Simulation Modelling Practice and Theory 130 (2024) 102861

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

A data-driven agent-based simulation of the public bicycle-sharing system in Sejong city

Jang Won Bae [a], Chun-Hee Lee [b], Jeong-Woo Lee [c], Seon Han Choi [d,e,*]

[a] School of Industrial Management, Korea University of Technology and Education, Cheonan 31253, South Korea
[b] Visual Intelligence Research Section, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea
[c] Intelligent Devices & Simulation Research Section, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea
[d] Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul 03760, South Korea
[e] Graduate Program in Smart Factory, Ewha Womans University, Seoul 03760, South Korea

ARTICLE INFO

Keywords:
Public bicycle-sharing system
Data-driven model
Agent-based simulation
Sejong city

ABSTRACT

In this study, we propose a data-driven agent-based model for simulating Eoulling, the public bicycle-sharing systems (PBSSs) in Sejong City, the administrative capital of South Korea. Most existing models for PBSSs based on top-down approaches have limitations in reflecting Eoulling users' behavioral characteristics and analyzing their convenience. Unlike these, the proposed model is based on a bottom-up approach of agent-based simulation. We model each user as an agent to capture their bicycle rental and return behaviors, and analyze user convenience through interactions with bicycle station agent models. To improve model fidelity, multiple parameters for determining agent behaviors are extracted from the actual operations data of Eoulling, along with the population and geographic information of Sejong City. The validation results showed that the proposed model accurately describes the behavioral characteristics. We provide a workable solution addressing multiple concerns with Eoulling by evaluating its utilization and user convenience in virtual scenarios via model simulations.

1. Introduction

Public bicycle-sharing systems (PBSSs) provide an eco-friendly means of transportation in countries around the world including South Korea [1–3]. PBSSs operate as bicycle stations installed throughout a city where users can rent the bicycles and return them when they are done. Previously, these bicycles were mounted in cradles at stations, and if a station had no empty cradle, users could not return bicycles to that station. To address this inconvenience, a new type of system that allows users to freely rent and return bicycles without cradles was recently developed. In Sejong City, the administrative capital of South Korea, a PBSS called Eoulling is currently operational (cf. https://www.sejongbike.kr/). The system currently operates both old-type stations that use cradles and new stations without them. Fig. 1 shows the 72 old-type and 511 new-type stations installed in Sejong City.

Eoulling is provided as a public service in Sejong City; its business is to maximize user convenience while minimizing operating costs, which are covered by taxes. Here, maximizing convenience refers to users' being able to rent and return bicycles at any time and station they want. Regarding this goal, the Eoulling managers have three major concerns:

* **Corresponding author at:** 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, South Korea.
E-mail addresses: jangwon_bae@koreatech.ac.kr (J.W. Bae), ch.lee@etri.re.kr (C.-H. Lee), jeongwoo@etri.re.kr (J.-W. Lee), seonhan.choi@ewha.ac.kr (S.H. Choi).

https://doi.org/10.1016/j.simpat.2023.102861
Received 18 March 2023; Received in revised form 25 October 2023; Accepted 6 November 2023
Available online 7 November 2023

1) the effects on convenience when the incompatible old system is eliminated to reduce costs;
2) the appropriate total number of bicycles to reduce costs without significantly impairing convenience; and
3) suitable locations for the new bicycle stations around the expanding city.

A PBSS incurs tremendous costs during its installation, and it is difficult to subsequently convert them. modeling and simulation (M&S) allow for effectively designing, analyzing, and optimizing PBSSs. Thus far, many achievements pertaining to the M&S of PBSSs have been reported; these can be classified according to the purposes of the M&S, as shown in Table 1 [4].

To solve managers' three concerns with Eoulling, we can apply the studies marked by the shaded parts in Table 1, which are classified under bicycle station design and fleet-sizing design. The bicycle station design focuses on determining the locations and capacities of stations based on a mathematical model considering various factors such as the demand, operator profit, investment cost, and interaction with other transportation systems. Scholars have studied cities such as Seoul, South Korea [3]; Nezahualcoyotl, Mexico [5]; Santander, Spain [6]; Coimbra, Portugal [7]; Istanbul, Turkey [8]; and Phoenix, Arizona, in the United States [9]. In addition, researchers have combined the geographic information system with mathematical models to increase the practicality of the station design [10,11]. Meanwhile, fleet-sizing design refers to determining the initial and total numbers of bicycles at each station, which is necessary because the initial investments in the bicycles are not negligible. In particular, in the cradleless systems, it is very important in terms of securing cost efficiency because the bicycles can be deployed freely. However, despite such importance, scholars have given it scant attention, unlike bicycle station design [4]. For instance, although Fricker and Gast [12] suggested a theoretical framework, they failed to apply it to a real-world scenario.

However, there are limitations in applying these existing studies to Eoulling. Earlier authors utilized top-down approaches that considered an entire PBSS as an abstracted single mathematical model; thus, they could not satisfactorily reflect the behavioral characteristics of users in Sejong City. Because of this, they were also limited to analyze user convenience such as the rental waiting time and rental failure rate. Instead, a bottom-up approach was necessary for addressing the three concerns with Eoulling by considering users' perspective and detailed actions; however, this remains scarce [4]. Additionally, these exiting studies were difficult to fully utilize the actual operations data of Eoulling, along with the population and geographic information of Sejong City.

As a bottom-up approach, agent-based simulation (ABS) is particularly appropriate for reflecting the behavior characteristics of Eoulling users in Sejong City [13]. ABS models each individual entity in the transportation system as an agent and simulates interactions between the agents, thereby analyzing the system-level characteristics. ABS has been applied to analyzing various aspects of transportation such as urban public transport [14], logistics [15], autonomous vehicles [16], and air transportation [17]. For this study, we proposed and built a data-driven agent-based model (ABM) for simulating Eoulling. We modeled each user as an agent to capture their behavioral characteristics and measured various indicators including user convenience through the interactions between user and bicycle station agent models. In addition, to improve model fidelity, we utilized actual operations data, such as the members' information, their bicycle rental-return history, and the population and road network information of Sejong City. Based on these data, we extracted multiple parameters for determining agent behaviors and reflected these in the model. The validation results showed that the proposed model accurately described users' behavioral characteristics. Using this model, we analyzed two of the aforementioned concerns pertaining to Eoulling: 1) the effect of halting the old system on user convenience and 2) reducing operations costs by determining the appropriate total number of bicycles required. Furthermore, in terms of addressing the third concern, the proposed model enables managers to analyze the effect of newly installed stations based on the features of the population around the station.

In summary, our contributions are as follows:

1) Utilizing a bottom-up approach, we proposed an ABM of PBSS to capture Eoulling users' behavioral characteristics and analyze their convenience.

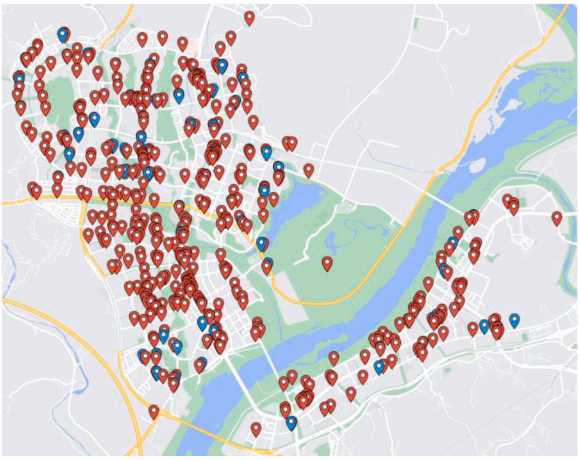

Fig. 1. Bicycle stations installed in Sejong City; blue and red flags indicate the old-type and new-type stations, respectively.

Table 1
Existing studies in M&S of PBSSs [4].

Purpose	Description
Bikeway network design	- Design a new bikeway network or improve an existing network with different bicycle paths, lanes, and routes.
Bicycle station design	- Determine the locations and capacities of bicycle stations.
Fleet-sizing design	- Determine the total number of bicycles in PBSSs and initial number of bicycles at each station.
Static bicycle relocation	- Design a static plan for relocating PBSS bicycles at night, when customer demand is negligible, considering the number of trucks, capacity of trucks, travel routes, etc.
Static demand management	- Design medium-term practices to manage customer demand, such as parking space reservation and different rental prices.
Inventory level management	- Determine the target number or range of usable and unusable bicycles at each station during the relocation operation throughout the daytime.
Dynamic bicycle relocation	- Design a dynamic plan for relocating PBSS bicycles during the daytime, when customers actively use bicycles, considering the real-time system usage.
Dynamic demand management	- Design short-term, real-time, and demand-responsive PBSSs policies to manage customer demand, such as dynamic pricing incentives and best-of-two regulation.

2) Based on the actual operations data, we extracted and reflected model parameters to describe the residents' characteristics in Sejong City and their bicycle rental-return patterns, thereby presenting a well-validated high-fidelity model.
3) By simulating the model in virtual scenarios, we provided a workable solution for the first and second concerns of Eoulling and an effective tool to address the third concern.

The remainder of this study is organized as follows. In Section II, we propose the data-driven ABM for simulating Eoulling, including the agent model designs, extraction of parameters from data, and their implementation and validation. Section III presents case studies for Eoulling based on the proposed model, and Section IV concludes the study.

2. Data-driven agent-based model development

In this section, we introduce the data-driven ABM for simulating Eoulling, describing 1) the design of the agent models, 2) the extraction of the model parameters from actual data, 3) the model implementation, and 4) the model validation.

2.1. Agent model design

The proposed ABM consists of two types of agents. The user agent model describes an Eoulling user and the station agent model represents an Eoulling station. Eoulling can be simulated through the interactions between these models.

2.1.1. User agent model design

We designed a scenario for an Eoulling user's bicycle rental and return guided by responses from interviews with residents of Sejong City. Based on this scenario, we derived eight states corresponding to the operations of the user agent model, as shown in Fig. 2. START

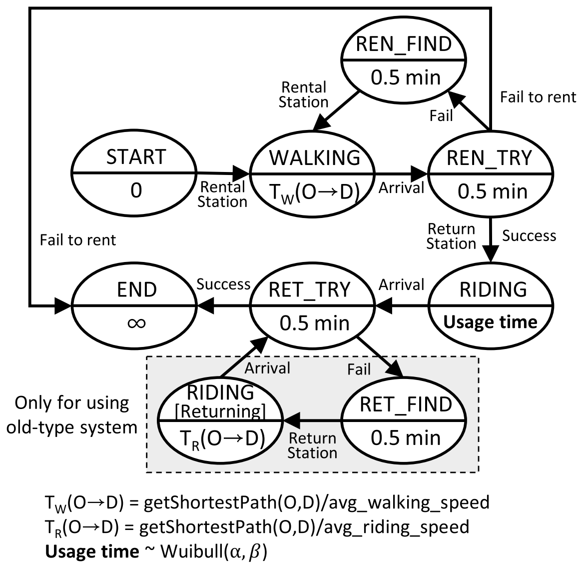

Fig. 2. Eight states of the user agent model.

is the initial state and WALKING is the state of walking to a station to rent a bicycle. REN_TRY is the state of trying to rent a bicycle after arriving at the station, and REN_FIND is the state of searching for possible nearby stations (users can actually check the location and status of bicycle stations using their Eoulling mobile application). RIDING indicates the state of moving on a bicycle toward a return destination, and RET_TRY indicates the state of attempting to return the bicycle after arriving at the station. END is the end state. The user model can also have the state RET_FIND, which applies to a user agent who rents a bicycle from an old-type station and attempts to return the bicycle to a station with no empty cradles. In this case, the user agent finds a possible nearby station through RET_FIND and rides the bicycle to that station in RIDING.

In Fig. 2, the time in each state is the time span required for the user agent to perform the corresponding operation; here, time is expressed in minutes. The time spans for REN_TRY, REN_FIND, RET_TRY, and RET_FIND were set to a constant of 0.5 min, according to the users' average usage time of the Eoulling kiosk or mobile application estimated from the connection histories. The time spans of WALKING and RIDING are dynamically determined according to the departure and destination of the user agent during the simulation. The time span of WALKING is obtained by dividing the shortest distance from the departure to the destination with the average walking speed that is determined depending on the user agent's gender and age [18]. By contrast, the time span of RIDING is a random value following a Weibull distribution estimated from the actual data; this is because owing to the nature of Sejong City, where there are many parks, bicycles are often used for leisure rather than transportation. In cases where there were insufficient data available for estimating the probability distribution, we determine the RIDING time span in the same way as that for WALKING except that we use the average riding speed (15 km/h). In addition, we calculate the time for moving around until the user agent returns the bicycle (i.e., the time span of RIDING[Returning] in Fig. 2) in this way.

Fig. 3 presents an example of the user agent behavior scenario for bicycle rental and return. First, a user agent starts from the start location in START and enters WALKING to station O1 to rent a bicycle. Here, the first-visit station O1 and the arrival time at O1 are randomly determined according to the probability distribution estimated from actual data, a procedure we detail in the next section. After arrival, in REN_TRY, the user agent attempts to rent a bicycle from station O1. Unfortunately, if the rental fails because there are no bicycles at station O1, the user agent searches for the nearest station with bicycles in REN_FIND; here, the type of station is not considered when searching for the nearest station. For instance, even if the first-arrival station O1 is a new-type station, the searched nearest station O2 can be an old-type station. After searching, the user agent walks to station O2 in WALKING and engages in REN_TRY to attempt to rent a bicycle. If the rental is successful, the user agent's state changes to RIDING, and the user agent rides the bicycle to station D1, where the agent wants to return the bicycle. Similar to station O1, station D1 and the arrival time at D1 are randomly determined based on the data. Meanwhile, the rental attempt at station O2 can fail because all the remaining bicycles at O2 were rented by other agents before the user reaches O2 (Eoulling does not have a reservation system). In this case, the user agent repeats the procedure with REN_FIND, WALKING, and REN_TRY to rent a bicycle.

The time span from arriving at station O1 to renting a bicycle is defined as the rental waiting time, as shown in Fig. 3. As the time span of REN_TRY is 0.5 min, the minimum waiting time is 0.5 min, which implies that the user agent rented a bicycle at the first arriving station. If the user agent fails to rent within the given maximum waiting time, it abandons the search and transfers to END. We set the maximum time to 5 min based on residents' interview responses.

After arriving at station D1, the user agent attempts to return the bicycle in RET_TRY. As mentioned earlier, bicycles in the new-type system can be returned without restrictions because the system has no cradle. If station O2 where the bicycle was rented is new-type (station D1 will also be new-type because the both systems are not compatible), the return is always successful, and the user agent will transfer to END. However, if station O2 is old-type (if so, station D1 will be also be old-type), the user agent can fail to return the bicycle

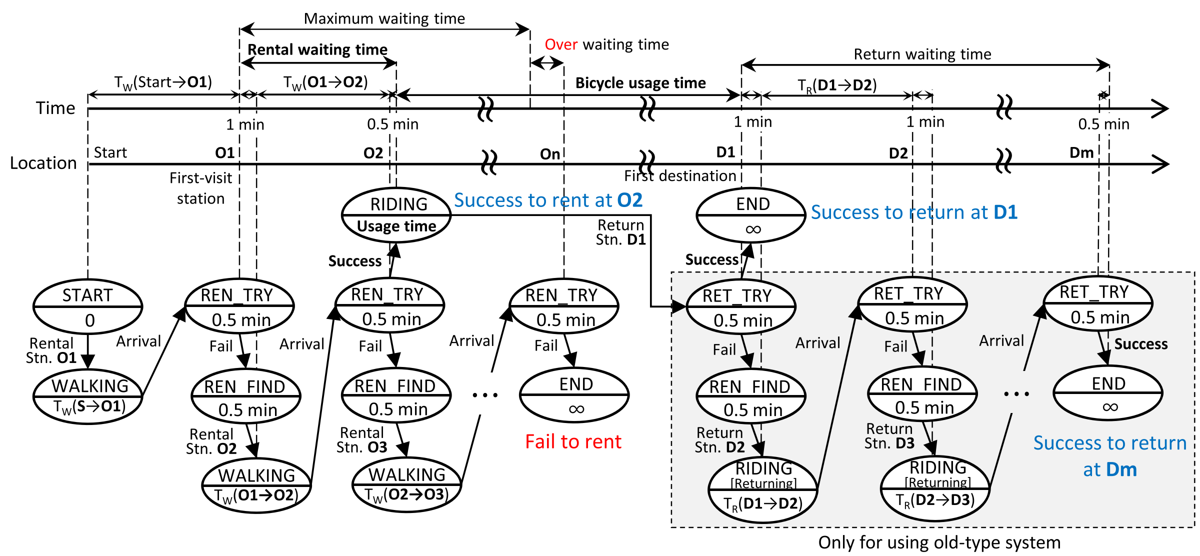

Fig. 3. An example of the user agent behavior scenario.

when there are no empty cradles at D1. In this case, the user agent finds the nearest old-type station with an empty cradle in RET_FIND. After searching, the agent moves to station D2 in RIDING[Returning] and attempts to return the bicycle at D2 in RET_TRY. Similar to the rental, the return attempt at station D2 can fail because all the remaining empty cradles at D2 were occupied by other agents before the user agent reaches D2. In this case, the user agent repeats the procedure with RET_FIND, RIDING[R*eturning*], and RET_TRY to return the bicycle. However, unlike the rental, this procedure is repeated until the bicycle is returned because the return must be unconditional.

The time span from arriving at the first station D1 to returning the bicycle is defined as the return waiting time, as shown in Fig. 3. User agents using the new-type system have the same waiting time (0.5 min) because the time span of RET_TRY is 0.5 min. However, user agents using the old-type system can have significantly longer return waiting times.

2.1.2. Station agent model design

Station agents operate according to the actions in the user agents' REN_TRY and RET_TRY. If a user agent attempts to rent a bicycle from the station agent in REN_TRY, the station agent delivers the bicycle if available; otherwise, it sends a rental failure message to the user agent. Similarly, when a user agent attempts to return the bicycle to the station agent in RET_TRY, the station agent receives and stores the bicycle. If it is an old-type station, it does so if a cradle is available; otherwise, it sends a return failure message to the user agent.

2.2. Model parameter extraction

To effectively analyze Eoulling using the proposed ABM, the two agent models needed to satisfactorily reflect the features of Eoulling. In particular, the station agent model had to reflect the locations of Eoulling stations, numbers of bicycles, and numbers of cradles, and the user agent model had to include the characteristics of Sejong residents and their bicycle rental-return patterns. To this end, we used one year of Eoulling operations data and the population and geographic information of Sejong City; Table 2 summarizes the data used in the proposed model. The rental-return history indicates the time and place from where a user rents a bicycle and the time and place where that bicycle is returned. The member information includes each user's ID, age, and gender. The station information includes the index number, name, type, installation location, and status of each station. The station monitoring history shows the number of bicycles at each station by date and time. The population information indicates the personal information of area residents aged between 10 and 70 years who can use Eoulling. The road network information includes the index number and length of each road, and information on intersections connected to both ends.

We used the rental-return history and member information to extract the probability distributions of the user agent model to capture the characteristics of residents and their rental-return patterns. We utilized the station information as a parameter for the station agent model and used the station monitoring history to derive the initial number of bicycles at each station. In addition, we used population information to characterize the populations around the stations and used these results when adding new stations. We used road network information to build an environment in which the user and station agent models could interact with each other and also to calculate a user agent's travel time in certain cases. In the following subsections, we describe the parameter extraction of the two agent models.

2.2.1. User agent model parameter extraction

Table 3 summarizes the parameters of the user agent model and the raw data from which they were extracted. Average walking speed, average riding speed, and maximum waiting time are deterministic parameters extracted directly from the data, whereas the others are stochastic parameters. These are determined randomly during the simulation and define the rental-return behavior of the user agent model specifically (e.g., at which station, when to rent a bicycle). They follow the four probability distributions that

Table 2
Raw data for Eoulling simulation model (period: 2019.05.01–2020.04.30).

Data	Fields	# of data	Usage in model
Rental–return history	ID, Rental date and time, Rental station no., Return date and time, Return station no., Usage time, Type	765,210	User agent model parameter extraction
Member information	ID, Gender, Age, Nationality, Status, Registration date, Withdrawal date	82,296	
Station information	Station no.[a], Station name, Type, Address, Location [lat, lon], Status	583 (Old: 72, New: 511)	Station agent model parameter extraction
Station monitoring history	Date, Time, # of bicycles in each station	8784	
Population information[b]	Gender, Age, Home location [lat, lon], Job	257,870	
Road network information	Road no., Length, Start junction no., End junction no.	7787	Environment and travel time calculation[c] of the user agent

[a] Old-type stations: NJ_001 – NJ_072, and new-type stations: SJ_001 – SJ_511.
[b] Population information of residents aged between 10 and 70 years who can use Eoulling.
[c] Calculating $T_W(O \rightarrow D)$ of WALKING and $T_R(O \rightarrow D)$ of RIDING[Returning] in Figs. 2 and 3.

Table 3
Parameters of the user agent model.

Parameters	Description	Raw data (or ref.)
First-visit station[a] and arrival time	- The station the user agent visits for the first time for rental from START and the time the agent visits this station (e.g., in Fig. 3, this is station O1 and the arrival time at O1). - The arrival time is determined using a Poisson process, where the interarrival time is exponentially distributed based on the **Poisson distribution** of the first-visit station.	Rental-return history + Member information
Gender and age	- User agent's gender and age - Follows the **gender-age categorical distribution** considering the first-visit station and arrival time.	
First destination[b]	- The station the user agent wants the bicycle returned to (e.g., in Fig. 3, this is station D1). - Follows the **destination categorical distribution** considering the first-visit station, arrival time, gender, and age.	
Bicycle usage time	- The amount of time the user agent uses the bicycle. - Follows the **Weibull distribution** considering the rental station (which may be different from the first-visit station), rental hours, gender, age, and first destination.	
Average walking speed	- Average walking speed of the user agent walking around the stations to rent a bicycle in WALKING. - Determined by the agent's gender and age based on [18].	[18]
Average riding speed	- Average riding speed of the user agent going around the stations to return the bicycle in RIDING (used only for the old-type system). - Set to 15 km/h based on the rental-return history.	Rental-return history
Maximum waiting time	- Maximum waiting time to rent a bicycle; if the user agents fail to rent within this waiting time, they abandon the search and transfer to END, as shown in Fig. 3. - Set to 5 min based on the residents' interview results.	Residents' interview results

[a] The first-visit station may be different from the actual rental station, as shown in Fig. 3.
[b] For the old-type system, the first destination station may be different from the actual return station, as shown in Fig. 3.

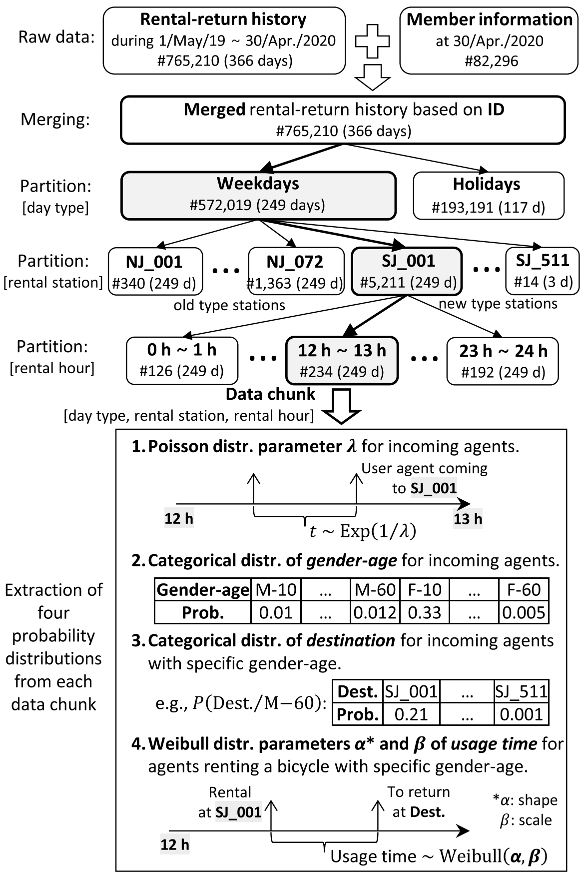

Fig. 4. Parameter extraction for the user agent model. Here, we considered 366 days because in 2020, February had 29 days.

modeled the characteristics of residents and their bicycle rental-return patterns. Fig. 4 shows the process of extracting these distributions from the raw data.

First, we merged the rental-return history and member information based on the user IDs. This enabled adding gender and age to the rental-return history, which enabled capturing the rental-return patterns in more detail according to the users' characteristics; the merged data were classified by rental day types, rental stations, and rental hours to create data chunks. The reason for classifying by day type is that the bicycle usage patterns in Sejong City on weekdays and holidays are highly different. On weekdays, bicycles are mainly used for commuting to work, whereas on holidays, they are used for leisure; indeed, the experimental results of the model validation confirmed this difference (see Figs. 7 and 8). In addition, because the patterns differed considerably for each station, we classified the data by station. We set the time unit for extracting parameters to an hour by appropriately considering the amount of data and variations in the patterns.

For each data chunk, we extracted the four probability distributions for the user agent model. First, assuming that users who visit a station to rent a bicycle follow a Poisson process [19], we extracted the parameter λ (rate), as shown in Fig. 4, which we calculated as the average number of rentals per day in a chunk; the calculations considered that some new-type stations were recently installed (e.g., SJ_511 had only three days of data, as shown in Fig. 4). Second, we extracted a categorical distribution to determine the characteristics of the user agent who come to rent a bicycle at the station. We created 12 categories based on gender and age (e.g., M-10 represents a teenage male) and calculated the average frequency of each category per day from the chunk. Third, we extracted a categorical distribution to determine the destination of the user agents who rented bicycles from the station based on their gender and age. As old- and new-type systems are not compatible, the rental and return had to take place at the same type of station; thus, if the data chunk was for an old-type station, there were 72 categories (because the rental and return station could be the same), whereas if it was for a new-type station, there were 511 categories. Considering gender and age, we calculated the average frequency of each category from the chunk.

Finally, because the users' bicycle usage time could be assumed to follow the Weibull distribution [20], we extracted the two parameters α (shape) and β (scale) of the distribution, as shown in Fig. 4. As mentioned previously, the reason for extracting the distribution from the data rather than calculating the usage time based on the road network information was that the purpose of bicycle usage in Sejong City is not limited to transportation alone; there are many parks in Sejong City, and bicycles are often also used for leisure. In addition, the usage time varies according to gender and age. To extract the distribution, we selected samples of the usage time considering both the destination and gender-age from the chunk. Subsequently, we fit α and β based on these samples, as shown in Fig. 5. If there were insufficient samples, we derived the parameters based on samples selected considering the destination alone. If the samples were insufficient even in this case, we calculated the usage time using the road network information.

The four distributions extracted from each data chunk determined the stochastic parameters of the user agent in Table 3 during the simulation by the following process. First, using a Poisson process, the arrival time of the user agent visiting this station as the first-visit station was determined, where the interarrival time follows the exponential distribution derived from the Poisson distribution of this station. Then, the agent's gender and age were determined according to the gender-age categorical distribution considering the first-visit station and arrival time. In addition, the first destination that the agent would visit to return the bicycle was determined based on the destination categorical distribution, considering the first station, the arrival time, and the agent's gender and age.

The user agent was then activated in START at a time that allowed the agent to arrive at the first station at the specified arrival time. After arrival, the agent would attempt to rent a bicycle at this station in REN_TRY, and depending on the behavior scenario in Fig. 3, if the user agent succeeded in renting a bicycle, the bicycle usage time was determined based on the Weibull distribution considering the rental station (which could be different from the first-visit station), rental hours, gender, age, and the predefined first destination. After the determined usage time in RIDING, the agent arrived at the first destination, and the bicycle was then returned according to the behavior scenario.

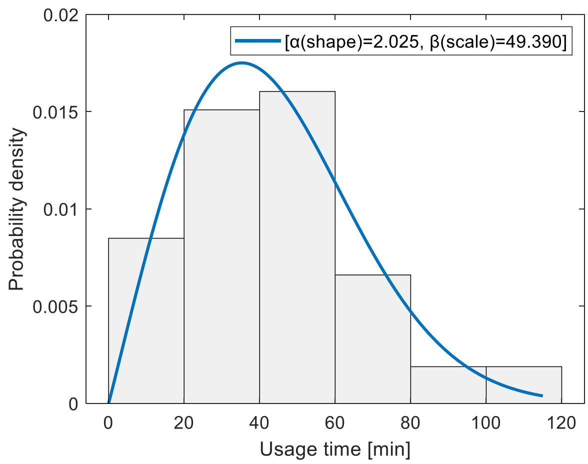

Fig. 5. Example of fitting Weibull distribution of bicycle usage time.

2.2.2. *Station agent model parameter extraction*

Table 4 summarizes the parameters of the station agent model and the raw data from which they were extracted. Unlike the user agent model, all parameters of the station agent model are deterministic. The parameters can be classified into two categories: specifications of the station and population characteristics. We directly obtained the specifications—index number, name, system type, location, and the number of cradles (for the old-type system only)—from the station information. By contrast, we extracted the initial number of bicycles for each station from the station monitoring history. Similar to the rental-return history, we also divided the station monitoring history into weekdays and holidays and estimated the average number of bicycles for each station and each hour. This was because the numbers of bicycles and the initial distributions between weekdays and holidays varied depending on the different rental-return patterns. We used average number of bicycles at each station at 3 a.m., when rentals are negligible, as the initial distribution of bicycles in the simulation, and used the sum of the number of bicycles at each station as the total number of bicycles. Table 5 shows the numbers of stations and bicycles extracted from the station monitoring history; we used this as the baseline.

The factors that affect the utilization of bicycle stations are diverse and complex; therefore, it is difficult to consider all the factors in the simulation. However, owing to the last-mile transportation of the PBSS, its utilization is significantly affected by the population around the station [21]. For example, stations near small, densely populated apartment complexes with large numbers of young people showed similar bicycle rental-return patterns, and these patterns differed from those at the stations near low-density residential areas with mostly older people. Therefore, in this study, using the population information, we extracted the population characteristics of stations such as the number of nearby residents and the proportions of 20- to 59-year-old, gender, and employed persons, as shown in Table 4. Based on this, when adding a new station to the Eoulling simulation, we estimated the user agent model parameters that reflected bicycle rental-return patterns at this station. In particular, because the area of Sejong City where new stations have been installed is mainly a residential area with apartment complexes, it was appropriate to extract and utilize the population characteristics of the stations accordingly. When a new station was added, we extracted the population characteristics around the station. By comparing these characteristics with those of the existing stations, the bicycle rental-return patterns of the most similar stations (Poisson distribution, gender-age distribution, etc.) could be applied to the user agents using this station. This allowed for a more rational analysis of the impact of new stations on the overall Eoulling system.

2.3. *Model implementation*

We implemented the proposed model using DTSim [22], a simulation engine created to efficiently develop a digital twin of Sejong City based on Repast for High Performance Computing [23] because the proposed simulation model is a subproject of developing Sejong City's digital twin; it should also be combined with other simulation models, such as for public transportation and commercial area analysis. Fig. 6 shows the overall structure of the implemented model. We implemented the user and station agent models as UserAgent and StationAgent, respectively, based on CityAgent in DTSim. UserAgentSpace and StationAgentSpace manage the location information of these agents during the simulation based on CitySpace. BikeRoadNetwork provides an environment in which UserAgent and StationAgent can interact with each other. DTSim not only provides a basis for model implementation but also schedules and executes events from each agent model during the simulation through CityScheduler.

As described in Table 2, the implemented model has six raw data points as the inputs. Among these data, the rental-return history and member information are processed in the preprocessor to extract four probability distributions as the parameters of the user agent model, as shown in Fig. 4. The extracted parameters and other data were applied to the model using DataManager. In addition, DataManager created outputs from the simulation of the Eoulling model. Specifically, it outputs the user agents' rental-return history in the same format as the input raw data and also various statistics for analyzing bicycle usage and user convenience (e.g., the number of rentals, average bicycle usage time, average rental waiting time, and rental failure rate).

2.4. *Model validation*

We compared the simulation results and actual data to validate the proposed model using two indicators: the number of rentals per

Table 4
Parameters of the station agent model.

Parameters	Description	Raw data
Index number	- Station number	Station information
Name	- Station name: NJ_[number] (old-type) and SJ_[number] (new-type)	
System type	- Station type: old-type or new-type	
Location	- Location of station	
Number of cradles	- Number of cradles (only for the old-type station)	
Initial number of bicycles	- Initial number of bicycles at station - Determined as average number of bicycles at 3 a.m. on weekdays and holidays	Station monitoring history
Number of nearby residents	- Total number of residents with home location within 500 m around the station.	Population information
20- to 59-year-old proportion	- Proportion of residents aged 20 to 59 who have a relatively high rate of bicycle usage among the nearby residents.	
Gender proportion	- Proportion of gender among the nearby residents.	
Worker proportion	- Proportion of employed persons among the nearby residents.	

Table 5
Numbers of stations and bicycles.

		Old-type	New-type	Overall
# of stations	–	72	511	583
# of bicycles[a]	Weekdays	533	1152	1685
	Holidays	530	1158	1688

[a] The number of bicycles is divided into weekdays and holidays because it can vary depending on their different rental-return patterns. Although not shown in this table, the initial distributions are also different.

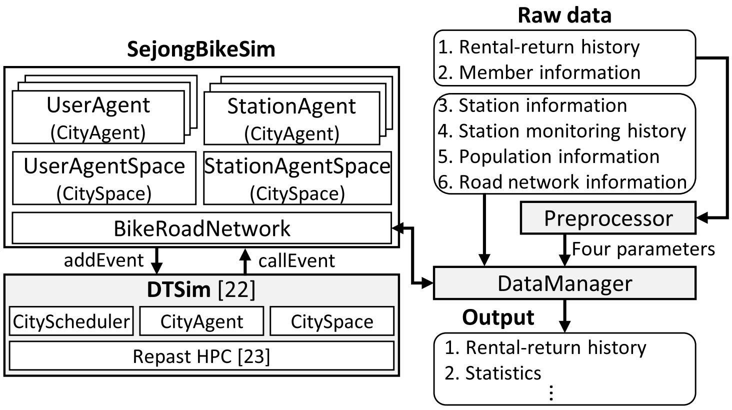

Fig. 6. Structure of implemented data-driven ABM.

hour and the average bicycle usage time per hour, which were actually used to measure the utilization of Eoulling in Sejong City. Since the time period between rental and return is the bicycle usage time, if these two indicators were validated, the number of returns could be also naturally validated; therefore, we did not consider the number of returns. We defined the number of rentals per hour as the total number of bicycles rented at each station per hour. We also defined the average bicycle usage time per hour as the sum of the usage time of each bicycle per hour divided by the total number of bicycles, that is, the average number of minutes a bicycle is used per hour.

As mentioned previously, in Sejong City, the bicycle usage patterns on weekdays and holidays are markedly different; therefore, we conducted comparisons for weekdays and holidays, respectively, in this study. Using the rental-return history data obtained for one year in Table 2, we estimated the averages of these two indicators for 249 days for weekdays and 117 days for holidays. However, in

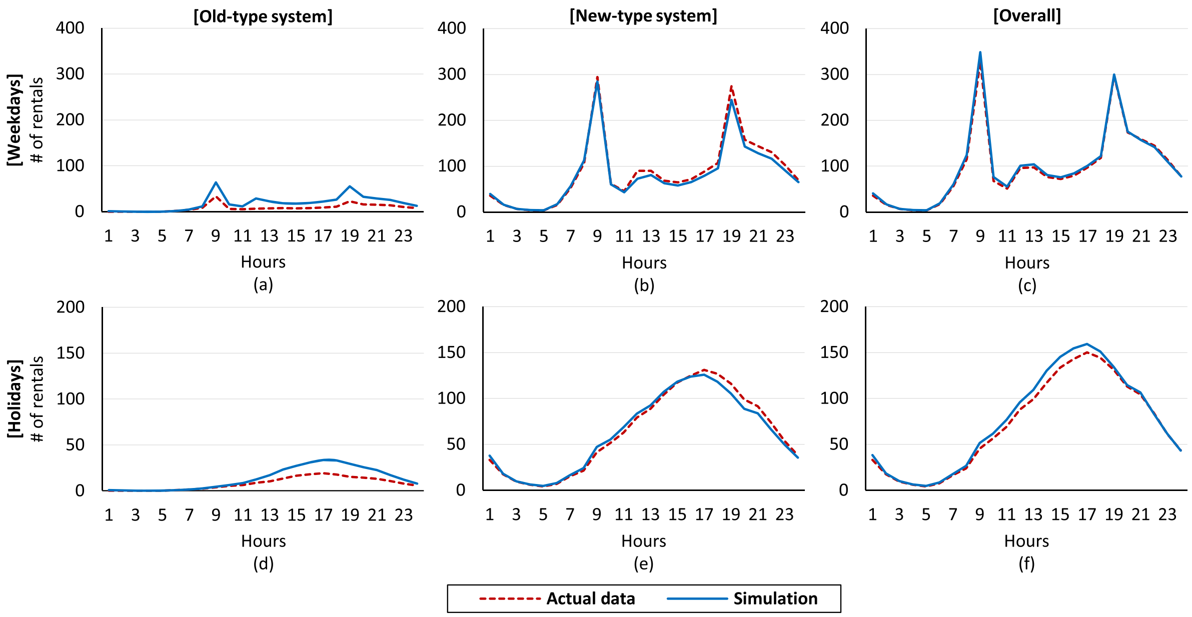

Fig. 7. Comparison of simulation results and actual data for number of rentals per hour.

the simulation, these averages were estimated through 1000 independently repeated simulations. For example, for weekdays, we applied the parameters extracted from the weekday data to the implemented model (e.g., the four probability distributions in Fig. 4, the number of stations and bicycles in Table 5, and the initial distribution of bicycles) and obtained the averages of the two indicators via 1000 independent simulations of this model. Figs. 7 and 8 show the comparison results for the two indicators.

The results indicate that the proposed model adequately describes the Eoulling users' behavior characteristics. On weekdays, because bicycles are mainly used for commuting to work, the number of rentals and average bicycle usage time showed peaks during the commuting hours, as shown in Figs. 7 and 8(a), (b), and (c). By contrast, as shown in Figs. 7 and 8(d), (e), and (f), these peaks occurred in the afternoon because bicycles are used for leisure on holidays. Although some differences exist, the two indicators of the simulation results were highly similar to those of the actual data. This implies that the behavior designs of the user agent model in Fig. 2, the assumptions on the probability distributions in Fig. 4, and the parameter extractions in Tables 3 and 4 are sufficiently accurate.

However, in the case of the old-type system, although the trend was similar, there were large differences in the compared values because there were significantly fewer rentals from the old-type stations, as shown in Figs. 7(a) and (d). Among the total of 765,210 rental-return histories in Table 2, the number of old-type rentals was 71,989, which is less than 10 %; thus, the accuracy of the extracted parameters for the old-type system is low, which leads to inaccurate values. Nevertheless, as shown in Figs. 7 and 8(c) and (f), there is very little difference in the overall system obtained by combining both these systems because the number of rentals for the old-type system was insignificant compared with that for the new-type system. Based on this validated model, we could perform in-depth analysis of users' behavioral characteristics and their convenience when using Eoulling, which could not be extracted from actual data. We describe this in the next section.

As shown in Figs. 8(b) and (e), the difference in the usage time for the new-type system is larger in the afternoon than in the morning, and this is because we did not include any bicycle relocation in the model. Eoulling workers relocate bicycles from stations where they are stacked to stations that have none, but in the model, as the user numbers increased with time, the imbalance increased in the numbers of bicycles at different stations. Consequently, the number of users who attempted to ride new-type bicycles but failed and eventually rode old-type bicycles increased (the old-type bicycles have fewer users and are, therefore, distributed relatively evenly). In the weekday simulation, this ratio was approximately 12 %, and for holidays, it was 8.5 %. Accordingly, as shown in Figs. 8 (a), (b), (d), and (e), the average afternoon bicycle usage time under the simulated old-type system exceeded the actual time, whereas the average for the new-type system was less than that in the data. However, the difference in the results of the overall system was low, as shown in Figs. 8(c) and (f). Improving the model by adding a relocation process will be considered in our future studies.

3. Case study

Using the developed data-driven ABM, we analyzed two concerns pertaining to Eoulling: 1) the effects on user convenience when Eoulling eliminates the old-type system and 2) the appropriate total number of bicycles to reduce operating costs.

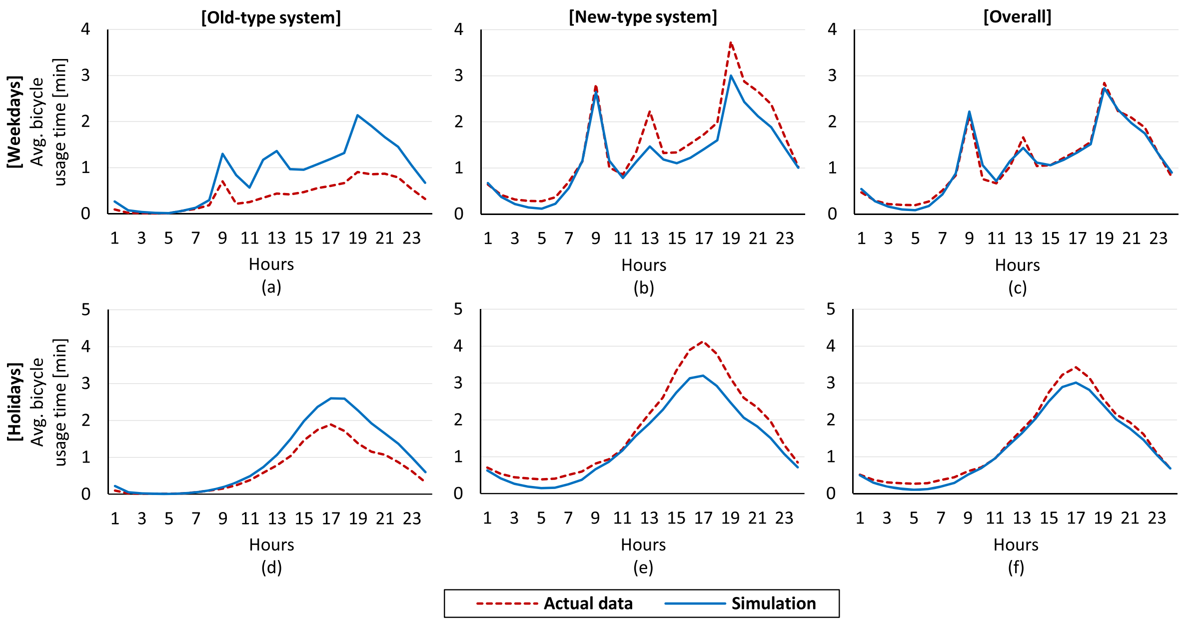

Fig. 8. Comparison of simulation results and actual data for average bicycle usage time per hour.

3.1. Disposal of the old-type system

The Sejong City municipality is still maintaining the old-type system for certain users' convenience even though it is not compatible with the new system and there are few rentals, as shown in Figs. 7(a) and (d). This results in wasted operating costs. Using the proposed model, we could analyze the changed utilization of Eoulling and the impacts on user convenience of discontinuing the old-type system, and our findings could be significantly helpful for managers' decision-making.

As mentioned before, we used the number of rentals and the average bicycle usage time to evaluate utilization, and we used three indicators for user convenience: the average rental waiting time, average return waiting time, and rental failure rate. In the model, the system type of the user agent's first rental attempt is fixed; however, if the user cannot borrow in the first attempt, the next attempt will be made at a nearby station with a bicycle regardless of the station type. Thus, the average rental waiting time was only defined for the overall system. By contrast, the average return waiting time could be defined for each system type, but it was only valid for the old-type system because the new-type system does not rely on cradles. The average return waiting time in the model is the average time period from when the user agents started returning their bicycles to when they completed this action. Finally, we defined the rental failure rate, which had the most direct impact on user convenience, as the proportion of user agents who failed to rent a bicycle within the maximum waiting time (see the "fail to rent" scenario in Fig. 3).

We used three scenarios for this analysis: 1) the present situation, 2) old-type system discarded, and 3) old-type system replaced; Table 6 summarizes the numbers of stations and bicycles for each scenario. The present-day scenario was identical to the model validation scenario, and we used it as a baseline to assess the effects of the other scenarios. In the disposal of the old-type scenario, every old-type station was eliminated, whereas in the old-type system replacement scenario, certain high-utility old-type stations were replaced with new-type stations and otherwise the old-type system was eliminated. For the three scenarios, we estimated the averages for each of the five indicators per day through 1000 independent simulations of the proposed model; Tables 7 and 8 present the simulation results for weekdays and holidays, respectively. In the present-day scenario, the results were divided because old- and new-type systems exist simultaneously, but the results for the other two scenarios are based solely on the new-type system. Fig. 9 shows the number of rentals per hour and average bicycle usage time per hour for the overall system under each scenario.

Comparing the present-day scenario with the disposal of old-type scenario, the number of rentals on both weekdays and holidays decreased slightly when the old-type system was discarded. By contrast, the average bicycle usage time increased significantly, as shown in Figs. 9(c) and (d) and in Tables 7 and 8, because there are fewer bicycles without the old-type system, as shown in Table 6. The fact that the usage time per bicycle increased whereas the number of rentals remained essentially unchanged indicates that the efficiency of utilizing Eoulling increased. However, in terms of convenience, the average rental waiting times increased by 23 % and 30 % on the weekdays and holidays, respectively. Moreover, the rental failure rate, which is the most detrimental factor for convenience, increased by 40 % and 49 %, also respectively, as shown in Tables 7 and 8. The average return waiting time decreased under the disposal scenario, but this value was meaningless because the old-type system was absent. Here, the time span of RET_TRY in the user agent model was 0.5 min. In summary, discarding the old-type system resulted in lower operating costs and greater efficiency, but the convenience of using Eoulling decreased considerably.

The third scenario, replacing of the old-type stations, can be a solution to this problem. Replacing the 59 most commonly used old-type stations among the 72 stations and their bicycles with new-type ones increased both utilization and convenience. As shown in Table 6, the total numbers of stations and bicycles decreased compared with the present-day scenario, but even so, the numbers of rentals on both weekdays and holidays increased significantly, and the rental failure rate decreased, as shown in Figs. 9(a) and (b) and in Tables 7 and 8. Although the average rental waiting time increased slightly, this increase did not significantly affect convenience. To summarize, rather than simply disposing of old-type stations, replacing some of them with new-type stations reduced the unnecessary costs of operating different systems, increased utilization, and secured convenience simultaneously. These simulation results offer a good solution to the first concern for Eoulling's managers. In addition, in the replacement scenario, we added new-type stations based on the population characteristics in the station agent model parameters in Table 4. This will help managers handle the third concern by enabling them to analyze the effects of newly installed stations in the expanding area of Sejong City.

3.2. Appropriate total number of bicycles

As shown in Tables 7 and 8, the average usage time per bicycle is very small; for example, 1685 bicycles of both old- and new-types

Table 6
Numbers of stations and bicycles for three scenarios.

		Scenarios	Old-type	New-type	Overall
# of stations	–	The present	72	511	583
		Disposal of old-type	0	511	511
		Replacement of old-type	0	570	570
# of bicycles	Weekdays	The present	533	1152	1685
		Disposal of old-type	0	1152	1152
		Replacement of old-type	0	1617	1617
	Holidays	The present	530	1158	1688
		Disposal of old-type	0	1158	1158
		Replacement of old-type	0	1627	1627

Table 7
Weekday averages of the five indicators for the three scenarios.

Scenarios	The present			Disposal of old-type	Replacement of old-type
	Old-type	New-type	Overall	Overall (=New-type)	Overall (=New-type)
# of rentals	439.54	1944.32	2383.86	2325.12	2726.68
Avg. bicycle usage time [min]	20.52	29.04	26.4	35.21	28.38
Avg. rental waiting time [min]	–	–	1.6	1.96	1.71
Avg. return waiting time [min]	0.76	0.5	0.55	0.5	0.5
Rental failure rate [%]	–	–	5.87	8.22	5.70

Table 8
Holiday averages of the five indicators for the three scenarios.

Scenarios	The present			Disposal of old-type	Replacement of old-type
	Old-type	New-type	Overall	Overall (=New-type)	Overall (=New-type)
# of rentals	315.53	1492.08	1807.61	1762.54	2084.97
Avg. bicycle usage time [min]	23.09	32.45	29.51	40.06	31.25
Avg. rental waiting time [min]	–	–	1.27	1.65	1.45
Avg. return waiting time [min]	0.66	0.5	0.53	0.5	0.5
Rental failure rate [%]	–	–	4.2	6.27	4.05

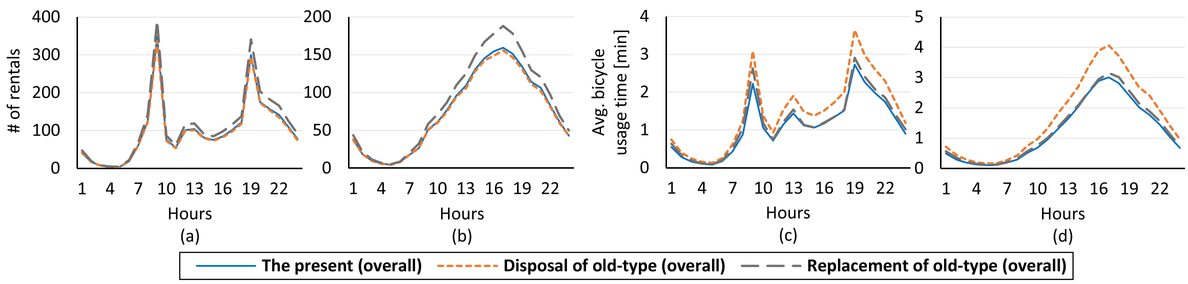

Fig. 9. Comparison results for three scenarios: (a) number of rentals on weekdays, (b) number of rentals on holidays, (c) average bicycle usage time on weekdays, and (d) average bicycle usage time on holidays; these results are for the overall system.

are used on average for only approximately 27 min on weekdays. This indicates that there are more bicycles than there is demand for them. Eoulling managers want to reduce the number of bicycles to decrease operating costs, but they are concerned about how this reduction will affect users' convenience; using the proposed model, they can estimate the demand for bicycles at each station and arrange for only the necessary number of bicycles lowering operating costs without significantly degenerating convenience.

To estimate the demand, we calculated the average number of the user agent's rental attempts (i.e., REN_TRY) per hour for each station through independently repeated simulations based on the present-day scenario. Instead of actual rentals, we considered rental attempts because the former might not reflect the demand; for example, the number of rentals for a station with high demand is zero if it does not have bicycles. We estimated the demand at each station on weekdays as the sum of the average rental attempts at each station until 12:00; similarly, for holidays, we used the average rental attempts until 15:00 to estimate the demand. We set the estimated demand at each station as the initial number of bicycles at each station. The number of rental attempts after 12:00 on weekdays and after 15:00 on holidays could also be used to estimate the demand; however, in that case, it would be difficult to obtain the effect of the operating cost reduction because it would not be possible to significantly reduce the number of bicycles. Because bicycles are arranged according to the estimated demand at each station, the total number of bicycles decreases by 402 and 487 on weekdays and holidays, respectively, reflected in the reduction of bicycles scenario presented in Table 9. Fig. 10 shows the differences in the numbers of bicycles at 583 stations overall under this reduction scenario compared with the present-day scenario as extracted from the station monitoring data in Table 2 for weekdays. As shown, the number of bicycles in the old-type system decreased.

To compare the two scenarios, we estimated the averages of the five indicators through 1000 independent simulations of the

Table 9
Number of bicycles for two scenarios.

		Scenarios	Old-type	New-type	Overall
# of bicycles	Weekdays	The present	533	1152	1685
		Reduction of bicycles	180	1127	1283
	Holidays	The present	530	1158	1688
		Reduction of bicycles	143	1058	1201

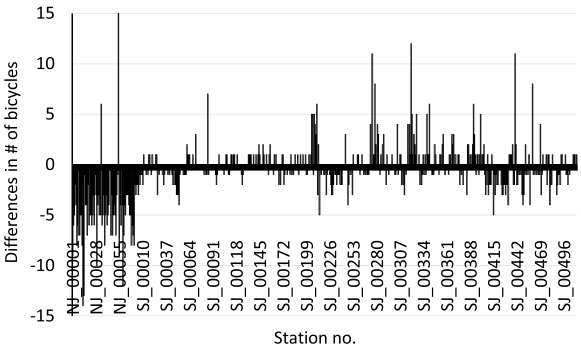

Fig. 10. Differences in number of bicycles of 583 stations overall in the reduction of bicycles scenario, relative to the present-day scenario, on weekdays.

proposed model. Tables 10 and 11 present the simulation results for weekdays and holidays, respectively, and Fig. 11 shows the number of rentals per hour and average bicycle usage time per hour for the two scenarios. As is shown, there was little change in the number of rentals, but the average usage time increased because there were fewer bicycles; in this way, the efficiency of utilizing Eoulling increased. Although the average rental waiting time and rental failure rate increased slightly, as shown in Tables 10 and 11, these increments did not significantly degrade the convenience; instead, the return waiting time for the old-type system decreased because the reduced number of old-type bicycles; this, in turn, positively affected convenience. In short, reducing the number of bicycles but optimizing their placement according to the demand can reduce operating costs without significantly affecting utilization and convenience. These simulation results provide a good solution to the second concern for Eoulling managers.

For all three Eoulling manager concerns, Table 12 summarizes the valuable conclusions drawn from simulating the proposed model in the case studies; these outline how the proposed model can be effectively used to simulate and analyze Eoulling in Sejong City.

4. Conclusion

M&S is an effective tool to design, analyze, and optimize PBSSs. Although many achievements pertaining to the M&S of PBSSs were reported, these studies based on top-down approaches had limitations in reflecting users' behavioral characteristics and analyzing their convenience. In addition, these existing studies had difficulty in fully utilizing the actual operations data of Eoulling, along with the population and geographic information of Sejong City. As a bottom-up approach, ABS is particularly appropriate for reflecting the behavioral characteristics of Eoulling users in Sejong City. In this study, we proposed and then built a data-driven ABM to simulate Eoulling, which consisted of user and station agents. The user agent model had eight states corresponding to the operations related to renting and returning bicycles, and it interacted with the station agent model using these operations to rent and return bicycles during the simulation. To increase model fidelity, the parameters of the two agent models, including the four probability distributions that modeled the residents' characteristics and their bicycle rental-return patterns, were extracted from the six types of actual data. In

Table 10
Weekday averages of the five indicators for two scenarios.

Scenarios	The present			Reduction of bicycles		
	Old-type	New-type	Overall	Old-type	New-type	Overall
# of rentals	439.54	1944.32	2383.86	335.46	2035.42	2370.88
Avg. bicycle usage time [min]	20.52	29.04	26.4	46.78	29.76	32.05
Avg. rental waiting time [min]	–	–	1.6	–	–	1.67
Avg. return waiting time [min]	0.76	0.5	0.55	0.59	0.5	0.51
Rental failure rate [%]	–	–	5.87	–	–	5.96

Table 11
Holiday averages of the five indicators for two scenarios.

Scenarios	The present			Reduction of bicycles		
	Old-type	New-type	Overall	Old-type	New-type	Overall
# of rentals	315.53	1492.08	1807.61	251.13	1552.56	1803.69
Avg. bicycle usage time [min]	23.09	32.45	29.51	70.09	36.58	40.57
Avg. rental waiting time [min]	–	–	1.27	–	–	1.35
Avg. return waiting time [min]	0.66	0.5	0.53	0.52	0.5	0.50
Rental failure rate [%]	–	–	4.2	–	–	4.33

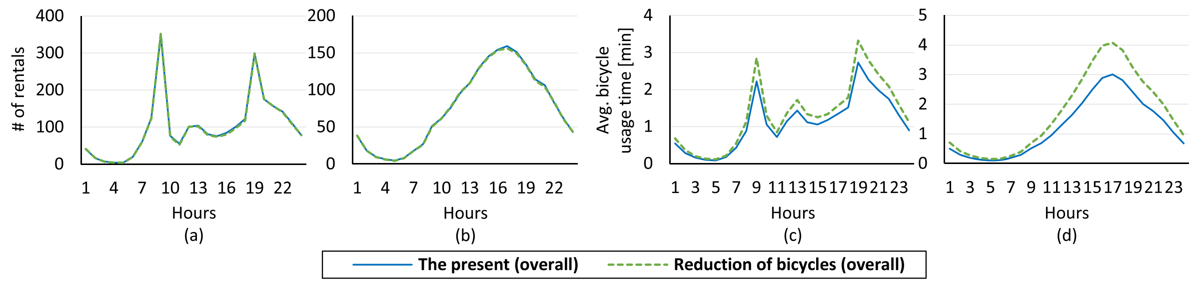

Fig. 11. Comparison results for two scenarios: (a) number of rentals on weekdays, (b) number of rentals on holidays, (c) average bicycle usage time on weekdays, and (d) average bicycle usage time on holidays; these results are for the overall system.

Table 12
Conclusions drawn from the simulation of the proposed model for the three Eoulling manager concerns.

No.	Concerns and conclusions
1	The effects on convenience when the incompatible old-type system is eliminated to reduce costs. ð If the old-type system is disposed of, the convenience of users is reduced considerably. ð Replacing popular old-type stations with new-type ones can secure convenience as well as reduce the unnecessary costs of operating different systems and increase utilization.
2	The appropriate total number of bicycles to reduce costs without significantly impairing convenience. ð Although the total number of bicycles is reduced, the utilization and convenience are not significantly degraded if the bicycles are appropriately arranged according to the demand estimated by simulation.
3	Suitable locations for the new-type stations around the expanding city. ð The proposed model can analyze the effect of newly installed stations owing to the function of adding a new station based on the population characteristics of stations, which enables managers to solve this concern.

addition, a road network environment was constructed for interactions between the two models. We implemented this model with DTSim based on Repast, and the validation results showed that the proposed model accurately describes users' behavioral characteristics when they use Eoulling in Sejong City.

Through simulations of the validated model, we evaluated user agents' utilization of Eoulling and their related convenience under several virtual scenarios. We expect these results to assist managers in addressing two concerns: disposing of the old-type system and establishing the appropriate total number of bicycles for the system in place. In addition, the proposed model can help managers handle the third concern by enabling them to analyze the effects of newly installed stations around the expanding city based on population characteristics. Although we designed the model to simulate the PBSS in Sejong City, it can be applied to PBSSs similar to Eoulling as long as their actual operations data are available. Thus, we plan to apply this model to analyze other systems in South Korea; we also aim to improve the model by incorporating dynamic and static bicycle relocation processes.

Availability of data and materials

Data available on request due to privacy restrictions of the Sejong City municipality

Funding

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Korean Government (Ministry of Science and ICT) under Grant 2022R1F1A1066267 and supported in part by Institute of Information & Communications Technology

Planning & Evaluation (IITP) grant funded by the Korean Government (MSIT) (No. 2018–0–00225, Development of City Interior Digital Twin Technology to Establish Scientific Policy).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the Sejong City municipality for providing various operations data and useful information regarding Eoulling. Additionally, the authors would like to thank an Editor, and Area Editor, and the anonymous reviewers for their valuable comments and constructive suggestions.

References

[1] Y. Jin, C. Ruiz, H. Liao, A simulation framework for optimizing bike rebalancing and maintenance in large-scale bike-sharing systems, Simul. Model. Pract. Theory 115 (2022), 102422.
[2] E. Eren, V.E. Uz, A review on bike-sharing: The factors affecting bike-sharing demand, Sustain. Cities Soc. 54 (2020) 1–12.
[3] C. Park, S.Y. Sohn, An optimization approach for the placement of bicycle-sharing stations to reduce short car trips: An application to the city of Seoul, Transp. Res. A Policy Pract. 105 (2017) 154–166.
[4] C.S. Shui, W.Y. Szeto, A review of bicycle-sharing service planning problems, Transp. Res. C Emerg. Technol. 117 (2020), 102648.
[5] J. Garcia-Gutierrez, J. Romero-Torres, J. Gaytan-Iniestra, Dimensioning of a bike sharing system (BSS): a study case in Nezahualcoyotl, Mexico, Procedia Soc. Behav. Sci. 162 (2014) 253–262.
[6] J. Romero, A. Ibeas, J. Moura, A simulation-optimization approach to design efficient systems of bike-sharing, Procedia Soc. Behav. Sci. 54 (2012) 646–655.
[7] I. Frade, A. Ribeiro, Bike-sharing stations: A maximal covering location approach, Transport. Res. A 82 (2015) 216–227.
[8] D. Çelebi, Bicycle sharing system design with capacity allocations, Transp. Res. B Methodol. 114 (2018) 86–98.
[9] L. Conrow, A.T. Murray, H.A. Fischer, An optimization approach for equitable bicycle share station siting, J. Transp. Geog. 69 (2018) 163–170.
[10] J.C. Garcia-Palomares, J. Gutiérrez, M. Latorre, Optimizing the location of stations in bike-sharing programs: A GIS approach, Appl. Geogr. 35 (2012) 235–246.
[11] J.H. Wang, C.H. Tsai, P.C. Lin, Applying spatial-temporal analysis and retail location theory to public bikes site selection in Taipei, Transp. Res. A Policy Pract. 94 (2016) 45–61.
[12] C. Fricker, N. Gast, Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity, Eur. J. Transp. Logistics 5 (2016) 261–291.
[13] K. Bernhardt, Agent-based modeling in transportation, Artif. Intell. Transp. 72 (2007) 72–80.
[14] B.N. Othman, et al., SUMMIT: A multi-modal agent-based co-simulation of urban public transport with applications in contingency planning, Simul. Model. Pract. Theory 126 (2023), 102760.
[15] A. Serrano-Hernandez, J. Faulin, P. Hirsch, C. Fikar, Agent-based simulation for horizontal cooperation in logistics and transportation: From the individual to the grand coalition, Simul. Model. Pract. Theory 85 (2018) 47–59.
[16] P. Jing, H. Hu, F. Zhan, Y. Chen, Y. Shi, Agent-based simulation of autonomous vehicles: A systematic literature review, IEEE Access 8 (2020) 79089–79103.
[17] S. Bouarfa, H.A. Blom, R. Curran, M.H. Everdij, Agent-based modeling and simulation of emergent behavior in air transportation, Complex Adapt. Syst. Model. 1 (2013) 15.
[18] M. Schimpl, et al., Association between walking speed and age in healthy, free-living individuals using mobile accelerometry—A cross-sectional study, PloS One 6 (2011) e23299.
[19] R.M.A. Swaszek, C.G. Cassandras, Receding horizon control for station inventory management in a bike-sharing system, IEEE Trans. Autom. Sci. Eng. 17 (2020) 407–417.
[20] S. Susilawati, M.A.P. Taylor, S.V.C. Somenahalli, Distributions of travel time variability on urban roads, J. Adv. Transp. 47 (2013) 720–736.
[21] P. Rietveld, The accessibility of railway stations: The role of the bicycle in The Netherlands, Transp. Res. D Transp. Environ. 5 (2000) 71–75.
[22] Y.C. Kim, K.S. Jin, S.M. Lee, Implementation of Simulation Platform for Smart City Digital Twin, in: Proc. Summer Annual Conf. IEIE, Jeju Island, South Korea, 2020, pp. 1869–1972.
[23] N. Collier, M. North, Repast HPC: A platform for large-scale agentbased modeling, in: W. Dubitzky, K. Kurowski, B. Schott (Eds.), Large-Scale Computing, Wiley, New York, 2011, pp. 81–109.