

## Location Optimization of Bicycle-sharing Stations Using Multiple-criteria Decision Making

Tae-Young Lee,<sup>1</sup> Myeong-Hun Jeong,<sup>1\*</sup> Seung-Bae Jeon,<sup>1</sup> and Jae-Myoung Cho<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, Chosun University,  
309 Pilmun-daero, Dong-gu, Gwangju 61452, Republic of Korea

<sup>2</sup>Department of Civil Engineering, Songwon University,  
Songha-dong, Nam-gu, Gwangju 503-742, Republic of Korea

(Received September 28, 2020; accepted December 9, 2020)

**Keywords:** bicycle-sharing system, location-allocation models, network analysis, multiple linear regression, MCDM

A bicycle-sharing system not only allows citizens to freely use bicycles installed in specific locations but is also a supplement to public transportation. In this study, we aim to improve the accessibility to public bicycles by finding the optimal locations of bicycle-sharing stations based on the history and spatial data of public bicycle operations. IoT sensors record the bicycles' movement information. Multiple linear regression analysis is used to select the most important criteria for a bicycle-sharing system. The selected criteria are then applied to multiple-criteria decision making (MCDM) to rank the potential locations of bicycle-sharing stations. The top-ranking locations are finally determined through an optimization stage.

### 1. Introduction

There is a growing interest globally in maintaining the environment, and bicycle-sharing systems have been launched as an alternative to motorized transportation. A bicycle-sharing system is not only eco-friendly but also a complementary means of transportation. In particular, a bicycle-sharing system complements the shortcomings of existing public transportation systems by increasing the mobility of users. Demand-driven transportation systems, such as bicycle-sharing systems, contribute to meeting the demand for highly convenient transportation.<sup>(1)</sup> A successful bicycle-sharing system should be accessible to citizens for use, and the optimal locations of bicycle-sharing stations are crucial.<sup>(2)</sup>

In this study, the optimal locations of bicycle-sharing stations were determined. As the experimental data, data records of the bicycle-sharing system called Fifteen provided by Goyang City, Gyeonggi Province, Republic of Korea, were used. The regional characteristics of the stations were statistically identified through multiple linear regression analysis to find the optimal locations. The statistically significant factors identified were applied as weights to multiple-criteria decision making (MCDM) to rank the potential station locations. We provide alternative locations for current stations that consider changes in urbanized areas such

\*Corresponding author: e-mail: mhjeong@chosun.ac.kr

<https://doi.org/10.18494/SAM.2020.3125>

as new housing developments, population movement between regions, and public transport infrastructure changes.

In Sect. 2, the features of our approach are reviewed on the basis of previous related studies. Section 3 presents the experimental data and the proposed method. The results are discussed in Sect. 4. Finally, Sect. 5 provides some conclusions along with a summary and suggestions for future research.

## 2. Background

The first bicycle-sharing system was the Velib scheme in Paris in 2007, and bicycle-sharing systems have since been established worldwide.<sup>(3)</sup> Previous studies have treated location optimization through various methods to increase the accessibility and use of systems. Researchers have applied mathematical models using a geographic information system (GIS) and MCDM to select station locations. Experts directly determined the weights of several factors and applied them to the analytic hierarchy process (AHP), which is one of the MCDM methods used to determine suitable locations.<sup>(4)</sup>

In addition, research has focused on location allocation using a GIS rather than ranking within the selected areas.<sup>(5,6)</sup> However, previous studies have estimated preferential locations by applying the same weights to the criteria. Bhuyan *et al.* grouped users by age and analyzed each group's demand and the station density to provide experts with geographic information for a bicycle-sharing system.<sup>(5)</sup> Loidl *et al.* performed a kernel density estimation (KDE) analysis to extract information on the potential of a bicycle-sharing system by forming and overlaying the layers of each criterion.<sup>(6)</sup>

Furthermore, researchers have applied the AHP to a GIS platform. Kabak *et al.* quantified the criteria influencing bicycle use based on Boolean relationships.<sup>(7)</sup> The authors and stakeholders set the importance of the criteria and applied them to the AHP. They created a suitability map based on the values obtained from the AHP. The suitable area and the existing stations were ranked through multi-objective optimization by ratio analysis (MOORA), and a comparative analysis was also conducted. Hoang *et al.* applied criteria calculated using a GIS to the AHP to create a set of weights and used principal component analysis (PCA) to find the correlations among the criteria to eliminate potential sites with less significant criteria.<sup>(8)</sup>

Studies focusing on origin-destination (OD) analysis have also been conducted for transportation planning. OD trajectory analysis can help to not only identify urban flow patterns but also provide a wealth of information on urban flow and transportation demand.<sup>(9,10)</sup>

In this study, we utilize multiple linear regression analysis to determine the critical criteria, in contrast to the previous research on manually selecting the criteria. We also use network analysis, such as that of the centrality, to consider the connectivity of bicycle-sharing stations. The selected criteria are then subjected to MOORA to determine the optimal locations of the stations.

### 3. Materials and Methods

#### 3.1 Area of study and data

The bicycle-sharing system data used were from the bicycle-sharing system called Fifteen provided by Goyang City, Gyeonggi Province, Republic of Korea. The data comprised departure stations, arrival stations, and times. Figure 1 presents temporal signatures for the data. Commuting on weekdays is widely seen as the key reason for the use of bicycles in the scheme.

Additional data (e.g., the number of people using buses or the subway, or the number of residents) were combined with the bicycle-sharing data. Various buffer sizes around the bicycle-sharing stations were tested for different data combinations. Finally, we set up a buffer of 400 m from each station as the spatial range for combining the data because the variables of the built environment are the most important in the model. The buffer size was based on a previous study.<sup>(11)</sup> Figure 2 shows the 155 bicycle-sharing stations currently installed in Goyang City, and Table 1 presents the experimental data used in this study.

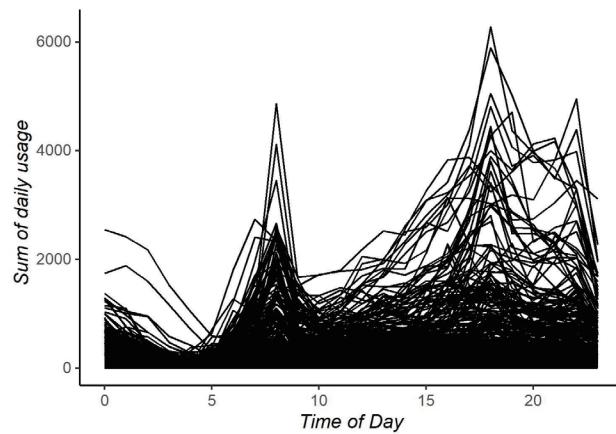


Fig. 1. Temporal signatures for Fifteen data.

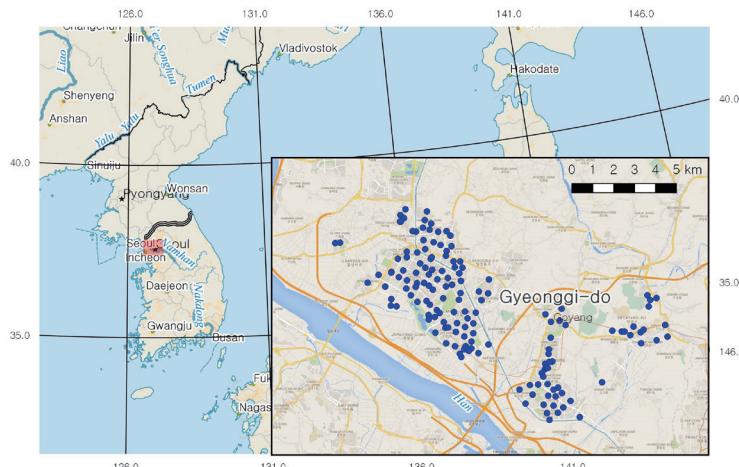


Fig. 2. (Color online) Current bicycle-sharing stations.

**Table 1**  
Overview of variables used in this study.

Variables	Description
Dependent variable	Number of bicycle rentals per day
	Number of passengers at subway stops
	Number of passengers at bus stops
	Total floor area of residences
	Total floor area of workplaces
Independent variables	Total floor area of buildings
	Length of bike lanes
	Number of schools
	Number of residents
	Number of residences
	Number of workplaces

### 3.2 Multiple linear regression

Multiple linear regression analysis involves two or more independent variables and one dependent variable. The influence of the independent variables is determined as a statistical value by comparing the absolute values of the regression analysis results. A positive value indicates a positive correlation with the dependent variable, whereas a negative value indicates a negative correlation.<sup>(12)</sup>

Using this multiple linear regression model, the effect of the environmental characteristics around the stations on the use of public bicycles was analyzed in this study. Although the environmental variables of bicycle-sharing stations were regarded as independent, the number of bicycle rentals at each station was regarded as dependent.

### 3.3 Centrality

In network analysis, the centrality can be used to judge the importance of a node (station). The purpose of using the centrality is to express the connectivity between stations and find the ones with the most influence. The weighted PageRank centrality indicates the importance of the nodes when applying weights to the strength of the connections but does not consider the connections between nodes equally.<sup>(13)</sup> A high connection strength between stations means that the movement between the stations is high. Thus, we used the centrality as one of the criteria for determining optimal locations.

### 3.4 MCDM

We applied MOORA to rank the potential sites for a station. The characteristics of the surrounding environment derived from multiple linear regression analysis and the centrality derived from network analysis were applied to the MOORA parameters. This method ranks the sum of each variable by applying negative numbers as the cost and positive numbers as the benefit based on the weight of the variable.<sup>(14)</sup>

## 4. Results

Figure 3 presents the top 20 OD trajectories in terms of the frequency of use obtained from the Fifteen data records. When the sum of each station usage was calculated, it was confirmed that the top 35 of the 155 stations accounted for most of the total usage. We determined the environmental characteristics of well-used stations and used them for location selection criteria.

An experiment was conducted for the 35 bicycle-sharing stations with the highest frequency of use to determine the conditions for successful locations of bicycle-sharing stations. The initial criteria were the 10 independent variables listed in Table 1. The experiment utilized backward elimination based on Akaike's information criterion (AIC) in multiple linear regression analysis to remove unnecessary parameters. Table 2 shows the five criteria remaining after backward elimination, which are statistically significant variables.

Equation (1) describes the multiple linear regression model. The adjusted R square value is 0.697, which indicates a relatively strong relationship. The *p*-value of the model is less than 0.05, which is highly significant.

$$y = 3662 - 0.87x_1 + 1.02x_2 + 0.44x_3 - 0.47x_4 + 1.34x_5. \quad (1)$$

Here,  $x_1$  is the number of passengers at bus stops,  $x_2$  is the number of passengers at subway

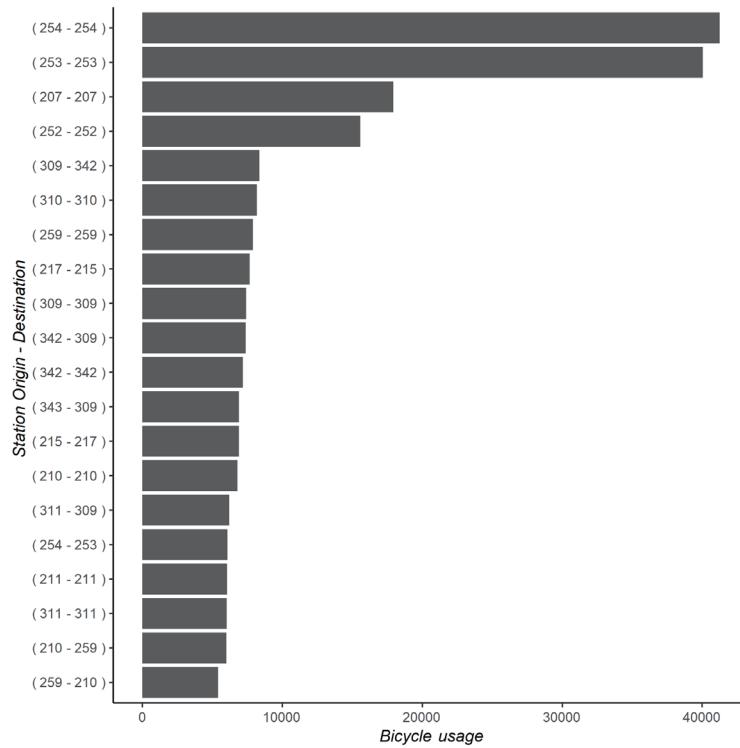


Fig. 3. Top 20 OD trajectories with station numbers.

Table 2  
Multiple linear regression coefficients.

Model	Coefficients			<i>t</i>	sig
	B	Std. Error	Beta		
(constant)	3662.36	588.01		6.228	0.000001
Number of passengers at bus stops	-0.013	0.003	-0.87	-3.852	0.0006
Number of passengers at subway stops	0.0008	0.0001	1.02	4.948	0.00003
Number of residences	2.873	0.935	0.44	3.071	0.004
Total floor area of workplaces	-0.004	-0.001	-0.47	-3.106	0.004
Total floor area of buildings	0.009	0.001	1.34	7.146	0.0000001

stops,  $x_3$  is the number of residences,  $x_4$  is the total floor area of workplaces, and  $x_5$  is the total floor area of buildings.

Equation (1) indicates that every 1% increase in the number of passengers at bus stops is associated with a 0.87% decrease in the number of bicycle rentals per day, and every 1% increase in the number of residences is associated with a 0.44% increase in the number of bicycle rentals per day.

Furthermore, we verified the assumptions of the multiple linear regression model. The non-constant variance (i.e., homogeneity of variance) was 0.23, which is greater than 0.05. No statistically significant difference was shown. The Durbin–Watson statistic (i.e., independence of observations) was 1.67, which is close to 2, and the variance inflation factor values were well below 10. The Shapiro–Wilk statistic for a normal distribution was 0.42, which is greater than 0.05. All tests show that the current multiple linear regression model satisfies all assumptions.

Figure 4 shows the centrality of the bicycle-sharing stations and the connectivity among the stations. The size of each point is proportional to the value of centrality, and the thickness of each connecting line between points is proportional to the number of movements among bicycle-sharing stations. Only stations with strong connectivity are presented in Fig. 4.

Nodes with the significant use of bicycles and strong connectivity, as shown in Fig. 4, will require additional bike stations when they are relocated in the future. In this study, we added the centrality to the MOORA criteria. Because the centrality is considered to be an important factor in location selection, the weight of the centrality criterion was set to 30% of the total weight. The total of the weights in MOORA must be 1. Thus, we defined the criteria weights of variables  $x_1$ – $x_5$  as shown in Eq. (2):

$$W = -0.148x_1 + 0.172x_2 + 0.074x_3 - 0.08x_4 + 0.226x_5 + 0.3x_6, \quad (2)$$

where  $x_1$  indicates the number of passengers at bus stops,  $x_2$  is the number of passengers at subway stops,  $x_3$  is the number of residences,  $x_4$  is the total floor area of workplaces,  $x_5$  is the total floor area of buildings, and  $x_6$  is the PageRank centrality.

The bicycle-sharing stations in Goyang City satisfy the minimum area covered by a bicycle-sharing system stipulated in the guidelines developed by the Institute for Transportation and Development Policy (ITDP).<sup>(15)</sup> According to the guidelines, an appropriate distance

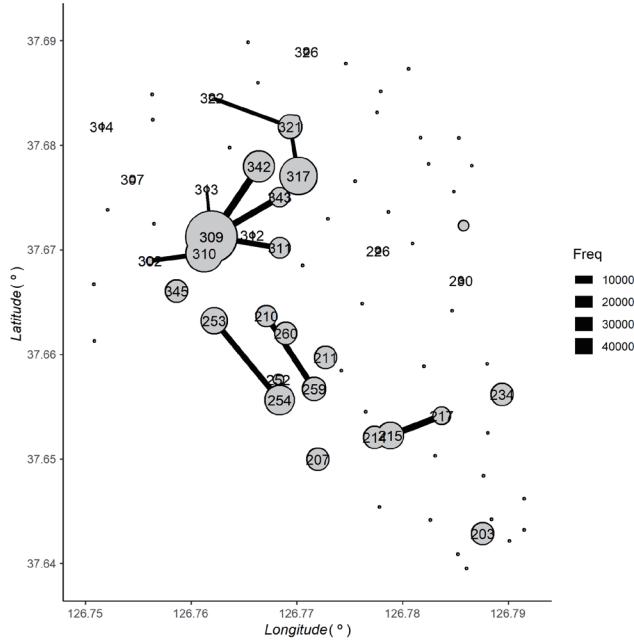


Fig. 4. Centrality plot. Note that frequencies of more than 30000 are hidden because the rental and return stations are the same.

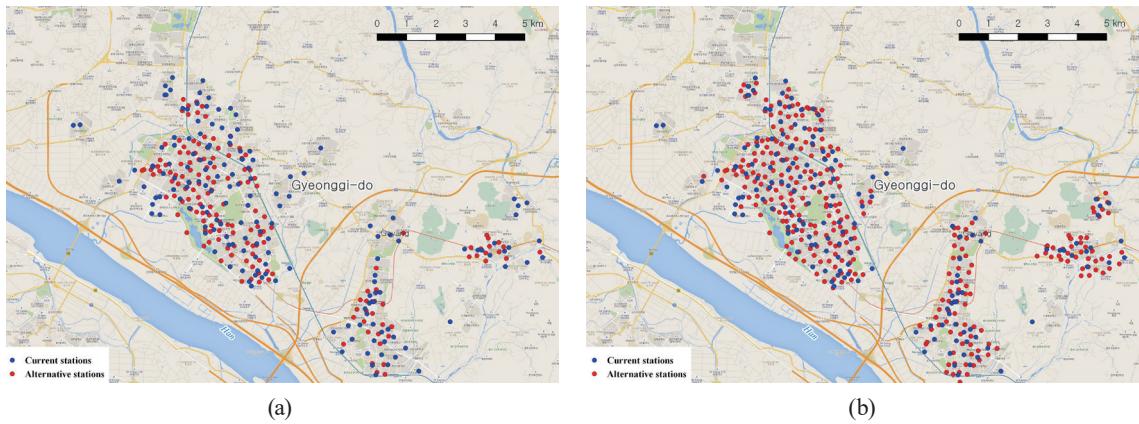


Fig. 5. (Color online) Current and alternative bicycle-sharing stations in the cases of (a) 155 and (b) 300 stations.

between stations is approximately 300 m, and thus 3038 stations can be installed in the entire area of Goyang City. Based on each point, the criteria in the 400 m buffer were calculated and applied to Eq. (2) to determine the 3038 rankings. The top 155 stations were selected for comparison with the existing stations. In addition, a bicycle station should be located on the sidewalk for better accessibility, and it is more valuable if it is located near a public transport station. We therefore applied optimization steps accordingly. Figure 5(a) shows a comparison between the current and alternative station locations for the same number of stations (i.e., 155), and Fig. 5(b) presents the alternative station locations when the number of stations is increased to 300.

## 5. Conclusions

The optimal utility of a bicycle-sharing system is closely related to its stations having good access for users. The environment surrounding bicycle-sharing stations was analyzed to determine the characteristics required for selecting the optimal locations. Multiple linear regression analysis was conducted to obtain the weights of the derived vital criteria, and significant correlations were confirmed. In addition, the centrality values of the network were calculated regarding the connectivity between stations. Finally, MOORA was performed to generate potential locations of bicycle-sharing stations.

This study provided alternative locations for existing stations to respond to changes in urbanized areas such as new housing developments, population movement between regions, and public transport infrastructure changes. We attempted to make a more actively used bicycle system by increasing citizens' accessibility and found some differences between the optimal locations of bicycle-sharing stations and their current locations. Future studies should focus on the appropriate number of bicycle stations and their allocation.

## Acknowledgments

This study was supported by research funding from Chosun University, 2018.

## References

- 1 S. A. Shaheen, S. Guzman, and H. Zhang: *Transp. Res. Rec.* **2143** (2010) 159. <https://doi.org/10.3141/2143-20>
- 2 C. M. de Chardon, G. Caruso, and I. Thomas: *Transp. Res Part A Policy Pract.* **100** (2017) 202. <https://doi.org/10.1016/j.tra.2017.04.020>
- 3 P. DeMaio: *J. Public Transport.* **12** (2009) 41. <http://doi.org/10.5038/2375-0901.12.4.3>
- 4 T. Kanjanakorn and M. Piantanakulchai: *ISAHP* (2013) 1. <https://doi.org/10.13033/isahp.y2013.066>
- 5 I. A. Bhuyan, C. Chavis, A. Nickkar, and P. Barnes: *Urban Sci.* **3** (2019) 42. <https://doi.org/10.3390/urbansci3020042>
- 6 M. Loidl, U. Witzmann-Müller, and B. Zagel: *Eur. Transp. Res. Rev.* **11** (2019) 9. <https://doi.org/10.1186/s12544-019-0347-7>
- 7 M. Kabak, M. Erbaş, C. Çetinkaya, and E. Özceylan: *J. Clean. Prod.* **201** (2018) 49. <https://doi.org/10.1016/j.jclepro.2018.08.033>
- 8 H. T. Hoang, Q. H. Truong, A. T. Nguyen, and L. Hens: *Sustainability* **10** (2018) 3097. <https://doi.org/10.3390/su10093097>
- 9 M.-H. Jeong, S.B. Jeon, S. Park, and S. Kang: *Sens. Mater.* **31** (2019) 3827. <https://doi.org/10.18494/sam.2019.2565>
- 10 M.-H. Jeong, J. Yin, and S. Wang: *GIScience* **114** (2018) 1. <https://doi.org/10.4230/LIPIcs.GISCIENCE.2018.6>
- 11 T. D. Tran, N. Ovtracht, and B. F. d'Arcier: *Procedia CIRP* **30** (2015) 293. <https://doi.org/10.1016/j.procir.2015.02.156>
- 12 L. S. Aiken, S. G. West, and R. R. Reno: *Multiple Regression: Testing and Interpreting Interactions* (Sage Publications, Newbury Park Calif, 1991).
- 13 W. Xing and A. Ghorbani: *Proc. 2nd Annu. Conf. Communication Networks and Services Research* (2004) 305. <https://doi.org/10.1109/dnsr.2004.1344743>
- 14 W. K. M. Brauers and E. K. Zavadskas: *Technol. Econ. Dev. Econ.* **16** (2010) 5. <https://doi.org/10.3846/tede.2010.01>
- 15 Institute for Transportation and Development Policy: <https://bikeshare.itdp.org> (accessed December 2013).