spatial equity since a uniform coverage is created, but in some areas, inefficiency would be generated due to low potential demand covered by selected stations. In addition, they concluded that the MCLP model is better in terms of efficiency, since this model maximizes the demand covered by selected stations.

To explain the design of this mathematical model, we first introduce the following notations for our models, expressing the variables and parameters.

Notations of models

Subscripts and sets	
i	index of demand points, $i = 1, \ldots, n$
j	index of supply points, $j = 1,\ldots, m$
Parameters	
d_{ij}	the Euclidean distance between site i and j
D	limited range within which demand site can be covered by supply site
N_i	a set of supply site j within distance D from demand site i
u_i	the number of floating populations in the i th demand site (amount of demand)
p	the number of supply sites to be selected as bicycle-sharing stations (predetermined)
Decision variable	
(*p*-median model)	
x_{ij}	1, if the demand site i is assigned to outpost j
	0, otherwise
y_{ij}	1, if the supply site j is selected as an outpost (bicycle-sharing station)
	0, otherwise
(*MCLP* model)	
x_i	1, if the demand site i is satisfied

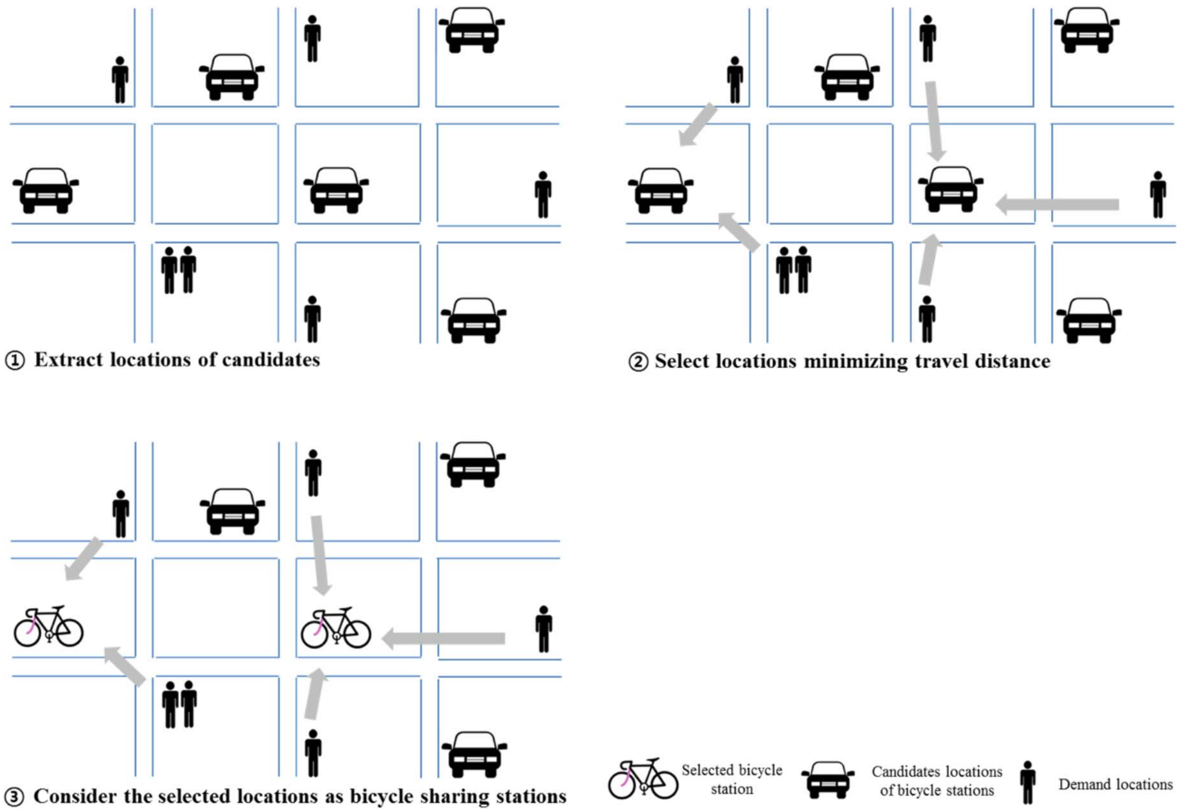

Fig. 2. Procedure for solving the minimum impedance model.

	0, otherwise
y_{ij}	1, if the supply site j is selected as an outpost (bicycle-sharing station)
	0, otherwise

As shown in Notation, this study uses the Euclidean distance instead of street distance in MCLP and *p-median* model. Most of the bicycle systems studied are based in urban areas, which typically have a high density of roads and intersections. This allows people to make journeys in any direction by using road, with the distance traveled unlikely to be far beyond the straight-line or Euclidean distance (O'brien et al., 2014). In addition, the bicycle user is less likely to take long detours than other transport users (e.g. car drivers and metro riders) who are more constrained within their road network in the urban area. For this reason O'brien et al. (2014) indicated that a Euclidean distance simplification in bicycle system is more likely to be valid for urban areas, with a higher road and intersection density. Therefore it is justified that both of Euclidean and street distance are directly proportional in urban area such as Seoul, that the street distances for bicycle trip are not significantly longer than Euclidean distance.

3.3.1. Minimum impedance (p-median) model

The minimum impedance (*p-median*) model aims to minimize the total distance (travel cost) between demand sites and their closest supplies or facilities (Bandyopadhyay and Singh, 2012). This factor is important because longer travel distance to stations makes public bicycles uncompetitive when compared to private cars (Romero et al., 2012). We considered the *p*- median model as follows. Given a set of sites of candidates for bicycle stations and the bicycle demand sites, we would like to know where we should locate the bicycle-sharing stations. The *p-median* model allocates the bicycle stations such that the sum of traveling distances from the demand sites to bicycle stations is minimized. Fig. 2 illustrates this procedure.

Based on the notation, the following mathematical models are formulated:

$$minimize:\ \sum_{i=1}^{n}\sum_{j=1}^{m}\mu_j d_{ij} x_{ij} \tag{1}$$

such that

$$\sum_{i=1}^{n} x_{ij} = 1\ forall\ i \tag{2}$$

$$x_{ij}, y_j = (0,1)$$

$$x_{ij} \leqslant y_j \text{ forall } i,j \tag{3}$$

$$\sum_{j=1}^{m} y_j = p \text{ forall } j \tag{4}$$

The objective in selecting the bicycle-sharing stations is to minimize the travel distance for each person to the demand sites, as shown in (1), where n is the number of candidate sites for bicycle-sharing stations in the cell and m is the number of potential demand sites. In other words, the objective function is the sum of bicycle users from all demand sites, weighted by the distance that they would have to travel to all possible sites of bicycle-sharing stations. The requirement that users in cell i must utilize specific one candidate site is formulated in constraint (2). Constraint (3) ensures that if the model solution does not assign users from demand site i to a supply site j, then anyone who is in demand site i should not use an bicycle station in supply site j.

The important factor in using the *p-median* model is to determine a reasonable number for p. Planners first consider how many bicycle stations are to be installed in a region. We suggest that it is reasonable to determine p by considering the number of residents in the region. We propose to determine p by multiplying the population of a specific region by the average ratio of bicycle stations per 1000 residents in other regions. The total number of stations (p) is used in constraint (4).

3.3.2. Maximum coverage location problem (MCLP)

The objective of the MCLP model is to maximize the demand covered by p facilities. The following mathematical model is formulated:

$$\text{Maximize: } \sum_{i=1}^{n} \mu_i x_i \tag{5}$$

such that

$$\sum_{j \in N_i} y_j \geqslant x_i \tag{6}$$

$$\sum_{j=1}^{m} y_j = p \tag{7}$$

$$x_i, y_j = (0,1)$$

$$N_i = \{j \in J | d_{ij} \leqslant D\} \tag{8}$$

To achieve its objective, the MCLP model seeks to maximize the coverage area of facilities, as expressed in (5), where n is the number of candidate sites for bicycle-sharing stations and m is the number of potential demand sites. At least one bicycle station must be within D from i to meet the demands of site i, as shown in constraint (6). The total number of stations (p) is calculated in constraint (7). Constraint (8) is related to parameter D. The distance between demand site i and supply site j that satisfies the demand site i is less than D, which is the maximum distance of potential demand to an supply site. In this paper, we consider D as 400 m because existing examples indicated that bicycle-sharing stations must be located within 300–500 m from major origins and destinations of traffic (Lin and Yang, 2011).

3.3.3. Determine the capacity of selected bicycle stations

By counting the frequency of taxi pick-ups at each station, we can determine the bicycle station capacity (i.e., the number of bicycles that a station can hold). Suppose that the government wants to introduce a total of h bicycles, considering budget constraints. Since the variation in frequency of taxi pick-ups can be extremely high between stations, it is necessary to reduce this variation.

To do so, we suggest clustering the selected stations by wth quantile in terms of frequency of taxi pick-up. Then we should determine a, the minimum number of bicycles in the wth cluster. This study suggests that the minimum number of bicycles should be more than 5. As the cluster quantile number decreases, the increment s is added to the prior cluster. In short, the formulation is as follows:

$$\frac{p}{w}\left\{\sum_{k=1}^{w} a + (k-1)s\right\} = h \tag{9}$$

where p is the number of selected stations and w is the quantile number. This indicates that p/w is equal to the number of bicycles per station in each cluster. In addition, a is the minimum number of bicycles in selected stations, s is the incremental number of bicycles, and h is the total number of bicycles proposed by the government. Through this equation, we can derive s, predetermining p, w, a, and h. p/w is the capacity of each bicycle station.

Fig. 3. Distribution of bicycle-sharing stations in Seoul.

4. An illustrative analysis

4.1. Features of Gangnam-gu, Seoul

In this research, we applied our method to an empirical analysis in Gangnam-*gu*, an administrative district of Seoul, Korea. Seoul is its country's capital and an economic hub of East Asia. The total surface area of Seoul is 605.2 km^2 and the city consists of 25 *gu* or administrative divisions. Seoul's population is about 10 million. Gangnam-*gu* is known around the world as a popular place for traveling and shopping, so large crowds concentrate there. There are 576,413 residents in Gangnam-*gu*, making it the fourth most populous district in Seoul. In addition, the average traffic volume per day in Gangnam-*gu* is about 1.89 million, or three times the average traffic volume of other administrative divisions and the average traffic speed is lower than the overall average speed in Seoul. Given these problems, it would be valuable to establish bicycle-sharing stations in Gangnam-*gu*, which currently has none. The size of the "Seoul-bike" sharing program has grown from its initial 200 bicycle stations to 650 and from 2000 bicycles to 5600 as of 2016, but as shown in Fig. 3, these stations are concentrated in the city's central zone.

Available data indicate that 3581 bicycles per day are rented in Seoul. The average ride covers about 2 miles and takes 27 min. These rentals include 416 bikes during the morning rush hour and 704 in the evening rush hour.

4.2. Select candidate sites for bicycle- sharing stations

We used a taxi trajectory dataset for February 2016, which is freely available from the Seoul Open Data Plaza (http://data.seoul.go.kr/). That source collects GPS data from taxis operating in Seoul every 30 min. The dataset used in this study is presented in Table 2. These data contained the coordinates of each taxi trajectory (i.e., the location of each pick-up and drop-off), the time, and the weather conditions on each 150 m of roadway. The unit size of 150 m is small enough to support the selection of optimum sites for the bicycle-sharing stations. In addition, since all taxi pick-up sites are near the roadside, they could be used as bicycle-sharing stations. Previous studies have had limitations with respect to the realistic feasibility of the suggested model.

In Seoul, there are many taxi stops, although taxis are also welcome to stop along most roadways (Lee and Sohn, 2017). Taxis can go to any place that is accessible by car. In addition, taxi fares in Seoul are relatively cheaper when compared with fares in large cities of other countries. The base fare of about $2.13 in 2016 ranked eighth among metro cities of the world. Therefore, taxis are very popular in Seoul, with an annual average of 6.8% (2014) of total traffic share according to Seoul Open Data Plaza. In Seoul, the average number of taxi rides used per day is 1,273,971, and the city's total population is 10,297,138 in 2015.

We selected the candidate sites using taxi trajectory data that exhibit similar patterns to the anticipated bicycle riding. As shown in Table 2, the taxi trajectory data used in this study include each traveler's pickup site and destination. We calculated the distance of

Table 2
Description of the taxi trajectory dataset in Seoul Open Data Plaza.

Data	Description
T_Link_ID	The location of the pick-up site (coordinates)
Destination	The location of the drop-off site (coordinates)
Time	The time when each taxi trajectory occurs
Weather	The weather conditions in which each taxi trajectory occurs

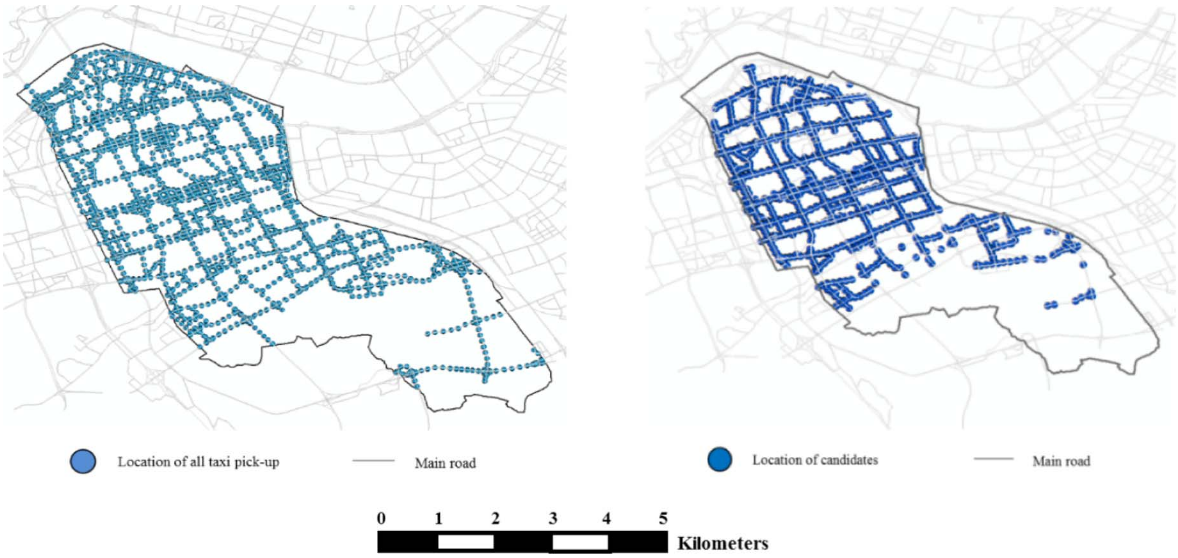

Fig. 4. The location of bicycle station candidates in Gangnam-*gu* (2664 reduced to 1461).

each travel by using these pick-up and drop-off sites, and then extracted trips of a distance no greater than three miles. In addition, we selected the trajectories in which the weather condition was not rainy and snowy and driving times between 6:00 a.m. and 10:00 p.m. to obtain data that have similar patterns to bicycle riding.

As a result, the number of candidate sites for bicycle-sharing stations was 2664. We calculated the frequency of pick-ups for each candidate and selected those with frequencies greater than 1000, reducing the number of final candidate locations to 1461. Fig. 4 illustrates the candidate locations. It shows that all candidate sites were on roadsides, making them sufficiently precise realistic potential locations for bicycle-sharing stations.

4.3. Extract potential demand sites

We used a dataset that indicates which includes the average number of floating populations counted by every 1 h in each unit of 50 m × 50 m, subdivided by age and gender. These data come from Geo-vision, a subsidiary of SK Telecom, which has the majority of subscribers in Korea. Geo-vision can report the number of people present in each 50 m × 50 m cell of area by sensing the location of mobile phone signals every hour. Here, a mobile signal means LTE and 3G transmission reception or telephone reception signals. We eliminated data on children and the elderly, because we assumed that people under age 10 and over age 80 would be unlikely to use a bicycle sharing system. The description of floating population data set used in this study is introduced in Table 3.

In Gangnam-*gu*, we extracted 126 demand sites: 86 residences (apartment complexes), 28 metro stations, 4 parks, and 8 shopping centers. This study focuses on apartment complexes because apartments are the primary type of residence in Gangnam-*gu*, as in all of Seoul. According to statistics maintained by the Korea National Statistical Office, 76.93% of the residential locations in Gangnam-*gu* are apartments and 67.27% of the total population resides in apartments. Therefore, we consider that large apartment complexes can reasonably be identified as a source of high demand for bicycles. Moreover, these residences are suitable for public facilities because each apartment complex contains at least 300 households. The residential types excluded from this study are single-family houses (4.5%) multi-family house (14.14%) and non-residential buildings (1.1%). These types of residences are unsuitable for public facilities because it tends to be widely spread out and has low population density. Therefore, public bicycle facilities already installed in other *gu* in Seoul are concentrated in large apartment complexes, central commercial districts, and university districts, all of which have high population density.

Table 3
Description of the floating population dataset in Geo-vision, by SK Telecom.

Data	Description
X_COORD	X coordinates of a specific 50 m × 50 m cell
Y_COORD	Y coordinates of a specific 50 m × 50 m cell
GNDR_CD	The number of females and males in a specific 50 m × 50 m cell
AGE_GR_SCTN_CD	The number of people age 10–80 in a specific 50 m × 50 m cell
WKDY_CD	The number of people at each day in a specific 50 m × 50 m cell
TMST_CD	The number of people counted every 30 min in a specific 50 m × 50 m cell
FLOW_POP_CNT	The total number of people in a specific 50 m × 50 m cell

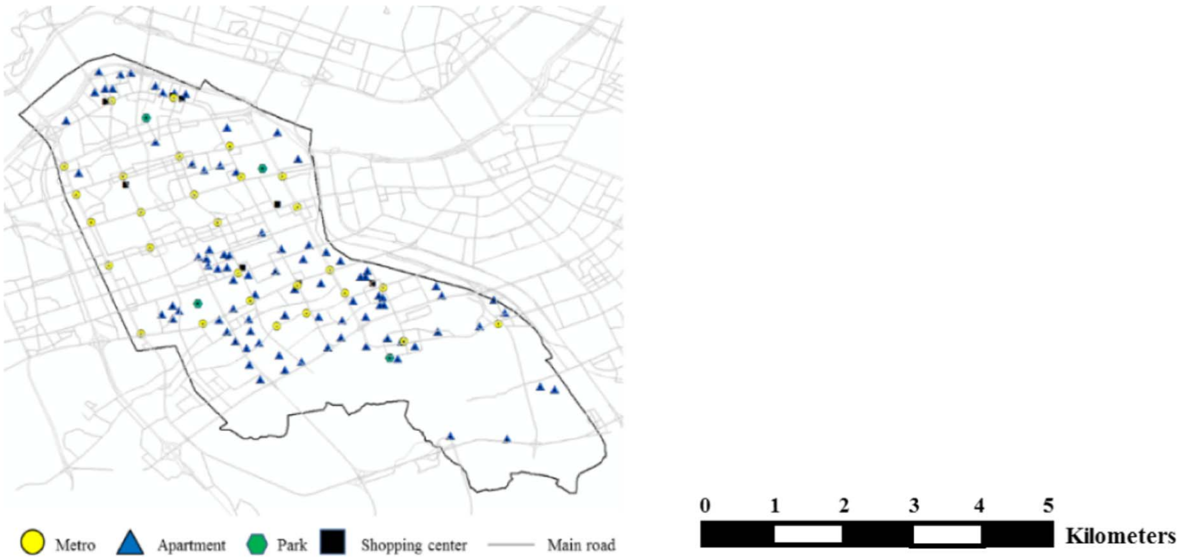

Fig. 5. Distribution of demand sites in Gangnam-*gu*.

Coordinate data for each demand site were obtained from the Seoul Big Data campus (https://bigdata.seoul.go.kr). Potential demand was calculated by creating a layer containing the number of floating population associated with each demand site. The layer sizes for each demand site were 100 m for metro stations and shopping centers, 200 m for residences, and 300 m for parks. Therefore, each demand site has an associated population value. Fig. 5 is illustrative map of demand sites in Gangnam-*gu*.

4.4. Solve the models

We solved the location-allocation problem using two models: the *p-median* model and the MCLP model. In addition, we calculated the optimum number for *p*, referencing the case of bicycle-sharing stations of other *gu*. We calculated the average ratio of bicycle-sharing stations to the resident population in each *gu* with an existing program (A/B), and we then multiplied it by the resident population of Gangnam-*gu* (Table 4). As a result, we concluded that *p* should be 80. To detect the method's sensitivity, we solved the location-allocation problem for three different values of *p*: 40, 80, and 120.

Fig. 6 shows the distribution of the selected bicycle stations in Gangnam-*gu*, with potential demand locations. As expected, in the case of the *p-median* model, the selected stations are more scattered over the whole area, whereas with the MCLP model, the selected stations are more concentrated in crowded central areas (Fig. 6 and Table 5). This means that the *p-median* model has an advantage in terms of spatial equity, whereas the maximum coverage model has greater efficiency to maximize the demand associated with all selected stations. This result can be justified by each model's nature.

Table 4
The number of bicycle-sharing stations (*p*) in Gangnam-*gu*.

Gu	Number of existing stations (A)	Resident populations (B)	A/B
Gu1	48	163,235	0.000294
Gu2	31	134,175	0.000231
Gu3	67	395,687	0.000169
Gu4	41	327,113	0.000125
Gu5	48	305,525	0.000157
Gu6	29	374,053	0.000078
Gu7	37	372,214	0.000099
Gu8	17	486,258	0.000035
Gu9	21	246,934	0.000085
Gu10	68	414,901	0.000164
Gu11	31	499,850	0.000062
Average	**40**	**338,177**	**0.000136 (D)**
	Total number of stations in Gangnam-*gu* (*p*) ***80 (= C × D)***	**Resident populations in Gangnam-*gu*** **576,413 (C)**	–

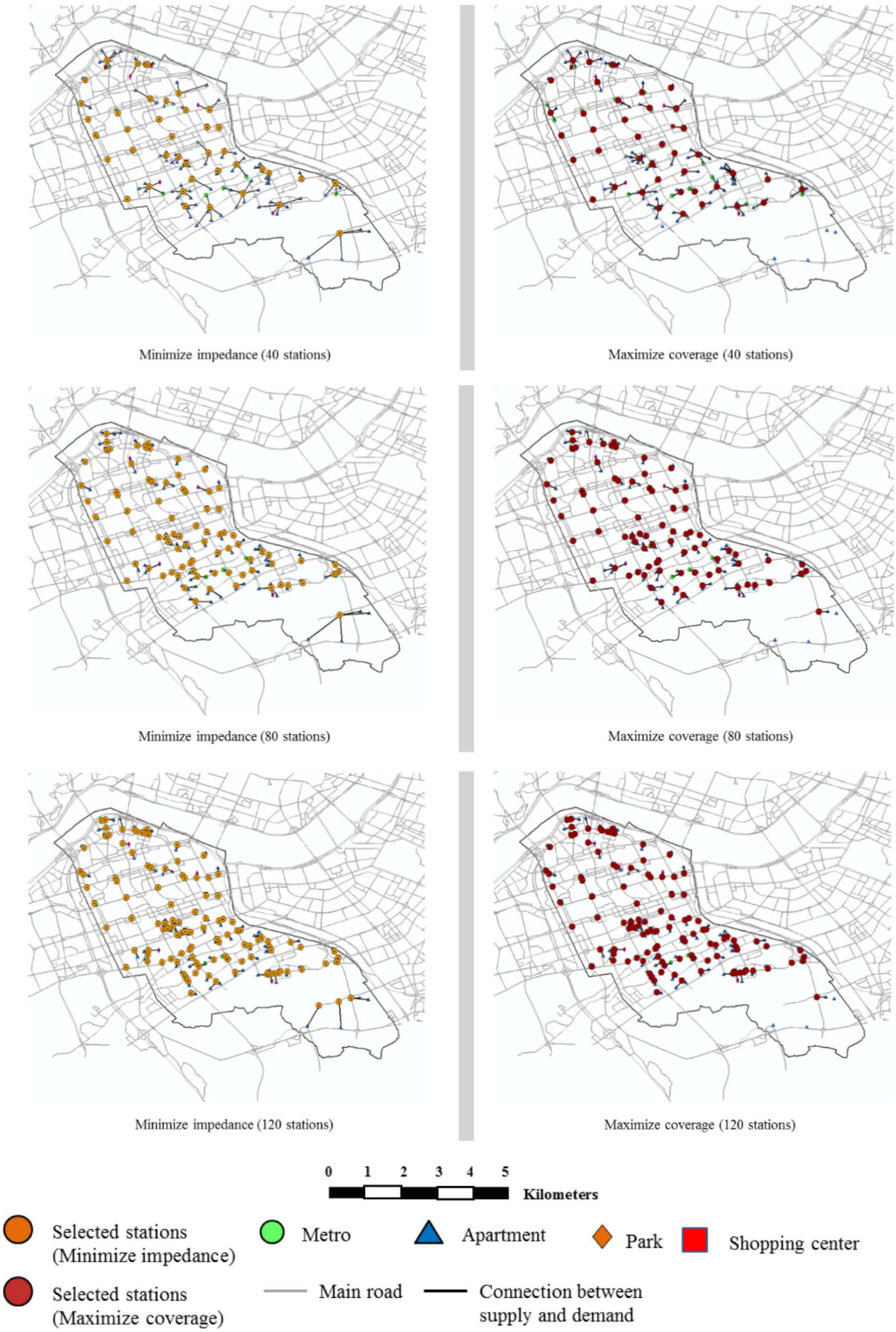

Fig. 6. Results of the location-allocation model.

Table 5
Distance between demand locations and selected stations using each model.

Number of stations	Minimize impedance (m)				Maximize coverage location problem (m)			
	Average	Std. dev	Max	Min	Average	Std. dev	Max	Min
40	248.40	218.00	1239.37	8.99	116.24	58.30	199.23	8.99
80	151.63	171.36	1239.37	7.25	93.96	56.03	198.97	7.25
120	121.47	126.14	917.78	7.25	82.67	51.38	197.15	7.25

4.5. Sensitivity analysis

We conducted sensitivity analysis to verify the influence of the parameter p, which represents the final number of bicycle-sharing stations. To illustrate how this parameter affects the demand covered by selected stations at 200 m and 400 m respectively, we set p at 40, 80, and 120 (Table 6).

The MCLP model solution covered a greater portion of demand when the maximum distance of pedestrian trips to stations was set at either 200 m or 400 m (Table 6). The differences are greater in cases with fewer stations and less substantial in cases with more stations. Both solutions showed diminishing returns in terms of demand covered at 200 m and 400 m as the number of bicycle stations increased.

As shown in Table 6, the test with $p = 40$ resulted in a significant difference in the percentage of demand covered in comparison to $p = 80$, whereas setting p at 120 provided little additional improvement in the percentage of demand covered over the case of $p = 80$ for both the *p-median* model and the MCLP model. Those results indicated that efficiency of demand coverage diminishes as p increases. From those results, we can determine the optimal value of p to be 80.

4.6. Determine the capacity of selected bicycle stations

By counting the frequency of taxi pick-ups at each station, we could determine the bicycle station capacity (i.e., the number of bicycles needed at each station). Supposed that the government wanted to introduce 1000 bicycles in view of budget constraints and that the minimum capacity of each selected station is 5. The number of clusters is 4 (since we are dividing the stations into quartiles). Then, we can identify the desired capacity of bicycle stations as shown in Table 7.

5. Conclusion

Bicycle-sharing systems have been in the spotlight recently, since they are seen as a way to replace short trips in private cars. They have this potential because short car trips and bicycle trips possess similar characteristics in terms of purposes and distance, bicycle use can reduce carbon dioxide emissions from motor vehicles, and bicycle travel helps to ease urban traffic congestion. Moreover, public bicycle systems help citizens to enjoy physically healthy lives. All these advantages can enable such systems not only to reduce short car trips but also to make urban life more pleasant.

In this study, we have suggested a framework for optimally locating bicycle-sharing stations to replace short trips by private car in the central city. We used the minimum impedance (*p-median*) model and the maximum coverage location problem (MCLP), two popular location-allocation methods. To select candidate bicycle-sharing stations, we extracted taxi trajectory data for driving patterns viewed as similar to bicycle riding patterns. We applied this methodology to one administrative division in Seoul, Gangnam-*gu*, and identified the optimum locations for bicycle-sharing stations.

Different results were obtained by the two location-allocation models. In the case of the *p-median* model, the selected stations were more scattered over the whole district, whereas with the MCLP model, the stations were more concentrated on central areas. The *p-median* model thus offers an advantage in terms of spatial equity, whereas the MCLP model appears to offer greater efficiency in maximizing the potential demand covered by the stations. Policymakers can choose to apply either model in accordance with their policy purpose.

According to the Seoul Development Institute, short-distance driving (trips of three miles or less) represent 44% of all driving in

Table 6
Sensitivity analysis.

Test problem	Description	% of potential demand covered by selected stations at 200 m		% of potential demand covered by selected stations at 400 m	
		Minimizing impedance	Maximizing coverage	Minimizing impedance	Maximizing coverage
1–1	Proposed example (80)	93.52	94.19	98.96	99.67
1–2	Lower p (40)	79.48	88.83	94.35	99.59
1–3	Higher p (1 2 0)	94.19	94.19	99.67	99.67

Table 7
The capacity of bicycle station (p = 80, MCLP).

Station ID	Frequency of taxi pick-up	Quantile cluster	Station capacity
Station 1	623,064	**1**	20
Station 2	63,805	**1**	20
Station 3	59,356	**1**	20
Station 4	26,438	**1**	20
Station 5	24,153	**1**	20
	…		
Station 21	9201	**2**	15
Station 22	8584	**2**	15
Station 23	8563	**2**	15
Station 24	8383	**2**	15
Station 25	8299	**2**	15
	…		
Station 41	5378	**3**	10
Station 42	5371	**3**	10
Station 43	5126	**3**	10
Station 44	4823	**3**	10
Station 45	4742	**3**	10
	…		
Station 76	1388	**4**	5
Station 77	1327	**4**	5
Station 78	1226	**4**	5
Station 79	1201	**4**	5
Station 80	1187	**4**	5
Sum			1000

Seoul, and trips of less than one mile represent 11% of the total. For these reasons, many experts in Korea have advocated for policies to restrict or control short-distance driving. Meanwhile, municipal authorities intend to expand further the “Seoul-bike” sharing system. The specific case examined by the present research provides guidance on how to locate bicycle-sharing stations in an administrative division that does not yet participate in the program. Moreover, although our case study was in Korea, our methodology can be adopted for developing a station location plan in any city that does not yet have a bicycle-sharing system.

In addition to identifying the optimal locations for public bicycle facilities, the following measures are required to make such a program successful. Primarily, extensive publicity would be essential. Smartphone applications are an effective method of both encouraging citizens to travel by bicycle and making them aware of a city’s bike-sharing program. In fact, smartphone applications for public bicycle systems have been developed and distributed in Seoul, Paris, and London. Second, the facilities should be installed in prominent places because public bicycle stations should be located where many people travel. This factor is reflected in our methodology for determining the locations of the public bicycle stations. Third, bicycle infrastructure such as bicycle-only roads and repair shops should be available. In Seoul, one can ride bicycles on most roadways and repair shops are distributed densely.

However, our study has several limitations. First, the taxi trajectories are not able to reflect accurately the travel patterns of private cars. Further studies incorporating real trajectories of private cars would be valuable. Second, our approach can result in relatively isolated stations, particularly in the MCLP model. Third, there are other types of potential demand sites that were not considered. Forth, this study uses the Euclidean distance instead of street distance in MCLP and *p-median* model. At present, Seoul is only providing road data that can be passed by cars. For this reason, the actual street distance between demand and supply sites cannot be calculated. In the end, it seems that setting the model with Euclidean distance is the most reasonable for now. Lastly, there are two types of trips in mobility studies, namely generated trips (defined by their starting point) and attracted trips (described in terms of their destination); however, the floating population data used in this study had no directional information, preventing us from distinguishing between generated and attracted trips. Those limitations should be addressed in future studies.

Acknowledgement

This research, ‘Geospatial Big Data Management, Analysis and Service Platform Technology Development’, was supported by the MOLIT (The Ministry of Land, Infrastructure and Transport), Korea, under the national spatial information research program supervised by the KAIA (Korea Agency for Infrastructure Technology Advancement)“ (16NSIP-B081011-03).

References

Bandyopadhyay, M., Singh, M., 2012. Spatial pattern analysis for finding weighted candidate set for *p-median* problem in locating emergency facilities. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2 (5), 69–74.

Braun, L.M., Rodriguez, D.A., Cole-Hunter, T., Ambros, A., Donaire-Gonzalez, D., Jerrett, M., Mendez, M.A., Nieuwenhuijsen, M.J., de Nazelle, A., 2016. Short-term planning and policy interventions to promote cycling in urban centers: findings from a commute mode choice analysis in Barcelona, Spain. Transport. Res. Part A: Policy Pract. 89, 164–183.

Cai, H., Zhan, X., Zhu, J., Jia, X., Chiu, A.S., Xu, M., 2016. Understanding taxi travel patterns. Phys. A: Stat. Mech. Appl. 457 (1), 590–597.

Dunlap, R.E., Jorgenson, A.K., 2012. Environmental Problems. The Wiley-Blackwell Encyclopedia of Globalization.
Erdoğan, G., Battarra, M., Calvo, R.W., 2015. An exact algorithm for the static rebalancing problem arising in bicycle sharing systems. Eur. J. Oper. Res. 245 (3), 667–679.
Flynn, B.S., Dana, G.S., Sears, J., Aultman-Hall, L., 2012. Weather factor impacts on commuting to work by bicycle. Prev. Med. 54 (2), 122–124.
Frade, I., Ribeiro, A., 2015. Bike-sharing stations: a maximal covering location approach. Transport. Res. Part A: Policy Pract. 82, 216–227.
García-Palomares, J.C., Gutiérrez, J., Latorre, M., 2012. Optimizing the location of stations in bike-sharing programs: a GIS approach. Appl. Geogr. 35 (1), 235–246.
Gwak, J.H., Lee, B.K., Lee, W.K., Sohn, S.Y., 2017. Optimal location selection for the installation of urban green roofs considering honeybee habitats along with socio-economic and environmental effects. J. Environ. Manage. 189, 125–133.
Jensen, P., Rouquier, J.B., Ovtracht, N., Robardet, C., 2010. Characterizing the speed and paths of shared bicycle use in Lyon. Transport. Res. Part D: Transp. Environ. 15 (8), 522–524.
Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R., 2010. Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system. Pervasive Mobile Comput. 6 (4), 455–466.
Kloimüllner, C., Papazek, P., Hu, B., Raidl, G.R., 2014. Balancing bicycle sharing systems: an approach for the dynamic case. Eur. Conf. Evol. Comput. Comb. Optimiz. 73–84.
Krykewycz, G., Puchalsky, C., Rocks, J., Bonnette, B., Jaskiewicz, F., 2010. Defining a primary market and estimating demand for major bicycle-sharing program in Philadelphia, Pennsylvania. Transport. Res. Rec.: J. Transport. Res. Board 2143, 117–124.
Lee, W.K., Sohn, S.Y., 2017. Taxi vacancy duration: a regression analysis. Transp. Plan. Technol. 1–25.
Li, H., Wang, Q., Shi, W., Deng, Z., Wang, H., 2015. Residential clustering and spatial access to public services in Shanghai. Habitat Int. 46, 119–129.
Lin, J.R., Yang, T.H., 2011. Strategic design of public bicycle sharing systems with service level constraints. Transport. Res. Part E: Logist. Transport. Rev. 47 (2), 284–294.
Lin, J.R., Yang, T.H., Chang, Y.C., 2013. A hub location inventory model for bicycle sharing system design: formulation and solution. Comput. Ind. Eng. 65 (1), 77–86.
Liu, Z., Jia, X., Cheng, W., 2012. Solving the last mile problem: ensure the success of public bicycle system in Beijing. Proc.-Soc. Behav. Sci. 43, 73–78.
Martinez, L.M., Caetano, L., Eiró, T., Cruz, F., 2012. An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon. Proc.-Soc. Behav. Sci. 54, 513–524.
Nankervis, M., 1999. The effect of weather and climate on bicycle commuting. Transport. Res. Part A: Policy Pract. 33 (6), 417–431.
O'brien, O., Cheshire, J., Batty, M., 2014. Mining bicycle sharing data for generating insights into sustainable transport systems. J. Transp. Geogr. 34, 262–273.
Papazek, P., Kloimüllner, C., Hu, B., Raidl, G.R., 2014. Balancing bicycle sharing systems: an analysis of path relinking and recombination within a GRASP hybrid. Int. Conf. Parall. Probl. Solv. Nat. 792–801.
Rahman, S.U., Smith, D.K., 2000. Use of location-allocation models in health service development planning in developing nations. Eur. J. Oper. Res. 123 (3), 437–452.
Rainer-Harbach, M., Papazek, P., Hu, B., Raidl, G.R., 2013. Balancing bicycle sharing systems: a variable neighborhood search approach. Eur. Conf. Evol. Comput. Comb. Optimiz. 121–132.
Romero, J.P., Ibeas, A., Moura, J.L., Benavente, J., Alonso, B., 2012. A simulation-optimization approach to design efficient systems of bike-sharing. Proc.-Soc. Behav. Sci. 54, 646–655.
Seoul Open Data Plaza. < http://data.seoul.go.kr/openinf/fileview.jsp?infId=OA-12066&tMenu=11 > .
Shu, J., Chou, M.C., Liu, Q., Teo, C.P., Wang, I.L., 2013. Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. Oper. Res. 61 (6), 1346–1359.
Zahabi, S.A.H., Chang, A., Miranda-Moreno, L.F., Patterson, Z., 2016. Exploring the link between the neighborhood typologies, bicycle infrastructure and commuting cycling over time and the potential impact on commuter GHG emissions. Transport. Res. Part D: Transp. Environ. 47, 89–103.