A GPS-Based Digital Dynamics Study of the Rugby World Cup 2019 Spectators

Masamichi Aihara¹, Koki Sakai² and Kenji Ono²
1. Keidai Sports & Culture Center, Faculty of Human Sciences, Osaka University of Economics, Osaka 533-8533, Japan
2. Outing Research Institute, Tokyo 104-0033, Japan

Abstract: This study’s dynamic survey of spectators at the HRS (Hanazono Rugby Stadium) in HOC (Higashi Osaka City) during the 2019 RWC (Rugby World Cup) tournament used location information big data to analyze nine items, including spectator attributes—60 min or more stay in HOC (excluding residents), more than 15 min stay in the HRS on match days (besides, the days before and after the match). To compare spectators, visitors to HOC during the matches were added to the target group. The results show that the RWC attracted a high number of male visitors aged 20, 40, and 50 years, mainly from the Kinki region, whose stopovers inside and outside the region were limited to Osaka City. Stopovers in tourist areas unrelated to the RWC were few, partly because it was possible to undertake a day trip solely to watch the game without any stopovers.

Key words: RWC, sports tourism, global positioning system, dynamic digital survey, positive and negative data.

1. Background to the Conception of this Research and the State of Preparation

The global tourism industry accounts for approximately 10% of the GDP (gross domestic product) (UNWTO, 2016). The expansion of the tourism industry is believed to contribute to economic revitalization and the creation of jobs, and it is an important industry for the development of the social economy. The tourism industry is one of the fastest growing industries in Japan. In the context of sports tourism, Japan hosted a series of mega-sporting events—the 2019 Rugby World Cup and the 2020 Tokyo Olympic and Paralympic Games—and will be hosting the 2021 World Masters Games. In June 2011, the Japan Tourism Agency formulated the basic policy to promote sports tourism, which considers sports a tourism resource and focuses on promoting and expanding domestic and inbound tourism. The key issue confronting Japan is how to increase the number of foreign tourists by adopting an approach that views sporting events as a tourism resource.

In this context, it must be noted that the fields of regional tourism, transportation, and urban planning surveys have explored the potential of the objective digital survey method that uses the smartphone location data held by telecommunication carriers. This approach is also expected to reduce or eliminate the problems faced by conventional sports tourism surveys and promote sports tourism. Hence, the current study applies this method to sports tourism surveys in order to ascertain with precision planned, tangible, and positive data as well as unplanned, intangible, and negative data. This study focuses on the spectators at the Rugby World Cup 2019 (hereafter Rugby WC)—the first event in Japan’s Golden Sports Years—held at the Higashi Osaka City Hanazono Rugby Stadium (hereafter the Hanazono Rugby Stadium) on September 22 and 28, 2019 and October 3 and 10, 2019. This study clarifies spectator attributes and behavioral characteristics by conducting visitor analysis, itinerary analysis, stopover analysis, accommodation analysis, transportation analysis, and excursion analysis.
2. Theoretical Background and Literature Review

Several sports tourism studies, in North America and other parts of the world, present several findings, including those on the motivations for volunteering at mega-sporting events (Alexander et al., 2015) and factors inhibiting participation in activities (Williams & Lattey, 1994). Gibson (1998) classifies sports tourism into three categories—active, nostalgia, and event sports tourism. Active sports tourism is synonymous with "participatory sport," and it involves some form of physical activity during the trip. Nostalgia sports tourism refers to visiting historical sites, such as stadiums hosting the Olympics. Event sports tourism refers to watching sporting events, such as the local events as well as the mega-sporting events such as the Olympic and Paralympic Games and the FIFA World Cup. Neirotti et al. (2001) study the foreign visitors at the Barcelona and Atlanta Olympic Games to determine their motivations for watching the games. The study shows that the spirit of athleticism and the cultural and historical experiences associated with these events and their hosting destinations motivate visitors to watch the Olympics. Studies have also been conducted to understand the behavioral patterns of foreign tourists during their stay in the hosting destination and to examine how these tourists affect the local economy (Irwin et al., 1996). Although several studies have been conducted on the economic impacts of foreign tourists, there have been difficulties in accurately estimating these effects. Studies have also focused on the psychological effect of tourism on the hosting destinations. The psychological effects include improving the image of the region and fostering local residents’ pride in the community (Harada, 2002), but it is understood that smaller-scale community sporting events are more likely to have an effect on local residents (Walo et al., 1996). Gratton et al. (2000) predict that large-scale mega-sporting events are more likely to be associated with tourism. In this study, we focused on large-scale mega-sporting events that attract tourists.

As per the International Olympic Committee, the literature records only the positive effects of the Olympic legacy on the host city and country. The candidacy files of cities that made past Olympic bids and the final reports of host cities are limited to positive topics such as nostalgia, sports, sustainability, culture, and economy (International Olympic Committee (IOC), 2014a, 2014b). Since the debate focuses on positive, tangible, and planned legacies, Gratton and Preuss (2008) indicate a likelihood of arbitrariness on the part of convention supporters is likely to occur.

The post-game consumption questionnaire survey and the inter industry analysis methods based on the survey’s results cannot take into account the negative effects caused by the substitution effects, concentration on one city, and congestion avoidance behavior of residents and tourists. We observed a limitation in the methods of impact research and effectiveness measurement before and after the mega-sporting events. Specifically, these methods cannot completely eliminate the arbitrariness of event supporters, who tend to emphasize positive effects. The survey methods of conventional sports events rely on official statistics and questionnaires. They have the following issues: lack of reliability in terms of sample size, difficulty in capturing visitor or participant behavior after or in an event, inability to conduct comparative surveys with other events separated geographically or temporally or run by different entities, the lack of basis for calculating pre-event estimates and difficulty in comparing and verifying the results, and large-scale cost and time requirements for surveys. Presently, the Japan Tourism Agency conducts questionnaire surveys in airports witnessing footfall of foreign tourists. However, analogue questionnaires relying on the foreign tourists’ memories are now reaching their limits.

In this context, it must be noted that sports events leave a legacy of benefits for the host city. Since it is
crucial to perform legacy planning from the bidding stage before the event, legacies are planned and manifest as tangible and positive elements. However, in reality, several unplanned legacies were not envisioned, and several intangible and negative legacies were invisible. Given that the Rugby World Cup was a mega-sporting event held in Japan for the first time, it was naturally infeasible to predict all the aspects. In this regard, Gratton and Preuss (2008) advocate measuring legacy effects on a three-dimensional cube comprising the following axes: planned-unplanned, negative-positive, and tangible-intangible. It is important to examine the presence of negative legacies and, accordingly, provide appropriate information for the bidding and hosting of future events. This examination can be conducted via the digital survey method, which uses the location data to collect unplanned positive and negative data.

3. Materials and Methods

3.1 Definitions of the Items of Analysis

First, we define “spectator,” from among the subjects to be studied. This study aims to understand the effects of sports tourism. Hence, we define spectators from the perspective of individuals who contributed to the expansion of the non-resident population of Higashi Osaka City. In this context, spectators were selected from among those who stayed in Higashi Osaka City for 60 min or more, excluding those who live in the city (Higashi Osaka City’s residents, commuters, long-term residents, and frequent visitors). Specifically, the survey included those who stayed in the Hanazono rugby stadium area (Fig. 1) for more than 15 min on the match days (September 22 and 28, 2019 and October 3 and 10, 2019), in addition to the
days before and after the match (September 21-23, 2019; September 27-29, 2019; October 2-4, 2019; and October 9-11, 2019). In order to compare the spectators, the visitors who came to Higashi Osaka City when the matches were not held (September 6-8, 2019; September 14-16, 2019; October 19-21, 2019; and October 24-26, 2019) were designated as “visitors during non-event period” and added to the target group.

Subsequently, we defined the target area for this study. We surveyed Higashi Osaka City from the perspective of contributing to the expansion of its non-resident population. To conduct a more detailed analysis, we divided the city into seven areas (Fig. 2: the Kusaka Park, Hiraoka Shrine, Buffer Green Park, Hanazono Rugby Stadium, Inada Hachiman Shrine, Kosaka Area, and Nagase Station areas), which we designated as Higashi Osaka City’s tourism areas. The other analysis definitions are shown in Table 1.

### 3.2 Details of Analysis

We conducted an analysis using various approaches that can be implemented with the location information big data (Table 2). To clarify the attributes and behavioral characteristics of spectators via a comparative analysis, we conducted each analysis separately for spectators and visitors during the non-event periods.

### Table 1 Other analysis definitions.

<table>
<thead>
<tr>
<th>Determination</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accommodation</td>
<td>The place where the person stayed the longest between 22:00 and 08:00 the next day is the municipality of accommodation. However, if the stay is less than 1 h, accommodation is assumed to be unknown.</td>
</tr>
<tr>
<td>Daytime</td>
<td>We sample visitors who stay between 09:00 and 20:59.</td>
</tr>
<tr>
<td>Stopover municipality</td>
<td>A city, town, or village where the person stays for 60 min or more during their itinerary is considered a stopover municipality.</td>
</tr>
</tbody>
</table>
Table 1 to be continued

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourism area visit</td>
<td>Visitors are defined as those who stay for 15 min or more in the seven sightseeing areas of Higashi Osaka City.</td>
</tr>
<tr>
<td>Aggregate results</td>
<td></td>
</tr>
<tr>
<td>Stay time</td>
<td>Average length of stay for visitors at Higashi Osaka City and tourist areas.</td>
</tr>
<tr>
<td>Average number of places visited</td>
<td>Average number of places visited by visitors to Higashi Osaka City when on excursions to tourist areas.</td>
</tr>
<tr>
<td>De-identification processing</td>
<td>Categories with a small number of visitors are displayed as blank or “****” to protect privacy.</td>
</tr>
</tbody>
</table>

Table 2  Details of analysis.

<table>
<thead>
<tr>
<th>Item</th>
<th>Details of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute analysis</td>
<td>Calculation of the number of visitors by gender and age group</td>
</tr>
<tr>
<td>Origin analysis</td>
<td>Heat map of the number of visitors by origin and municipality</td>
</tr>
<tr>
<td>Itinerary analysis</td>
<td>Calculation of the ratios of visitors’ itineraries</td>
</tr>
<tr>
<td>Lodging location analysis</td>
<td>Heat map of the number of night-lodging units by municipality</td>
</tr>
<tr>
<td>Stopover analysis</td>
<td>250 m-mesh heat map of locations visited by municipality</td>
</tr>
<tr>
<td>Tourism area analysis</td>
<td>Calculation of the number of visitors and visitation rate by tourism area</td>
</tr>
<tr>
<td>Analysis of incomers, outgoers, and stayers by time zone</td>
<td>Graphical display of trends in the number of incomers, outgoers, and stayers by time zone Analysis of only the Hanazono Rugby Stadium and Hiraoka Shrine areas, which have high spectator visitation rates</td>
</tr>
<tr>
<td>Analysis of average daytime duration of stay</td>
<td>Calculation of the average daytime duration of stay in each tourist area</td>
</tr>
<tr>
<td>Excursion analysis</td>
<td>Calculation of the rate of touring multiple areas and the composition ratio by each excursion route.</td>
</tr>
</tbody>
</table>

4. Results and Discussion

4.1 Attribute Analysis

We used the KDDI Corporation’s location information big data, which helped in linking the location information with the gender and age information that KDDI obtains at the time of concluding a contract for a communication line with the consent of the individual user. This provides an understanding of the attributes of visitors to Higashi Osaka City. It should be noted that the survey did not include spectators below 15 years of age; owing to the characteristics of smartphone users, we had a very small number of licensees over 70 years of age.

Fig. 3 shows the results of the analysis. The percentage of male spectators was higher than that of the male visitors during the non-event periods. There were more males aged 20 years and those between 40 and 50 years. The findings also indicated that the Rugby World Cup may have brought people with different attributes to Higashi Osaka City.

4.2 Origin Analysis

We ascertained the distribution of the origin—the place of residence of the visitors to Higashi Osaka City—and captured the characteristics of spectators. Based on the movement trends, the KDDI location information big data define the municipality of the licensee as that which includes the place where the licensee spent the largest number of nights for 60 days (estimated place of residence), including 15 days before and after each month.
The results of the origin analysis showed that the ratio of people originating from the Kinki region, other than the Osaka Prefecture, was approximately 90%, with 88.2% and 91.0% spectators and visitors, respectively, during the non-event periods. However, the ratios of people originating from the Osaka Prefecture were 58.8% and 68.1% for spectators and visitors, respectively, during non-event periods, indicating that a higher percentage of spectators came from prefectures in the Kinki region, other than Osaka.

The breakdown by municipality in Figs. 1 and 4 shows that the numbers of spectators originating from Kyoto, Kobe, Sakai, and other large neighboring cities were more than the visitors during the non-event periods. This indicates that the host destination of the Rugby World Cup attracted visitors from further regions.

### 4.3 Itinerary Analysis

We determined the itinerary (day trip, overnight stay, or more than one night stay) according to the duration between the departure from the estimated place of residence and return. In relation to these aspects, we ascertained the characteristics of spectators.

Fig. 5 shows the results of the analysis. In the case of spectators, 86.8% were on a day trip, which is an increase of 5.2% over the proportion of visitors during the non-event periods. While the analysis in the previous section confirmed that the visitors came from more distant places, this showed that the itineraries were shorter and that spectators were less likely to stay overnight when they came to Higashi Osaka City.

### 4.4 Lodging Location Analysis

We ascertained the characteristics of spectators based on the distribution trend of the lodging location. We define lodging location as a location other than home where a person spends the most time at night; however, if the stay is less than 1 h, the lodging location is considered to be unknown. Lodging locations were analyzed on a municipality basis.

Regarding spectators, the analysis showed that Osaka City accounted for the highest percentage of lodgings at 28.9%, followed by Higashi Osaka City at...
28.2%. Conversely, for visitors during the non-event periods, Higashi Osaka City accounted for 44.8% of lodging, followed by Osaka City at 20.2%. This shows that the spectators tend to choose Osaka City as a lodging location as it has good transportation connectivity to distant places (Figs. 6 and 7).

Fig. 4  Results of origin analysis (visitors during non-event periods).
Source: KDDI × COLOPL Location Trends.

Fig. 5  Itinerary analysis results.
Source: KDDI × COLOPL “Location Trends”.
Fig. 6  Results of analysis of number of nights lodging (spectators).
Source: KDDI × COLOPL “Location Trends”.

Fig. 7  Results of analysis of number of nights lodging (visitors during non-event periods).
Source: KDDI × COLOPL “Location Trends”.
4.5 Stopover Analysis

We analyzed the tendency of stopovers in the same itinerary in order to ascertain the characteristics of spectators. A stopover is defined as a stay of 60 min or more in cities and towns other than Higashi Osaka City. In the case of Higashi Osaka City, we use a 250 m-mesh basis to analyze places where these spectators did a stopover of 15 min or more.

The results of the analysis showed that the stopover rate for Osaka City was the highest at 15.6%, for both spectators and visitors, during the non-event periods, (Figs. 8 and 9). Unlike the spectators whose stopovers were concentrated in the areas around the Hanazono rugby stadium and the Higashi Hanazono station, the stopovers of visitors during the non-event periods were distributed in the areas around the Fuse Station, Kinki University, Osaka University of Commerce, AEON Higashi Osaka, CAINZ Higashi Osaka, Frespo Higashi Osaka, HOS Bowling Space hit, and the Higashi Osaka City Culture Creation Hall, among others, in addition to the Hanazono Rugby Stadium, indicating diverse purposes of visits (Figs. 10 and 11).

4.6 Tourism Area Analysis

We ascertained the characteristics of spectators based on the number of visitors and visitation rate by tourist area. In this analysis, a stay of 15 min or more in a tourist area has been defined as a visit, and the ratio of the number of visitors to the total number of spectators or the total number of visitors during non-event periods is defined as the visitation rate (Figs. 12 and 13).
Fig. 9  Results of analysis of stopover municipalities (visitors during non-event periods).
Source: KDDI × COLOPL “Location Trends”.

Fig. 10  Results of stopover locations (spectators).
Source: KDDI × COLOPL “Location Trends”.
The analysis showed that the Hiraoka Shrine area had the highest percentage of spectators at 28.5% (Fig.12). This may be attributed to the presence of the Higashi Hanazono Station, the station closest to the Hanazono Rugby Stadium. During the non-event periods, the Kosaka area had the highest percentage of visitors at 33.5%, followed by the Buffer Green Park area at 20.7%, indicating that the visitors had more diverse purposes of visit than that of the spectators.

4.7 Analysis of Incomers, Outgoers, and Stayers by Time Zone

We analyzed the trends in the number of incomers, outgoers, and stayers by time zone for each tourist area in order to ascertain the characteristics of spectators. We analyzed the trends in the number of people in each tourist area in terms of the units of time, with the timings at which they first visited and left and that between the arrival and departure defined as the incoming time, outgoing time, and duration of stay, respectively. From the perspective of the number of samples, the tourist areas analyzed were the two areas with the highest spectator visitation rates, namely, the Hanazono Rugby Stadium and Hiraoka Shrine areas, including the Higashi Hanazono Station. The Hanazono Rugby Stadium area (excluding the Hanazono Rugby Stadium) and the Hanazono Rugby Stadium were established separately for spectators.

The analysis showed that the spectators had the tendency to enter the Hanazono Rugby Stadium between 11:00 and 14:00, and their outflow periods concentrated at 16:00. In the surrounding Hanazono Rugby Stadium area, the outflow and inflow occurred at 13:00 and 16:00, respectively. Conversely, the stays of visitors during the non-event periods peaked at 14:00, and the inflow and outflow timings were widely dispersed (Fig. 14).
Fig. 12  Results of analysis of visitation rate by area (spectators).
Source: KDDI × COLOPL “Location Trends”.

Fig. 13  Results of analysis of visitation rate by area (visitors during non-event periods).
Source: KDDI × COLOPL “Location Trends”.
Relative to the Hanazono Rugby Stadium area, the Makino Shrine area showed a strong tendency for spectators’ inflow and outflow to be concentrated after the end of the game. Conversely, the inflow/outflow timings before attending the game were a little earlier, around 11:00-12:00. The reason for this is that the Higashi Hanazono Station, which is included in the Makino Shrine area, is the closest station. While spectators come to the venue early to avoid a possible congestion before a game, they leave the venue immediately after watching the game, and hence the inflow/outflow timing is concentrated (Fig. 15).
4.8 Analysis of Average Daytime Duration of Stay

We analyzed the average length of stay of visitors in each tourist area during the daytime (09:00-21:00) in order to ascertain the characteristics of spectators. By focusing on the daytime hours, we provided an understanding of the duration of visits not linked to lodging stays and provided suggestions regarding the purpose and details of visits to each tourist area.

The analysis showed that the spectators spent the longest time at the Hanazono Rugby Stadium, followed by the Hanamaki Rugby area (excluding the Hanazono Rugby Stadium). However, they stayed at the Kosaka and Hanamaki Shrine areas, which have the highest visitation rates, for less than 1 h. This finding indicates that their purpose of visit was solely to watch a match at the Hanamaki Rugby Stadium (Fig. 16). However, the average daytime duration of stay by visitors during the non-event periods was more than 2 h for all tourist areas, indicating that visitors had their own reason for visit to each tourist area (Fig. 17).

4.9 Excursion Analysis

We analyzed the average number of places visited and the composition ratio by the excursion route to understand the characteristics of spectators. An excursion is determined the same way as a visit to the tourist area—stay at each tourist area for more than 15 min. The composition ratio by the excursion route indicates the composition ratio of the excursion route selected by visitors to multiple areas.
Fig. 16  Results of analysis of average daytime duration of stay by area (spectators).
Source: KDDI × COLOPL “Location Trends”.

Fig. 17  Results of analysis of average daytime duration of stay by area (visitors during non-event periods).
Source: KDDI × COLOPL “Location Trends”.
The analysis revealed that 54% of the spectators visited only the Hanazono Rugby Stadium and did not stay in other tourist areas, and the rate of multiple-area excursions was 20.1% (60 visitors to multiple areas, out of the 298 visitors to any tourist area). Conversely, 20.4% of the visitors (870 visitors to multiple areas, out of the 4,269 visitors to any tourist area) went to multiple areas, during the non-event periods, which was almost at the same level as the spectators.

The excursion routes indicated were concentrated between the “Hanazono Rugby Stadium Area <= Hiraoka Shrine Area” and the “Kosaka Area <= Hiraoka Shrine Area” for spectators. However, they were dispersed mainly in the west of Higashi Osaka City for visitors during the non-event periods (Figs. 18 and 19).

![Spectator Tour route Top 5 Map](image)

Fig. 18  Results of analysis of the top 5 excursion routes (spectators).

Source: KDDI × COLOPL “Location Trends”.

5. Conclusion

This study analyzed nine items using the location information big data in order to ascertain the attributes and behavioral characteristics of spectators from various perspectives. Table 3 presents the results of analysis. The Rugby World Cup brought an increasing number of male visitors aged 20, 40, and 50 years, mainly from the Kinki region. Conversely, the purpose of the spectators was only to watch the rugby game, and their stopovers inside and outside the region were limited to Osaka City, which provides convenient regional transportation and is near transportation nodes such as the Higashi Hanamaki Station. The percentage of visitors from outside the Kinki region was small, and hence there were few stopovers in tourist areas not related to the Rugby World Cup. This is partly attributed to the fact that it was possible to make a day trip if the purpose was only to watch the game and not make any other stopover. Organizing a Rugby World Cup match in the city has an extremely limited effect on the surrounding tourist areas, and hence does not contribute toward expanding the non-resident population of Higashi Osaka City. In addition to increasing the proportion of spectators from distant areas who need to make itineraries that include overnight stays, it is also necessary to hold events in conjunction with the Rugby World Cup and take measures targeting the Rugby World Cup audience, in order to increase overnight stays and excursions in the city. This approach can positively create new economic effects. When planning such a project, the target audience, the location of the event, and the entire event duration can be selected based on
the attributes and behavioral characteristics identified in this study, in order to increase the probability of success. It would also be appropriate to set quantitative targets for the measures, such as the number of nights spent in the region and the number of places visited on excursions. The quantitative understanding of the dynamics of event spectators using the location-based big data may be useful for planning and evaluation aimed at expanding the non-resident population of the region and may contribute to increasing the probability of success of measures through continuous efforts. At the same time, the utilization of objective data may be useful for fulfilling accountability to the community and citizens when implementing municipal initiatives.

The results of this research can contribute to a new research approach for the survey and analysis of the Tokyo Olympic and Paralympic Games and the Kansai World Masters Games. The results of these digital dynamic surveys in Japan will be used to conduct comparative research, with the 2023 Rugby World Cup in France and the 2024 Paris Olympic and Paralympic Games.

References

Appendix

Table A1

Comparison of Existing Statistical and Smartphone Location Data

<table>
<thead>
<tr>
<th>Item</th>
<th>PT survey</th>
<th>Metropolitan transportation census survey of railway users</th>
<th>Smartphone location data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigation target</td>
<td>Residents of urban areas (a small percentage selected as a sample)</td>
<td>Distributed to passengers who get off at the target stations</td>
<td>Specific smartphone app users (millions of people across Japan)</td>
</tr>
<tr>
<td>Survey frequency</td>
<td>Roughly once every 10 years</td>
<td>Once every 5 years</td>
<td>Continuous (365 days a year)</td>
</tr>
<tr>
<td>Survey area</td>
<td>Each urban area</td>
<td>Tokyo metropolitan, Chukyo, and Kinki area</td>
<td>All of Japan</td>
</tr>
<tr>
<td>Attributes (examples)</td>
<td>Gender, age, the places of residence and work, and household composition</td>
<td>Gender, age, place of residence</td>
<td>Gender, age&lt;sup&gt;a&lt;/sup&gt; Place of residence&lt;sup&gt;b&lt;/sup&gt;, place of work&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Time resolution</td>
<td>Minute units</td>
<td>Minute units</td>
<td>Minute units</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>Many zone units</td>
<td>Each station on each line</td>
<td>Facility units (minimum is 10 m-mesh)</td>
</tr>
<tr>
<td>Means of transportation</td>
<td>Identification of means of transportation possible (can be set according to the purpose of the survey)</td>
<td>Rail users only</td>
<td>Identification of rail, aircraft, ship, car, and walking/biking possible (Estimation may be difficult according to the conditions)</td>
</tr>
<tr>
<td>Purpose of travel</td>
<td>Identification by the purpose of travel possible (can be set according to the purpose of the survey)</td>
<td>Identification by the purpose possible (2015 survey: commuting to work and school, business-related, private, and returning home)</td>
<td>Commuting to work and school, business-related, tourism, and shopping (Estimation may be difficult according to the conditions)</td>
</tr>
<tr>
<td>Route of travel</td>
<td>Identifies the boarding and alighting stations and bus stops</td>
<td>Identification of rail route possible</td>
<td>Identification of routes possible with data at intervals of several minutes</td>
</tr>
</tbody>
</table>

<sup>a</sup> Difficult to grasp for younger and older age groups with low smartphone usage rates.  
<sup>b</sup> Information estimated from the movement history. Source: Ishii et al. (2018).

**Population ratio by municipality**

*Fig. A1 Correlation between smartphone location data and census results. Source: Ishii et al. (2018).*
Overview and Characteristics of Smartphone Location Data

Overview of Data

The smartphone location data utilized in this research include the current terminal location GPS information obtained by telecommunication carriers through specific applications after the prior individual consent of smartphone users. The data have been accumulated in a continuous and sustainable manner from millions of active users since October 2014. They can be used for retrospective research and analysis. As the data are acquired 24 h a day and 365 days a year, it is expected to be used for the continuous monitoring of the city and for understanding actual traffic conditions during the events. In addition, the apps acquiring the above location information include official carrier portal apps. Note that the attribute bias in the population is relatively small, showing a higher correlation than that of the population in municipalities in the 2015 census (Fig. A1). However, owing to the nature of smartphone applications, some attributes are skewed, such as high smartphone usage rates by age group. Hence, the analysis was weighted to match the distribution of prefectures of residence and the sex and age group composition ratios from the 2015 National Census.

As smartphone location data contain high privacy risks, we performed sufficient anonymity processing in accordance with the location information privacy report of the Ministry of Internal Affairs and Communications. The processing entailed ID hashing, temporal and spatial meshing of location information, and de-identification processing for small numbers of people, masking the values of small numbers of people (less than 10) when providing the data externally. The 2014 Ministry of Internal Affairs and Communications’ white paper on information and communications details this service provision system, and it was handled appropriately in accordance with the Act on the Protection of Personal Information. There has been no case of controversy in the past. This study does not analyze anonymized data. However, as it utilizes data that have been sufficiently statistically processed by the telecommunication carrier, it will not lead to the personal identification of the research collaborators.
Data Characteristics

The smartphone location data used in this research are statistical data, and the data for analysis held by the telecommunication carrier include the location information acquisition time and latitude/longitude information in a form that is linked to hashed ID information to prevent the identification of individuals (hashed ID). The hashed ID is accompanied by gender and age group information obtained from the communication line contractee information. The residential and work area information is estimated on the basis of the trend of location information over a certain period. In addition, the GPS location information is acquired at a minimum resolution of 10 m$^2$ and at minimum intervals of a few minutes, and background positioning is performed for service provision even when the app is not running directly, which supports the analysis of means and purposes of travel under certain conditions.

Ishii et al. (2018) compared and verified the results of the existing official statistics, such as the person trip survey (PT survey), to conduct research on the reliability of smartphone location data handled in this study for use in the field of urban transportation. The next section describes the main comparative verification results, and Table A1 lists the characteristics of the smartphone location data handled in this study.

Data Reliability

As action logs are obtained at intervals of a minimum of several minutes in the smartphone location data handled in this study, it is possible to define a trip as the movement from one stay log to the next stay log, after strictly determining whether it is a movement or a stay based on a large number of action logs. In addition, as the spatial analysis granularity can be set at a minimum of a 10 m-mesh, it is possible to flexibly design the threshold of spatial distance, such that it can detect a stay at the same location. Relative to the Tokyo PT survey’s data, the stay/move determination logic implemented for smartphone location data has a significant impact on the number and length of trips—the survey results. Accordingly, Ishii et al. (2018) developed a trip-determination logic that is more sophisticated, as Fig. A2 shows.

(1) The first action log is set as the reference point.
(2) If an action log occurs more than 100 m away from the reference point, it is set as a new reference point.
(3) If the total time difference between the reference point and the action log within 100 m of the reference point is 15 min or more, the user is considered to have stayed at that reference point.
(4) The last action log occurring within 100 m of the reference point determined to be a stay is set as the starting point of the trip.
(5) Thereafter, (2) and (3) are repeated, and the reference point defined as the stay is set as the end point of the trip. (4), (2), and (3) are repeated.

Ishii et al. (2018) calculated the origin-destination (OD) traffic volumes for trips, implementing the above trip determination method by bundling the OD locations into municipal units. Subsequently, they compared the results with the Tokyo PT survey data in terms of the number of trips and inter-municipal OD traffic volumes. Consequently, they identified the following points:

(i) The smartphone location data captured greater number of trips per person. Given that many short-distance trips within a municipality or between neighboring municipalities were captured, the time of the day was mostly after 10:00, and the gender was mainly male, it can be inferred that a greater number of short-distance trips for business purposes or private reasons, such as lunch, were captured than in the PT survey.

(ii) Examining the inter-municipal OD volume, the trips by distance band show a similar trend as that of the PT survey for medium distance and above, and the OD pattern is generally correlated with the PT survey. However, overall, the number of movements over medium distances was lower than that in the PT survey data. In addition, concerning daytime, there were fewer trips between 05:00 and 09:00 than in the PT survey data. The inference is that the long-distance travel for commuting to work or school declines because it is hidden.
The results of the above comparison and verification confirm the possibility that smartphone location data can be used to a certain extent for wide-area travel beyond municipal boundaries. In future, with advancements in the estimation methods of the purpose of travel and means of transportation, it is assumed that smartphone data will be used as basic data for the study of wide-area transportation networks and wide-area cooperation across municipalities. It was also confirmed that data might have been obtained for short-distance travel in more detail than those in the existing surveys. As verifications confirm the data acquired on the kind of movement, this technique is expected to be used for understanding the flow of people within districts in the future.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP19K12568 and the 2019 Joint Research Fund from the Osaka University of Economics.