

## Article

# Analysis of Mobile Phone Data to Compare Mobility Flows and Hotspots Before and After the Opening of High-Speed Railway: Case Study of Honam KTX in Korea

Jin Ki Eom , Kwang-Sub Lee, Ji Young Song and Jun Lee

Railroad Policy Research Team, Future Transport Policy Research Division, Korea Railroad Research Institute, Uiwang 16105, Korea; leeks33@krri.re.kr (K.-S.L.); jysong@krri.re.kr (J.Y.S.); leejun11@krri.re.kr (J.L.)

\* Correspondence: jkom00@krri.re.kr; Tel.: +82-31-460-5467

Received: 26 June 2020; Accepted: 18 July 2020; Published: 21 July 2020



**Abstract:** Mobile phone data provides information, such as the home (origin) and current locations of people. The data can be used to analyze the impact of new high-speed railway (HSR) openings. This study examined the population observed in stations and cities of the Honam HSR line in Korea, based on mobile phone data recorded one year before and after its opening. We analyzed the volume of the population observed at each railway station, density of the distance between home and station, and activity hotspots in a city. The results show that the number of people and travel distance increased after the opening of the HSR. The distance to access railway stations increased, as the HSR saves travel time. Moreover, the activity hotspots in a city increased after the opening of the HSR, as more people gathered near the station area. The findings show that the mobility measures enhanced after the opening of the HSR for regional travel and local activities. These measures can help transit agencies and planners in providing better intercity travel.

**Keywords:** high-speed railway (HSR); mobile phone data; travel behavior; hotspots; compactness

## 1. Introduction

Opening a new high-speed railway (HSR), as a transport-mobility enhancement strategy, has been approved by policymakers in many countries. Since the introduction of the first HSR in the world (in Japan, 1964), several HSRs have been built, and are still proposed worldwide. The HSR network is expected to reach 25,000 km by 2020, even though the cost of construction, maintenance, and operation of HSRs is high [1].

Because of the remarkable increase in accessibility owing to the HSR, cities connected to the railway line have realized new opportunities and changes of all aspects in our society. For example, urban regeneration opportunities, increasing diverse individual activities owing to the reduced travel time while increasing personal time, providing higher quality of service than competing modes, and/or influencing local economics through individuals changing homes and jobs. Many countries have conducted extensive studies on the impact of HSR from various perspectives, such as regional socioeconomics, changes in transport mode shares, and travel patterns [2,3]. The results of the studies had a positive effect on the regional economic growth [4–6]. On the other hand, a recent study in Japan showed that the “straw effect,” defined as a negative economic externality of new transport accessibility was observed, wherein a large city absorbs the commercial and industrial activities from smaller cities connected by the HSR.

Studies on the impact of the new HSR largely focused on the changes in travel demand, mode share, and travel patterns using travel survey data obtained at sites, such as stations and terminals.

The results showed that the HSR helped in increasing passenger volume and mode share ratio, particularly with respect to air transportation. However, despite a growing concern about the impact of new HSR construction, there have been limited studies because of a lack of data; studies were only restricted to travel surveys and passenger statistics. Recently, large amounts of data have been collected via mobile communication; these data include the spatiotemporal locations of phone users, as well as their gender, age, and home location. This is a reliable source for investigating urban issues related to mobility patterns and spatiotemporal demographic dispersion of a city. Thus, it is expected that mobile phone data allow for the provision of disaggregated, microscopic information without incurring significant costs involved in data collection.

The mobile phone data offered by SK Telecom (SKT), Korea's largest telecommunications company with more than 29 million mobile subscribers; this accounts for around 50% of the total market in Korea (IT STAT 2015). In order to represent the total mobile phone users, SKT expanded its mobile phone records by adjusting the number of SKT users by their market share in each region. Therefore, the mobile phone data represents the total population of mobile users (hereafter, we use 'mobile population' for short). One of major benefits using SKT mobile phone data is that the dataset contains all mobile phone signals unless users turn off their phone. That is because a cellular base tower sends a push-alarm signal to the user's phone every 15 min, the mobile signal is recorded as long as the phone is turned on. According to SKT, the turn-off ratio is less than 7% of total users. Due to concerns of privacy, the provided data is not available at an individual person level and does not include location trajectories of each user. Instead, SKT aggregated the individual record to a grid cell ( $50 \times 50$  m) and provided to the public. Therefore, these valuable data that can be analyzed represent a broad and significant sample; these data will be able to help in comprehensively understanding spatiotemporal patterns of mobile phone users.

The main purpose of this study is to demonstrate the research potential of mobile phone data for analyzing the impact the HSR has had, as a case study. We evaluate the impact of the HSR based on the mobile phone data recorded at an HSR station one year before and after its opening. Mobile phone data provide information, such as home location and all spatial locations of mobile phone users, as well as their gender and age group, regardless of whether they make a phone call. This information is used in the impact analysis of the HSR in terms of the volume of mobile phone users, geographical expansion of access areas, and urban demographic dispersion. In our study, we investigate the Honam Korea Train Express (KTX) line introduced on April 2, 2015. We focus on three cities—Iksan, Jeongeup, and Mokpo—which are located at the south approximately 200 km, 250 km, and 330 km, respectively, away from Seoul. We geographically explore the impact of influence area with respect to the distance from the capital city. We extracted all mobile phone data recorded at the three Honam HSR stations for a one-week period in March 2015 and March 2016 for a before-and-after study.

We perform basic statistical analysis of mobile phone data in terms of the gender and age group to investigate the volume changes and test whether the volume of age group by gender changes statistically. We then calculate the access distance for each HSR station based on the home (of the phone users) and HSR station locations. We then compare the geographical distribution of access locations. Further, we compare hotspot changes and compactness of hotspots, distance from station, to investigate how the HSR changes the hotspots in a city and attracts people near the stations. This study will be useful as it shows the potential use of mobile phone data in conducting the impact study of HSRs.

## 2. Literature Review

The impact HSRs have had have been studied from different perspectives worldwide. Because of the different environment in terms of connectivity, technology, and socioeconomic status, it is difficult to compare each case study and to draw a systematic and homogeneous conclusion [5]. Most studies on the impact of the HSR have focused on the regional economy and transport accessibility of the cities along the HSR line.

Garmendia et al. [7] reviewed the current situation and future challenges of the HSR in Europe in terms of inter-city relationships, wider spatial implications, and the role of HSR stations. Urena et al. [8] developed a multilevel analysis (national, regional, and local levels) focusing on big intermediate cities along HSR lines: Cordoba and Zaragoza in Spain and Lille in France. Brocker et al. [9] developed a spatial computable general equilibrium model to evaluate the spatial distribution of user benefits. They applied the model to 22 infrastructure projects of the TEN-T priority list and found that 12 projects have a yearly rate of return of above 5% for the EU, while the remaining projects may be unprofitable. Feiyang and Yuri [10] used panel survey data in China to investigate the direct and spillover impact on urban and rural regions. Zhenhua [11] studied the impact of HSR investment with respect to the economy and environment in China and found that the HSR stimulated the regional economy, and HSRs have had a significant positive impact on CO<sub>2</sub> emissions. Takeshi and Lingling [12] studied the impact of the Shinkansen network extension on tourism development using survey data on tourism demand and tourist behavior. Garmendia et al. [13] analyzed the different territorial performances of a national highway and an HSR line. They conducted a mobility survey and compared the modal split, travel frequencies, and travel purposes to understand long-distance mobility patterns.

Although most previous studies found that the HSR had a positive effect on the regional economic growth, some researchers argued that there is a differential impact on the local economy, a negative effect on the local economic growth, and/or a mixed result [14,15]. For example, Gutierrez et al. [14] estimated the future impact of the proposed European high-speed train network using the spatial distribution of the accessibility indicator based on a weighted average of the travel times, taking the gross domestic product (GDP) of the centers as the weights. They found that the proposed high-speed train network could increase the imbalances between the main cities and the nearby regions. Kim et al. [15] analyzed the negative impact of the Gyeongbu KTX on regional economies, termed the “straw effect.” Straw effect refers to a negative economic impact, wherein local human resources and material goods are concentrated near metropolitan areas, whereas infrastructure construction negatively affects the local economy. They compared the impact of the KTX in 16 cities and argued that the introduction of the KTX negatively affected the regional economic growth. However, there is still no consensus regarding the existence of straw effects. For example, Hur [16] investigated whether the KTX contributes to increasing population concentration in the Seoul metropolitan area, or whether the KTX leads to a depression in the market area in non-metropolitan areas. However, no concrete evidence was found regarding the straw effects of the introduction of the KTX. Jo and Woo [17] empirically analyzed the effect of the KTX on the regional economy and investigated the existence of straw effects. They found mixed results; an unbalanced regional economic growth was observed in terms of the gross regional domestic product (GRDP) and population distribution, such as the straw effect of the KTX, whereas balanced regional economic effects were observed in terms of employment. Cho et al. [18] conducted a passenger survey to identify whether the first introduced Korean high-speed train (KTX) affects the local shopping markets. They concluded that there is no negative impact of KTX on the local shopping activities at least in the first year of its introduction, although the respondents in local areas expected more visitors from the Seoul metropolitan area rather than those from local cities because of the reduced travel times. Park and Kim [19] investigated the effect of Gyeongbu KTX on the regional economy. They focused on the department store and regional medical records to calculate the economic influence of the KTX. The results showed that the effect of KTX on the local department store market is positive, except in the Daegu metropolitan area. The regional medical records are statistically significant only in the Busan metropolitan area.

Several studies have been proposed to analyze the increased accessibility via the HSR. The most common indicator of the accessibility is an aggregate spatial interaction model, which is combined with the transport impedance and locational attractiveness that is based on the Newton’s law of gravitational force [20]. Gutierrez [21] evaluated the accessibility impact of the future Madrid-Barcelona-French border HSR. Three different accessibility indicators were employed: weighted average travel time (location indicator), economic potential (a gravity-based measure), and daily accessibility.

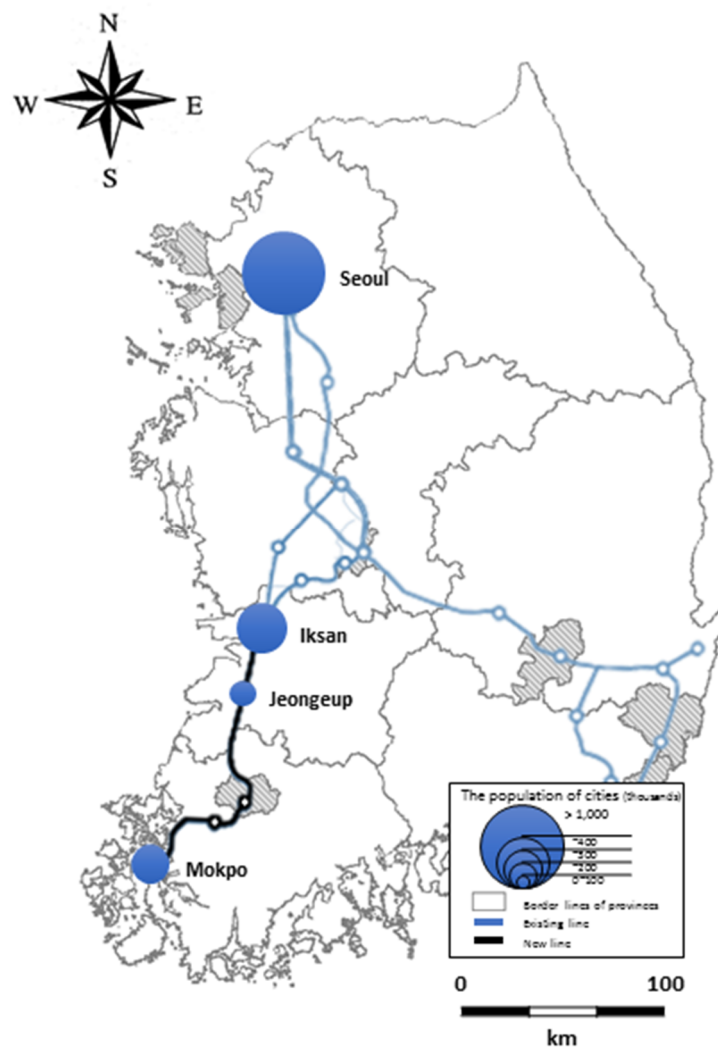
Chang and Lee [22] highlighted the importance of improving the accessibility of the Korea HSR (KTX) stations to increase the ridership. They performed an accessibility analysis of the KTX focusing on the Seoul metropolitan area. Using a Hansen-type accessibility measure and survey data, they identified areas with varying degrees of KTX impact boundary and zonal accessibility. Monzon et al. [23] proposed an assessment methodology to analyze the combined effect of the efficiency and equity of the accessibility improvements via a HSR project. Wu et al. [24] evaluated the effects of the rail transport network in China on the local accessibility from 2006 to 2014 using the accessibility-based market potential methodology. They found that new intercity railway investments strongly affect the accessibility levels in both the core regions and periphery regions. They also empirically found that the rail network expansions could have a differential impact on the accessibility dynamics between the core and periphery regions. Shih-Lung [25] studied the overall accessibility changes with respect to the travel time, travel cost, and distance accessibility by taking a timetable-based accessibility evaluation approach for each of the four main stages of the HSR development in China.

Previous studies have relied on basic statistics provided by the observed passenger ridership and on-site interview survey data for HSR passengers. In addition, socioeconomic data for the HSR impact analysis were widely used and partially implemented to a certain purpose of travel at a city level. Unlike conventional studies on the impacts of HSR, this research focuses on the microscopic and massive spatial unit, 50 m X 50 m, of mobile phone data, which provide user home locations and local locations of phone users. Every location where phone users drop by was recorded as a snapshot representing the number of users in a day. Interestingly, we can extract the data recorded at the HSR station and analyze the constitution of gender and age group in terms of their home location and spatial changes in the hot-spot locations. Therefore, this study addresses the potential use of mobile phone data to analyze the effect of HSR on the cities along the HSR line.

### 3. Case Study Area

Figure 1 shows the HSR routes of both existing and newly opened Honam HSR lines (bold) in April 2015. Several cities had an HSR line that directly connected to the capital city, Seoul. We focus on three cities—Iksan, Jeongeup, and Mokpo—for evaluating the impact of the HSR since these cities are expected to directly benefit from the HSR. Among the three cities, Mokpo is expected to see regional economic growth as the long travel time to Seoul will decrease because it is the last train station in the HSR line (330 km in length). The dot size represents the population of each city—Iksan (320,000), Jeongeup (130,000), and Mokpo (250,000). Since there were no special changes on transport systems and events that cause irregular travel demand within a one-year analysis period, we would like to see empirical evidence of the impact of the HSR via mobile phone data, one year before and after the opening of the HSR.

Table 1 presents before and after conditions of the Honam HSR. Before the opening of the Honam HSR line, there were two segments: a high-speed segment (300 km/h between Seoul and Daejeon) and a regular-speed segment (150 km/h between Daejeon to Mokpo) since there was no infrastructure in place for the Honam line. According to the statistics from the Korea Railroad Corporation in 2006, the New Honam line decreased the travel time by 53 min from Seoul to Mokpo and increased the passenger volume by 281(million/year), a 42% increase, although the fare was increased by approximately 30%.



**Figure 1.** Honam high-speed railway (HSR) line and cities for a case study in Korea.

**Table 1.** Comparison Before and After the Introduction of the Honam HSR line.

Conditions	Existing Line	New Line
Route Info.	Seoul-(Daejeon*)-Iksan-Jeongeup-Mokpo	Seoul-Iksan-Jeongeup-(Kwangju*)-Mokpo
Distance (km)	352	330
Travel Time (min)	202	149 (−26%)
Operation speed (km/h)	(Seoul)-(Daejeon): 300, Iksan-Mokpo:150	Seoul-Iksan-Jeongeup - (Kwangju)-Mokpo: 300
Fare (\$, one-way)	\$42 (Seoul-Mokpo)	\$50 (Seoul-Mokpo)
No. of passengers (million/year)	669	950(+42%)

Note: (\*) represents a city not analyzed in this study.

#### 4. Mobile Phone Data

In 2015, there were about 57 million mobile phone users in Korea, which is equal to 1.13 mobile phones per person on average. SKT, one of three wireless telecommunications operators in Korea, has held about half of the country's total mobile phone accounts. Therefore, SKT expanded mobile phone data to include the total users in Korea, using the country's market share rate. In this study, we used SKT's complete mobile phone dataset, representing the total population of mobile phone users.

The mobile phone data include information about the number of mobile users as well as locations of these users, including their home location and existing location of the phone user. The existing

location of the phone user is pointed on a grid cell of 50 × 50 m, called a P-cell. Table 2 presents information on mobile phone data. The data are the daily mobile phone records, grouped by age and gender and consist of 16 columns, including dates; locations (X- and Y- coordinates) in a 50 × 50 cell. In a P-cell, a user exists at any time of the day; the number of mobile phone users that are classified by six age groups by gender; and the home location of the user. It is a daily-based record and double counts are not allowed in the same cell and in the same day. If he/she exists in the same cell in the same day even at different times in a day, he/she is counted as one. It should be noted that the number of records is not equal to the mobile population. Each record is constructed based on a combination of two locations, including a 50-m grid cell and a home location; that is, one record represents the number of people, who are in the same 50-m grid cell in each hour, as well as the same home location. For instance, even if two mobile phone users exist at the same grid cell in the same hour of the day, their records appear in separate rows if their home locations are different.

As the Honam HSR opened in April 2, 2015, we use the mobile phone records of one week—from March 16 (Monday) to March 22 (Sunday) in 2015 and from March 14 (Monday) to March 20 (Sunday) in 2016.

**Table 2.** Mobile Phone Data Structure.

Information	Description
Date	Date of data -2015: March 16–22 (7 days): before period
X	-2016: March 14–20 (7 days): after period Position of the center of a cell based on the UTM-K coordinate system
Y	
Number of men by age groups	Daily mobile phone users (men) by 6 age groups (10 s, 20 s, 30 s, 40 s, 50 s, over 60 s)
Number of women by age groups	Daily mobile phone users (men) by 6 age groups (10 s, 20 s, 30 s, 40 s, 50 s, over 60 s)
Home	Code of home location of users

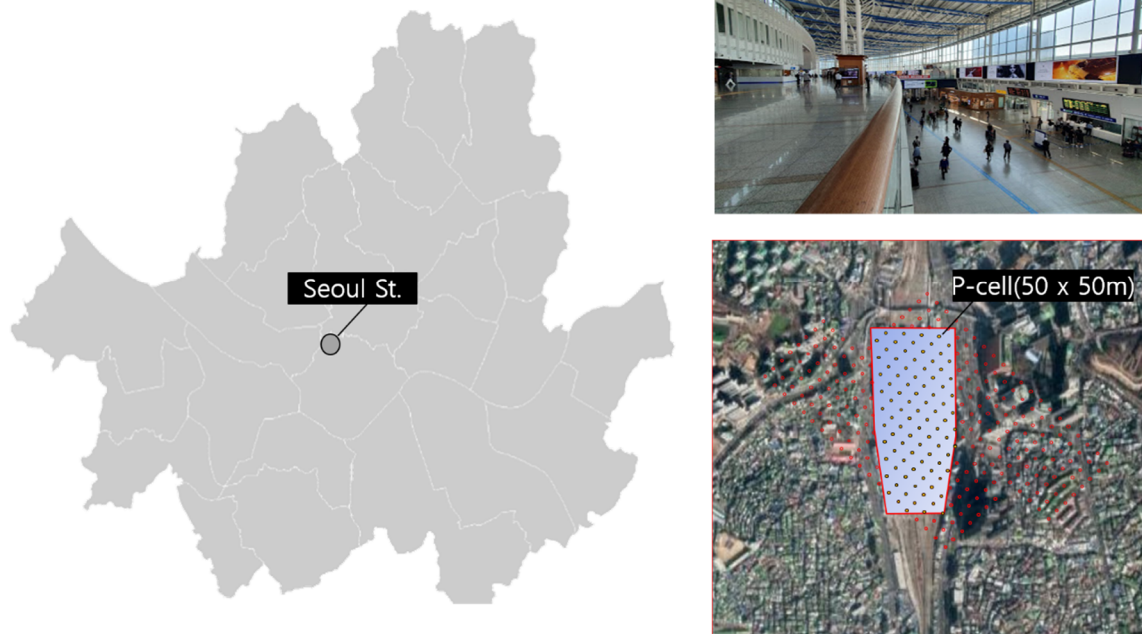
In order to analyze the diverse geographical impact of the HSR, we classify the mobile phone data into two levels. One is the station level, wherein mobile phone data are observed at a station area used for analyzing the changes in the volume of mobile population and access distance from an individual user's home to the HSR station. The other level is the city level, wherein the change in hotspots after the opening of the HSR is investigated; their compactness at the station periphery is also investigated. For this analysis, we aggregated the mobile population at a P-cell to those at a station (as shown in Figure 2) and city boundary using ArcGIS. Then, we averaged the daily data by weekday and weekend. Table 3 shows that the final extracted numbers of mobile phone records differ from each station.

**Table 3.** Number of records of mobile phone data (averaged).

Station	Population (Person)	Before		After	
		Weekday	Weekend	Weekday	Weekend
Seoul	10,125,000	181,869	134,471	188,744	125,252
Iksan	307,000	21,724	24,672	21,637	26,014
Jeongeup	114,370	10,343	10,977	9714	10,687
Mokpo	235,516	11,383	14,931	13,042	19,334



## Seoul



**Figure 2.** Example of the selected mobile phone centroid of p-cell data within Seoul St.

## 5. Analysis and Results

In the analysis, we first compared the volume changes of mobile population recorded at HSR stations before and after their installation according to gender and age group for both weekday and weekend, and then performed a statistical analysis by using non-parametric statistics for evaluating statistical differences of observed mobile population before and after HSR opening. Second, we calculate the access distance between a phone user's home and each HSR station and developed distance-density graphs. Then, we compared the changes of the geographical distribution of the access locations. Finally, we developed hotspot maps of four cities along with the volume level of mobile population to investigate whether the HSR changes at the hotspots in these cities. In addition, we compared changes in the number of hotspots and the compactness of hotspots at the periphery of an HSR station.

## 6. Comparison of Basic Statistics

In this section, we explain the outcomes obtained from the exploratory analysis in Table 4, which focuses on the impact of the HSR on mobile population growth at individual stations. The table shows the differences in mobile population recorded at each HSR station according to gender and age group. For example, Jeongeup city has a total of 172,597 (weekday average) mobile population records and the data observed at station is 13,117 which is about 7.6% of city total. Of 13,117, the number of records for age group of 60 is 2476 (18.9%) at before the opening of HSR. Moreover, the total records of city increase as 178,120 (+3.2%), but the number of over 60 is decreased as 1603 after the opening of HSR. The number differences are only presented in Table 4.

Overall, the mobile population increased at all four cities during the weekdays, and the growth is shown to increase with the travel distance from Seoul. However, the mobile population and growth on weekends at Seoul are seen to decrease because of the decrease in the number of people over the age of 40.

Interestingly, the HSR line was found to attract more people, especially the older people, to local cities than to Seoul. We would expect that the HSR stations at local cities form new hotspots due to the development of multi-functional facilities at the station periphery.

Mokpo, the outermost city from Seoul, has seen an increased mobile population growth for all age groups, with an almost 50% growth during the weekdays and over 60% during the weekends. This result shows that we can expect the people in Mokpo city to largely benefit from HSR installation.

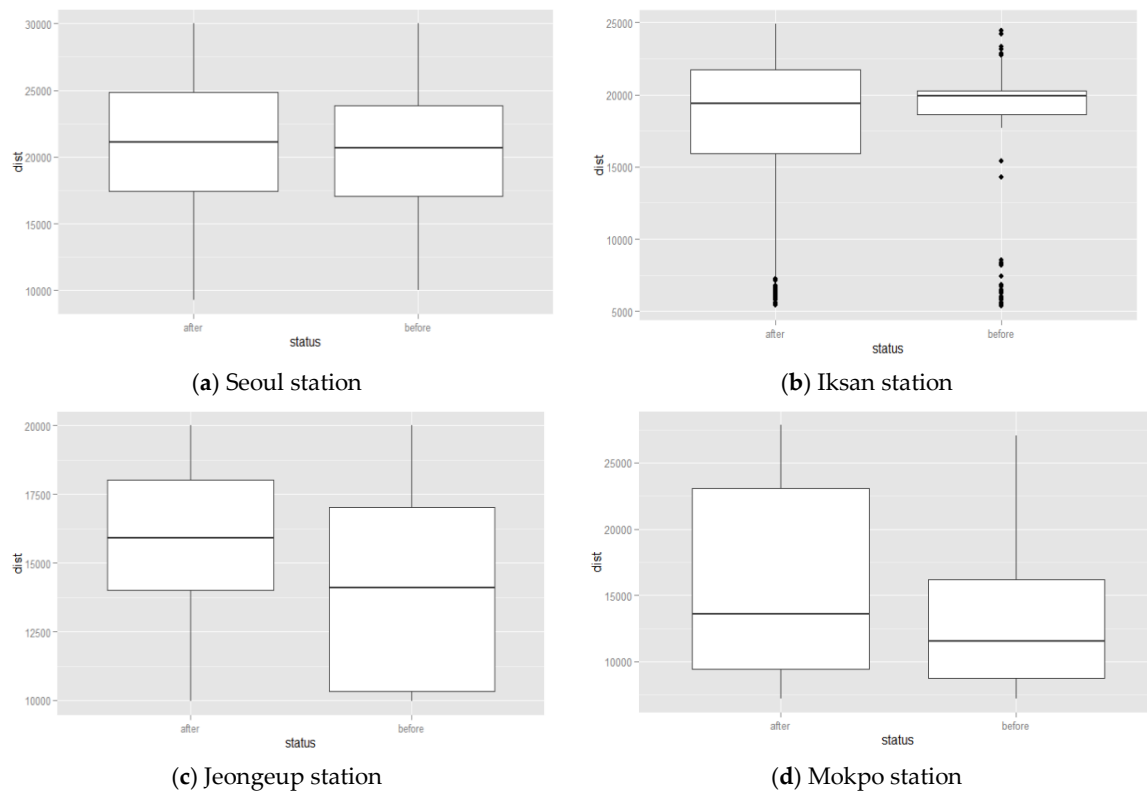
Figure 3 shows the box-plot of access distance range of mobile population. The box-plot represents the ranges of access distance to reach HSR station. We expect that if someone benefits with a short travelling time by using the HSR, then he/she would access an HSR station even if he/she resided far away from the HSR station. Therefore, Iksan and Mokpo stations were determined to have wider spectra of access distance after the launch of the HSR as the lower-and-upper quartile ranged wider compared to before condition. This illustrates that HSR attracts more people from a distance of approximately 10–15 km to 10–22 km at both stations. Jeongeup station shows an increase in the median access distance, although the access distance range is slightly decreased. Thus, the access distance range may vary depending on the city size, socio-demographic characteristics, and income level, etc. These results show that the HSR attracted more people living in widely from the area near the station, while the range of lower-and-upper quantiles varies at stations. This is especially true for Jeongeup, which is the smallest city along the HSR line with a population of approximately 114,000, consisting of over 30% of the age group of 60. However, it is difficult to determine the reason of decrease in the access distance at Jeongeup station. In contrast, Seoul station shows similar lower-and-upper quantile range of access distance although the median distance is barely increased. Therefore, we understand that the opening of the HSR had a relatively small effect on the mega city in terms of access distance range.

**Table 4.** Differences of mobile population before and after the opening of HSR at each station.

Category	Weekday				Weekend			
	Seoul	Jeongeup	Iksan	Mokpo	Seoul	Jeongeup	Iksan	Mokpo
Man (10+)	1957 (14.8%)	243 (27.3%)	1359 (40.2%)	911 (82.5%)	3310 (21.1%)	300 (21.3%)	1063 (26.7%)	1613 (88.5%)
Man (20+)	6185 (11.7%)	430 (33.2%)	1369 (32.3%)	1103 (51.7%)	5559 (10%)	469 (28.8%)	1388 (23.8%)	1685 (65.1%)
Man (30+)	13,474 (11.0%)	455 (13.7%)	1171 (13.2%)	2060 (49.3%)	713 (1.0%)	429 (15.0%)	1047 (12.3%)	2804 (64.0%)
Man (40+)	11,511 (8.3%)	904 (16.8%)	2924 (26.3%)	2864 (55.6%)	−2834 (−4.0%)	1102 (24.6%)	2750 (28.9%)	3772 (75.3%)
Man (50+)	−6018 (−5.5%)	1312 (26.3%)	3531 (34.8%)	3136 (57.0%)	−14,258 (−18.5%)	1497 (35.2%)	3445 (37.5%)	3990 (69.8%)
Man (60+)	−43,196 (−45.3%)	−364 (−10.1%)	142 (1.8%)	1162 (22.3%)	−37,368 (−49.6%)	−196 (−6.3%)	617 (8.4%)	1799 (35.0%)
Woman (10+)	4078 (30.9%)	319 (30.0%)	1412 (35.9%)	684 (28.6%)	3417 (22.5%)	208 (14.3%)	759 (19.2%)	1151 (48.6%)
Woman (20+)	11,102 (18.5%)	288 (23.9%)	1293 (29.0%)	931 (47.7%)	10,427 (18.9%)	461 (29.0%)	1378 (24.4%)	1603 (65.7%)
Woman (30+)	11,041 (13.5%)	384 (16.5%)	1337 (17.1%)	1647 (49.3%)	4357 (7.2%)	443 (18.5%)	1303 (17.4%)	2473 (69.2%)
Woman (40+)	12,196 (16.7%)	943 (29.0%)	2547 (31.4%)	2148 (56.1%)	5970 (10.8%)	1113 (36.3%)	2362 (30.8%)	3025 (77.1%)
Woman (50+)	9809 (16.2%)	1116 (37.8%)	2754 (32.9%)	3154 (77.6%)	6221 (11.8%)	1161 (40.3%)	2813 (34.8%)	4134 (94.6%)
Woman (60+)	−3897 (−9.1%)	−509 (−19.6%)	−382 (−5.6%)	807 (19.2%)	−2437 (−6.9%)	−222 (−9.5%)	9 (0.1%)	1375 (32.8%)
Total	28,242 (3.3%)	5523 (16.8%)	19,457 (22.8%)	20,607 (47.8%)	−16,922 (−2.6%)	6765 (21.5%)	18,934 (22.7%)	29,425 (64.6%)

Note: the number represents difference of the volume of mobile population before and after the opening of HSR (e.g., value = after volume-before volume). () represents % of changes of the volume of mobile population before and after the opening of HSR.





**Figure 3.** Comparison of mobile population volume by access distance (km) from each station before and after the opening of HSR (Note: left-side of  $y$ -axis: after, right-side: before).

In order to evaluate whether the changes of the mobile population before and after the opening of the HSR are statistically meaningful, we employed the nonparametric Wilcoxon signed ranks test. The paired  $t$ -test compares the mean of the two populations by using repeated measures, in which it is assumed that the data are measured on an interval or a ratio scale, and are normally distributed. However, the mobile phone data do not support these assumptions. The Wilcoxon signed ranks test requires that the differences are approximately symmetric and that the data are measured on an ordinal, interval, or ratio scale. As the assumptions for the Wilcoxon signed ranks test are met, unlike those of the  $t$ -test, which are violated, the Wilcoxon signed ranks test is said to be usually more powerful in detecting a difference between the two populations [26]. Even under conditions appropriate to the paired  $t$ -test, the Wilcoxon signed ranks test is almost as powerful and uses the test statistic  $W$ , as shown in Equation (1).

$$W = \sum_{i=1}^n R_i^{(+)} \quad (1)$$

where  $n$ : sample size (nonzero absolute difference);  $R$ : assign ranks  $R_i^{(+)}$  from 1 to  $n$  (based on the absolute difference score  $|D_i|$ , between the two paired values (e.g.,  $P$ -cell values observed before & after opening of the HSR);  $W$ : sum of the positive ranks.

Table 5 shows the result of Wilcoxon signed ranks test. All the  $p$ -values of the signed rank test are seen to be statistically significant at a 95% confidence level. Overall, all stations show a positive  $W$  value, implying that the mobile population increases, as expected, after the opening of the HSR. Without considering the gender and age classification, the opening of the HSR positively affects the increase in passenger population.

**Table 5.** Wilcoxon’s signed-rank test(W(T)) of differences between mobile population before and after the HSR on weekdays.

Station		Iksan				Jeongeup				Mokpo			
Group		mean	Std.	W(T)	p-Value	Mean	Std.	W(T)	p-Value	Mean	Std.	W(T)	p-Value
Age (difference)	Male (10 s)	0.52	6.99	14,4026	<0.0001	0.21	5.04	56,024	<0.0001	0.83	8.91	116,083.5	<0.0001
	Male (20 s)	0.47	6.75	128,913.5	<0.0001	0.43	5.68	68,322	<0.0001	0.90	6.37	95,911.5	<0.0001
	Male (30 s)	−1.00	20.55	66,120.5	<0.0001	0.25	10.80	46,944	<0.0001	1.64	11.84	105,441	<0.0001
	Male (40 s)	−0.60	29.32	126,105.5	<0.0001	0.55	22.60	64,043	<0.0001	2.69	17.42	142,288	<0.0001
	Male (50 s)	0.35	27.59	150,861	<0.0001	0.79	29.78	89,578.5	<0.0001	1.70	18.46	106,674.5	<0.0001
	Male (60 s+)	−2.31	24.28	−42,512	<0.0001	−0.72	23.54	−59,502.5	<0.0001	−0.82	18.14	20,536	<0.0099
	Female (10 s)	0.17	11.71	149,036	<0.0001	0.29	5.70	70,785.5	<0.0001	1.10	1.10	84,894	<0.0001
	Female (20 s)	−0.04	9.28	102,034	<0.0001	0.20	5.56	46,194.5	<0.0001	0.74	0.74	107,051.5	<0.0001
	Female (30 s)	−0.99	20.87	55,393.5	<0.0001	0.28	10.93	41,680.5	<0.0001	1.32	1.32	111,349.5	<0.0001
	Female (40 s)	−0.32	22.25	124,325	<0.0001	0.70	19.21	87,972.5	<0.0001	2.09	2.09	137,593.5	<0.0001
	Female (50 s)	−0.57	22.77	109,914.5	<0.0001	0.94	19.50	93,954	<0.0001	2.22	2.22	134,428	<0.0001
	Female (60 s+)	−2.56	18.87	−105,326	<0.0001	−0.71	12.37	−83,217.5	<0.0001	−1.15	14.15	−44,984	<0.0001
Total gender (difference)	Male	−2.58	108.40	114,328	<0.0001	1.50	92.31	55,278.5	<0.0001	6.94	75.85	110,134.5	<0.0001
	Female	−4.31	99.40	81,155	<0.0001	1.69	71.23	48,944.5	<0.0001	6.31	80.40	96,197	<0.0001
Total (difference)	All pop	−6.88	206.29	103,147	<0.0001	3.19	160.04	54,131	<0.0001	13.25	155.20	105,829.5	<0.0001

However, interestingly, the statistical result shows a different impact based on age and gender group. That is, the female group of over 60 shows negative W-statistics at all three cities, Iksan, Jeongeup, and Mokpo, while the male group of over 60 only shows negative W statistic in Iksan and Jeongeup. Unlike other cities, Mokpo shows a positive W-statistic for the male group of over 60 years because of the shorter travelling time of over an hour. Thus, more people prefer to use the HSR to save the travel time, which is worthier than saving the fare.

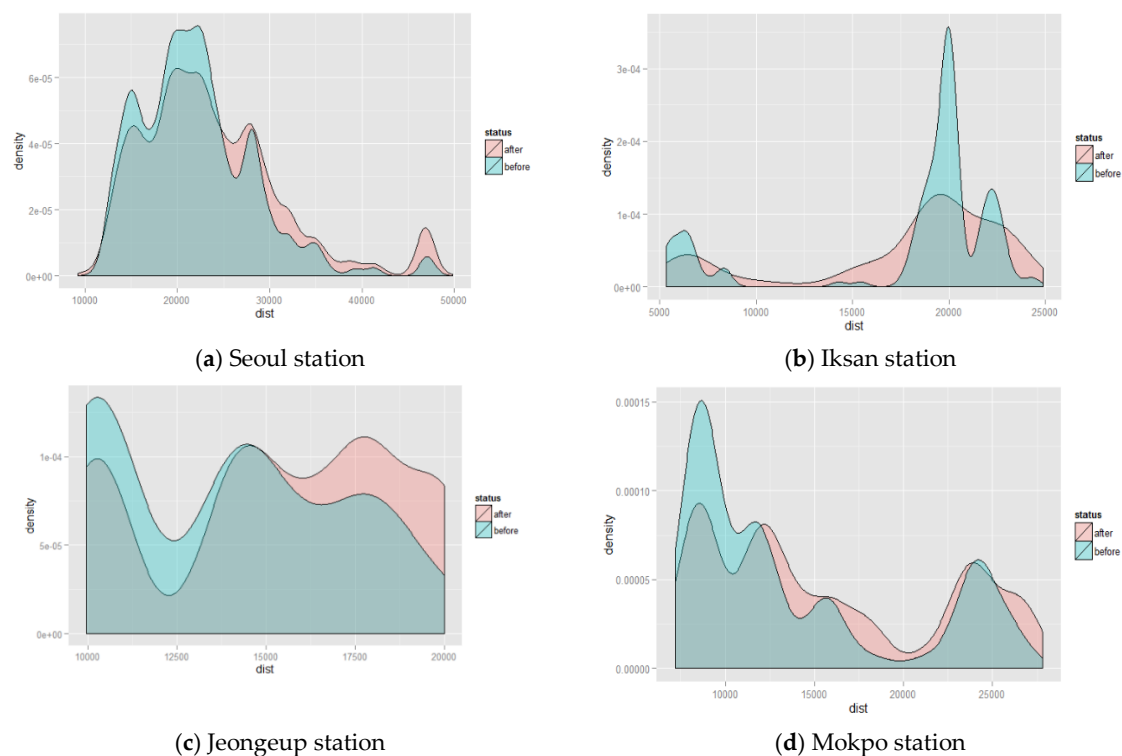
This illustrates that the HSR did not attract more people over 60 s, except at Mokpo. One of the main reasons would be the increase in the travel fare by approximately 30%. The older might choose a cheaper mode than the HSR as they are not likely to pay more money to save travel time owing to their life style, which comprises surplus time. However, from the W-statistic of Mokpo, HSR attracted more people because of travel time reduction by an hour.

## 7. Access Distance from Home Location to Station

To evaluate the impact of access distance from the HSR, we calculate Euclidian distance between the HSR station and origin (Home), and develop density figures based on the mobile phone data, which only includes origins inside the city boundary. The density of mobile population against the Euclidian distance is calculated according to the location of each HSR station and individual location of mobile phone data observed. The density is denoted as seen in Equation (2).

$$D_i(p/m^2) = \sum p/d_i \quad (2)$$

where,  $D_i$  is the density of mobile population at distance (i) from the station,  $p$  is the observed mobile population within cells of  $50 \times 50$  m, and  $d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2}$ , which is the Euclidian distance between the HSR station and individual origin location of the mobile user. The x-y location of each station is denoted as  $(x_s, y_s)$ , and each origin location of the p-cell  $(x_i, y_i)$  is used to calculate the distance from the HSR station. The density distribution of the number of observed instances of mobile data with respect to distance at each HSR station is shown in Figure 4.



**Figure 4.** Comparison of density of access distance before and after the opening of the HSR.

From the figure, we can see that most of the access trips in Seoul were made within 50 km, and majority of trips ranged between 20 km and 25 km from the station. Although the density is slightly decreased within a radius of 25 km from the station, Seoul is megacity, and various transit modes are available to access an HSR station. Therefore, the HSR seems to cover a wider area in Seoul than in other cities. However, the density distribution is similar to that of the before condition, which means that HSR had little effect on the access conditions, because the station for Seoul–Busan HSR line opened in 2004, and is shared with the Honam HSR line.

Interestingly, it is found that the access distance distribution varies depending on city size (small or medium). Iksan station shows that the access population is observed in wider area after HSR, as major access trips were generated approximately 15 km–25 km away from the station. This phenomenon can be explained by the shape of the city center and location of Iksan station. Iksan station is located at the southern section of the city boundary, so that people living in the outer side of the city center can gain access to the station (refer to Figure 4b). Another interesting point is that the density of mobile population observed approximately 20 km away from the station was the highest at before the HSR opened, and it plateaued in the after condition. Although the reason cannot be clearly explained, the expensive fare of the HSR would make people choose cheaper modes of travel, such as express buses and the existing railway.

Jeongeup city, as a small city, also shows that the HSR attracted more people living farther from station, compared to access distance in the before condition. However, the density of access distance within 15 km was shrunk after the opening of the HSR, which would be caused by the decreased number of older people (over 60 years old).

Mokpo, as a medium city, presents high access density close to the HSR station, mainly within 15 km. From the results, it can be seen that the distribution of access distance from the station is affected by the geographical structure of cities, socio-demographic conditions, and transport systems. We also found that as access distance increases, the density also increases. Therefore, the HSR clearly provides an opportunity for travelers to use high-quality transport services in a wider area.

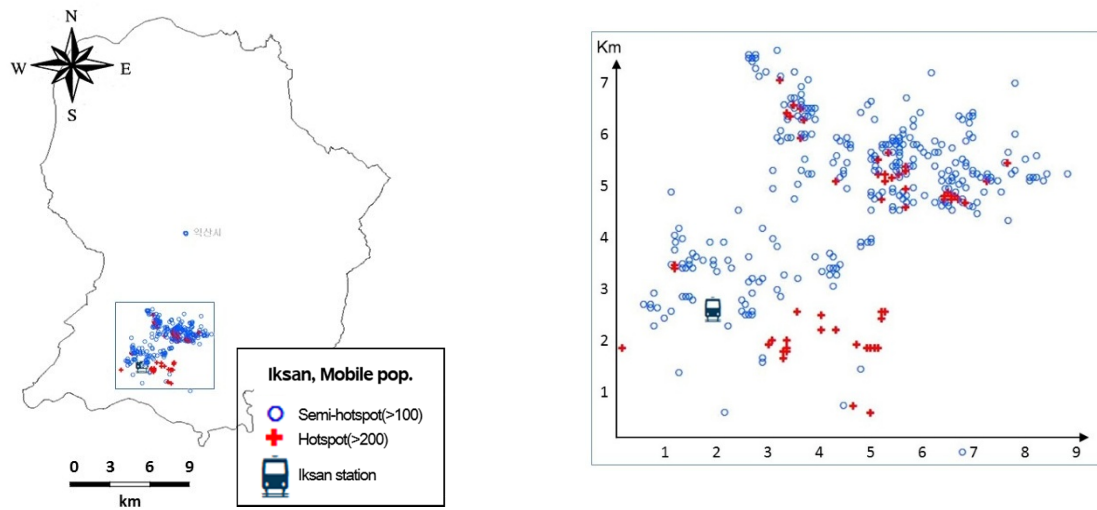
## 8. Hot Spots Around Station

Hotspots are locations where large number of people converge, for undertaking activities such as work, shopping, recreation, etc. Various methods and definitions have been proposed to identify activity or employment sub-centers by McDonald [27], McMillen [28], McMillen and Smith [29], Redfearn [30], and Redfearn [31]. In this study, using mobile phone data, based on a 50 m × 50 m P-cell in each city level allows us to identify hotspots to consider the density of mobile population spread over spaces.

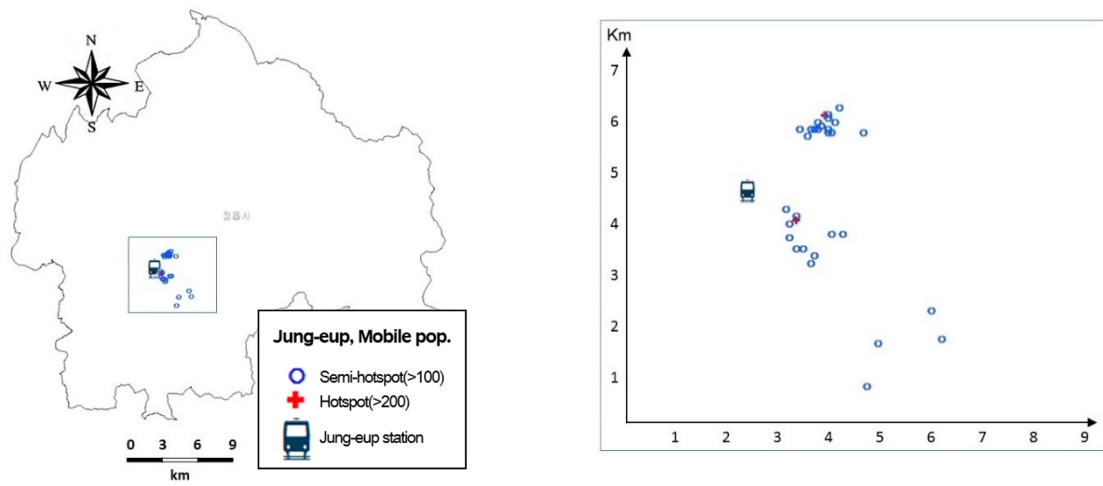
Figure 5 shows the hotspots changes caused by HSR in the three cities. We developed a map to identify the hotspot locations, showing the difference of mobile population before and after the opening of the HSR. In the figure, we define the hotspot and semi-hotspot with the increased mobile population over 200 and 100, respectively. As shown in Figure 5a, Iksan city has a population of approximately 300,000 people, with an area of 506 km<sup>2</sup>. There are many spots that changed to semi-hotspots and hotspots due to the HSR, and the hotspots are shown to be appear within 10 km of Iksan station.

Although the hotspots are spread out over the city boundary, more hotspots are located at the HSR station periphery, within 5 km. The appearance of many hotspots provides general evidence of the positive impact of the HSR, which attracts more people at Iksan city and the station catchment area.

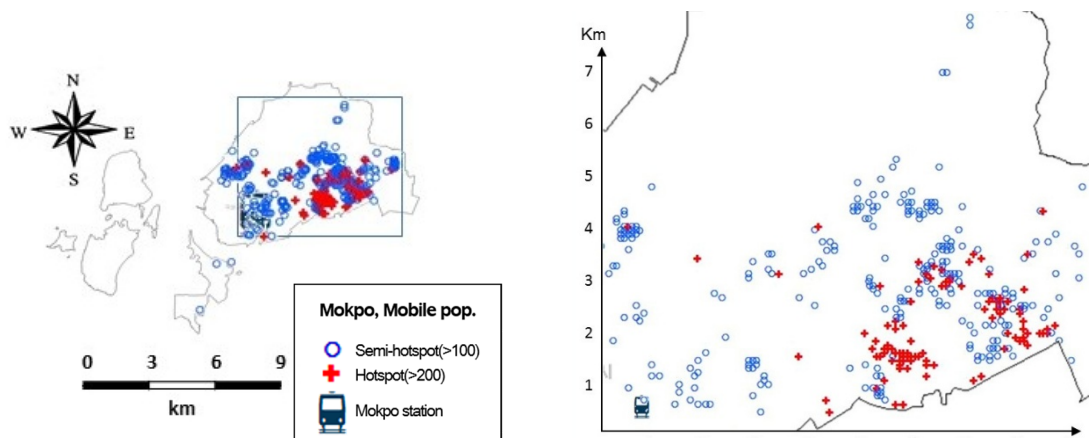
Jeongeup has a small population of 114,000 people, with large area of 700 km<sup>2</sup>. The area comprises small farms and mountains. Therefore, no significant difference in the hotspots can be seen before and after the opening of the HSR. However, the number of hotspots is slightly increased, and they appear in the northeast and southeast region from Jeongeup station, because the agricultural and industrial complexes are located northeast, and the city hall is located southeast from the station.



(a) Hotspots in Iksan city



(b) Hotspots in Jeongeup city



(c) Hotspots in Mokpo city

Figure 5. Maps of hotspot changes after the opening of the HSR. (Note: x and y-axis unit is km).



Mokpo has many hotspots after the opening of the HSR, which spread out farther from the station due to the impact of the HSR. Interestingly, it is found that the hotspots at the intercity bus terminal periphery decreased, which means that a transport mode shift occurred, from intercity bus to HSR. By comparing the number of hot spots in the three cities, we expect that the attractive places such as shopping centers or convention centers may have been built station periphery since the opening of HSR. This may explain the low number of hot spots in Jeongeup and the high number of hotspots in Iksan and Mokpo.

According to ticket sale statistics, bus passengers diminished by as much as 20% after the opening of the HSR. This was reflected by the hotspots at the intercity bus terminal located northeast from Mokpo station, as seen in Figure 5c, with a negative value of difference (denoted as “o”).

It is also interesting to examine the compactness of hotspots from the HSR station, revealing how close their spatial distributions are to the HSR station. The compactness of hotspots can also be explained by how land use is formed and distributed around the station. If hotspots in a city are located spatially close to the HSR station, it would indicate that the hotspots are seen to converge at the station. The spatial compactness of hotspots,  $C_i$ , is calculated by Equation (3) in this study [32].

$$C_i = \frac{H_{hotspot}}{N_i \sqrt{A_i}} \times 100 \quad (3)$$

where,  $H_{hotspot}$  is the average distance between station,  $N_i$  is the number of hotspots (mobile population greater than 50) in a city  $i$ , and  $A_i$  is the area of city  $i$ . Therefore,  $C_i$  is the compactness, the average distance between the HSR station and hotspots, weighted by the number of hotspots and area of the city. A lower  $C_i$  means that the spatial structure of hotspots is compact around the station; that is, they are spatially close to station. Table 6 illustrates the changes in the number of hotspots and compactness before and after the opening of the HSR. We observed that all cities have more hotspots and high compactness after the opening of the HSR.

**Table 6.** Number of hotspots and compactness in the HSR station catchment area.

City	Before		After	
	Number of Hotspots (pop. > 50)	Compactness	Number of Hotspots	Compactness
Iksan	127	0.186	224 ( $\Delta 76.4\%$ )	0.101 ( $\Delta 45.5\%$ )
Jeongeup	6	0.987	9 ( $\Delta 50.0\%$ )	0.831 ( $\Delta 15.8\%$ )
Mokpo	119	0.762	229 ( $\Delta 92.4\%$ )	0.379 ( $\Delta 50.2\%$ )

Iksan station has very low  $C_i$  value, 0.186 (very high compactness of hotspots from the HSR station), and the compactness is even higher, at approximately 45% after the opening of the HSR. Considering the large number of hotspots in the two cities—Iksan and Mokpo—the very high compactness indicates that it is very convenient for undertaking some activities in the station catchment area.

Mokpo station has the highest enhancement of compactness (50%) after the opening of the HSR. However, hotspots are not as close to the Jeongeup station periphery as other stations, although the compactness is slightly higher (15.8%) after the opening of the HSR.

## 9. Conclusions

The large amount of data collected from mobile communications provides the spatial-temporal locations of phone users, which acts as a source for investigating urban issues related to mobility behavior and the spatial-temporal demographic dispersion of a city. This study examined the potential use of mobile phone data for empirical analysis of the impact of HSR in terms of mobile population changes by gender and age group, travel distance, density of access distance, location of hotspot changes, and compactness of hotspots from the HSR station. In order to accomplish this, we choose

a case study of the Honam HSR line, introduced on April 2, 2015, which directly connects the capital city, Seoul, to several cities, namely, Iksan, Jeongeup, and Mokpo, located 200–300 km south of Seoul. We use the mobile phone records of one week of data, from March 16 (Monday) to March 22 (Sunday), 2015; and from March 14 (Monday) to March 20 (Sunday), 2016, for a before-and-after study.

From the basic statistical analysis, we found that the mobile population is seen to increase in Seoul and all three cities during weekdays. All age groups show an increased mobile population; however, those over 60 years old showed a decrease. Older people usually make a relatively short trip in Korea, and use other intercity modes of travel, as they are sensitive to the increased fare of HSR and have a surplus of time. Among three cities, Mokpo, which is 300 km away from Seoul, only has a positive value of the Wilcoxon signed ranks test statistic  $W$  for males over 60, and the value of test statistic  $W$  for women over their 60 s is negative. However, the magnitude of the value is smaller than that of other cities. Therefore, we found that the longer the travel distance, the higher the number of people who would like to take the HSR, no matter what age group.

By developing the map of mobile population changes with respect to travel distance, we can easily evaluate whether the mobile population increases due to the HSR. From the results, we found that the mobile population within 200 km of Seoul is not shown to increase, because the operation speed of the HSR line is almost the same as that of the existing line connecting Seoul to Daejeon, close to Iksan city, approximately 200 km away from Seoul. This illustrates that a resident within 200 km of Seoul did not change their travel pattern due to the HSR in Korea, because the travel time saved is not enough to warrant a change in transport mode. The HSR fare went up by approximately 30% and hindered the attraction of the mobile population. However, more people from Jeongeup and Mokpo are observed in Seoul, which is seen to be a “straw effect”. However, we cannot determine whether this effect will continue. Unlike Seoul, Iksan and Jeongeup have an increased mobile population within 150 km, which decreased at over 200 km and 150 km–200 km, respectively. Iksan did not attract more people from the northern part of Seoul, Gyeonggi-do, while Jeongeup attracted less people from Seoul after HSR. The reason for this cannot be explained, but the attraction may vary depending on the size of city and business structure. However, Mokpo shows increased mobile population at all distance levels, which indicates that the HSR line had a positive influence on the relatively short-distance travel between the neighboring cities, and long-distance travel as well. The increased population were expected to shift from bus, railway, and automobile modes of transport.

With respect to the density of access distance to the HSR station, the origins (home) of mobile population are within 40 km, and a majority of them ranged between 10–25 km from Seoul station; however, the distribution of density before and after is similar. Interestingly, it is found that the distribution of access distance at Iksan station mainly ranges between 20–25 km, and it is smoother than before distribution, which is different from Jeongeup and Mokpo, where the distribution mainly ranged within 20 km. Jeongeup station is shown to expand the access distance to 15–20 km after the opening of the HSR.

We also develop a map of activity spots, and analyze the changes in the number of hotspots in three local cities, and their compactness from the station before and after the opening of the HSR. Among the three cities, Mokpo station has the highest enhancement of compactness (50%) after the opening of the HSR. However, hotspots are not as close to the station periphery in Jeongeup as other stations, although the compactness is slightly higher (15.8%) after the opening of the HSR. We observed that all cities have more hotspots and higher compactness after the opening of the HSR, and that the geographic locations of hotspots were changed. This evidence indicates that the HSR changes land use and business positively. Further, the development of land use and business facilities is focused at the station periphery.

Although the mobile phone data provide detailed information regarding the volume of phone users by age and gender observed at the HSR station, we cannot specify the full impact of the opening of the HSR and evaluate the reason for the increase or decrease of the mobile population, as a lot of studies or reports in other countries are based on interviews or other surveys. For a more detailed

analysis in the future, it is necessary to acquire the mobile phone data for a long-term period and to explore how the mobility flows and hotspots of cities move over time. This study, nevertheless, is meaningful as it is the first study to analyze the impact of HSR using mobile phone data. We would like to show the potential use of mobile phone data in the field of transport studies. Even though the data are limited-access for the public, this study is informative for transit agencies to evaluate the impact of transport systems and to create better transit services for the public.

As a result of innovative data collection technologies and advanced analytic tools, transport agencies can expect to see more extensive studies and applications utilizing big data. This study can provide a foundation for future work. For example, future study topics may include evaluating current railway network connecting to cities and hotspots by time of day, implementing flexible railway operations depending on demand (i.e., based on the hourly amount of population in station). Analyses of mobile phone data can also help urban planners to identify and prioritize hotspot locations where more public transport services are required.

**Author Contributions:** In this study, all of the authors contributed to the writing of the manuscript. J.K.E. designed the overall modeling framework and performed the data analysis. K.-S.L. helped with the data analysis and drafted the initial version of the manuscript. J.Y.S. contributed to develop the maps and interpretation. J.L. advised on the modeling process and coordinated the overall research. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant from R&D Program of the Korea Railroad Research Institute, Republic of Korea.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Campos, J.; De Rus, G. Some stylized facts about high-speed rail: A review of HSR experiences around the world. *Transp. Policy* **2009**, *16*, 19–28. [[CrossRef](#)]
2. David, H.; Zheng, L.; Corinne, M. The impact of High speed rail on Land and Property values. *Australas. Railw. Assoc.* **2012**, *A03958*, 1–20.
3. Clewlow, R.; MIT USA. Impacts of High-speed Rail on Air Transportation in Europe: An Analysis of Demand and Emissions. *Assoc. Eur. Transp.* **2011**, 1–14.
4. Chen, C.L.; Hall, P. The wider spatial-economic impacts of high-speed trains: A comparative case study of Manchester and Lille sub-regions. *J. Transp. Geogr.* **2012**, *24*, 89–110. [[CrossRef](#)]
5. Levinson, D.M. Accessibility impacts of high-speed rail. *J. Transp. Geogr.* **2012**, *22*, 288–291. [[CrossRef](#)]
6. Han, J.; Hayashi, Y.; Jia, P.; Yuan, Q. Economic Effect of High-Speed Rail: Empirical Analysis of Shinkansen's Impact on Industrial Location. *J. Transp. Eng.* **2012**, *138*, 1551–1557. [[CrossRef](#)]
7. Garmendia, M.; Ribalaygua, C.; Urena, J.M. High speed rail: Implication for cities. *Cities* **2012**, *29*, 26–31. [[CrossRef](#)]
8. Urena, J.M.; Menerault, P.; Garmendia, M. The high-speed rail challenge for big intermediate cities: A national, regional and local perspective. *Cities* **2009**, *26*, 266–279. [[CrossRef](#)]
9. Bocker, J.; Korzhenevych, A.; Schurmann, C. Assessing spatial equity and efficiency impacts of transport infrastructure projects. *Transp. Res. Part B* **2010**, *44*, 795–811. [[CrossRef](#)]
10. Sun, F.; Mansury, S.Y. Economic Impact of High-Speed Rail on Household Income in China. *J. Transp. Res. Rec.* **2012**, *2581*, 71–78. [[CrossRef](#)]
11. Chen, Z.H. The Impact of High-Speed Rail Investment on Economic and Environmental Change in China: A Dynamic CGE Analysis. *Transp. Res. A* **2015**, *92*, 232–245. [[CrossRef](#)]
12. Kurihara, T.; Wu, L.L. The Impact of High Speed Rail on Tourism Development: A Case Study of Japan. *Open Transp. J.* **2016**, *10*, 35–44. [[CrossRef](#)]
13. Garmendia, M.; Urena, J.M.; Coronado, J.M. Long-distance trips in a sparsely populated region: The impact of high-speed infrastructures. *J. Transp. Geogr.* **2010**, *19*, 537–551. [[CrossRef](#)]
14. Gutierrez, J.; Gonzalez, R.; Gomez, G. The European high-speed train network: Predicted effects on accessibility patterns. *J. Transp. Geogr.* **1996**, *4*, 227–238. [[CrossRef](#)]

15. Kim, H.W.; Lee, D.H.; Park, H.S. The impact of Gyeongbu high speed rail construction on regional economic growth. *Ksce J. Civ. Eng.* **2013**, *17*, 1206–1212. [\[CrossRef\]](#)
16. Hur, J.W. A critical review on the straw effects of high speed train. *J. Korean Urban Manag. Assoc.* **2010**, *23*, 59–74. (In Korean)
17. Jo, J.U.; Woo, M. The impacts of high speed rail on regional economy and balanced development: Focused on Gyeongbu and Gyeongjeon lines of KTX. *J. Korea Plan. Assoc.* **2014**, *49*, 263–278. [\[CrossRef\]](#)
18. Cho, N.; Lee, H.; Jin, S. Analysis on the straw effect by the HSR (High Speed Rail): Focusing on the shopping trips. *Korea Spat. Plan. Rev.* **2005**, *47*, e123.
19. Park, M.S.; Kim, Y. The impacts of high speed train on the regional economy of Korea. *Korean J. Appl. Stat.* **2016**, *29*, 13–25. [\[CrossRef\]](#)
20. Lopez, E.; Gutierrez, J.; Gomez, G. Measuring regional cohesion effects of large-scale transport infrastructure investments: An accessibility approach. *Eur. Plan. Studies* **2008**, *16*, 277–301. [\[CrossRef\]](#)
21. Gutierrez, J. Location, economic potential and daily accessibility: An analysis of the accessibility impact of the high-speed line Madrid-Barcelona-French border. *J. Transp. Geogr.* **2001**, *9*, 229–242. [\[CrossRef\]](#)
22. Chang, J.S.; Lee, J.H. Accessibility analysis of Korean high-speed rail: A case study of the Seoul Metropolitan Area. *Transp. Rev.* **2008**, *28*, 87–103. [\[CrossRef\]](#)
23. Monzon, A.; Ortega, E.; Lopez, E. Efficiency and spatial equity impacts of high-speed rail extensions in urban areas. *Cities* **2013**, *30*, 18–30. [\[CrossRef\]](#)
24. Wu, W.; Liang, Y.; Wu, D. Evaluating the impact of China's rail network expansions on local accessibility: A market potential approach. *Sustainability* **2016**, *8*, 512. [\[CrossRef\]](#)
25. Shaw, S.-L. Impacts of high speed rail on railroad network accessibility in China. *J. Transp. Geogr.* **2014**, *40*, 112–122. [\[CrossRef\]](#)
26. McDonald, J.H. *Handbook of Biological Statistics*, 3rd ed.; Sparky House Publishing: Baltimore, Maryland, 2014.
27. McDonald, J.F. The identification of urban employment sub centers. *J. Urban Econ.* **1987**, *21*, 242–258. [\[CrossRef\]](#)
28. McMillen, D.P. Nonparametric employment subcenter identification. *J. Urban Econ.* **2001**, *50*, 448–473. [\[CrossRef\]](#)
29. McMillen, D.P.; Smith, S.C. The number of subcenters in large urban areas. *J. Urban Econ.* **2003**, *53*, 321–338. [\[CrossRef\]](#)
30. Redfearn, C.L. The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *J. Urban Econ.* **2007**, *61*, 519–541. [\[CrossRef\]](#)
31. Redfearn, C.L. Persistence in urban form: The long-run durability of employment centers in metropolitan area. *Reg. Sci. Urban Econ.* **2009**, *39*, 224–232. [\[CrossRef\]](#)
32. Lee, K.S.; You, S.Y.; Eom, J.K.; Song, J.; Min, J.H. Urban spatiotemporal analysis using mobile phone data: Case study of medium-and-large-sized Korean cities. *Habitat Int.* **2018**, *73*, 6–15. [\[CrossRef\]](#)

