

## Research Article

# Development of a Model for Evaluating the Coverage Area of Transit Center Using Smart Card Data

Hyunjun Hwang <sup>1,2</sup> Shin-Hyung Cho <sup>3,4</sup> Dong-Kyu Kim <sup>1,4</sup>  
and Seung-Young Kho <sup>1,4</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea

<sup>2</sup>Department of Transportation System Research, The Seoul Institute, Seoul 06756, Republic of Korea

<sup>3</sup>School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

<sup>4</sup>Institute of Construction and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea

Correspondence should be addressed to Seung-Young Kho; sykho@snu.ac.kr

Received 23 April 2020; Revised 18 October 2020; Accepted 29 October 2020; Published 19 November 2020

Academic Editor: Hong Yang

Copyright © 2020 Hyunjun Hwang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since metropolitan cities are broadening as a result of urban sprawl, multimodal transportation systems have been adopted to fulfill the connection between the suburban and urban areas. The transportation system is being revamped around the transit center in the urban area to facilitate access to the downtown area from the suburbs. Studies are being conducted to improve the accessibility of public transportation by using the concept of hub-and-spoke. In this study, we develop a coverage area index (CAI) to assess the impact of a transit center on access to urban areas from the suburbs quantitatively. The concept of network centrality and the kernel density function is used to evaluate the extent of the influence of a transit center. The smart card data in the Seoul metropolitan area are used to analyze the CAI. Six transit centers in the Seoul metropolitan area are investigated to compare the coverage area to the transit center. The bandwidth of the kernel density function is set as 2 km considering the size and influence of each region. We evaluate six transit centers using the CAIs in Seoul compared to the index characteristics with transit accessibility (TA) index from previous studies. The CAI is possible to identify the incompetent centers, alternative routes to solve the problems of overcrowding on the centers, and areas with insufficient supplies of regional transit.

## 1. Introduction

As the population increases, urban sprawl can occur rapidly, and it can impinge on the suburban areas. A hub-and-spoke structure of the integrated multimodal public transit system is adequate for covering large suburban areas with restricted infrastructure and operation costs. Travelers who wish to go downtown take regional express transit from suburban areas to access a transit center near the downtown area, where they can transfer to a city bus or an urban railway to take them downtown. This system is cost-effective, but passengers may be uncomfortable due to transfers if transit modes are not well coordinated.

Transit centers serve as coordination points between downtown and the suburban areas in a hub-and-spoke structure

network. To evaluate the functionality of transit centers, operators should investigate the area influenced by the centers. The investigation will provide operators with information concerning the expected boundary of influence when some actions are implemented at the transit centers. Conversely, poor connectivity of a suburban area to the transit centers means that trips from the suburban area to downtown are inefficient.

Several studies have evaluated the effectiveness of transit centers using a data envelope analysis (DEA) model [1, 2]. Some studies have developed an index of connectivity and accessibility with variables like headway and transfer time [3, 4]. In the planning phase, other studies have used accessibility and connectivity indices derived from the socioeconomic indices of neighborhoods to evaluate transit centers [5, 6], and they have used heuristic models to

evaluate the locations of transit centers [7]. However, such models have limitations that they do not consider the origins and destinations of transferred travelers. To overcome this limitation, we suggested developing an index using smart card data to know the origin and destination of transit center users.

This study aims to develop a coverage area index (CAI) for transit centers considering the center's role which is to facilitate the connection in the hub-and-spoke structure of the regional transit system that connects the suburban areas with the downtown area. The concept of betweenness centrality is applied to the origin and destination of actual trips, aiming to indicate the travelers' preferences. The CAI is defined as the proportion of the number of trips through a transit center to the number of total trips taken from the suburban area to downtown. The kernel density function is employed to aggregate discrete trip chain data to traffic analysis zones. As the index usage, we discuss a comparison of transit centers, transit supply, and structure of the transit network.

Research on the assessment of transit centers includes relative evaluation between centers with empirical data, defining and measuring indices, using socioeconomic factors to identify optimal hub locations during the planning phase and developing indices using a network centrality concept from social science. There have been several studies on the data-based evaluation of transit centers using the DEA model. The relative efficiency of individual centers was calculated through mutual comparisons of the centers' defined inputs and outputs. DEA model was employed with the inputs, e.g., the areas of the hubs, the cost of operation, the number of staff members required, and capacities of the hubs. The model's outputs were the number of transfers per day, average, and deviations of the transfer times [1]. Another research employed DEA model with the inputs, e.g., transfer time, number of subway and bus lines, and waiting time of the hubs and outputs of the model were number of transfer trips and transfer rate [8]. In Kochi City, Japan, the same DEA model was used with smart card data to evaluate transfers between buses and trams. Headways and the number of lines were used as input, and the number of transfers and the time required for transfers was the desired outputs of the model [2]. The DEA model was useful for assessing the efficiency of the transit centers, but it was difficult to evaluate them in conjunction with the surrounding area.

Accessibility and connectivity were common indices used to evaluate the transit centers, and there were various detailed definitions in each concept. From the first perspective, connectivity was regarded as an index for evaluating the availability of transit. A study evaluated the connectivity of nodes, lines, transfer centers, and networks using the connecting power calculated from the headway and capacity of the transit lines [3]. From the second perspective, connectivity was regarded as an index for evaluating travelers' experiences. There was a study that defined connectivity as the transfer time between modes and that measured the index with empirical data from a transit corridor in Auckland, New Zealand [4].

Both views treated connectivity as a function of capacity and time, but not considered the spatial distribution of trips.

Accessibility of the transit facility was evaluated based on the population or the number of jobs within a small distance from the facility [9, 10]. In the Transit Capacity and Quality of Service Manual (TCQSM), the traffic analysis zones' accessibility index was defined as the percentage of an area within a certain accessing distance of stations among the transit-supportive area [11]. A study measured the accessibility of zones with a weighted average of several indices to find services gaps [12]. Accessing distance and coverage of a facility were often considered as a walking distance to access the station, but some studies expanded the concepts of accessing distance and coverage to other modes. One study measured the accessibility of a station with a total area that could be reached in a reasonable time by walking, riding a bicycle, and using local bus services [13]. Another study showed a transit center's access zone with proper access time to reach the center with multimodal transit [14]. Another accessibility measure that considers the origin-destination (OD) of trips was motivated by the gravity model of demand forecasting [15]. There have been many studies that measured accessibility using the supply of transit service, distance, and time. In contrast, there have been only a few studies that focused on the transit center and the actual OD pair.

At the planning level, the locations of transit centers were optimized to maximize the network's performance considering the center of mass as the optimal location of the hub. This approach had a limit in that it assumed that distance was the only variable in the relationship between a transit center and the coverage area [5]. In another approach, it was assumed that the accessibility index is the neighborhood population not overlapping with other neighborhoods. The connectivity index was viewed as the number of connections between nodes and areas with large numbers of travelers to optimize the overall transit network [6].

Some approaches have used the network centrality concept from network theory to evaluate transportation networks. Several centrality measures were used to evaluate traffic analysis zones in Seoul, Korea, with a kernel density function to aggregate network nodes to zones [16]. Specifically, closeness centrality was used to represent the accessibility of transit nodes [17].

Studies used smart card data to assess the accessibility of transit and acquire total ridership of the transit network. There was a study that developed a platform using smart card data and vehicle GPS data. The platform was designed to visualize transit performance measures in several levels—network, route, and stop level [18]. Another approach of using smart card data was finding a relationship between the data and demographics, land use, and transportation factors to predict ridership or purposes of trips [19, 20]. Various data other than ridership can be obtained from smart card data in Seoul where the departure and destination time and location were specified. There were studies used the travel time obtained from the data to assess the competitiveness of transit network compared to other

modes [21, 22]. There was another approach using mobile phone data to get the origin and destination of trips with unknown modes. The origins and destinations, travel time, and distance of the private car and public transit were compared to define the accessibility index [23]. These studies presented some ways to assess accessibility by data-driven approach, but they did not use the origin and destination of transit trips. Evaluating transit centers, not only origin and destination but also the transfer locations of transit trips, also needed to be considered.

As indicated above, there are several ways to evaluate transit centers, but there is no specific measure that can be used to evaluate the pivotal role of transit centers as connecting nodes between suburban areas and the downtown area with data from actual trips. Thus, if we understand the connections in the transit center, our understanding of the overall distribution of trips in the metropolitan area can be improved. At the network level, centrality can be one of the concepts that address connecting the role of a transit center, but previous research does not concentrate on transit centers.

## 2. Model Formulation

**2.1. Definition of Coverage Area for a Transit Center.** The transit center was defined as a meeting point of different modes and several routes and lines that function as effective hubs working as a regional collection-distribution function [7, 11]. The original concept of coverage was that “something can be understood as an area in which the influence of such a thing can be perceived” [24]. The coverage of public transportation facilities refers to the range of space accessible by walking or riding a bicycle. For a rail station in the suburban area, the coverage can be extended to an accessible zone with automobile access, such as drop off or kiss-and-ride locations in a particular case [11]. In this study, because the transit center is the location where many travelers transfer from fast regional transit to city transit, the access modes to stipulate the coverage of transit centers were expanded to every mode, including regional buses. From this perspective, the coverage area for each transit center is the set of suburban zones which are easily accessible and strongly dependent on the transit center to approach the central city area. The expanded concept of a coverage area in a metropolitan area is shown in Figure 1. Travelers departing from the suburban zones go through transit centers in the route of a trip to the central city. Thus, with a well-constructed hub-and-spoke structure, a large portion of travelers transfers from the regional transit to other modes at the transit centers. This phenomenon is similar to the “betweenness” concept in network theory. To analyze the expanded coverage, we developed a coverage area index using betweenness centrality concepts. OD trips were used to measure centrality instead of a transit network topology because the topology was too complicated to reflect travelers’ preferences. The kernel density function was used to obtain correctly aggregated OD trip data from smart card data.

**2.2. Network Centrality.** Network centrality is a measure to evaluate the nodes of networks in social science and network theory. There are several kinds of centrality, e.g., betweenness centrality, closeness centrality, and straightness centrality [25–27]. Betweenness centrality is defined as how a zone is located between all pairs of other zones. It is measured as the ratio of the number of shortest paths through the zone to the total number of the shortest paths between all pairs of zones. It is a suitable centrality measure to evaluate the coverage area of transit centers because these centers’ role in transit network is the connection node between zones in the coverage area and the downtown zones. Betweenness centrality can be expressed mathematically as follows [16]:

$$C_B(i) = \frac{1}{(N-1)(N-2)} \sum_{\forall j,k | j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \quad (1)$$

where  $C_B(i)$  means betweenness centrality of zone  $i$ ,  $N$  means the number of zones,  $n_{jk}(i)$  means the number of the shortest paths from zone  $j$  to zone  $k$  through zone  $i$ , and  $n_{jk}$  means the number of the shortest path from zone  $j$  to zone  $k$ .

To apply the definition for the coverage area, trip chain data are used instead of the shortest-path data. The term of “number of the shortest paths” is substituted with the number of trips due to traveler choice behavior to define CAI.

**2.3. Kernel Density Function (KDF).** When aggregating actual trip data, boarded or alighted zones recorded in the smart card data are not the actual origin or destination zones of the trips because travelers walked or took other modes without records to the neighbor zones to access transit. The kernel density function (KDF) is used to apply this neighborhood effect. The value of KDF is set by the type of function and the bandwidth. In this paper, we use the simple triangular kernel function. The influence of neighbor areas decreases linearly with distance, and when the distance exceeds the bandwidth, the influence became zero. KDF can be formulated as

$$K(d_{ik} | w) = \begin{cases} 1 - \frac{d_{ik}}{w}, & \text{if } d_{ik} < w, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where  $d_{ik}$  means the Euclidean distance between the centroids of zone  $i$  and  $k$  (km) and  $w$  means the kernel bandwidth (km).

The bandwidth of KDF is the only variable that affects the power of the neighborhood effect, but the optimal bandwidth is not defined. In this research, we set the bandwidth to 2 km, which is the median of the Euclidean distance between adjacent zones in the Seoul metropolitan area. This bandwidth is justified based on the case study.

**2.4. Coverage Area Index (CAI).** Inevitably, suburban zones in a metropolitan area can be connected directly to only some points of the central city because of the limitation in

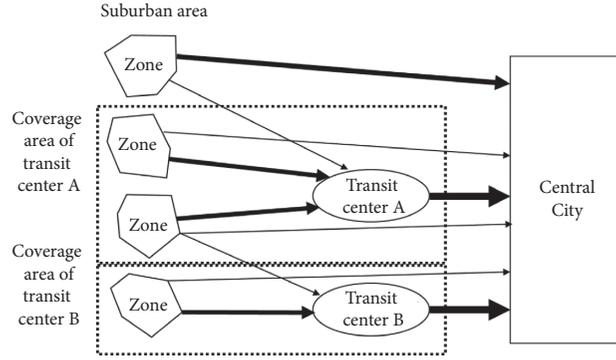


FIGURE 1: An expanded concept of a coverage area of the transit center.

the infrastructure or the running cost. Given this constraint, travelers have to transfer from regional transit to city transit, such as an urban bus or the metro, and transit network should be formed as a hub-and-spoke structure. Existing accessibility indices for a transit center have rarely been focused on the complexity of the transit network and travelers' choice behaviors caused by the structure of the network. Thus, this study uses actual trip data that reflects all travelers' personal preferences rather than using the shortest path in betweenness centrality. Actual trip data allow us to analyze the dependence of travelers from each suburban zone on specific transit points by using the definition of betweenness centrality inversely. Since actual trip data represent current travelers' travel behavior, it would be suitable to analyze the current coverage rather than predict the future or potential coverage.

From equation (1), the central city was considered as zone  $j$ , and each zone in the suburban area was considered as zone  $k$ . The number of shortest paths between zones  $j$  and  $k$  was transformed into the number of trips between the central city and each zone. The number of shortest paths through a transit center,  $i$ , was transformed to the number of trips between the central city and the zone that included a transfer at the transit center. This was reasonable because a route that is chosen for a trip indicates that the route had the smallest general cost among the possible choice set.

The coverage area index of zone  $i$  for transit center  $j$  ( $CAI_{ij}$ ) is defined as the ratio of the modified number of trips through the transit center  $j$  to the total number of trips taken from zone  $i$  to the central city. KDF is used to account for the neighborhood effect and modify the original number of trips. The definition of CAI is

$$CAI_{ij} = \frac{T'_{ijC} + T'_{Cji}}{O'_{iC} + D'_{iC}}, \quad (3)$$

$$\begin{aligned} O'_{iC} &= O_{iC} + \sum_{k \neq i} K(d_{ik}|w)O_{kC}, \\ D'_{iC} &= D_{iC} + \sum_{k \neq i} K(d_{ik}|w)D_{kC}, \\ T'_{ijC} &= T_{ijC} + \sum_{k \neq i} K(d_{ik}|w)T_{kjC}, \end{aligned} \quad (4)$$

where  $CAI_{ij}$  means the coverage area index of zone  $i$  for transit center  $j$ ,  $O_{iC}$  means the total number of trips to the central city from zone  $i$ ,  $D_{iC}$  means the total number of trips from the central city to zone  $i$ ,  $d_{ik}$  means the Euclidian distance between the centroid of zones  $i$  and  $k$ ,  $T_{ijC}$  means the number of trips from zone  $i$  to the central city through transit center  $j$ ,  $T_{Cji}$  means the number of trips from the central city to zone  $i$  through transit center  $j$ , and  $K(d_{ik}|w)$  means the kernel density function with distance  $d_{ik}$  and bandwidth  $w$ .

CAI is the ratio, so it has a value between 0 and 1. When the value is close to 1, most travelers from the zone to the central city transfer to city transit at the center. This indicates that the zone is in the area covered by the transit center. When the index is close to 0, almost none of trips from the zone to the central city passed through the transit center. If the central city is big enough to have more than one transit center, the index can be measured individually for each transit center.

We employ the two methods for network evaluation, which are network centrality and kernel density function. To evaluate the transit center, betweenness centrality is used to define the geographical location from other zones, and kernel density function is applied to catch the effective area from the transit center. While betweenness centrality is focused on one zone or node, CAI considers the trips between transit center and destination compared to other transit centers. It is shown that comparing equations (1) and (3), betweenness centrality is defined as  $C_B(i)$  and CAI is defined as  $CAI_{ij}$ . This characteristic gives CAI a unique usage that not only evaluates transit centers but also tests transit network of each zone with spatial distribution of CAI.

### 2.5. Evaluation of CAI Compared to Accessibility Index.

The CAI usage was evaluated by the contrast of the CAI and the accessibility measure presented in the previous study. The accessibility measure used in the study was defined as the product of attractiveness at destinations and friction between zones and destinations. This interpretation came from the demand forecasting model with socioeconomic conditions and transit networks [15], using parameters and

variables from Sacramento, CA [28]. Because this study was focused on Seoul, Korea, to reflect local characteristics, the parameters and variables from Korea's national demand forecasting model were used [29]. The model was derived

from a national survey and used a gravity model to forecast. To calculate TA with the model, some of the variables needed to be modified while maintaining the TA's fundamental meaning. The definition of TA is

$$\begin{aligned} TA_{ij} &= (CATT_j) * (ZFRIC_{ij}), \\ CATT_j &= 2.296 * EMP_j + 0.536 * POP_j + 0.443 * EMPS_j + 0.801 * STU_j, \\ ZFRIC_{ij} &= \begin{cases} 0.1093 * MTD_{ij}^{1.58808} * \exp(-0.2437 * MTD_{ij}), & \text{if } MTD_{ij} < 13.0 \text{ km,} \\ 2.4043 * MTD_{ij}^{-0.73839} * \exp(-0.0190 * MTD_{ij}), & \text{if } MTD_{ij} > 13.0 \text{ km,} \end{cases} \end{aligned} \quad (5)$$

where  $TA_{ij}$  means the transit accessibility between zone  $i$  and center  $j$ ;  $CATT_j$  means the attraction of the transit center  $j$ ;  $ZFRIC_{ij}$  means the friction of access between zone  $i$  and the center  $j$ , which was defined as a combined model;  $EMP_j$  means the total number of employers in the zone of transit center  $j$  (people);  $POP_j$  means the population of the zone of transit center  $j$  (people);  $EMPS_j$  means the total number of people working in service in the zone of transit center  $j$  (people);  $STU_j$  means the total number of students in the zone of transit center  $j$  (people); and  $MTD_{ij}$  means median travel distance by transit between zones  $i$  and  $j$ .

### 3. Case Study

**3.1. Data Description.** Smart card data in the Seoul metropolitan area were used to obtain public transit trip chain data, and the data covered almost 99% of trips by public transit in Seoul. Travelers' boarding time and station, arrival time and station, modes used, the number of transfers, a unique ID for each smart card, and a transaction ID for each trip chain are recorded because users have to tag their card both when they get on and off transit modes in Seoul. With appropriate treatment using the transaction ID recorded in the data, all trip chains between the suburban area and Seoul and locations they passed through can be obtained. In this research, 180 million trip data for five weekdays (May 16–19 and May 22, 2017) were used. Data for 2.5 million trips (2.33% of total trips) were not used because they did not include arrival time and station. After treatment, 15.77 million trip chains were obtained between suburban areas and Seoul.

Six transit centers were selected for the case study. They serve as the destinations for most of the regional express buses from the Seoul suburban area because they are accessible to the highway, and they are connected to the metro lines in Seoul. Three of them (Gangnam, Sadang, and Jamsil) are in the southern part of Seoul near the Gangnam business district, subcenters of Seoul. The other three (Seoul Station, Cheongnyangni, and Hapjeong) are near the central business district, the historical downtown of Seoul, in Figure 2.

About two million travelers transferred their modes from regional bus to the city bus and subway or vice versa at the centers each day. An average number of trips between each zone in the suburban area and Seoul was 4,526.5 per

day, and 13.8% of the total trips (626.8 trips per day for each zone) passed through the transit centers. According to the percentage of trips, it could be shown that the transit centers were prominent in the regional transit network of Seoul. The average travel time and the travel distance of trips through the transit centers were 49.4 minutes and 26.9 km. The average travel speed of trips through the transit centers was faster than the speed of total trips between the suburban area and Seoul. The basic statistics of the result of the data treatment is shown in Table 1.

**3.2. Effect of the Bandwidth of KDF on CAI.** Stations often were located near the boundaries of the zones, and such stations only counted as one zone even though people from neighboring zones used them. In addition to this, smart card data might not reflect the actual origin of the trip because people often cross the neighboring zone of their origin to access transit modes. This effect was adjusted by applying the neighborhood effect with KDF when aggregating discrete smart card data to spatial zones.

The median distance between centroids of neighbor zones in the study area is 2.1 km, and the average zone size is 16 km<sup>2</sup>. We assumed that people would not cross more than one zone to access transit. Sensitivity analysis on the bandwidth was conducted based on value of 2 km. The bandwidth of 0 km—without KDF, 2 km, and 5 km was used. The measured result of CAI of Gangnam in part of the region is shown in Figure 3. Without applying KDF, some adjacent zones had significant differences in CAI. When the bandwidth of the kernel function was 5 km, the index flattened in a large area, and the power of adjacent zones increased. This resulted in smaller differences in the index between adjacent zones.

According to the sensitivity analysis, the proper bandwidth of KDF depends on the purpose of the evaluation. Short bandwidth is suitable for observing the differences in a small view, such as the zone level. Conversely, long bandwidth is suitable for observing differences in a broad view, such as the regional or in-county levels. In this study, we used 2 km for the bandwidth to observe a difference between adjacent zones because we wanted to compare the difference in CAI in zone level.

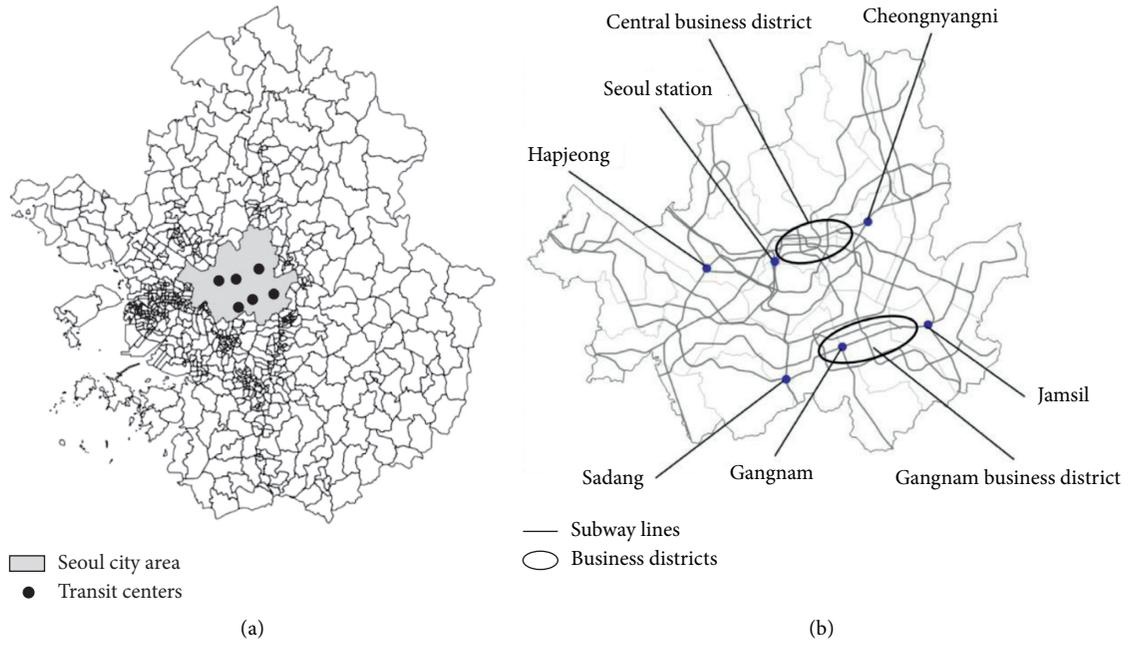


FIGURE 2: Study area: (a) outline of the Seoul metropolitan area; (b) transit centers used in the analysis.

TABLE 1: Basic statistics of the obtained trip chain data between suburban zones and Seoul.

Division	Total # of trips	# of regional bus lines	Travel time (min)	Distance (km)	Travel speed (km/h)
Total (697 zones)	15,775,018		51.4	23.0	26.8
Trips through the transit centers	2,184,397		49.4	26.9	32.7
Gangnam	741,220	59	50.9	30.3	35.7
Jamsil	389,536	35	45.4	23.3	30.8
Sadang	442,237	15	39.7	23.0	34.8
Seoul Station	367,146	34	58.7	29.6	30.3
Cheongnyangni	114,217	17	63.6	26.6	25.1
Hapjeong	130,041	18	47.8	24.4	30.6

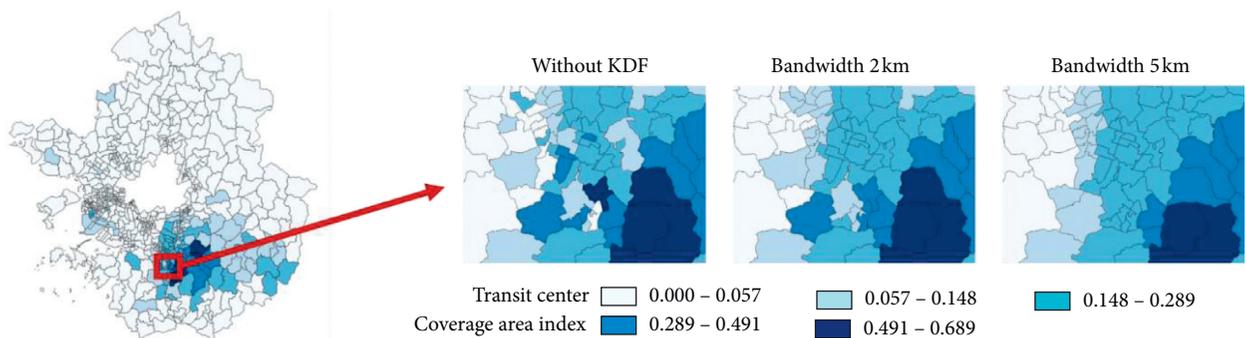


FIGURE 3: Effect of kernel bandwidth to CAI of Gangnam in the southern part of the Seoul suburban area.

3.3. Coverage Area of Transit Center Using CAI. The overall coverage area of each transit center is shown in Figure 4. The coverage area was defined as the zone in which the biggest CAI value was bigger than 0.05 and had more than 100 trips to Seoul. Sadang covered the area in the southwest part of the Seoul suburban area. The CAI for Sadang had the highest value of 0.922 at Jangan. This indicates that 92.2% of the trips from the Jangan zone to Seoul passed through the Sadang transit center. The Gangnam transit center is close to the

most important highway in Seoul, and it is the transit center that is closest to the Gangnam district. These characteristics of the location led to the broad area that it covers. Cheongnyangni and Jamsil are located in the eastern part of Seoul, so their coverage areas are shown in the eastern part of suburban Seoul. Seoul Station and Hapjeong covered the west part of Seoul suburban area, but CAIs of each zone to them were smaller than the value of other centers. This tendency means that Seoul Station and Hapjeong were not

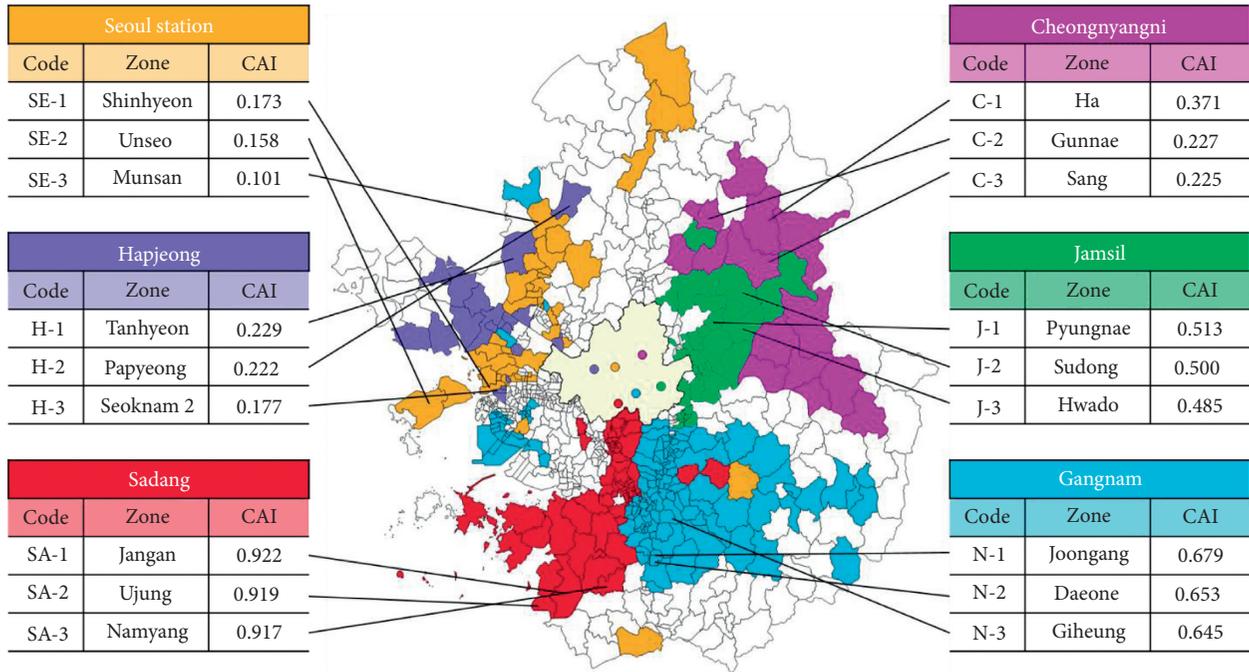


FIGURE 4: Areas influenced by each center and the three regions in each center that have the highest indices.

functional as other centers and less important than other centers.

Seoul Station and Gangnam show the most significant difference in the public transportation network. As shown in Figure 5, they had a similar distribution of TA, and they both located in major business districts of Seoul, but their CAI was very different from the centers. Gangnam had a highly concentrated CAI in the southern part of Seoul. It means that the transit network in the southern part of Seoul had a hub-and-spoke structure centering on Gangnam. In contrast, the coverage of Seoul Station located in both southern and northwestern parts of Seoul and CAI values of the area was low. It means that Seoul Station was connected to the board part of the Seoul metropolitan area, but the power of the center was not strong enough to constitute the hub-and-spoke structure transit network. Through this analysis, we could discuss that transit lines connecting Seoul Station and the southern part of Seoul should be adjusted to connecting the center and northeastern part of Seoul to achieve the effective hub-and-spoke structure of transit network centering Seoul Station. In terms of transit in downtown, it is necessary to examine whether travelers arriving at Seoul Station from a suburban area can travel to other parts of the city adequately.

Consequently, CAI can identify transit centers with not enough power as a hub in a hub-and-spoke structure; it means the functionality of the transit center in the overall network would not be sufficient. With Figure 4, three centers with a lower CAI over their own coverage—Seoul Station, Hapjeong, Cheongnyangni—would correspond to the case. The centers should be upgraded to be attractive to travelers, such as speeding up regional buses with bus rapid transit or

making a convenient environment to transfer with a high-quality in-door bus shelter and information system.

From a broad perspective rather than an individual center, the coverage zones' CAI of the transit centers in the central business district (Seoul Station, Hapjeong, and Cheongnyangni) were smaller than the index of the centers in Gangnam district (Sadang, Jamsil, and Gangnam). We can infer that the transit network near the central business district was not forming a hub-and-spoke structure. Therefore, to make the transit network of the district efficient, it can be considered that the government should reform the transit network or transit centers of the district to forming a hub-and-spoke structure.

**3.4. Evaluation of Index Performance.** We compared CAI and TA for three centers—Sadang, Gangnam, and Seoul Station—of total six centers. Sadang was a hub that performs well, and it had a well-defined coverage. Gangnam and Seoul Station were located in two different subcenters of Seoul. They had a similar distribution of TA, but their CAI shows the difference between the two hubs. The trend shown in Figure 5 indicated a significant difference between CAI and TA. There were some core zones in the suburban area with high CAI values, but CAI decreased rapidly when the distance from the core increased. TA was shown to have its highest value near the central city, and it decreased slowly as the distance from the central city increased. The reason for this difference in trends was that the friction of the TA index was affected strongly by the Euclidian distance from the transit center. In contrast, CAI was not strongly affected by the distance between the zone and the central city. Due to

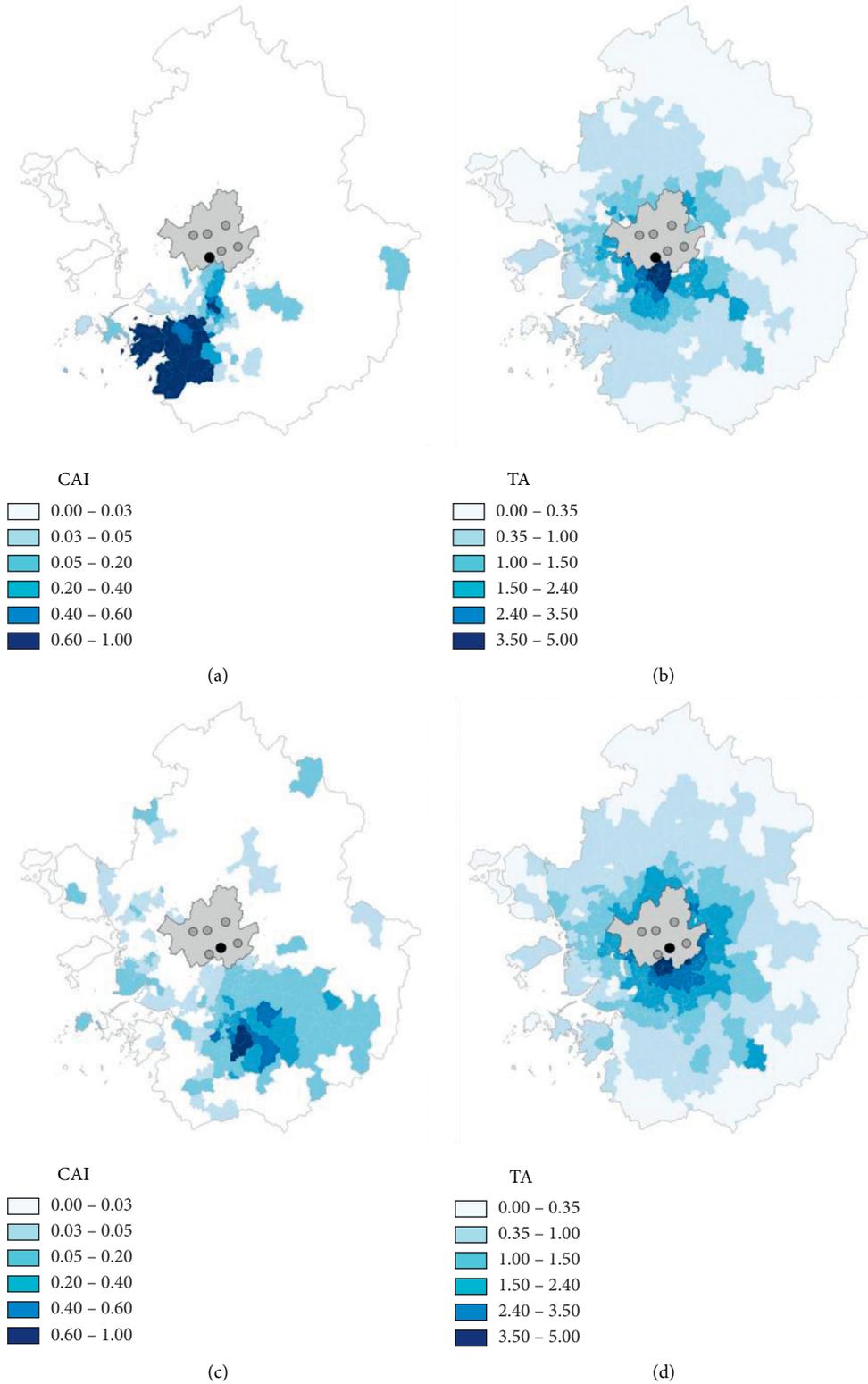


FIGURE 5: Continued.

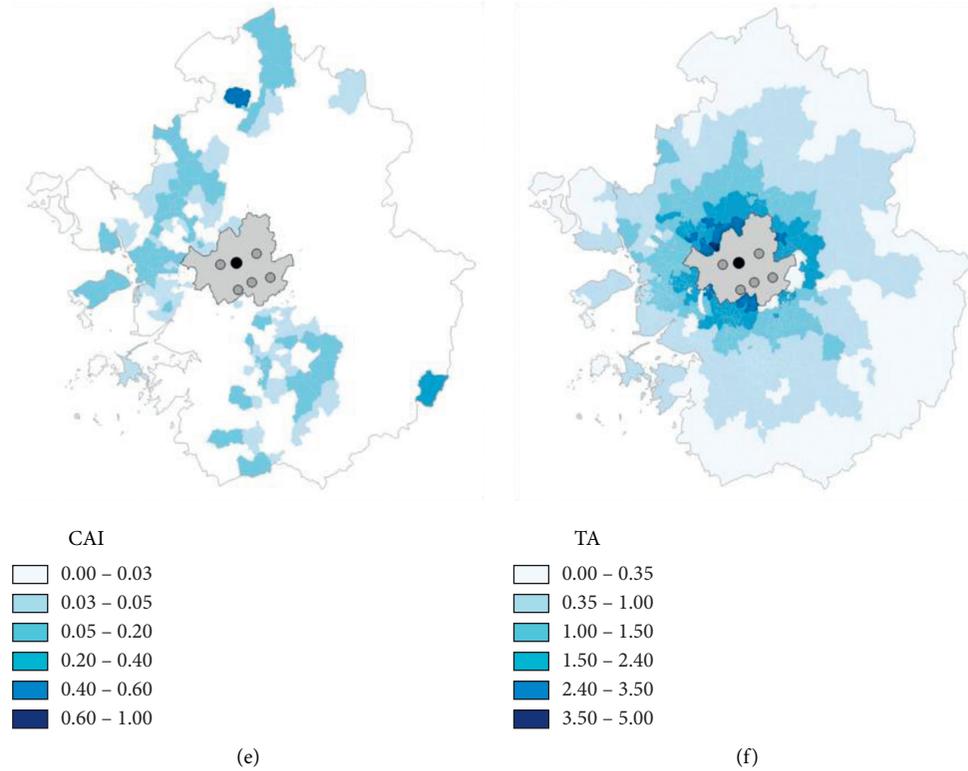


FIGURE 5: Comparison between CAI and TA of three transit centers: (a) CAI of Sadang; (b) TA of Sadang; (c) CAI of Gangnam; (d) TA of Gangnam; (e) CAI of Seoul Station; (f) TA of Seoul Station.

CAI's definition, the ratio of the number of trips was not affected directly by the distance.

This result showed the contribution of CAI. Almost all zones near the central city are connected to city transit. Travelers from those zones do not have to transfer at the transit centers. Zones connected to a transit center with a regional express bus line would have high CAI to the center and would be affected significantly by the center's renovation even though they are far from the city. If the bus line were terminated, the level of service would decrease dramatically. This kind of spatial dependency analysis can offer new motives and insights into planners and operators. The CAI and other accessibility indices, such as TA, have complementary relationships.

## 4. Discussion

**4.1. Measure of Transit Supply and User Preference.** A zone with high CAI of a transit center means a high percentage of travelers between the zone and the central city was integrated at the center. Except for the case where the total number of trips was small, a route passing through the center was the route most preferred by travelers when they chose a route between the zone and the city. This means that the zone is connected to the center with a high level of regional services, such as an express bus line in the case of Seoul. If a zone has a small CAI of a transit center, travelers from the zone to the city will transfer to city transit at another center

or out of the city. The latter case implements a longer trip by slower transit modes, such as the metro or a city bus.

Observing the result of CAI with the existing regional transit network, especially the main corridor of express bus lines, it was apparent why the index is what it is. Figure 6 shows that the zones with high CAI were connected closely to existing express bus routes. Figure 6(a) shows that zones in the southwestern part of the study area had a high CAI to Sadang, i.e., 0.92. They were connected to Seoul by express bus lines to Sadang. Many express lines used highways, so they were regarded as the coverage of Sadang. Figure 6(b) shows that the Gangnam transit center had a broad coverage in the southern part of the study area because Gangnam was the nearest transit center to the Gyungbu Expressway, a central corridor in the Seoul metropolitan area.

Accordingly, CAI can identify transit centers with too broad coverage that can cause overcrowding of travelers and vehicles in the center. For example, Gangnam of Seoul had extensive coverage in the southern part of Seoul, as seen in Figure 6(b). Contrarily, Jamsil did not adequately cover the southeastern part of the area in Figure 6(c). In another view, the coverage area of Sadang was too dependent on one transit center, as seen in Figure 6(a). If there are severe incidents on the road to Sadang, the areas would be isolated. The government should have other options to connect the area to Seoul with other routes or modes. This observation can suggest a solution for overcrowding in Gangnam, such as changing the destinations of regional buses from the

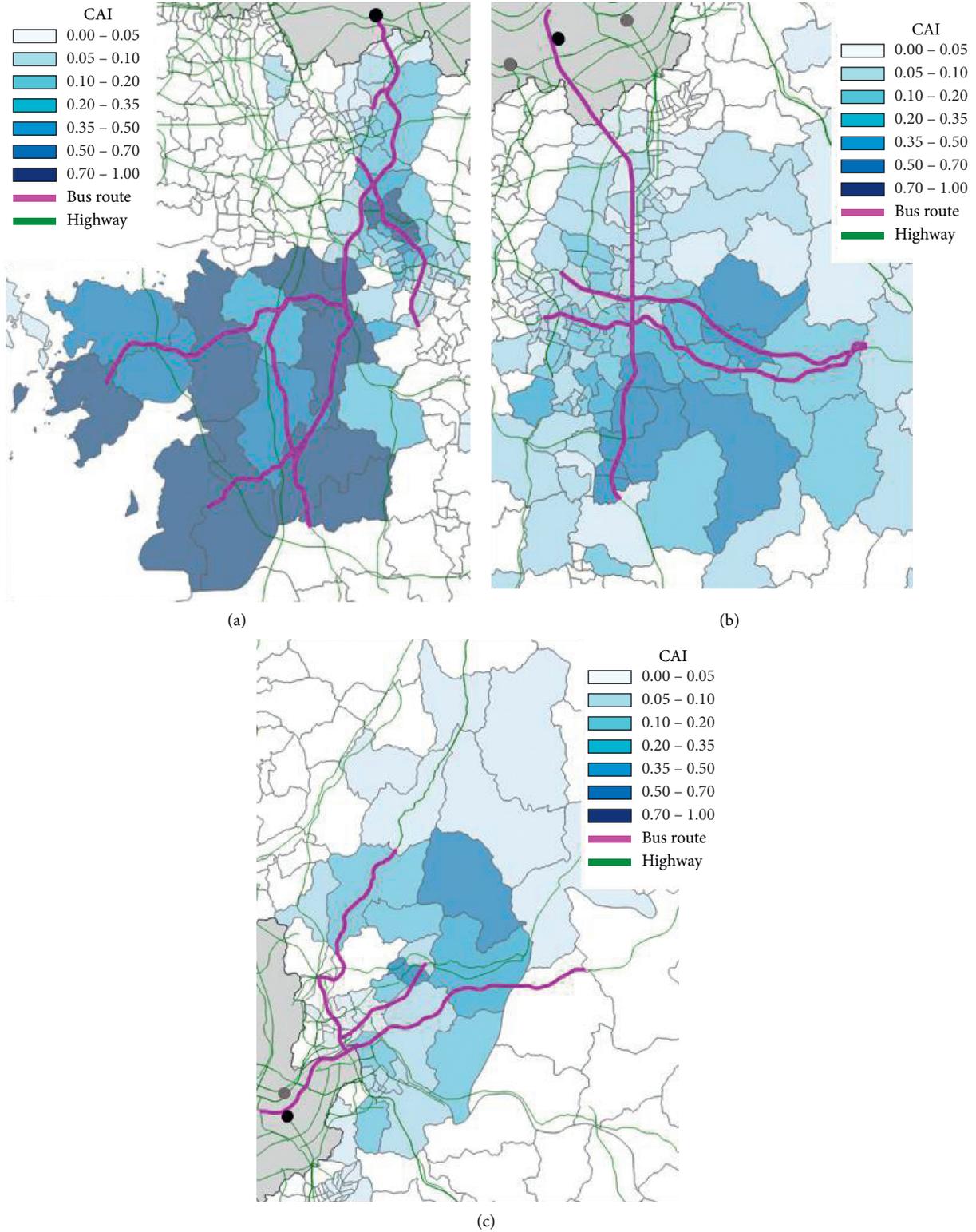


FIGURE 6: A result of CAI with regional transportation (bus routes and highway): (a) Sadang; (b) Gangnam; (c) Jamsil.

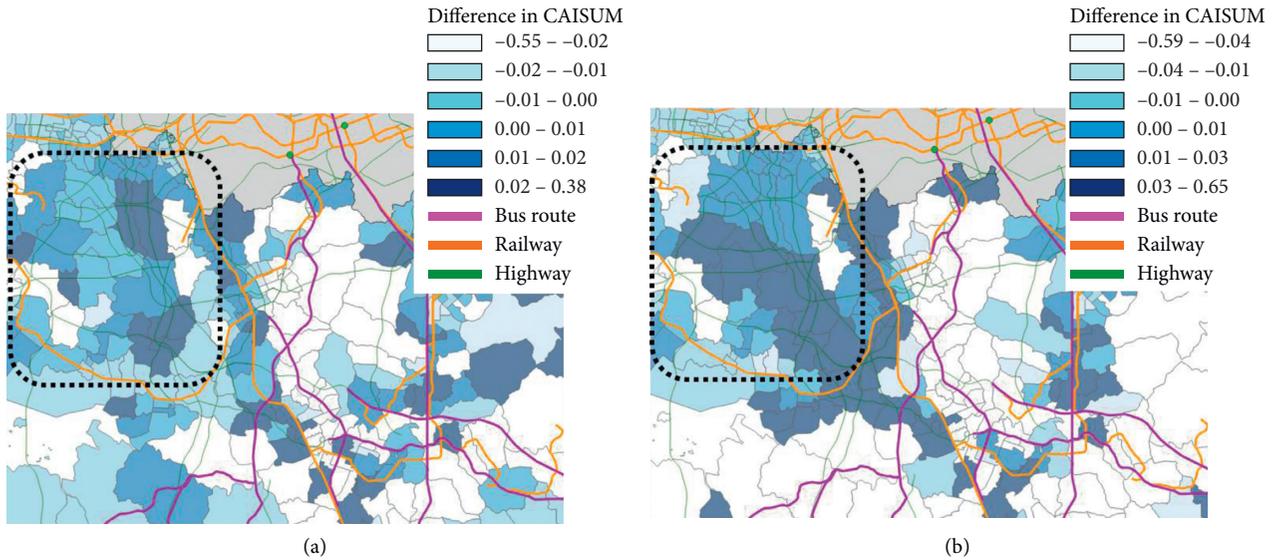


FIGURE 7: A difference in the CAISUMs between neighboring zones with regional transits (express bus routes and railway): (a) compared with zones in 5 km; (b) compared with zones in 10 km.

zones to Jamsil. With a proper supply of express bus lines, trips from the zones to the central city can be dispersed.

**4.2. Connectivity between the Centers and Suburban Area.** The CAIs also can be used to evaluate the regional transit in suburban zones by obtaining the sum of the CAIs for all six centers (CAISUMs). If the CAISUMs of a zone are small, it indicates that the zone is connected to the downtown through another transit center or that there is no advanced regional transit connection to the transit centers, meaning that the transit of the zone does not form hub-and-spoke structure to downtown. To distinguish between the two cases, the CAISUMs of a zone should be compared with the CAISUMs of the neighboring zones. When a zone and its neighboring zones together have small CAISUMs, it can be inferred that all of them are connected to other centers. So, comparing the CAISUM of a zone with those of its neighboring zones can be used to identify where the structure of the transit network is inefficient.

Figure 7 shows differences in the CAISUM of a zone and the CAISUMs of the neighboring zones. We used various bandwidths of KDFs to compare the CAISUMs. Differences in the CAISUM with bandwidths of 2, 5, and 10 km can show the difference in the CAISUM compared to adjusting the zones with a number. If a zone has a positive value of the difference, it had a smaller CAISUM than neighboring zones. When the railroad passes zones, the CAISUMs of the zones were calculated to be a small value. The zones were observed as having a lack of regional transit because trips from the zones to the central city by railway did not require transferring at the transit center, and the transfer between metro lines is not counted in the smart card data. The contribution of this analysis can be derived from the dotted boxes in Figure 7. These zones had lower CAISUMs than the

neighboring zones because they did not have any railways or bus routes connecting to the transit centers. Trips between the zones and the city were on city transit, which had a low level of service and not forming hub-and-spoke structure through transit centers. The contrast between Figures 7(a) and 7(b) was the size of the view. Figure 7(a) was suitable for finding the zone with an inefficient transit network in the individual zone level, and Figure 7(b) was suitable for finding the zones at the county level.

With this approach, TAZs with an insufficient service of regional transit would be identified. Travelers from the zones would experience low speed and a poor environment for transfer when they travel to downtown. The government should supply a regional express bus from the zones to existing transit centers. If the area's transit demand is too big to handle in the centers, opening a new transit center to connect subway and city buses with regional transit should be considered.

## 5. Conclusion

In this research, the concept of coverage was expanded to evaluate the integration and coordination of multimodal public transit system. The coverage area index (CAI) was defined as the ratio of the number of trips through the transit center to the total number of trips from the zones to the central city. Smart card data were used to conduct the case study, and the kernel density function was applied to aggregate the point-to-point origin-destination data into zones. Six transit centers that connect Seoul and the southern part of Seoul metropolitan area were selected to measure the CAIs. The peak value of CAI varied between the centers, from 0.173 for Seoul Station up to 0.92 for Sadang. CAI can be used to compare the importance of transit centers, to measure transit supply and travelers' preference and evaluate whether transit from a suburban zone to the

central city was integrated into intercity transit network of the city.

Transit centers have the role of integrating transit modes between downtown and the suburban area in the complicated hub-and-spoke transit network in big cities. Some researchers have considered the effectiveness, connectivity, and accessibility of the transit centers, but they have rarely dealt with the integration or coordination of the overall transit network. We developed CAI using the concept of betweenness centrality to evaluate the role of the centers. With the CAI, a dependence of suburban areas on the centers could be visualized. By comparing CAI and TA, we were able to evaluate CAI having the independence of the distance from the city center. Since only OD trip data were needed to measure CAI, it was not necessary to directly compare the various alternative routes and modes. The characteristics of CAI had advantages when transit networks are too complicated to analyze in the metropolitan area.

CAI could be used for three purposes: finding coverage of transit centers, measuring transit supply, and measuring connections between the centers and suburban areas. With the coverage of each center, we could find incompetent centers that had narrow and vague coverage. It meant that they could not function well as a hub, so the centers should be more attractive to travelers. With the measured supply of regional transit, alternative routes to solve overcrowding problems of centers were implicated. With the measured connection, areas with insufficient supplies of regional transit were identified, and strategies to improve the connection in the areas were suggested.

Given our conclusions and observations, future research should take into account the rarity of Seoul's smart card dataset, which provided the trip chain data we used that included the exact boarding and alighting stations for trips. The definitions suggested for coverage and CAI can be applied to other types of trip data, such as the OD matrix and transit ridership data. However, these characteristics could be limitations of the index because the index could not be used to test scenarios in the planning level due to the absence of future trip data. It can be further work to use the index with predicted future trip data. By expanding the index with the data, potential coverage areas that travelers can be users of a transit center can be identified. This result would be helpful for planners of a transit center and a transit network. In future research, analysis of coverage of transit centers could be assigned to different metropolitan areas to identify zones that have inadequate regional transit. It can be further work to use the index with the data of predicted future trips.

## Data Availability

The smart card data used to support the findings of this study are restricted by the Ministry of Land, Infrastructure, and Transport of Korea in order to protect the privacy. The data are available from TRLab in Seoul National University for researchers who meet the criteria for access to confidential data.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2020R1F1A1061802).

## References

- [1] L. Sun, J. Rong, and L. Yao, "Measuring transfer efficiency of urban public transportation terminals by data envelopment analysis," *Journal of Urban Planning and Development*, vol. 136, no. 4, pp. 314–319, 2010.
- [2] H. Nishiuchi, T. Todoroki, and Y. Kishi, "A fundamental study on evaluation of public transport transfer nodes by data envelop analysis approach using smart card data," *Transportation Research Procedia*, vol. 6, pp. 391–401, 2015.
- [3] S. Mishra, T. F. Welch, and M. K. Jha, "Performance indicators for public transit connectivity in multi-modal transportation networks," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 7, pp. 1066–1085, 2012.
- [4] A. Ceder, Y. Net, and C. Coriat, "Measuring public transport connectivity performance applied in Auckland, New Zealand," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2111, no. 2111, pp. 139–147, 2009.
- [5] Z. Wang and Y. Chen, "Development of location method for urban public transit networks based on hub-and-spoke network structure," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2276, no. 1, pp. 17–25, 2012.
- [6] B. Yu, H. Zhu, W. Cai, N. Ma, Q. Kuang, and B. Yao, "Two-phase optimization approach to transit hub location - the case of Dalian," *Journal of Transport Geography*, vol. 33, pp. 62–71, 2013.
- [7] M.-C. Shish, H. Mahamassani, and M. Baaj, "Planning and design model for transit route networks with coordinated operations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1623, no. 1, pp. 16–23, 1998.
- [8] E. H. Lee, H. Lee, S. Y. Kho, and D. K. Kim, "Evaluation of transfer efficiency between bus and subway based on data envelopment analysis using smart card data," *KSCE Journal of Civil Engineering*, vol. 23, no. 9, pp. 1–12, 2019.
- [9] J. G. Koenig, "Indicators of urban accessibility: theory and application," *Transportation*, vol. 9, no. 2, pp. 145–172, 1980.
- [10] S. K. Sathisan and N. Srinivasan, "Evaluation of accessibility of urban transportation networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1617, no. 1, pp. 78–83, 1998.
- [11] Kittelson & Associates, USFT Administration, TCR Program, and TD Corporation, *Transit Capacity and Quality of Service Manual*, Transportation Research Board, Washington, DC, USA, 2003.
- [12] S. A. Mamun and N. E. Lownes, "Measuring service gaps: accessibility-based transit need index," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2217, no. 1, pp. 153–161, 2011.
- [13] I. Tasic, C. Bozic, and E. Hanss, "Towards true multimodal transportation accessibility: data, measures, and methods," in *Proceedings of the 96th Annual Meeting of the Transportation Research Board (TRB)*, Washington, DC, USA, January 2017.

- [14] C.-L. Cheng and A. W. Agrawal, "TTSAT: a new approach to mapping transit accessibility," *Journal of Public Transportation*, vol. 13, no. 1, pp. 55–72, 2010.
- [15] B. M. Alam, G. L. Thompson, and J. R. Brown, "Estimating transit accessibility with an alternative method: evidence from Broward County, Florida," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2144, no. 1, pp. 62–71, 2010.
- [16] K. Sohn and D. Kim, "Zonal centrality measures and the neighborhood effect," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 9, pp. 733–743, 2010.
- [17] S. B. Chung, S. H. Hur, J. H. Baek, D. Kim, and K. H. Song, "Development of accessibility index using closeness centrality," in *Proceedings of the Eastern Asia Society for Transportation Studies the 9th International Conference of Eastern Asia Society for Transportation Studies*, Eastern Asia Society for Transportation Studies, Jeju, Korea, 2011.
- [18] X. Ma and Y. Wang, "Development of a data-driven platform for transit performance measures using smart card and GPS data," *Journal of Transportation Engineering*, vol. 140, no. 12, 2014.
- [19] E.-J. Kim, Y. Kim, and D.-K. Kim, "Interpretable machine-learning models for estimating trip purpose in smart card data," in *Proceedings of the Institution of Civil Engineers—Municipal Engineer*, Thomas Telford Ltd, pp. 1–22, 2020.
- [20] W. Tu, R. Cao, Y. Yue, B. Zhou, Q. Li, and Q. Li, "Spatial variations in urban public ridership derived from GPS trajectories and smart card data," *Journal of Transport Geography*, vol. 69, pp. 45–57, 2018.
- [21] H. Lee, H. Park, S. Kho, and D. Kim, "Assessing transit competitiveness in Seoul considering actual transit times based on smart card data," *Journal of Transport Geography*, vol. 80, Article ID 102546, 2019.
- [22] C. Kapuku, S.-Y. Kho, D.-K. Kim, and S.-H. Cho, "Modeling the competitiveness of a bike-sharing system using bicycle GPS and transit smartcard data," *Transportation Letters*, pp. 1–5, 2020.
- [23] Z. Cai, D. Wang, and X. M. Chen, "A novel trip coverage index for transit accessibility assessment using mobile phone data," *Journal of Advanced Transportation*, vol. 2017, Article ID 9754508, 14 pages, 2017.
- [24] M. T. Magalhães, "Spatial coverage index for assessing national and regional transportation infrastructures," *Journal of Transport Geography*, vol. 56, pp. 53–61, 2016.
- [25] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
- [26] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [27] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, 2001.
- [28] G. L. Thompson, "Achieving suburban transit potential: sacramento revisited," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1571, no. 1, pp. 151–160, 1997.
- [29] Korea Transport DataBase, *Establishing Network for Transport Analysis 2018*, The Korea Transport Institute and Ministry of Land, Infrastructure and Transport of Korea, Sejong City, South Korea, 2018.