IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 January 2024, accepted 8 January 2024, date of publication 15 January 2024,
date of current version 19 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3353780

RESEARCH ARTICLE

Differential Dynamics of Transit Use Resilience During the COVID-19 Pandemic Using Multivariate Two-Dimensional Functional Data Analysis

WON GYUN CHOI[1], SEUNGHEE RYU[1], PAUL H. JUNG[2,3], AND SEUNGMO KANG[1], (Member, IEEE)

[1]Department of Architectural, Civil, and Environmental Engineering, Korea University, Seoul 02841, South Korea
[2]Asia Pacific School of Logistics, Inha University, Incheon 22212, South Korea
[3]School of Public Policy, University of California at Riverside, Riverside, CA 92521, USA

Corresponding authors: Seungmo Kang (s_kang@korea.ac.kr) and Paul H. Jung (Paul.Jung@inha.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education and the Ministry of Science and ICT under Grant 2020R1A6A1A03045059 and Grant 2021R1A2C2004125, and in part by the Inha University New Faculty Research Grant 71350-1.

ABSTRACT This study delves into the nuanced patterns of shock and recovery in transit ridership during and after the COVID-19 pandemic, aiming to illuminate the resilience exhibited by various geographic areas. This resilience is measured by the ability of transportation systems to withstand, adapt to, and bounce back from unforeseen shocks. In this research, smart card big data were exploited to track real-time mobility dynamics and economic activity within the city of Seoul, Korea. The approach employed multivariate two-dimensional functional data analysis and a hierarchical clustering method to examine both boarding and alighting patterns, taking into account multi-scalar temporal units, monthly and hourly demand fluctuations. The findings present distinct varied shock-and-recovery patterns across areas in transit ridership based on the socioeconomic characteristics of specific areas. These characteristics encompass factors such as industry and land-use composition, income levels, population density, and proximity to points of interest. Additionally, this methodology proves effective in identifying abnormal surges in demand linked to local large-scale development projects.

INDEX TERMS COVID-19, functional data analysis, MFPCA, public transit ridership, resilience.

I. INTRODUCTION

The unforeseen aftermath of the COVID-19 pandemic imposed a formidable stress test on different socioeconomic systems across countries. The transportation sector was one of the sectors that experienced harsh damage from the COVID-19 pandemic. As it unfolded, the pandemic revealed fluctuating patterns in transportation systems, and underscored the diverse challenges in mobility. Amidst the unprecedented situation, transportation researchers and policymakers embarked on inquiries to delineate and comprehend the patterns of shock and recovery. They sought to frame these patterns within the broader concept of ''resilience'', aiming to understand the capacity of transportation systems to endure, adapt and recover from unforeseen shocks.

The majority of the relevant research focuses on the impact of the spread of the pandemic itself and consequent reservation towards crowded locales and various policy interventions on transportation demand and ridership patterns. Major interest was placed on the drop and recovery of passenger flows as the pandemic trembled transportation systems, and the manner in which these effects manifested themselves differently depending on local characteristics. One strand of research has focused on the influence of lockdowns and

The associate editor coordinating the review of this manuscript and approving it for publication was Jesus Felez.

interpersonal distancing measures as policy tools, and the bounceback of passenger movements immediately following the relaxation of such impositions [1], [2]. Persistent shifts in economic activity and transportation usage necessitates the examination of patterns over a longer period of time in order to achieve a comprehensive analysis of the new equilibria. Another strand of literature primarily explores the overall impact of the pandemic on transportation demand, specifically its drop and recovery at the system level [3], [4], [5]. The heterogeneity of such impacts by location requires further study to accurately measure and analyze the socioeconomic and policy implications of the pandemic.

In this study, the concept of ''differential resilience'' [6] is adopted to enhance the discussion and applied on different timescales. Using transit ridership as a particularly compelling measure of neighborhood economic activity, the varied responses and adaptive and restorative capacities exhibited by urban neighborhoods in the face of large-scale external shocks are explored. Specifically, focus is placed on the following questions: 1) Do neighborhoods respond to the unexpected shock in varied patterns?; 2) Which socioeconomic and spatial characteristics contribute to the heterogeneity of resilience in the metropolis? By analyzing the variations in transit ridership, this research aims to uncover the underlying factors and characteristics that contributed to the divergent responses exhibited by each neighborhood and shed light on the multifaceted nature of differential resilience at a broader temporal and spatial scale.

This research posits a novel approach to effectively answer these intricate complexities as thus. By employing a considerably wider spatiotemporal range of data from January 2019 to May 2023, both the macroscopic shifts before and after the pandemic can be effectively captured, and a lens through which differential resilience can be measured and analyzed is formulated. Smart card big data were utilized to trace the real-time dynamics of human mobility and economic activity during and after the COVID-19 pandemic. Moreover, the analytical powers of functional data analysis (FDA) and the hierarchical clustering method allow for an assessment of the varied spatiotemporal changes in Seoul's metro systems during the analysis period. This innovative multivariate two-dimensional approach is able to capture the changes in hourly and monthly transit demand for both boarding and alighting simultaneously.

This research presents notable differential resilience patterns among stations in Seoul. Among others, industry and land-use makeup, income level, population density, and proximity to points of interest (POIs) are identified as relevant factors for explaining this differential resilience. Using this knowledge, the resilience of neighborhoods vis-Ã -vis their respective characteristics and functions within the urban sphere is discussed, and valuable insights that can contribute to the ex-ante establishment of policies to reinforce resilience and equity in view of future similar shocks are offered. The potential of FDA in spotting outliers which display unique and distinguishable patterns are also confirmed in the context of passenger flows.

The rest of this paper is structured as follows: first, a review of past literature on the COVID-19 pandemic, the resilience of transportation systems, and the applications of FDA in transportation studies is provided. Then this research's main methodological approach is introduced and presented as a model to examine differential dynamics of transit use resilience across neighborhoods in Seoul. The next section discusses the results of the model. The final section concludes with a discussion on the limitations of this research and opportunities for future research.

II. LITERATURE REVIEW

A. THE COVID-19 PANDEMIC AND DIFFERENTIAL RESILIENCE IN TRANSPORTATION SYSTEMS

The COVID-19 pandemic functioned as a devastating shock in our society, and brought about unexpected changes in the transportation sector. In addition to lockdown restrictions suppressing the overall level of travel demand, the impact of the COVID-19 pandemic on traffic demand displayed significant variability contingent upon transportation modes, with public transportation notably experiencing the most substantial decline [7]. This marked contrast in demand reduction across modes triggered a noteworthy shift in modal share dynamics away from public transportation to private cars and more active modes [8], [9], [10]. The pandemic instilled a sense of reservation among passengers regarding the use of public transit due to concerns over potential transmission vectors, particularly dense crowds and shared surfaces, leading individuals to opt for private vehicles as a means of reducing contact with others [11]. For example, more than half of the passengers studied by Nikolaidou et al. [12] reduced their reliance on public transportation or abstained from using it altogether during the pandemic due to concerns with being in crowded, potentially unsanitary vehicles for extended periods of time. Notably, in the case of Budapest, the share of other modes rose, while only the modal share of public transportation plummeted from 43% to 18%, a shift which began before proper lockdown measures were implemented [13]. The preference for ride-hailing and ride-sharing services also tended to decrease, primarily owing to the heightened risk of transmission in enclosed vehicles, even with reduced passenger interactions [14]. The decrease in passenger movement was also varied depending on the distance and purpose of travel, with passengers travelling significantly shorter distances and making less trips for shopping purposes [15].

The transition from public transit to private cars during the pandemic exhibited varying patterns based on socioeconomic characteristics. Lower-income populations, in particular, continued to rely on public transportation and paratransit services, given the financial challenges associated with owning and operating personal vehicles, thereby facing a

higher risk of infectious disease transmission compared to their higher-income counterparts [1], [11], [16]. During crises involving nationwide travel restrictions, individuals typically curtailed less-essential trips first [17]. In China, as confirmed by Zhou and Lee [18], recreational activities and social visits experienced the most significant declines during the peak of the epidemic compared to the pre-COVID-19 period. Conversely, shopping trips, characterized by shorter trip distances and durations and foci on acquiring food and essential items, did not witness a decline and even exhibited slight increases, becoming the primary purpose of trips during the pandemic [11], [15]. However, the specific patterns of change varied from one country to another, contingent upon the implementation of distancing measures and travel restrictions as well as cultural attitudes and responses [19].

As the spread of the COVID-19 slowed down and lockdown measures were subsequently lifted, economic activity began to return to a new ''normal'', and travel volumes and public transportation ridership levels were observed to follow the recovery in hand in many cases. Notably, this provided an incredibly unique setting where one can observe how a bounce-back of the transportation system proceeds over time. Even though there is just a limited number of studies on this yet, they have presented a surprisingly common finding that the pace and magnitude of the recovery are revealed to vary across regions serving different functions within the city [16]. Fernández Pozo et al. [20] manually divided the eight-month peak pandemic period in Madrid, Spain into seven distinct phases, from the pre-pandemic to the end of lockdown measures, and analyzed public transport ticket validation data by mode, station, ticket type, and income to find that there exists a variation in daily total ridership recovery patterns depending on those factors. Sharma et al. [3] created a counterfactual scenario to estimate demand without the influence of the pandemic and performed a mixture of spatial lag and geographically weighted regression to confirm the existence of spatial clusters and identify elements in the built environment, socioeconomic conditions, and public risk perception levels whose spatial distributions can account for the spatial heterogeneity of changes in ridership over time. Ammoury et al. [21] analyzed ridership patterns from both transit supply and demand perspectives and explored the influence of socioeconomic factors such as income, age, and education level, as well as locational factors including adjacency to a local university, hospitals, and supermarkets. While these studies investigated the extent to which the effect of relaxation of lockdown policies and interpersonal distancing regulations manifested themselves in different socioeconomic groups, the coverage was limited to only the initial stages of the post-pandemic period, usually no later than 2022. However, this paper examines the entire trends and patterns of the long-term recovery of economic activity and passenger flows as global systems fully settle into the ''new normality'' throughout all stages of the COVID-19 pandemic.

In the literature of differential resilience, researchers acknowledge and measure the spatial diversity in resilience, which refers to a system's inherent capacity to maintain, revert to, or adapt its functionalities in the aftermath of an external shock [6]. The series of disruptions and recoveries observed in the transportation sector during the COVID-19 provided an opportunity to capture the temporal dynamics of the regional resilience of transportation systems. The focus of existing research on regional resilience includes economic activity in view of external shocks. Giannakis and Bruggeman [22] and Sensier et al. [23] analyzed regional disparities in the impact of the Great Recession on economic activity and emphasized the comparison of local resilience levels relative to the overall average. Moreover, they explored the existence of spatial autocorrelation and the influence of the interplay of a range of factors such as industry makeup, demography, and migration patterns using multilevel regression. Research on the regional resilience of transportation movements merit particular attention. Dobruszkes and Van Hamme [24] quantified through multiple regression the relationship between changes in economic output and air travel movements in view of the Great Recession at the country level to find globally shared patterns, and offered nuanced explanations for outlying cases whose trajectories diverged from it. Potter et al. [25] identified differences in the resilience of rail freight flows in response to exogenous disturbances in the infrastructure, market, and macroeconomy that occurred during a multi-year period. Ziedan et al. [5] aggregated monthly transit ridership trajectories in the United States at the metropolitan level and compared trends among them using changepoint analysis to find that metropolitan areas where the return of ridership levels was most prominent were those that offered fare-free transit for extended periods. This paper exemplifies a comprehensive analysis, assessing the differential resilience to COVID-19 across sub-sections of a large city by comparing travel demand changes before, during, and after the event.

B. APPLICATIONS OF FDA IN TRANSPORTATION STUDIES

Even though the newly emerged phenomena of varied recovery patterns in transportation systems during the COVID-19 pandemic are consistent with the concept of differential resilience, researchers still lack a methodological framework that can quantify and measure the temporal patterns of recovery over time. In the specific context of the COVID-19 pandemic, numerous research endeavors have been undertaken to assess the changes in public transit ridership. As can be seen in Table 1, data from various sources have been employed, including surveys [8], mobile location data [2], [26], counting stations [13], and online service usage information [4]. Another key method through which data-based travel pattern analysis is performed is automatic fare collection (AFC) records. For example, Luo [27] aggregated AFC transactions to the journey level and extracted key behavioral characteristics from their trajectories. Based on this data, k-means clustering

was applied to identify passenger groups which exhibited similar shifts in behavioral patterns since the pandemic began. Jiao et al. [28] analyzed the regional discrepancies in the impact of the pandemic on public transit (PT) ridership in the Greater Austin area by using multivariate k-means clustering and geographically weighted regression (GWR) to compare percentage changes in ridership with demographic characteristics, distance from the city center and PT stop density. Liu et al. [29] used transit navigation app data and built a logistic model to find that communities with higher proportions of essential workers, vulnerable populations, and more coronavirus internet searches maintained higher levels of minimal demand during COVID-19.

As the process of resilience proceeds on time, FDA can serve as an effective methodological framework through which the varied patterns of resilience across places can be examined. Seya et al. [30] employed an approach that harnessed the capabilities of FDA by transforming multi-year land price data into functional data. This allowed for the identification of the areas benefiting from railway line projects, leveraging the unique attributes of FDA. Jung and Song [31] applied the FDA approach to analyze and cluster the time dynamics of neighborhood changes. Furthermore, an extended multivariate FDA was employed, allowing for the simultaneous consideration of multiple socioeconomic variables. The current body of knowledge in FDA-based approaches to travel behavior and public transit demand analysis is limited. Building upon previous studies that used discrete data in its raw form [32], [33], Park et al. [34] and Wang and Tsung [35] applied FDA methodology to cluster metro stations. Their approach analyzed average annual hourly demand without considering extended shifts measured over longer-term periods. Jung et al. [36] defined the average travel time between metro stations as transit accessibility, and clustered them to elucidate shared characteristics using FDA. Furthermore, Galvani et al. [37] and Roy et al. [38] applied functional data approaches to analyze the spatiotemporal patterns of bicycle ridership.

A summary of relevant literature is presented in Table 1.

This paper makes several significant contributions to the existing literature. Firstly, it investigates the evolution of travel demand throughout all stages of the COVID-19 pandemic, from the initial stage to the ''new normal'' stage after recovery—around May 2023. Secondly, it employs a two-dimensional analysis, encompassing both monthly demand variations and hourly fluctuations for both boarding and alighting. Thirdly, the implementation of the FDA method enables the functionalization of temporal demand changes, adding a nuanced perspective to the analysis. Lastly, the paper utilizes hierarchical clustering methods to effectively group, categorize, and compare different regions within the city and identify the relevant socioeconomic and geographical factors for each cluster. To the best of the authors' knowledge, this paper represents the first comprehensive and intensive analysis covering these aspects in the field.

III. DATA DESCRIPTION

A. FUNCTIONAL DATA: SEOUL METRO PASSENGER VOLUME

To identify changes in transit demand over time, the boarding and alighting passenger volume data by station provided by the smart card operator (''T-money'') were utilized. Specifically, this study focused on stations in the Seoul Metropolitan metro system located within the boundary of Seoul proper. To maintain consistency, stations which are connected to lines which began service during the analysis period were excluded. The data are anonymized to exhibit no personal information, and aggregated at the station level by hour.

One of the concerns raised in the literature regarding the use of smart card data is the issue of under- or mis-representation of transit demand and passenger movement due to: (1) cash payments, (2) the absence of automated gates at some stations or exits, (3) fare evasion, and (4) measurement error [40], [41], [42]. However, those cases are not likely to take place on the Seoul metro system. Smart cards are the only payment method available for the Seoul metro [43]. Even passengers seeking to pay by cash are issued temporary smart cards to use the system. All stations in the system are fully gated at both entrances and exits, so fare evasion is exceptionally rare. According to the Seoul Metropolitan Government [44], the overall incidence number of wrongful entries at metro gates in 2015 stood at only 0.0014% and 0.0021% for metro lines 1–4 and 5–8, respectively. Among those cases, only 33.21% stem from not tapping the smart cards, and the remainder are from misuse of discount cards intended for seniors or students, which does not alter the tagging statistics. Measurement errors resulting in unrecorded transactions are also very uncommon. A study on the smart card data of Seoul from 1 October 2015 to 12 November 2015 [45] reported negligible transaction errors, approaching zero percent for boardings whereas the incidence of missing alighting taggings was measured at only 0.003%. Given the circumstances of widespread smart card use for fare collection in the Seoul metro system, patterns found in the smart card data would well represent the overall transit demand and passenger movement.

In total, the dataset comprised 294 stations whose locations are depicted in the Appendix (Fig. 43). The data encompass attributes such as date (year and month), line, station ID, station name, and boarding and alighting passenger volume. Passenger volume is the average of the cumulative count of passengers throughout the respective month during each hour. The analysis period spans from 1 January 2019 to 31 May 2023, with the examination focusing on the hours between 06:00 and 23:00. To analyze the impact of COVID-19 on passenger travel patterns, volume level is defined as the percentage ratio of the passenger volume for each month relative to the passenger volume recorded during the same month in 2019, which represents the baseline pre-pandemic values. The utilization of this ratio can also mitigate temporal

TABLE 1. Summary of relevant literature on public transit ridership analysis of COVID-19 impact.

Literature	Transit Data Source	Temporal Unit	Explanatory data	Analysis Scale	Geographical Scope	Analysis Method	Long-term Recovery Included
Bucsky [13]	Vehicle & bicycle counts, Mobility app data, PT ridership count, Google mobility report	Daily	-	City level	Single city	Individual comparisons	No
Liu *et al.* [29]	Transit app data	Daily	Population data, Economic data, COVID-19 awareness, PT dependency	District level	Nation-wide	Linear regression	No
Anke *et al.* [8]	Online survey	One-time	Demographic data, Urban structure, PT subscription, Vehicle ownership, Bicycle ownership	Survey group	Nation-wide	Applied chi-square test	No
Luan *et al.* [14]	Online survey	One-time	Demographic data Economic data Vehicle ownership	Survey group	Multiple cities	Random Utility Maximization, Random Regret Minimization, Generalized Regret Minimization	No
Luo [27]	PT ridership count	Daily	Passenger characteristics	Survey group	Single city	Manual profiling, K-means clustering	No
Parker *et al.* [2]	Online survey, GPS tracking data	Weekly	Demographic data, Economic data, Urban structure, Political leanings, Household makeup	Survey group	Nation-wide	Binomial regression	No
Fernández Pozo *et al.* [20]	PT ridership count	Daily	Passenger characteristics	Survey group	Single city	Linear regression	No
Li *et al.* [39]	PT ridership count	Annual	Industry makeup, Household makeup, Vehicle ownership, PT density	Local level	Single city	Hist gradient boosting regressor	Yes
Marra *et al.* [26]	GPS tracking data	Weekly	Demographic data, Economic data, Household makeup	Survey group	Single city	Mixed path size logit	No
Ammoury *et al.* [21]	PT ridership count	Daily	Population data, Economic data	District level	Single city	Pearson correlation coefficient	No
Jiao *et al.* [28]	PT ridership count	Daily	Population data, Economic data, Vehicle ownership, Urban structure, Transportation time	Local level	Single city	GWR, K-means clustering	No
Nikolaidou *et al.* [12]	Mobility app data	Annual	Population data, Economic data, COVID-19 situation	City level	Worldwide	Linear regression	No
Sharma *et al.* [3]	PT ridership count	Weekly	Economic data, Vehicle ownership, COVID-19 situation, Infrastructure density	Local level	Single city	Multivariate spatial regression, Moran's I clustering	No
Wilbur *et al.* [16]	PT ridership count, Paratransit usage data	Monthly	Population data, Economic data	Local level	Multiple cities	Pearson correlation coefficient, Linear regression	Yes
Ziedan *et al.* [5]	PT ridership count	Monthly	-	City level	Nation-wide	Changepoint analysis	Yes

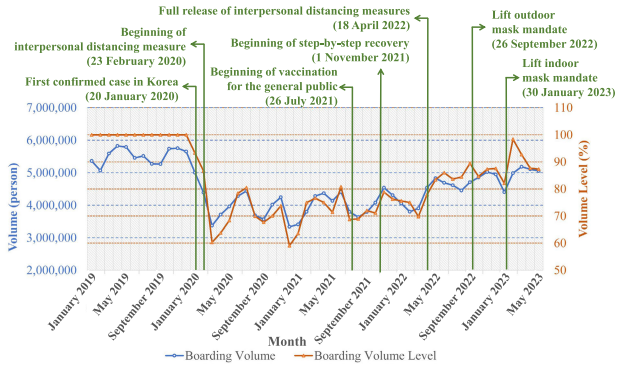

FIGURE 1. Average daily metro passenger boarding volume and volume level by month.

fluctuations in the time series data, enhancing the robustness of the analysis.

As can be seen in Fig. 1, during the pre-pandemic era (before February 2020), the average daily ridership, calculated across all analyzed stations, amounted to approximately 5.5 million passengers. In the early stage of the pandemic (February 2020–October 2021), as confirmed cases of COVID-19 began to rise dramatically and interpersonal distancing policies were implemented, ridership declined significantly to around 3.4 million passengers, denoting a reduction of approximately 40% compared to the pre-pandemic period. Subsequently, throughout the COVID-19 period, the transportation system witnessed recurrent patterns of transient recuperation and subsequent decline in response to the progression of infectious transmission and alterations in distancing directives. With the beginning of step-by-step recovery in November 2021 and the complete lifting of interpersonal distancing policies in April 2022, metro passenger volume gradually rose, recovering to volume levels exceeding 80% during the late stages of the pandemic (November 2021–December 2022). Notably, commencing from February 2023, following the relaxation of mask mandates in public spaces, passenger volumes exhibited stabilization, surmounting the threshold of 5 million passengers. This resurgence signifies an impressive rebound to nearly 90% of pre-pandemic passenger volume in the post-pandemic era (after January 2023). This observed trend is broadly evident across Seoul's metro stations, albeit subject to notable divergences contingent on the neighborhood attributes encompassing the vicinity of each station.

Table 2 shows the average boarding volume and volume level during different times of day. Compared to the pre-pandemic period, early-stage reduction showed a pronounced decline, hovering around 70% volume level. This reduction can be attributed to the rigorous implementation of interpersonal distancing measures and lockdown protocols. Subsequently, a modest recovery ensued during the late stage of the pandemic, restoring the volume level to approximately 80%. After the pandemic, a substantial rebound ranging from 86.79% to 96.74% was observed. This overall trend was evident; however, variations in the degree to which it was observed varied across different times of day. During the morning peak hours (07:00–09:00) and evening peak hours (17:00–19:00), elevated boarding volumes and volume levels were observed compared to other time segments post the early stage of COVID-19. These peak hours exhibited a milder decline and expedited recuperation, achieving volume levels of 96.63% to 96.74% following the pandemic. In the daytime non-peak hours (09:00–17:00), the volume level fell to 70.21%, which is about eight percentage points lower than the peak-time values and corresponds to a reduction of approximately 86,500 passengers per hour. This value recovered to 94.15% in the post-pandemic months. At nighttime (20:00–22:00), the initial reduction mirrored that of the non-peak hours; however, recovery remained sluggish, and the volume level remained below 90% even in the post-pandemic period. The substantial reduction in non-peak hours and slow recovery show that the pandemic's impact on non-commuter trips, including leisure and recreational trips, is much greater than that on commuter-oriented trips.

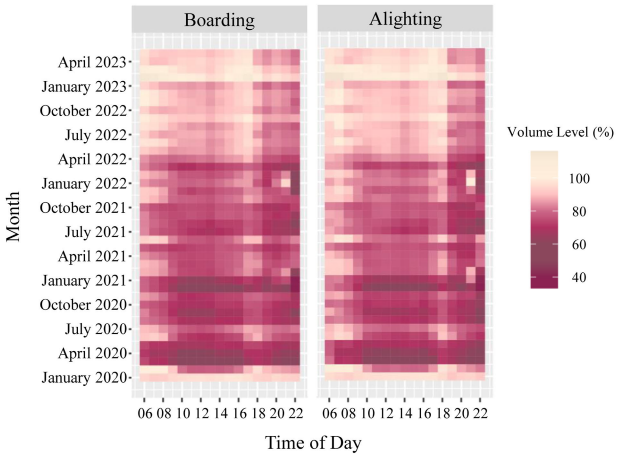

FIGURE 2. Heatmap of city-wide average volume level.

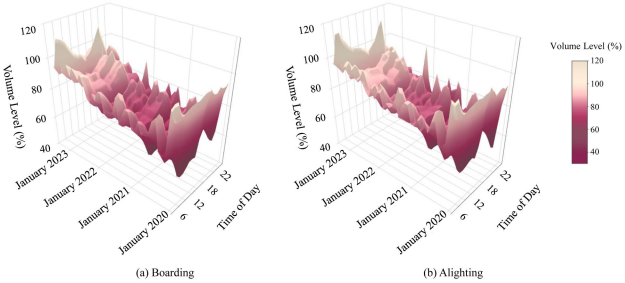

FIGURE 3. Surface plot of city-wide average volume level.

Fig. 2 and Fig. 3 depict the average volume level of all stations, representing the general ridership changes over time compared to the pre-pandemic period. The initial pandemic wave, commencing in February 2020, coincided with a notable decrease in volume level during March

TABLE 2. Average metro passenger boarding volume and volume level changes by time of day.

Period	Morning Peak		Non-Peak		Evening Peak		Nighttime		All Day	
	Volume (persons/hr)	Volume Level (%)	Volume (persons/hr)	Volume Level (%)	Volume (persons/hr)	Volume Level (%)	Volume (persons/hr)	Volume Level (%)	Volume (persons/hr)	Volume Level (%)
Pre-Pandemic (January 2019 to January 2020)	414,633	99.44	281,208	99.74	529,253	99.62	277,795	99.53	316,625	99.63
Early Stage of COVID-19 (February 2020 to October 2021)	332,404	79.90	193,103	71.21	413,104	78.50	187,854	70.64	229,434	73.68
Late Stage of COVID-19 (November 2021 to December 2022)	362,256	89.15	229,908	84.13	464,529	88.07	211,343	78.47	262,438	84.38
Post-Pandemic (January 2023 to May 2023)	384,621	96.74	253,891	94.15	501,536	96.63	234,250	86.79	286,560	93.40

2020. Subsequent pandemic waves resulted in lower peaks in August and December 2020, and July 2021. Only in February 2023 did the system-wide average volume levels exceed 100%, stabilizing to values between 90% and 100% thereafter. As shown in Table 2, significant fluctuations were evident in hourly patterns. During the morning and evening peak hours, the decline in ridership remained comparatively modest, rebounding to volume levels ranging from 104.12% to 116.23% in February 2023. Conversely, a substantial reduction to 50% and almost 30% in comparison to previous passenger volumes was recorded during the non-peak hours and the nighttime hours respectively. The degree of recuperation remains modest, with only 85% of the pre-existing passenger flow having been reinstated in nighttime hours. Noteworthy recovery in early 2023 was observed during morning and evening peak hours. In addition, except for the difference of about an hour due to travel time, the average boarding and alighting pattern exhibited a similar trend across the entire network.

B. REGIONAL EXPLANATORY VARIABLES

Regional explanatory variables include land use, socioeconomic characteristics, and POIs. Land use data[1] provided by the Korean Ministry of Land, Infrastructure, and Transport includes five categories for zones: high density residential (*ResHigh* variable), low density residential (*ResLow* variable), commercial (*Comm* variable), industrial (*Indust* variable), and greenspace. In line with previous research [46], [47], [48], a buffer of 500 meters was applied to each metro station[2] to represent the direct catchment area. The proportion of each zone category within the respective catchment area for each station was calculated (Fig. 4). The data are current as of March 2022.

Regional socioeconomic characteristics were also included, namely household income (*Income* variable), population density (*PopDen* variable), and the density of foodservice establishments (*Foodservice* variable). Household income data are available on the Korean Water Resources Corporation data platform,[3] and population density is derived from the latest census report.[4] As both are gathered at the dong level (smallest level administrative divisions), they were converted to the station level by calculating average of dong level values in the direct catchment area weighted by area. The density of hospitality venues is calculated by counting the number of relevant businesses including restaurants, cafes, bakeries, and bars within the catchment area of each station and dividing that figure by the area of the catchment area. The business location data[5] are provided by the Korean Ministry of the Interior and Safety. Table 3 displays the units and value ranges of each continuous variable included.

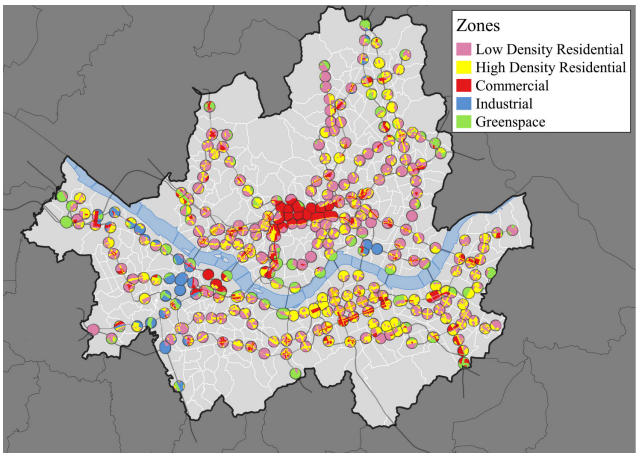

FIGURE 4. Land use zones makeup of metro station catchment areas.

To integrate the POI data as binary variables, a total of 36 stations are identified to be adjacent to major universities with campus sizes exceeding seven thousand students (*Univ* variable). Three Intercity bus stations (Express Bus Terminal, Nambu Terminal, and Dong Seoul Terminal) and six high-speed rail stations (Seoul, Yongsan, Cheongnyangni, Suseo, Yeongdeungpo, Sangbong), all connected to the

[1] http://vworld.kr/dtmk/dtmk_ntads_s002.do?dsId=30300
[2] http://data.seoul.go.kr/dataList/OA-21232/S/1/datasetView.do
[3] https://www.bigdata-environment.kr/user/data_market/detail.do?id=8cee0160-2dff-11ea-9713-eb3e5186fb38
[4] https://data.seoul.go.kr/dataList/10584/S/2/datasetView.do
[5] https://www.localdata.go.kr/

TABLE 3. Summary of continuous explanatory variables.

Variable	Unit	Min	Mean	Max
ResHigh	% of catchment area	0.00	27.29	88.12
ResLow	% of catchment area	0.00	46.65	100.00
Comm	% of catchment area	0.00	12.49	100.00
Indust	% of catchment area	0.00	4.76	100.00
Greenspace	% of catchment area	0.00	8.81	87.53
PopDen	persons/km^2	2,600	21,178	54,309
Foodservice	thousand locations/km^2	0.00	0.59	3.01
Income	10,000KRW/year	2,764	4,237	20,000

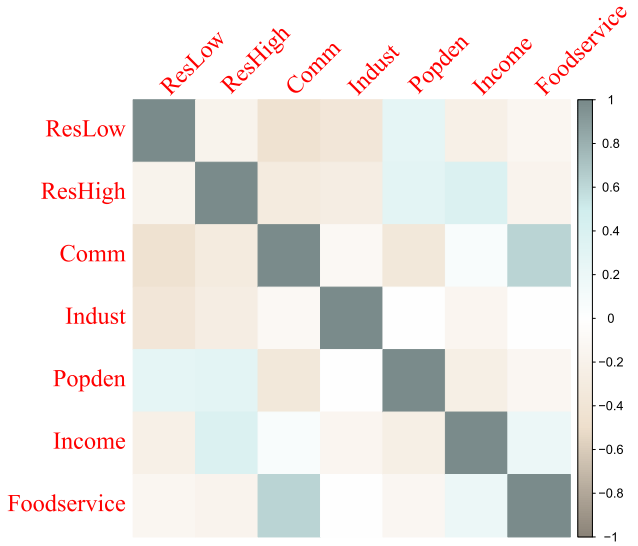

FIGURE 5. Correlation plot between continuous explanatory variables.

metro, were also marked as intercity transportation hubs (*Tr_hub* variable).

Before proceeding with the modeling process, an assessment of multicollinearity was conducted for the socioeconomic and land use variables. Given that the sum of the ratio of all land use zones always amounts to 100%, the *Greenspace* variable was excluded from further analysis. Fig. 5 presents a correlation plot depicting the relationships among all the variables used in the study. All correlation values between variables lie between −0.44 and 0.63.

IV. METHODOLOGY

A. FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

FDA is a statistical analysis method that operates on data represented as functions, rather than discrete values. Unlike conventional statistical techniques, FDA can effectively analyze complex and high-dimensional data by leveraging functional data representations, enabling the recognition of patterns within voluminous and intricate datasets [49]. Given recent advancements in data collection, processing, and storage technologies, FDA emerges as a promising data analysis approach, particularly in contexts where time series data are prevalent. This study aims to analyze the trend of change over time—when and how much the values decrease

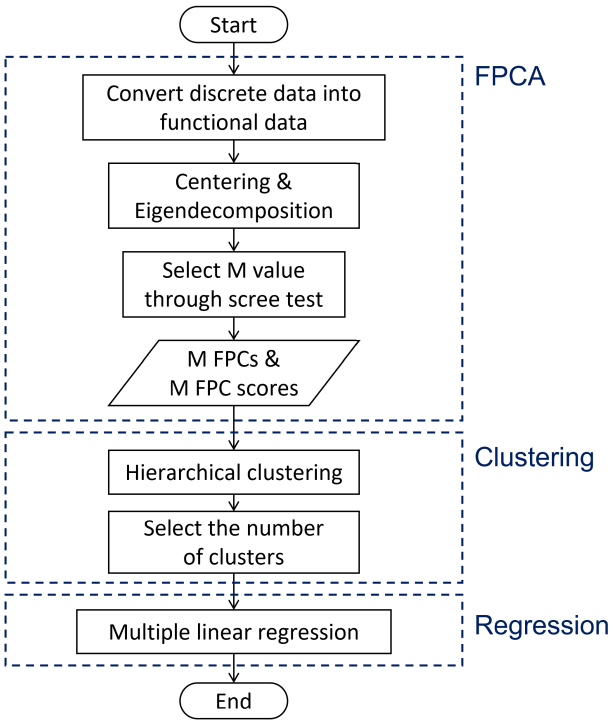

FIGURE 6. Flow chart of analysis.

and recover based on discretized data aggregated by month and hour, so an FDA-based approach is appropriate.

FDA encompasses various methods, including functional regression, functional clustering, and time warping. Among them, Functional Principal Component Analysis (FPCA) serves as a dimension reduction technique that transforms infinite-dimensional functional data into finite-dimensional vector data. Most existing studies employed FPCA for one-dimensional analysis. Extending FPCA to two dimensions presents computational complexities. Nevertheless, it offers the advantage of analyzing data from a broader perspective by simultaneously considering two axes. In this study, analysis was performed both on the month-to-month scale to see the long-term demand reduction caused by COVID-19 and subsequent recovery, and the time of day to identify how such changes differ in the peak, non-peak, and morning and afternoon.

Before delving into the methodology, Table. 4 presents the notations used in this paper and an overall flowchart is provided in Fig. 6. The initial step in FPCA involves converting discrete data with continuous characteristics, such as time series and longitudinal data, into functions. The volume level Y_{ijk} for 1 hour at $t_{jk} = (t_j^1, t_k^2) \in \Omega$ (j^{th} time in the time of day axis, in k^{th} month in the month axis), for the i^{th} metro station can be expressed as

$$Y_{ijk} = X_i(t_{jk}) + \epsilon_{ijk}, \quad \text{for} \quad \begin{matrix} i = 1, 2, \ldots, N, \\ j = 1, 2, \ldots, U, \\ k = 1, 2, \ldots, V \end{matrix} \qquad (1)$$

TABLE 4. Notations.

Sets	
Ω	Set of coordinates with two indices (time of day and month)
Indices	
i	Index representing a metro station
j	Index representing a time on the time of day axis
k	Index representing a month on the time of day axis
m	Index representing an eigenfunction (functional principal component) or an eigenvalue
p	Index representing a spline basis function on the time of day axis
q	Index representing a spline basis function on the month axis
Parameters	
N	Total number of metro stations
U	Total number of time of day indices
V	Total number of months
M	Total number of functional principal components
θ_1	Total number of splines on the time of day axis
θ_2	Total number of splines on the month axis
Variables	
$\mathbf{t}_{jk}, \mathbf{t}, \mathbf{s}$	Vector of two time axes (j^{th} time of day and k^{th}month)
λ_m	m^{th} eigenvalue
Y_{ijk}	Volume level of the i^{th} station on j^{th} time of day and k^{th} month
$\mathbf{a_i}$	FPC score vector of the i^{th} station
$\mathbf{C_m}$	Matrix of spline coefficients of m^{th} eigenfunction (FPC) ϕ_m
γ_{mpq}	p^{th} row, q^{th} column value of $\mathbf{C_m}$
$\gamma_{\mathbf{m}}$	Vector of $\mathbf{C_m}$
$b_p^{(1)}(t^1)$	p^{th} spline basis function on the time of day axis
$b_q^{(2)}(t^2)$	q^{th} spline basis function on the month axis
$\mathbf{b_1}$	Vector of spline basis functions on the time of day axis
$\mathbf{b_2}$	Vector of spline basis functions on the month axis
$\mathbf{b}(t^1, t^2)$	Vector of tensor product of time of day, month spline basis functions
α_{im}	m^{th} FPC score of the i^{th} station
G_1, G_2	Cluster
n_G	Number of metro stations in the cluster G
$\bar{\eta}_G$	Average of FPC score vectors belonging to the cluster G
Functions	
$X_i(\mathbf{t})$	Volume level function on time $\mathbf{t}$ of the i^{th} station
$\mu(\mathbf{t})$	Average volume level on time $\mathbf{t}$ of all stations
$X^c(\mathbf{t})$	Volume level function on time $\mathbf{t}$ subtracted by the overall average $\mu(\mathbf{t})$
$\Sigma(\mathbf{s}, \mathbf{t})$	Covariance function on time $\mathbf{s}$ and $\mathbf{t}$
$\phi_m(\mathbf{t})$	m^{th} two-dimensional eigenfunction on time $\mathbf{t}$
$\Delta(G_1, G_2)$	Ward's distance between two clusters G_1 and G_2

where $X_i(t)$ denotes a function as the dimension of observation extends to infinity, and ϵ represents random error satisfying the mean-zero condition and uncorrelated across time. Subsequently, centering is performed by subtracting the overall average function $E(X_i(t)) = \mu(t)$, where $t = (t^1, t^2)$ across all stations, resulting in the function $X(t) - \mu(t)$ denoted as $X^c(t)$. Eigenanalysis is then conducted on the covariance function to estimate the functional principal components. As per Mercer's theorem, the covariance function can be decomposed into the form

$$\Sigma(s,t) = Cov\left(X^c(s), X^c(t)\right) = \sum_{m=1}^{\infty} \lambda_m \phi_m(s)\phi_m(t), \quad (2)$$
$$s = (s^1, s^2), s, t \in \Omega,$$

where λ_m denotes non-negative eigenvalues and $\phi_m(t)$ signifies the corresponding orthonormal eigenfunctions arranged in descending order. The eigenfunctions are often formulated using basis functions such as B-splines when dealing with discretely observed data [50]. In two-dimensional FPCA, the introduction of an additional dimension necessitates the establishment of a new basis system for the basis functions used in functional data smoothing. A suitable approach for this purpose is the tensor-product of B-spline bases, especially when all data share the same sampling points and have regular domain structures, such as rectangular shapes [51]. It is important to note that the following descriptions are limited to cases where the value Y_{ijk} is available for all $j = 1, 2, \ldots, U$ and $k = 1, 2, \ldots, V$ without any missing data points; this condition holds true for the ridership data used in this study.

The m^{th} eigenfunction or $\phi_m(t)$, also called the functional principal component (FPC), is calculated through eigendecomposition as follows.

$$\begin{aligned}\phi_m(t^1, t^2) &= \sum_{p=1}^{\theta_1}\sum_{q=1}^{\theta_2} \gamma_{mpq} b_p^{(1)}(t^1) b_q^{(2)}(t^2) \\ &= \mathbf{b_1}(t^1)^\top \mathbf{C}_m \mathbf{b_2}(t^2) \\ &= \gamma_{\mathbf{m}}{}^\top \mathbf{b}(t^1, t^2), \end{aligned} \quad (3)$$

where

$$\mathbf{b}(t^1, t^2) = \text{vec}\left(\mathbf{b_1}(t^1) \otimes \mathbf{b_2}(t^2)\right),$$
$$\mathbf{C_m} = (\gamma_{mpq}) : \text{matrix of spline coefficients of } m^{th} \text{ eigenfunction (FPC) } \phi_m,$$
$$\gamma_\mathbf{m} = \text{vec}(\mathbf{C_m}).$$

The tensor-product of $\mathbf{b_1} = (b_1^{(1)}, \ldots, b_{\theta_1}^{(1)})^\top$, and $\mathbf{b_2} = (b_1^{(2)}, \ldots, b_{\theta_2}^{(2)})^\top$ are basis functions on the time of day axis, and month axis with θ_1, and θ_2 number of splines (usually given values), respectively.

The function $X_i(t)$ for each metro can then be expressed as

$$X_i(t^1, t^2) = \mu(t^1, t^2) + \sum_{m=1}^{\infty} \alpha_{im}\phi_m(t^1, t^2), \tag{4}$$

where

$$\alpha_{im} = \int_\Omega (X_i(t) - \mu(t))\, \phi_m(t)dt \tag{5}$$

refers to the FPC score representing the weight of each eigenfunction in the linear combination.

If $X_i(t^1, t^2)$ is approximated by finding the M value, which is the number of top-few eigenfunctions that capture most of the variation (i.e., the number of FPCs), it can be expressed as

$$X_i(t^1, t^2) \approx \mu(t^1, t^2) + \sum_{m=1}^{M} \alpha_{im}\phi_m(t^1, t^2). \tag{6}$$

By determining the appropriate value of M, the data can be effectively reduced to an M-dimensional vector $\mathbf{a_i}$, retaining as much of the data variance as possible. Two common methods for selecting M are: 1) using the minimum M that accounts for over 90% of the data variance, and 2) employing the scree test, which analyzes the eigenvalues plotted against their order to identify the "elbow point" where further eigenvalue reduction decreases significantly compared to previous ones [35], [52].

Since this study focuses on both boarding and alighting data for analysis, a multivariate functional principal component analysis (MFPCA) methodology was adopted. Performing FPCA separately on each data can result in the potential correlation between the two elements being misrepresented in the subsequent phases of analysis [53]. Considering each ridership data entry as a vector of two items, namely boarding and alighting volume levels, the resulting functional principal components of MFPCA are also in a vector form of the same dimensions, with each element being defined as a function. The two functions share the same loading value for each entry, allowing the MFPCA to capture the relationship between the two variables and enabling a more nuanced and rigorous understanding of metro ridership resilience.

B. HIERARCHICAL CLUSTERING

Functional clustering is conducted to group together stations with similar patterns of ridership fluctuations, and by using vector-type data consisting of M FPC scores for each station calculated through FPCA. Clustering is a technique that groups data into similar clusters, employing various methodologies such as centroid-based, density-based, and connectivity-based approaches. Hierarchical clustering is a form of connectivity-based clustering which uses an agglomerative method that reduces the number of clusters in a bottom-up fashion by successively merging two clusters. Initially, each station constitutes its own cluster, resulting in as many clusters as the number of metro stations. As the process iterates, clusters are combined until only one cluster containing all individuals remains. To facilitate the clustering process, a distance matrix is calculated to determine the dissimilarity between each cluster. Two clusters with the greatest similarity (shortest distance) are merged into one, and this process is repeated by continually updating the distance matrix and integrating the closest cluster pair. Various methods can be employed to measure the distance between clusters, such as the distance between centroids, or the shortest, longest, or average distances between stations in the two clusters. In this study, Ward's (minimum variation) method was applied, which minimizes the error sum of squares within groups that increase when two clusters are combined.

The distance between two clusters G_1 and G_2, $\Delta(G_1, G_2)$, can be expressed using Equation (7) [54]. Vector $\bar{\eta}_G$ is the average of FPC score vector $\mathbf{a_i}$ of station i belonging to the cluster G, and n_G signifies the number of metro stations in cluster G.

$$\begin{aligned}\Delta(G_1, G_2) = &\sum_{i\in G_1\cup G_2} \| \mathbf{a_i} - \bar{\eta}_{\mathbf{G_1}\cup\mathbf{G_2}} \|^2 \\ &- \sum_{i\in G_1} \| \mathbf{a_i} - \bar{\eta}_{G_1} \|^2 \\ &- \sum_{i\in G_2} \| \mathbf{a_i} - \bar{\eta}_{G_2} \|^2 \\ &= \frac{n_{G_1} n_{G_2}}{n_{G_1} + n_{G_2}} \| \bar{\eta}_{G_1} - \bar{\eta}_{G_2} \|^2 \end{aligned} \tag{7}$$

Upon completing the cluster formation process, a dendrogram is constructed to visually represent the sequence of cluster formations and the interrelationships between each cluster in a tree structure. The clustering outcome of this study is depicted as a dendrogram in Fig. 9. The x-axis corresponds to each metro station in this analysis, and the y-axis indicates $\Delta(G_1, G_2)$. The height of the branches on the dendrogram reflects the similarity between the connected clusters. Lower branch heights indicate earlier-formed clusters, whereas higher y-axis values denote clusters that have merged despite significant differences. As the last step, the data are divided into several clusters by selecting a cutting height, whereby the entire stations can be partitioned into the desired number of clusters. This process allows for the creation of distinct

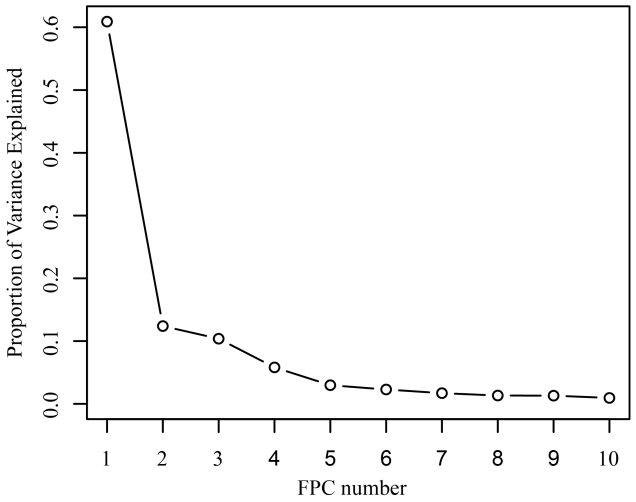

FIGURE 7. Scree plot for determining the appropriate number of FPCs.

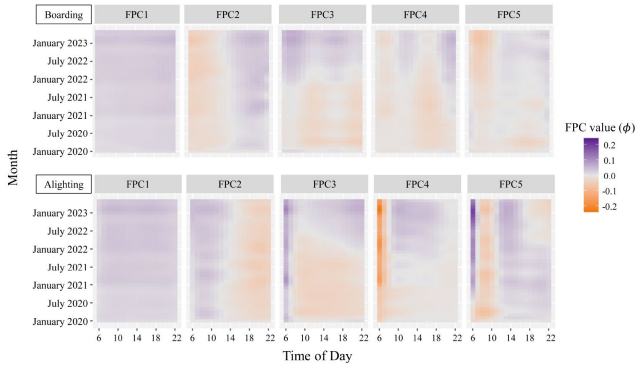

FIGURE 8. Heatmap of five FPC values (ϕ).

groups based on the underlying patterns and relationships identified during the clustering analysis.

V. ANALYSIS RESULTS

A. FPC PROFILES

The process of centering in FPCA yields the average across all metro stations, as illustrated in Fig. 2 and Fig. 3 in the previous section. After centering, FPCs were computed through eigendecomposition. To determine the appropriate number of FPCs M, a scree test was performed. Fig. 7 portrays the scree plot, indicating an evident elbow point at 5 FPCs. These FPCs collectively account for a cumulative variance of 84%, distributed in ascending order of FPC as follows: 55.34%, 11.26%, 9.43%, 5.28%, and 2.70%.

Each FPC is visually depicted in Fig. 8, with distinct roles attributed to each. FPC1, boasting the highest variance explanatory power, exhibits a slight positive trend throughout the entire period with minimal temporal variation. FPC2 primarily captures the divergence between morning and afternoon patterns, showcasing contrasting boarding and alighting behaviors. FPC3 clearly shows a significant reduction of ridership following the initial drop at the breakout of COVID-19 in early 2020 and the recovery pattern after the end of the pandemic (i.e., new normal). Meanwhile, FPC4 and FPC5 complexly express aspects according to time and period, and also play a role in distinguishing between peak and non-peak hours. Remarkably, an unusual observation pertains to early morning alighting patterns, a pattern elaborated further in the forthcoming segments. This observation linked to the influence of site-specific events surrounding the stations, rather than being attributable to the pandemic's effects.

B. CLUSTER GROUPS

The classification framework was established based on the assessment of five distinct FPC scores. Subsequently, hierarchical clustering was employed to analyze these score values. The dendrogram illustrated in Fig. 9 provided a guiding structure for segmenting the data. Initially, the cluster group was divided into four primary partitions, which were further refined into eleven subgroups to achieve a more intricate clustering. The spatial distribution of stations across these eleven clusters is visually represented in the map provided in Fig. 44 (in the Appendix).

Upon analyzing the individual FPC scores depicted in Fig. 10, where each scatter plot portrays two of the five FPC score axes, the demarcation of cluster group boundaries is predominantly influenced by FPC1 as shown in the first row and first column. Notably, Group B exhibits the highest FPC1 scores, followed by Group A, C, and D in descending order. It can be deduced that cluster groups with elevated FPC1 scores encompass stations that showcase a swifter recovery of volume levels, signifying a higher level of overall resilience.

Whereas Groups A and B share comparable FPC1 scores, Group A demonstrates slightly higher FPC2 values, as demonstrated in the FPC1 vs. FPC2 graph (top-left panel) in Fig. 10. Recalling that FPC2 elucidates divergent boarding and alighting behavior across different hours of the day, Group A encompasses stations where the recovery or increase in morning alightings and evening boardings has outpaced other hours of the day, relative to Group B. These patterns are also reflected in the average volume levels for each cluster group, as depicted in Fig. 11 and Fig. 12. A more comprehensive elaboration on the distinct characteristics of each cluster group is provided below.

1) GROUP A

Stations belonging to Group A saw an overall decline of ridership in both peak and non-peak hours during the early stages of the pandemic, which was then followed by a recovery of peak-time travel beginning in late 2022. This recovery was most prominent in the morning peak for alightings and the evening peak for boardings, peaking at a volume level greater than 140%. The location of these stations corresponds to areas where businesses and hospitality venues are primarily located, specifically those with high recovery or increases in commuter flows during the analysis period.

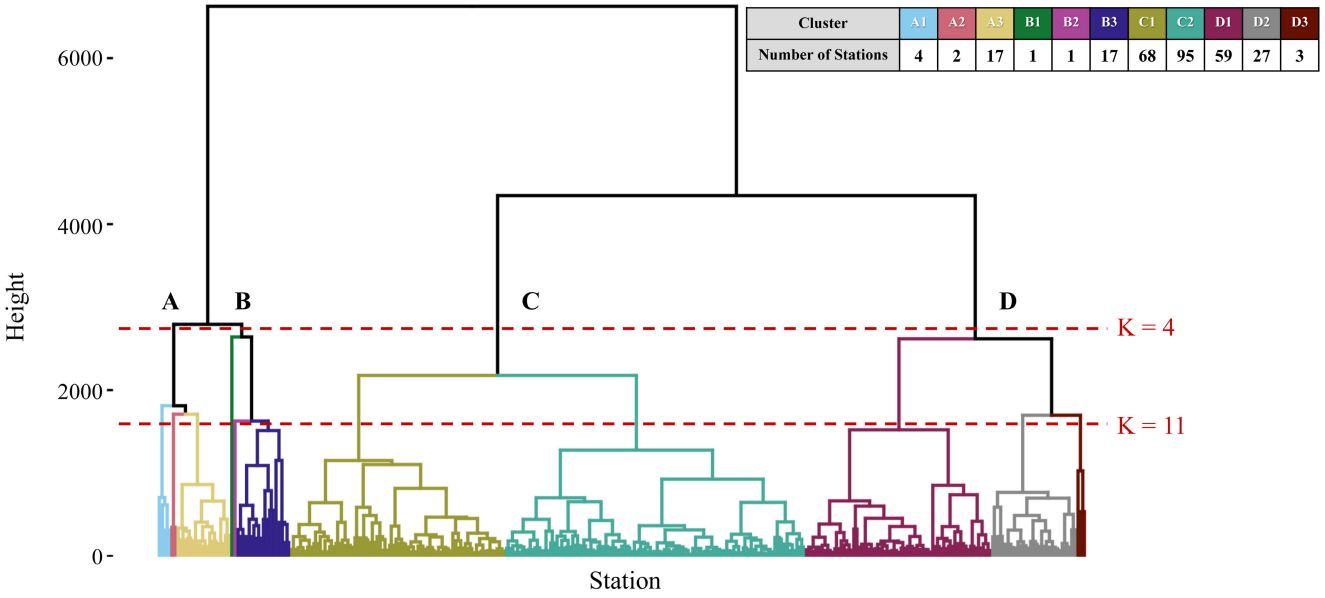

FIGURE 9. Dendrogram of hierarchical clustering.

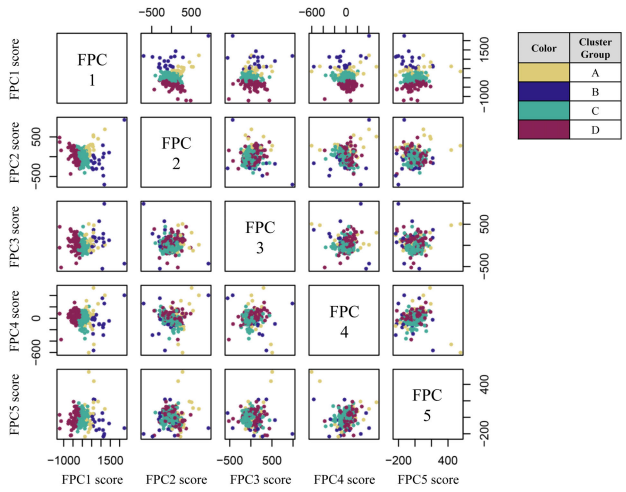

FIGURE 10. Scatter plots of FPC pairs.

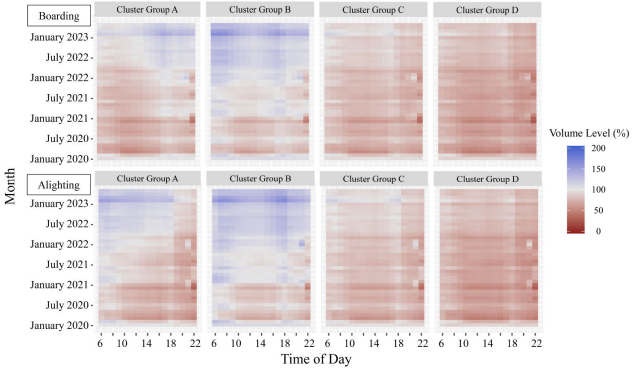

FIGURE 11. Average volume level heatmaps of cluster groups.

2) GROUP B

Cluster B pertains to stations that exhibited a marginal decline in ridership during the initial phases of the study duration. Remarkably, the recovery trajectory for this group outpaced that of other groups. By early- to mid-2022, these stations had successfully surpassed the pre-pandemic ridership benchmarks. Furthermore, by February 2023, these stations achieved a noteworthy upsurge, with volume level exceeding 160% during both morning and evening peak hours. This cluster group's remarkable resilience, characterized by a minor decrease in ridership and a swift resurgence likely stems from specific localized conditions, such as ongoing residential and commercial development initiatives or the introduction of new transit lines and services.

3) GROUP C

The volume levels of stations in Group C experienced a more pronounced collapse at the onset of COVID-19 and failed to return to 2019 levels as of early 2023. Although peak-time volume levels retained higher values compared to the rest of the day, it is not until early 2023 that this is achieved.

4) GROUP D

Group D comprises stations hardest hit by the pandemic, characterized by smaller resilience of volume levels at peak hours compared to other groups and extremely low minimum values. At no period of time does the volume of passenger flows exceed that of 2019 for all hours of the day.

C. CLUSTERS

Learning from the cluster group traits identified, the stations were further categorized into eleven clusters. Specifically, Groups A, B, and D were divided into three clusters, and

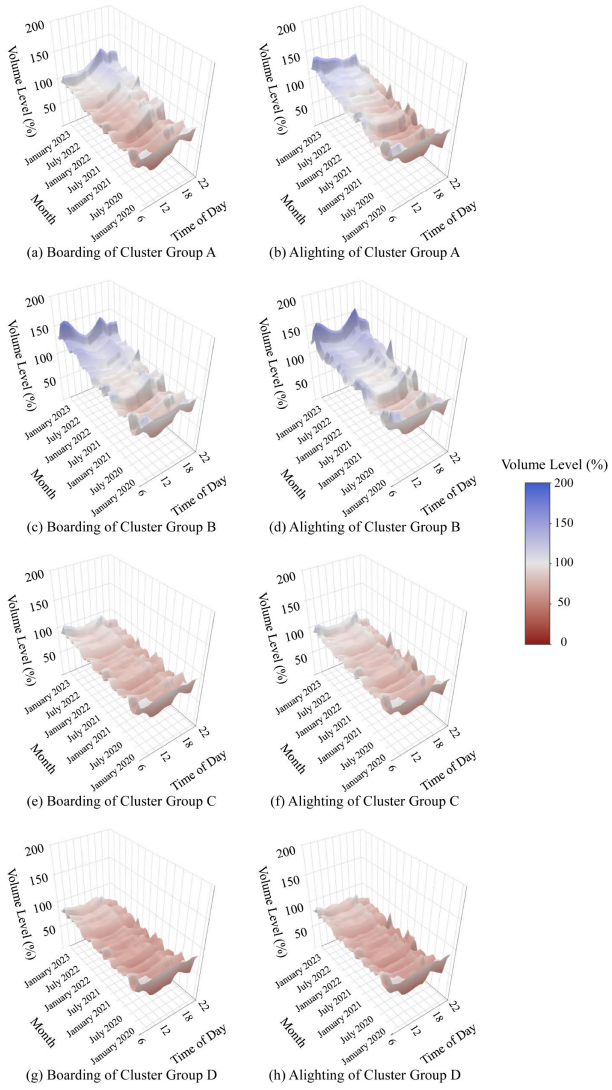

FIGURE 12. Average volume level surface plots of cluster groups.

Group C was split into two clusters. Fig. 45 in the Appendix depicts the average volume level of each cluster. To explore the relevant socioeconomic and locational factors associated with each cluster and its ridership shift patterns, a multiple linear regression model was constructed in line with existing literature on neighborhood characteristics and cluster analysis [46], [55]. For each cluster, a distinct linear regression model was formulated, where the explanatory variables are the socioeconomic and locational variables explored above and the response variable is a binary value indicating whether a station belonged to that cluster. All regression analyses were performed using the R *stats* package. It needs to be noted that because of the nature in which linear regression models are fitted, clusters with very few members are difficult to explain. When one or few stations exhibited highly distinctive ridership patterns and were thus given its own cluster, exogenous factors that led to such unique patterns including urban

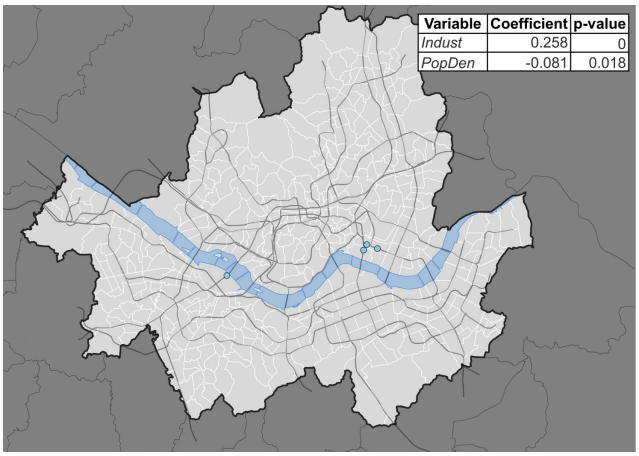

FIGURE 13. Location of Cluster A1.

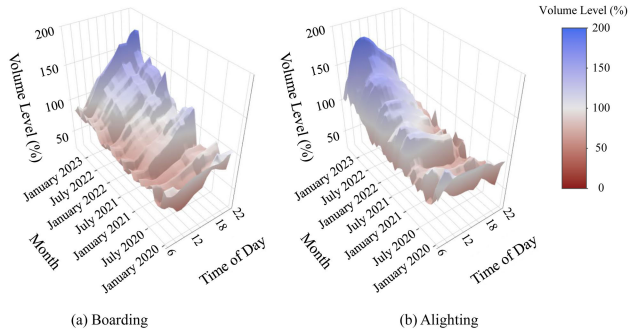

FIGURE 14. Volume level surface plot of Cluster A1.

development and PT service changes were identified and described.

1) CLUSTER A1

The first cluster consists of three stations in the Seongsu area in the eastern part of the city, and one in Yeouido, an island on the Han river (Fig. 13). The regression results identify the proportion of industrial zones as the primarily significant variable and indicate a low positive correlation with a coefficient of 0.258. Volume levels of both boardings and alightings at these stations plummeted in early 2020, and increased considerably throughout the analysis period to values over 160% for evening-peak boardings and morning-peak alightings, respectively (Fig. 14 and Fig. 15). This correctly suggests a growth of these areas as hubs of economic activity and subsequent increases in commuter movements. During the analysis period multiple large corporations and start-ups moved their headquarters into these areas. Despite the COVID-19 pandemic's general suppression of passenger flows, the development around stations in Cluster A1 resulted in consistent increases in ridership to well above 2019 levels.

2) CLUSTER A2

Cluster A2 includes two stations in the southern part of Seoul (Fig. 16). These stations underwent a distinctively

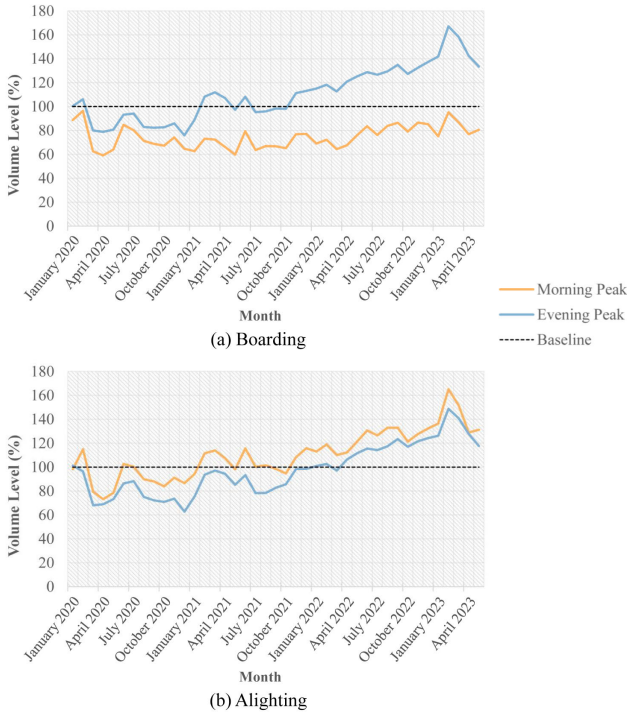

FIGURE 15. Monthly volume level graph for Cluster A1.

FIGURE 16. Location of Cluster A2.

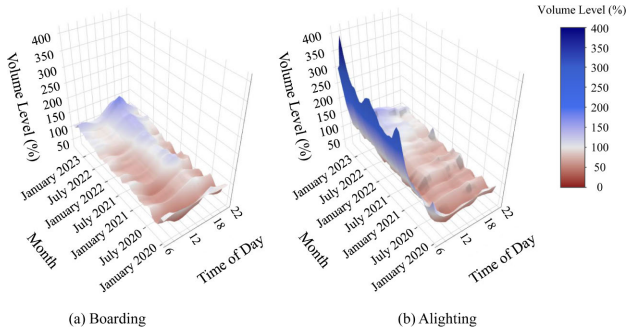

FIGURE 17. Volume level surface plot of Cluster A2.

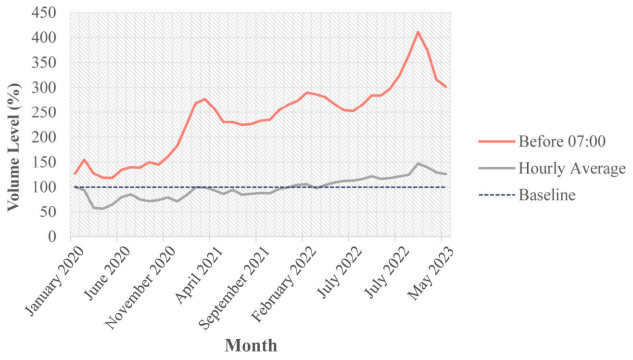

FIGURE 18. Monthly alighting volume level graph for Cluster A2.

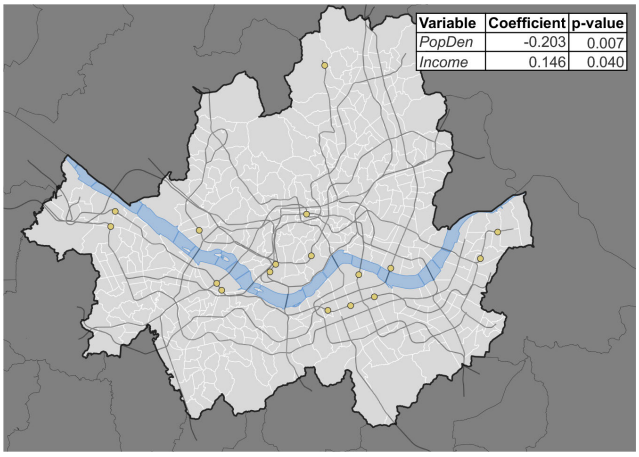

FIGURE 19. Location of Cluster A3.

sharp rise in the number of alighting passengers in the early morning hours, before 07:00 (Fig. 17 and Fig. 18). These are highly unique patterns not found in other clusters, and the factors contributing to such a shift are found outside the regression model. At both of these stations, large-scale housing developments of high-density residential complexes commenced during the analysis period, and the beginnings of their early-morning ridership hikes match those of the dates when the developments started. Around Sinbanpo station, located in the central southern part of the city, housing development of a 3,000-household residential complex began in April of 2020 and was still ongoing as of May 2023. Likewise, near Guryong station, housing development of a 6,700-household complex started in June 2020. This result confirms that multivariate multidimensional FDA is an effective tool to identify unique cases of exceptional volume level changes driven by urban redevelopment.

3) CLUSTER A3

Cluster A3 encompasses a total of seventeen stations, making up the majority of Group A. The spatial distribution of these stations (Fig. 19), along with their alignment to the population density and income variables, points to their positioning within the sub-centers of the city of Seoul or local hubs

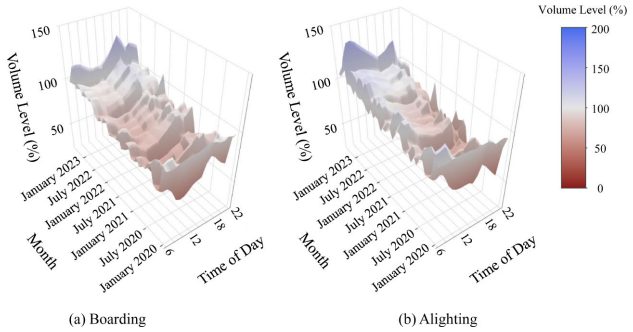

FIGURE 20. Volume level surface plot of Cluster A3.

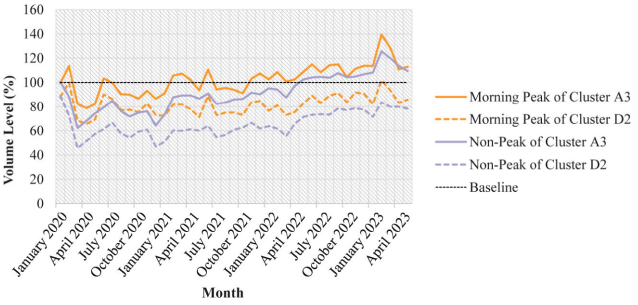

FIGURE 21. Monthly alighting volume level graph for Clusters A3 and D2.

FIGURE 22. Location of Clusters B1 and B2.

of commercial and service functions, often characterized by substantial office spaces.

Significantly, the bounceback of ridership in these areas has exceeded that of stations located in central business districts (CBDs) of the city (e.g., Cluster D2), extending not only to peak hours but also encompassing non-peak hours (Fig. 20 and Fig. 21). These outcomes suggest that considering the enduring recuperation trends observed in recent months, the presence of secondary cores as essential activity centers may assume a more prominent role than prior to the pandemic.

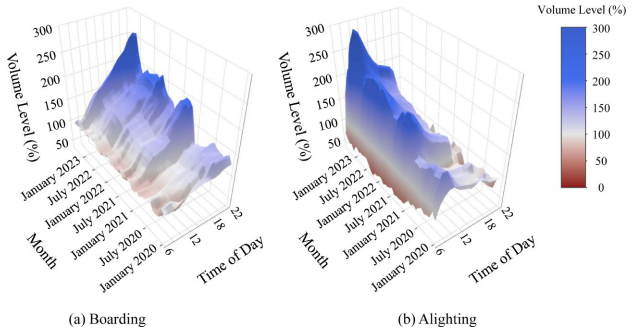

FIGURE 23. Volume level surface plot of Cluster B1.

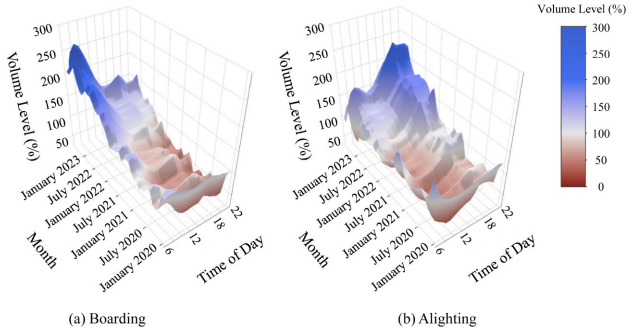

FIGURE 24. Volume level surface plot of Cluster B2.

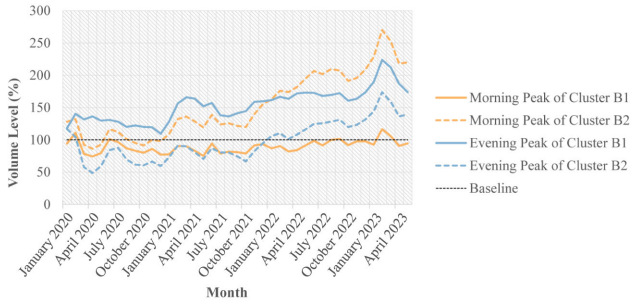

FIGURE 25. Monthly alighting volume level graph for Clusters B1 and B2.

4) CLUSTERS B1 AND B2

Clusters B1 and B2 experienced significant hikes in peak travel volumes throughout the analysis period in opposite passenger flow directions. Cluster B1, which consists solely of Magok station at the western end of the city, saw a large rise in alightings in the morning peak hours and boardings in the evening peak hours (Fig. 23) alike Cluster A1, whereas Cluster B2, whose only member is Yangwon station located at the eastern end of the city (Fig. 22), experienced increases in morning-peak boardings and evening-peak alightings (Fig. 24). In both cases the volume levels reached apices of over 270%, indicating nearly triple the hourly passenger flow of the same month in 2019 (Fig. 25).

The area surrounding Cluster B1 forms the commercial and business center of an ongoing large-scale development project encompassing 33.6 hectares which includes research

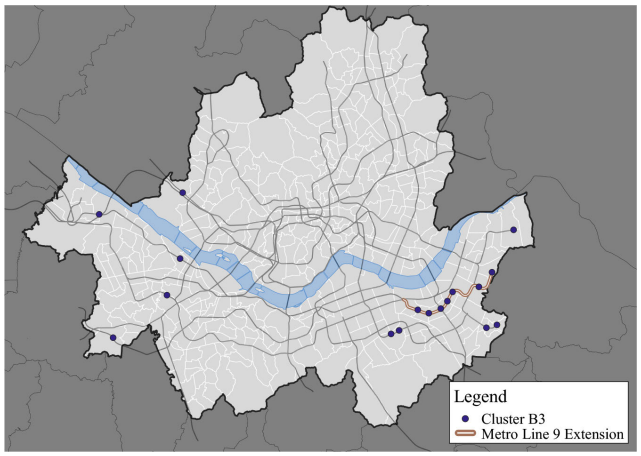

FIGURE 26. Location of Cluster B3 and Metro Line 9 Extension.

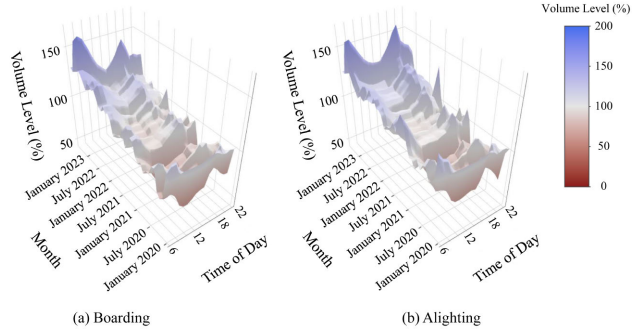

FIGURE 27. Volume level surface plot of Cluster B3.

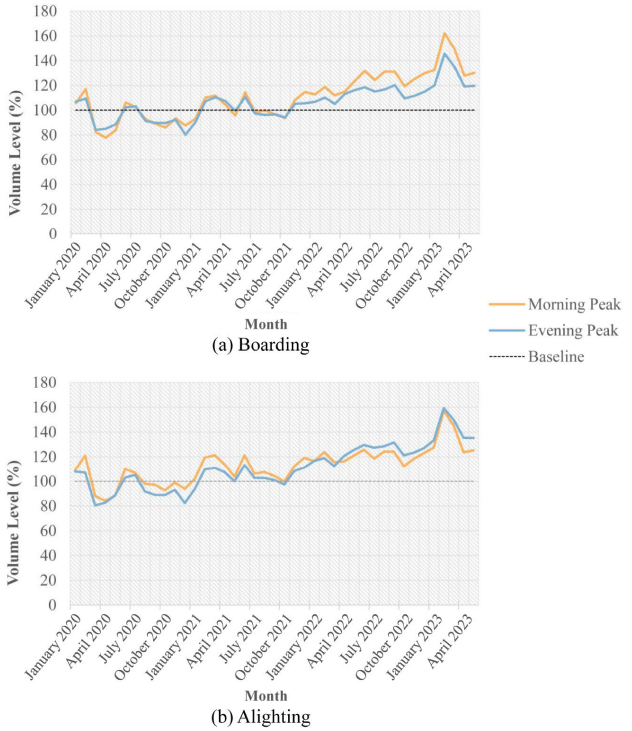

FIGURE 28. Monthly alighting volume level graph for Cluster B3.

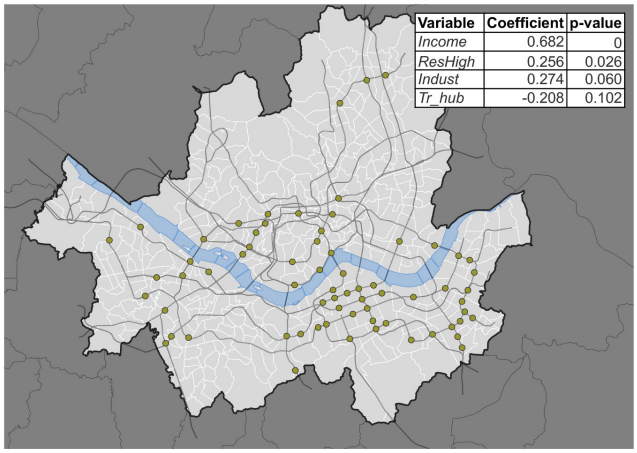

FIGURE 29. Location of Cluster C1.

complexes and high-density commercial zones. The analysis period aligns with the period when much of these commercial buildings completed construction and businesses began to establish themselves in them, leading to a consistent and substantial increase in commuter flows at this station. On the other hand, Cluster B2 is adjacent to a major new public housing project area with more than three thousand households, into which residents began to move in August 2021. The comparison of these clusters clearly shows the distinction between commercial and residential development areas.

5) CLUSTER B3

Cluster B3 showcases a minimal decrease in overall volume levels. Instead, the volume levels exceed 100% during morning and afternoon peak times throughout most of the analysis period, indicating that demand during the COVID-19 pandemic exceeded that of the year 2019 (Fig. 27 and Fig. 28). The selected stations within this cluster can be categorized into two distinct groups. The first group comprises seven stations situated along the eastern portion of metro line 9. This segment forms the third phase of the line's construction, and only commenced operation in December 2018. Consequently, the observed increase in demand is largely attributed to the low usage in the initial stage of operation of this newly introduced segment in 2019. The alignment of this extension project is highlighted in Fig. 26.

The second group of stations pertains to the new adjacent residential and commercial development projects positioned in the southeastern and western parts of the city. These factors have notably contributed to the increased demand observed in these stations.

6) CLUSTERS C1 AND C2

Comparison of these two clusters highlights the difference between residential areas with high density and high incomes and those with low density and low incomes. Cluster C1

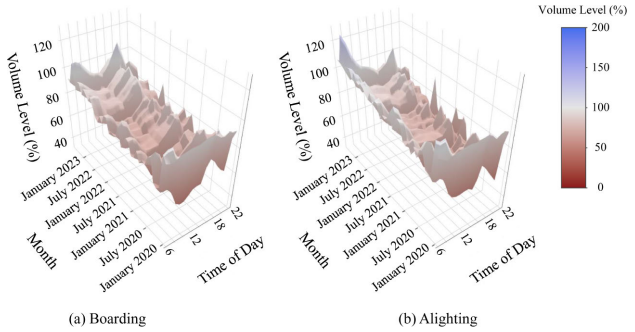

FIGURE 30. Volume level surface plot of Cluster C1.

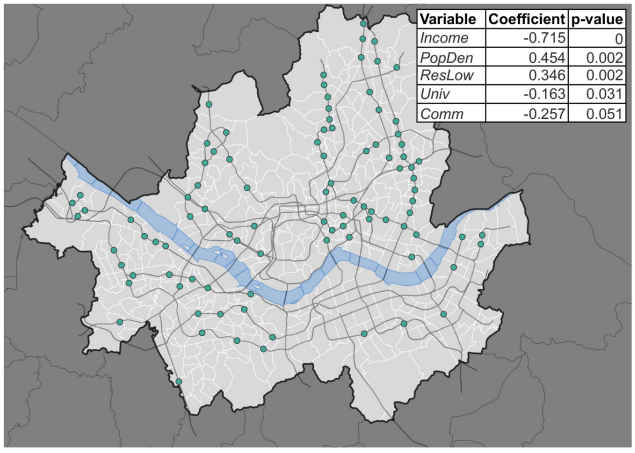

FIGURE 31. Location of Cluster C2.

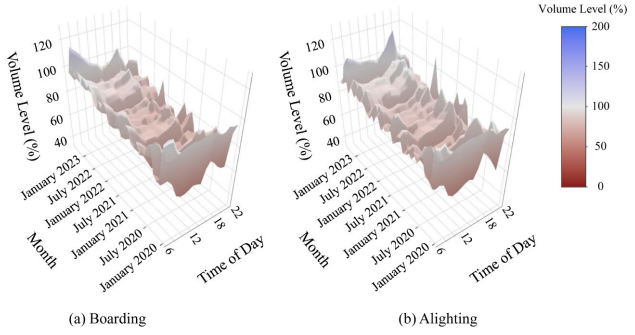

FIGURE 32. Volume level surface plot of Cluster C2.

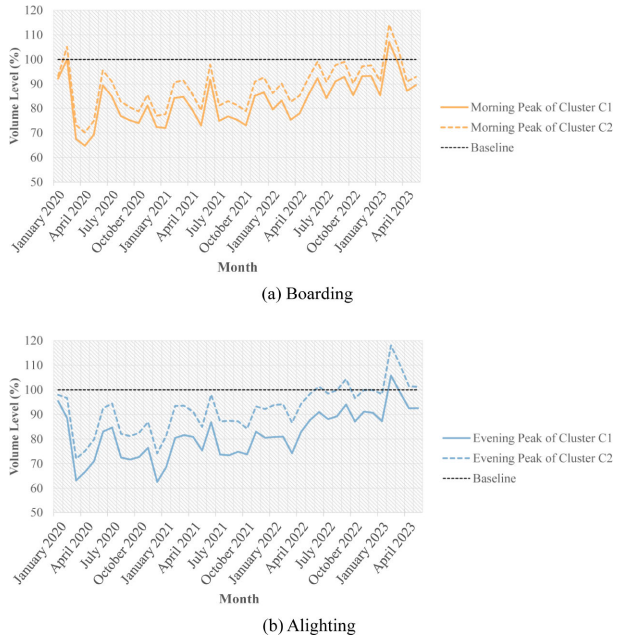

FIGURE 33. Monthly volume level graph for Clusters C1 and C2.

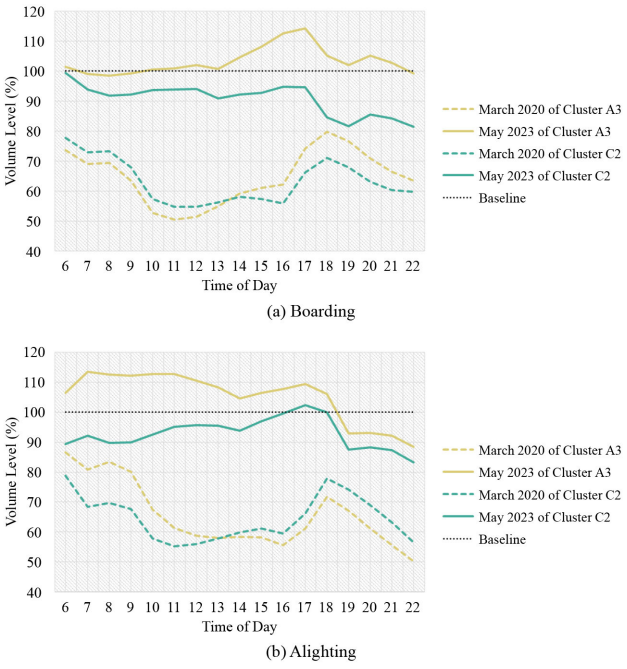

FIGURE 34. Hourly volume level graph for Clusters A3 and C2.

corresponds with the Gangnam district, where the largest agglomeration of businesses and residences in Seoul can be found (Fig. 29). Deviating from previous clusters examined, volume level trends for boardings and alightings in this cluster exhibit resilience in both morning and evening peak hours, figures for which recovered to 100% by the end of the analysis period after having previously dipped to 50%. On the contrary, non-peak volume levels have yet to recover their 2019 values, suggesting that travel unrelated to business activity was much less resilient (Fig. 30). This may be attributed to the large number of both residents and businesses in this area as is further evidenced by the regression results where the high density residential and income variables show high correlation coefficients of 0.256 and 0.682 respectively with p-values of less than 0.05.

Cluster C2 aligns most closely with the income variable, followed by the population density and low-density residential variables. Stations in this cluster are located in areas where low-rise residences are tightly packed together, primarily in the northeast of the city (Fig. 31). As with Cluster C1, the initial impact at the onset of the pandemic is

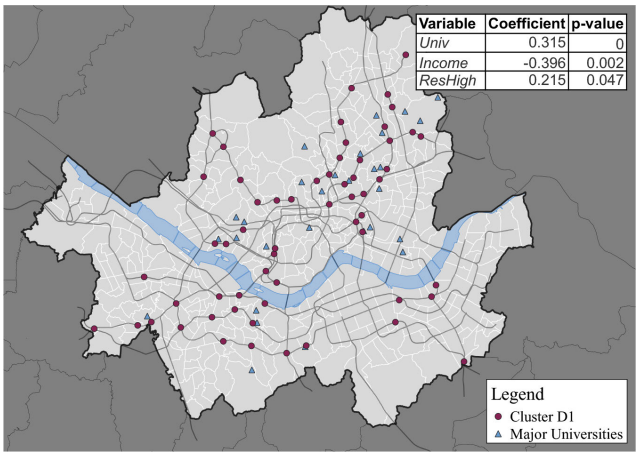

FIGURE 35. Location of Cluster D1.

noticeable, especially during the non-peak hours. However, the recovery of morning alightings and evening boardings is not as swift and significant, owing to the predominantly residential nature of these areas (Fig. 32 and Fig. 33).

Fig. 34 illustrates a comparative analysis between Cluster A2 (secondary commercial cores) and Cluster C2 (residential areas). Both regions exhibit an overall decrease attributed to the pandemic, with notably larger reductions during non-peak hours (40–50%) compared to peak hours. Moreover, the reduction in morning peak boarding and afternoon peak alightings in residential areas (Cluster C2), as well as morning peak alightings and afternoon peak boardings in commercial areas (Cluster A3), is relatively modest. This suggests that, during the analyzed timeframe, individuals engaged in minimal travel activities beyond their usual commuting routines. The demand recovery in residential areas (Cluster C2) is observed exclusively in relation to commuting activities, whereas Cluster A3, representing secondary centers within the city, demonstrated a robust recovery exceeding 100% volume levels for most hours of the day. Despite the recovery, both areas experienced substantial reductions in nighttime volume levels compared to pre-pandemic levels, and the recuperation process has been characterized by a gradual pace.

7) CLUSTER D1

Assignment into Cluster D1 strongly correlates with the *Univ* variable. Fig. 35 provides a comparison of the placement of Cluster D1 and the locations of major universities in Seoul. These stations served a significant portion of university students and local residents who sought low-price residential options in proximity to educational institutions. The demand for regular commuting from these areas, represented in the data as morning boardings and evening alightings, closely mirrors the city-wide average. However, morning alightings and evening boardings notably fall below the average and have not witnessed substantial recovery post-pandemic (Fig. 36 and Fig. 37). This observation suggests a

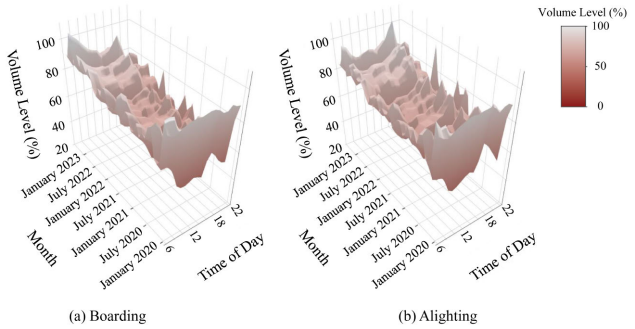

FIGURE 36. Volume level surface plot of Cluster D1.

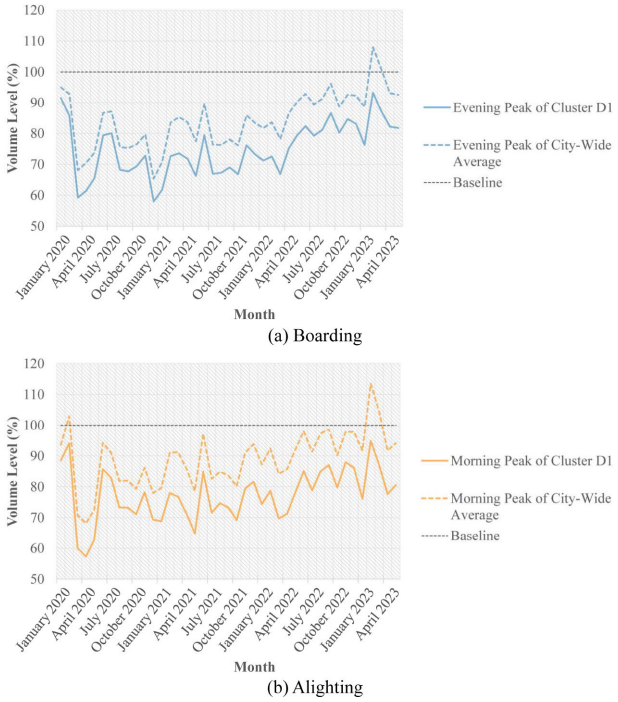

FIGURE 37. Monthly volume level graph for Cluster D1 and city-wide average.

decline in the number of students attending in-person classes on a daily basis, attributable to the widespread adoption of remote classes and a reduction in face-to-face extracurricular activities.

8) CLUSTER D2

Cluster D2's distribution aligns itself closely with the central Jongno area and parts of the Gangnam area (Fig. 38). These are Seoul's primary CBDs, as is evidenced by the positive correlation with the *Comm* and *Foodservice* variables, whereas there exists a negative correlation between assignment to this cluster and the *PopDen* variable. This is an area dominated by commercial zones with extremely low resident population density. The introduction of work-from-home and interpersonal distancing policies greatly lowered commuter demand in these areas initially. As interpersonal

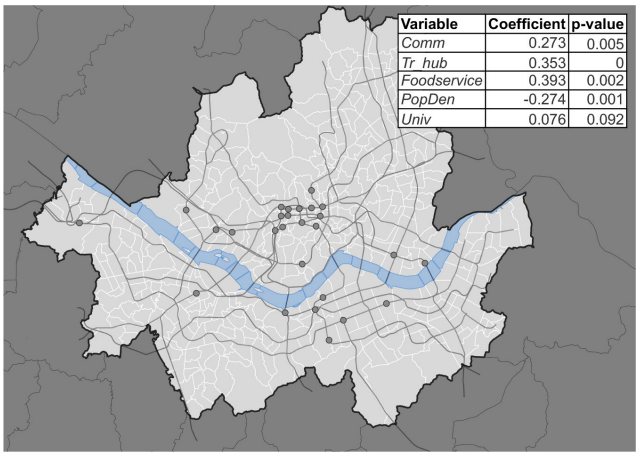

FIGURE 38. Location of Cluster D2.

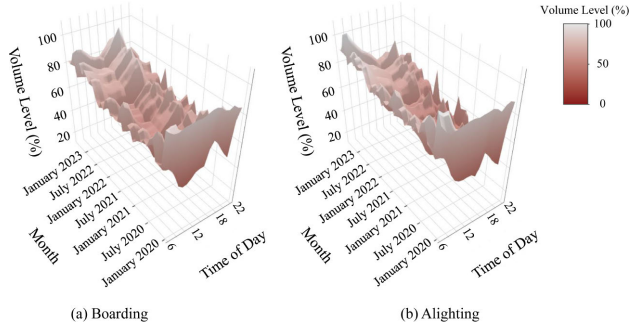

FIGURE 39. Volume level surface plot of Cluster D2.

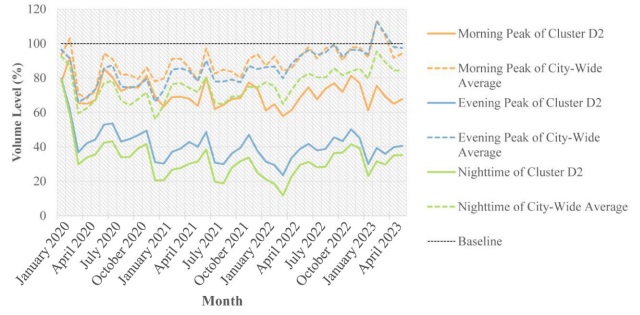

FIGURE 40. Monthly alighting volume level graph for Cluster D2 and city-wide average.

distancing restrictions were loosened commuters in these areas slowly began to return, up to volume levels of 105% for morning arrivals and 95% for evening departures by February 2023. On the contrary, such patterns are not observed during the afternoon and night hours, as non-work travel exhibits a much weaker resilience (Fig. 40).

In addition, the major intercity transportation hubs are assigned to this cluster, including Seoul station and the three intercity bus stations. Demand for long distance travel fell sharply during the pandemic period. Note the slower pace of recovery especially in non-peak hours compared to sub-centers of the city in Cluster A3 (Fig. 39).

FIGURE 41. Location of Cluster D3.

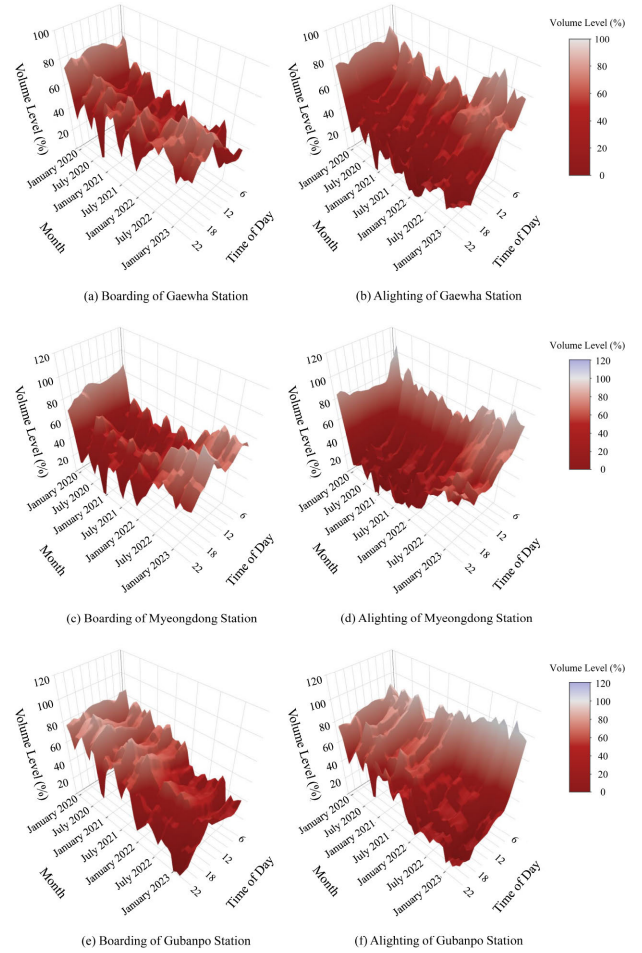

FIGURE 42. Volume level surface plot of stations in Cluster D3.

9) CLUSTER D3

All three stations in Cluster D3 (Fig. 41) exhibited consistent and substantial declines in ridership throughout the analysis period, deviating from the typical resilience pattern observed in other clusters. The unique circumstances surrounding each station shed light on these divergent trends (Fig. 42).

In the case of Gaewha station, situated at the western extremity of the city, the introduction of a new light metro played a pivotal role in reshaping its ridership dynamics. This new transport option displaced metropolitan buses connected to Gaewha station as the primary means through which residents from western suburbs accessed central Seoul. Consequently, this station witnessed a decline in ridership as the new metro system gained prominence.

Myeongdong station, located in the heart of the city and in close proximity to one of Seoul's most renowned shopping districts, had been a magnet for foreign tourists prior to the pandemic. As global border closures imposed an unprecedented hindrance to foreign tourism, Myeongdong station experienced a drastic shift in its ranking within the transit system, plummeting from being the twenty-seventh most frequented station in 2019 to dropping below the top one hundred during the pandemic. Unfortunately, the resurgence of foreign tourist inflows to its pre-pandemic levels has yet to be realized.

Gubanpo station, located in the southern part of Seoul, underwent a distinct ridership decline due to this unique set of circumstances. After a large-scale residential redevelopment began there in mid-2021, a substantial number of former residents of the 3,500-household neighborhood were relocated during the analysis period. This sudden outflow of resident population coincides with the period during which the station's ridership dropped significantly.

In summary, these three stations in Cluster D3 experienced ridership decreases due to specific factors such as the introduction of new transportation modes, the absence of foreign tourism, and the impact of residential redevelopment.

D. DISCUSSIONS

Through an analysis of volume level patterns since the onset of the COVID-19 pandemic employing multivariate two-dimensional functional data analysis, a spectrum of valuable insights emerges regarding the resilience of neighborhoods to such shocks. The pandemic's overarching influence on metro ridership becomes evident through a significant reduction during the initial months of the analysis period, followed by a gradual recovery over subsequent months.

Furthermore, commuting trips have exhibited a propensity to return to pre-pandemic patterns as remote work trends waned and more individuals reintegrated into traditional office environments. In contrast, non-peak hours have not witnessed a commensurate level of recovery in volume levels. Moreover, areas with differing densities of residentials demonstrate varying rates of peak-time volume level recovery due to disparities in the concentration.

In comparison, business and commercial districts experienced a notably sharper decline in volume levels starting in early 2020. This decline can be traced back to the adoption of remote work arrangements and public apprehension about densely populated regions. It is noteworthy that the recuperation of non-peak demand within these areas has been substantially limited even into the year 2023. This is especially the case in CBDs where the impact of the decreases in international and intercity travel are more prevalent, and the recovery of regional sub-centers has proven to be more pronounced than that of the CBDs.

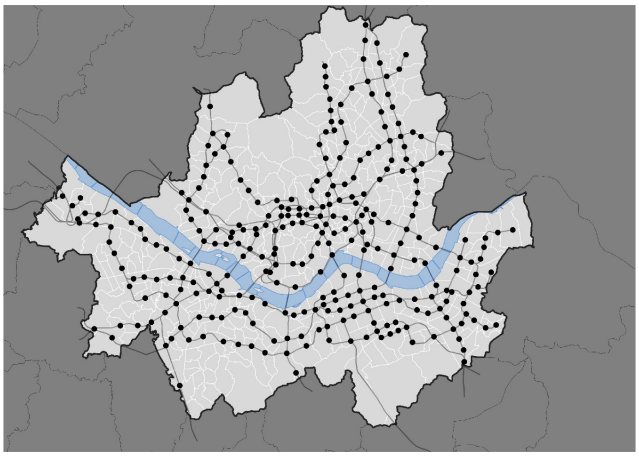

FIGURE 43. Boundary of Seoul and locations of target stations.

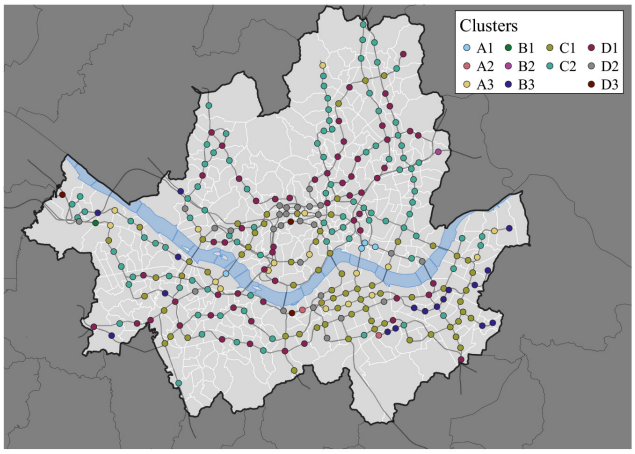

FIGURE 44. Location and assigned cluster of metro stations.

Remarkably, unique circumstances, such as the initiation or completion of large-scale development projects, have led to a significant surge in travel demand, surpassing pre-pandemic levels, despite the prevailing challenges. The process of delineating and grouping stations that showcase such exceptional patterns within the clustering framework, enables a more profound and nuanced exploration of the underlying factors that contribute to these distinctive trends.

VI. CONCLUSION

In this paper, the shock-and-recovery process of transit ridership before, during, and after the COVID-19 pandemic was examined. Adopting the analytical framework of multivariate FDA, the varied patterns of resilience in transit ridership from 2020 to 2023 in Seoul were investigated. The

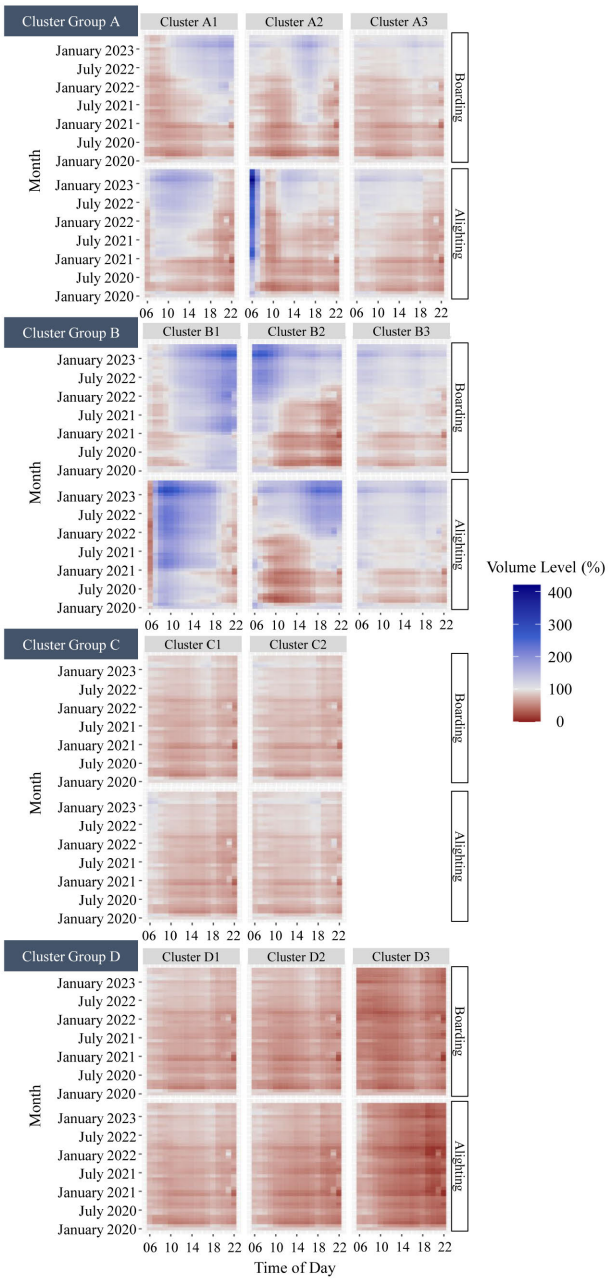

FIGURE 45. Volume level heatmap of each cluster.

findings in this paper illuminate the heterogeneous nature of transit ridership patterns across neighborhoods during the analysis period. The analysis model was formulated by transforming ridership data into a two-dimensional format consisting of month and time-of-day axes and exploring concurrent boardings and alightings together in a multivariate context. Through the application of FPCA on smart card data and augmenting these results with hierarchical clustering, distinctive shared characteristics of each cluster were able to be identified. The clustering process serves to categorize stations exhibiting similar passenger flow patterns, and the regression analysis assists in pinpointing the underlying socioeconomic and land use factors that link each cluster. Further, the model was able to effectively identify stations with significantly divergent ridership patterns caused by extraordinary circumstances as outliers.

The insights gained from this research have illuminated regional disparities in resilience, providing a foundation for the development of targeted and contextually relevant ex-ante mitigation strategies. The findings derived from this study would help urban planners and policymakers proactively design measures that foster resilient neighborhoods and contingency plans for unexpected extreme events like the COVID-19 pandemic and other forms.

A handful of opportunities for further research present themselves:

1) The expansion of the geographical scope of research to the metropolitan level with the maximum commuting range may offer more comprehensive insights with regards to peak-time commuter flow patterns.
2) The inclusion of different modes of transportation such as buses, automobiles, and shared mobility services could also provide a finer analysis that takes into account nuances such as modal shifts incurred by the pandemic.
3) Whereas the COVID-19 pandemic was a universal phenomenon, its impact was not necessarily geographically uniform, especially in the early stages where infection mainly occurred in clusters in limited areas. As such, the deployment of spatial regression models could be used to capture these geographical aspects of resilience.
4) More sophisticated FDA techniques such as functional regression could allow the researcher to explore the interactions between socioeconomic factors and PT ridership.

APPENDIX

See Figures 43–45.

ACKNOWLEDGMENT

(Won Gyun Choi and Seunghee Ryu are co-first authors.)

REFERENCES

[1] M. J. Beck and D. A. Hensher, "Insights into the impact of COVID-19 on household travel and activities in Australia—The early days under restrictions," *Transp. Policy*, vol. 96, pp. 76–93, Sep. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X20304200

[2] M. E. G. Parker, M. Li, M. A. Bouzaghrane, H. Obeid, D. Hayes, K. T. Frick, D. A. Rodríguez, R. Sengupta, J. Walker, and D. G. Chatman, "Public transit use in the United States in the era of COVID-19: Transit riders' travel behavior in the COVID-19 impact and recovery period," *Transp. Policy*, vol. 111, pp. 53–62, Sep. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X21002067

[3] D. Sharma, C. Zhong, and H. Wong, "Lockdown lifted: Measuring spatial resilience from London's public transport demand recovery," *Geo-Spatial Inf. Sci.*, vol. 26, no. 4, pp. 685–702, Feb. 2023, doi: 10.1080/10095020.2022.2156300.

[4] O. Manout, L. Bouzouina, K. Kourtit, and P. Nijkamp, "On the bumpy road to recovery: Resilience of public transport ridership during COVID-19 in 15 European cities," *Lett. Spatial Resource Sci.*, vol. 16, no. 1, p. 14, Apr. 2023, doi: 10.1007/s12076-023-00338-8.

[5] A. Ziedan, C. Brakewood, and K. Watkins, "Will transit recover? A retrospective study of nationwide ridership in the United States during the COVID-19 pandemic," *J. Public Transp.*, vol. 25, 2023, Art. no. 100046. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077291X23000073

[6] R. Martin and P. Sunley, "On the notion of regional economic resilience: Conceptualization and explanation," *J. Econ. Geogr.*, vol. 15, no. 1, pp. 1–42, Jan. 2015, doi: 10.1093/jeg/lbu015.

[7] M. Vagliasindi, *The Impact of COVID-19 on Mobility and Congestion* (Policy Research Working Papers). Washington, DC, USA: The World Bank, Mar. 2023, doi: 10.1596/1813-9450-10351.

[8] J. Anke, A. Francke, L.-M. Schaefer, and T. Petzoldt, "Impact of SARS-CoV-2 on the mobility behaviour in Germany," *Eur. Transp. Res. Rev.*, vol. 13, no. 1, p. 10, Jan. 2021, doi: 10.1186/s12544-021-00469-3.

[9] M. J. Beck and D. A. Hensher, "Insights into the impact of COVID-19 on household travel and activities in Australia—The early days of easing restrictions," *Transp. Policy*, vol. 99, pp. 95–119, Dec. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X20306880

[10] M. de Haas, R. Faber, and M. Hamersma, "How COVID-19 and the Dutch 'intelligent lockdown' change activities, work and travel behaviour: Evidence from longitudinal data in The Netherlands," *Transp. Res. Interdiscipl. Perspect.*, vol. 6, Jul. 2020, Art. no. 100150. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590198220300610

[11] T. Paul, R. Chakraborty, and N. Anwari, "Impact of COVID-19 on daily travel behaviour: A literature review," *Transp. Saf. Environ.*, vol. 4, no. 2, Jun. 2022, Art. no. tdac013, doi: 10.1093/tse/tdac013/6613419.

[12] A. Nikolaidou, A. Kopsacheilis, G. Georgiadis, T. Noutsias, I. Politis, and I. Fyrogenis, "Factors affecting public transport performance due to the COVID-19 outbreak: A worldwide analysis," *Cities*, vol. 134, Mar. 2023, Art. no. 104206. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0264275123000185

[13] P. Bucsky, "Modal share changes due to COVID-19: The case of Budapest," *Transp. Res. Interdiscipl. Perspect.*, vol. 8, Nov. 2020, Art. no. 100141. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S259019822030052X

[14] S. Luan, Q. Yang, Z. Jiang, and W. Wang, "Exploring the impact of COVID-19 on individual's travel mode choice in China," *Transp. Policy*, vol. 106, pp. 271–280, Jun. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X21001062

[15] M. Abdullah, C. Dias, D. Muley, and M. Shahin, "Exploring the impacts of COVID-19 on travel behavior and mode preferences," *Transp. Res. Interdiscipl. Perspect.*, vol. 8, Nov. 2020, Art. no. 100255. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590198220301664

[16] M. Wilbur, A. Ayman, A. Sivagnanam, A. Ouyang, V. Poon, R. Kabir, A. Vadali, P. Pugliese, D. Freudberg, A. Laszka, and A. Dubey, "Impact of COVID-19 on public transit accessibility and ridership," *Transp. Res. Rec.*, vol. 2677, no. 4, pp. 531–546, Apr. 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10107069/

[17] J. Zhang, "People's responses to the COVID-19 pandemic during its early stages and factors affecting those responses," *Humanities Social Sci. Commun.*, vol. 8, no. 1, pp. 1–13, Feb. 2021. [Online]. Available: https://www.nature.com/articles/s41599-021-00720-1

[18] E. Zhou and J. Lee, "How has COVID-19 changed trip patterns by purpose in China?" *Transp. Saf. Environ.*, vol. 4, no. 4, Nov. 2022, Art. no. tdac030, doi: 10.1093/tse/tdac030.

[19] J. Zhang, Y. Hayashi, and L. D. Frank, "COVID-19 and transport: Findings from a world-wide expert survey," *Transp. Policy*, vol. 103, pp. 68–85, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X21000172

[20] R. Fernández Pozo, M. R. Wilby, J. J. Vinagre Díaz, and A. B. R. González, "Data-driven analysis of the impact of COVID-19 on Madrid's public transport during each phase of the pandemic," *Cities*, vol. 127, Aug. 2022, Art. no. 103723. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0264275122001627

[21] M. Ammoury, B. Salman, C. E. Caicedo Bastidas, and S. Kumar, "Impacts of COVID-19 on bus ridership and recovery trends in syracuse, new York," *J. Transp. Eng., A, Syst.*, vol. 149, no. 2, Feb. 2023, Art. no. 05022009, doi: 10.1061/JTEPBS.TEENG-7498.

[22] E. Giannakis and A. Bruggeman, "Regional disparities in economic resilience in the European union across the urban–rural divide," *Regional Stud.*, vol. 54, no. 9, pp. 1200–1213, Sep. 2020, doi: 10.1080/00343404.2019.1698720.

[23] M. Sensier, G. Bristow, and A. Healy, "Measuring regional economic resilience across Europe: Operationalizing a complex concept," *Spatial Econ. Anal.*, vol. 11, no. 2, pp. 128–151, Apr. 2016, doi: 10.1080/17421772.2016.1129435.

[24] F. Dobruszkes and G. Van Hamme, "The impact of the current economic crisis on the geography of air traffic volumes: An empirical analysis," *J. Transp. Geogr.*, vol. 19, no. 6, pp. 1387–1398, Nov. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966692311001220

[25] A. Potter, A. Soroka, and M. Naim, "Regional resilience for rail freight transport," *J. Transp. Geogr.*, vol. 104, Oct. 2022, Art. no. 103448. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966692322001715

[26] A. D. Marra, L. Sun, and F. Corman, "The impact of COVID-19 pandemic on public transport usage and route choice: Evidences from a long-term tracking study in urban area," *Transp. Policy*, vol. 116, pp. 258–268, Feb. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X21003620

[27] R. L.-J. Luo, "Data-driven customer segmentation: Assessing disparities in COVID impact on public transit user groups and recovery," M.S. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, Jun. 2021. [Online]. Available: https://dspace.mit.edu/handle/1721.1/138908

[28] J. Jiao, K. Hansen, and A. Azimian, "Disparities in the impacts of the COVID-19 pandemic on public transit ridership in Austin, Texas, U.S.A," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2677, no. 4, pp. 287–297, Apr. 2023, doi: 10.1177/03611981231159906.

[29] L. Liu, H. J. Miller, and J. Scheff, "The impacts of COVID-19 pandemic on public transit demand in the United States," *PLoS ONE*, vol. 15, no. 11, Nov. 2020, Art. no. e0242476. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242476

[30] H. Seya, T. Yoshida, and M. Tsutsumi, "Ex-post identification of geographical extent of benefited area by a transportation project: Functional data analysis method," *J. Transp. Geogr.*, vol. 55, pp. 1–10, Jul. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966692316303763

[31] P. H. Jung and J. Song, "Multivariate neighborhood trajectory analysis: An exploration of the functional data analysis approach," *Geographical Anal.*, vol. 54, no. 4, pp. 789–819, Oct. 2022, doi: 10.1111/gean.12298.

[32] J. Wang, X. Kong, A. Rahim, F. Xia, A. Tolba, and Z. Al-Makhadmeh, "IS2Fun: Identification of subway station functions using massive urban data," *IEEE Access*, vol. 5, pp. 27103–27113, 2017.

[33] T. Tang, X. Kong, M. Li, J. Wang, G. Shen, and X. Wang, "VISOS: A visual interactive system for spatial–temporal exploring station importance based on subway data," *IEEE Access*, vol. 6, pp. 42131–42141, 2018.

[34] Y. Park, Y. Choi, K. Kim, and J. K. Yoo, "Machine learning approach for study on subway passenger flow," *Sci. Rep.*, vol. 12, no. 1, p. 2754, Feb. 2022. [Online]. Available: https://www.nature.com/articles/s41598-022-06767-7

[35] K. Wang and F. Tsung, "Sparse and robust multivariate functional principal component analysis for passenger flow pattern discovery in metro systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8367–8379, Jul. 2022.

[36] P. H. Jung, H. Kim, K. Lee, and Y. Song, "Examining day-to-day dynamic transit accessibility using functional data analysis," *Prof. Geographer*, vol. 74, no. 3, pp. 503–515, Jul. 2022, doi: 10.1080/00330124.2021.2007493.

[37] M. Galvani, A. Torti, A. Menafoglio, and S. Vantini, "A novel spatio-temporal clustering technique to study the bike sharing system in Lyon," in *Proc. EDBT/ICDT Workshops*, 2020, pp. 1–10.

[38] A. Roy, T. Nelson, and P. Turaga, "Functional data analysis approach for mapping change in time series: A case study using bicycle ridership patterns," *Transp. Res. Interdiscipl. Perspect.*, vol. 17, Jan. 2023, Art. no. 100752. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590198222002135

[39] X. Li, J. Ha, and S. Lee, "Mobility resilience of commute trips during the COVID-19 pandemic in Seoul, Korea," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 135–142, Oct. 2022. [Online]. Available: https://isprs-annals.copernicus.org/articles/X-4-W3-2022/135/2022/isprs-annals-X-4-W3-2022-135-2022.html

[40] B. Barabino, M. Di Francesco, and R. Ventura, "Evaluating fare evasion risk in bus transit networks," *Transp. Res. Interdiscipl. Perspect.*, vol. 20, Jul. 2023, Art. no. 100854. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S259019822300101X
[41] B. Barabino, C. Lai, and A. Olivo, "Fare evasion in public transport systems: A review of the literature," *Public Transp.*, vol. 12, no. 1, pp. 27–88, Mar. 2020, doi: 10.1007/s12469-019-00225-w.
[42] A. V. Reddy, J. Kuhls, and A. Lu, "Measuring and controlling subway fare evasion: Improving safety and security at New York City transit authority," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2216, no. 1, pp. 85–99, Jan. 2011, doi: 10.3141/2216-10.
[43] E. H. Lee, "Exploring transit use during COVID-19 based on XGB and SHAP using smart card data," *J. Adv. Transp.*, vol. 2022, pp. 1–12, Sep. 2022. [Online]. Available: https://www.hindawi.com/journals/jat/2022/6458371/
[44] Seoul Metropolitan Government. (Mar. 2016). *1350 Wone Beorin Yangsim. Bujeongseungcha Subeopdo Gajigaji [Conscience Thrown Away for 1,350 Won.. Various Methods of Unauthorized Boarding]*. [Online]. Available: https://opengov.seoul.go.kr/mediahub/7865064
[45] J. Lee and J. Lee, "Spatial clustering of Seoul's elderly captive riders using smart card spatial autocorrelation analysis," *Sensors Mater.*, vol. 31, no. 11, p. 3859, Nov. 2019. [Online]. Available: http://myukk.org/SM2017/article.php?ss=2589
[46] X. Zhao, Y.-P. Wu, G. Ren, K. Ji, and W.-W. Qian, "Clustering analysis of ridership patterns at subway stations: A case in Nanjing, China," *J. Urban Planning Develop.*, vol. 145, no. 2, Jun. 2019, Art. no. 04019005.
[47] D. Kwon, S. E. S. Oh, S. Choi, and B. H. S. Kim, "Viability of compact cities in the post-COVID-19 era: Subway ridership variations in Seoul Korea," *Ann. Regional Sci.*, vol. 71, no. 1, pp. 175–203, Mar. 2022, doi: 10.1007/s00168-022-01119-9.
[48] Z. Chen, P. Li, Y. Jin, S. Bharule, N. Jia, W. Li, X. Song, R. Shibasaki, and H. Zhang, "Using mobile phone big data to identify inequity of aging groups in transit-oriented development station usage: A case of Tokyo," *Transp. Policy*, vol. 132, pp. 65–75, Mar. 2023.
[49] J. O. Ramsay and C. J. Dalzell, "Some tools for functional data analysis," *J. Roy. Stat. Soc., B, Methodolog.*, vol. 53, no. 3, pp. 539–561, Jul. 1991. [Online]. Available: https://www.jstor.org/stable/2345586
[50] S. Ullah and C. F. Finch, "Applications of functional data analysis: A systematic review," *BMC Med. Res. Methodol.*, vol. 13, no. 1, p. 43, Mar. 2013, doi: 10.1186/1471-2288-13-43.
[51] L. Zhou and H. Pan, "Principal component analysis of two-dimensional functional data," *J. Comput. Graph. Statist.*, vol. 23, no. 3, pp. 779–801, Jul. 2014, doi: 10.1080/10618600.2013.827986.
[52] R. B. Cattell, "The scree test for the number of factors," *Multivariate Behav. Res.*, vol. 1, no. 2, pp. 245–276, Apr. 1966, doi: 10.1207/s15327906mbr0102_10.
[53] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains," *J. Amer. Stat. Assoc.*, vol. 113, no. 522, pp. 649–659, Apr. 2018, doi: 10.1080/01621459.2016.1273115.
[54] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, Mar. 1963. [Online]. Available: https://www.jstor.org/stable/2282967
[55] M.-K. Kim, S.-P. Kim, J. Heo, and H.-G. Sohn, "Ridership patterns at subway stations of Seoul capital area and characteristics of station influence area," *KSCE J. Civil Eng.*, vol. 21, no. 3, pp. 964–975, Mar. 2017, doi: 10.1007/s12205-016-1099-8.

WON GYUN CHOI received the bachelor's degree in international studies and architectural, civil, and environmental engineering from Korea University, Seoul, South Korea, in 2021, where he is currently pursuing the Ph.D. degree with the Transportation and Logistics Systems Laboratory. His research interests include public transit planning, transportation engineering, and international development cooperation.

SEUNGHEE RYU received the bachelor's degree in civil, environmental and architectural engineering and mechanical engineering from Korea University, Seoul, South Korea, in 2021, where she is currently pursuing the Ph.D. degree with the Transportation and Logistics Systems Laboratory. Her research interests include mobility data intelligence, public transportation, and optimization of transport systems.

PAUL H. JUNG received the B.S. degree in urban planning and engineering from Yonsei University, in 2007, the M.C.P. degree in urban and regional planning from Seoul National University, in 2013, and the Ph.D. degree in geography and urban regional analysis from the University of North Carolina at Charlotte, in 2021. He is currently an Assistant Professor in international trade logistics with the Asia Pacific School of Logistics, Inha University, Incheon, South Korea, and has a joint appointment as an Assistant Professional Researcher with the UC Riverside School of Public Policy. His research interests include economic geography, transport geography, and spatial data science.

SEUNGMO KANG (Member, IEEE) received the B.S. and M.S. degrees in civil engineering from Seoul National University, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree in civil engineering from the University of Illinois at Urbana–Champaign, in 2008. He is currently an Associate Professor with the School of Civil, Environmental and Architectural Engineering, Korea University. His research interest includes developing mathematical models to address challenging problems that arise in the fields of transportation planning and logistics.