Computers, Environment and Urban Systems 86 (2021) 101593

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Towards a multidimensional view of tourist mobility patterns in cities: A mobile phone data perspective

Yang Xu[a,*], Jiaying Xue[a], Sangwon Park[b], Yang Yue[c]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[b] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong
[c] Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University, China

ARTICLE INFO

Keywords:
Tourist mobility
Travel behavior
Mobile phone data
Activity space
Eigendecomposition

ABSTRACT

The last decade has witnessed a wealth of studies on characterizing human mobility patterns using movement datasets. Such efforts have highlighted a few salient dimensions of individual travel behavior relevant to urban planning and policy analysis. Despite the fruitful research outcomes, most of the findings are drawn upon urban residents. The behavioral characteristics of other population groups, such as tourists, remain underexplored. In this study, we introduce an analytical framework to gain insights into tourist mobility patterns. By analyzing mobile phone trajectories of international travelers to three different cities in South Korea, we introduce nine mobility indicators to capture different facets of tourist travel behavior (e.g., duration of stay in a city, spatial extent of activities, location visited and trips conducted, and mobility diversity), and examine their statistical properties across cities. An eigendecomposition approach is then introduced to better understand the interdependency of these mobility indicators and inherent variations among individual travelers. Based on the eigendecomposition results, we further employ a dimension reduction technique to describe the key characteristics of each traveler. Since the mobile phone dataset captures the nationality of tourists, we use such information to quantify the behavioral heterogeneity of travelers across countries and regions. Finally, we select a few traveler groups with distinctive mobility patterns in each city and examine the spatial patterns of their activities. Substantial differences are observed among traveler groups in their spatial preferences. The implications for location recommendation and deployment of tourism services (e.g., transportation) are discussed. We hope the study brings a synergy between classic human mobility analysis and the emerging field of tourism big data. The framework can be applied or extended to compatible datasets to understand travel behavior of tourists, residents, and special population groups in cities.

1. Introduction

For many cities around the world, tourism has been an important industry to their social and economic developments. An improved understanding of tourist travel behavior is essential to the management and planning of cities. Despite such importance, obtaining useful information about tourist activities is not easy. Research on tourist travel behavior used to reply on surveys and questionnaires (Lau & McKercher, 2006; Mckercher & Lau, 2008). Such data could capture sociodemographic attributes of travelers and activities conducted by them, thus providing rich contextual information for tourism analysis. However, collection of tourist surveys is usually costly and time-consuming, therefore limiting the scale and scope of these studies.

The recent two decades have witnessed an increasing adoption of new technologies for tourism research (Li, Xu, Tang, Wang, & Li, 2018; Shoval & Ahas, 2016). These technologies, such as Global Positioning System (GPS) (Birenboim, Anton-Clavé, Russo, & Shoval, 2013; De Cantis, Ferrante, Kahani, & Shoval, 2016; Grinberger, Shoval, & McKercher, 2014; Tchetchik, Fleischer, & Shoval, 2009), mobile tracking (Ahas, Aasa, Mark, Pae, & Kull, 2007; Raun, Ahas, & Tiru, 2016; Saluveer et al., 2020; Xu et al., 2015), and geocoded social media (Kim, Kim, Lee, Lee, & Andrada, 2019; Mou et al., 2020; Vu, Li, Law, & Ye, 2015; Wood, Guerry, Silver, & Lacayo, 2013; Zhou, Xu, & Kimmons, 2015), have enabled new ways for studying tourist mobility patterns. As a result, we are seeing mobility studies over larger tourist populations that can complement findings from small-scale, survey based research.

* Corresponding author.
E-mail address: yang.ls.xu@polyu.edu.hk (Y. Xu).

https://doi.org/10.1016/j.compenvurbsys.2020.101593
Received 3 July 2020; Received in revised form 16 November 2020; Accepted 30 December 2020
Available online 15 January 2021

At the meantime, the dimensions of tourist mobility that can be derived and analyzed are augmented as big data proliferate. For instance, how long tourists stay in a city, how many places they visit, and how frequently they travel — which are important research questions in tourism studies (Lau & McKercher, 2006; Mckercher & Lau, 2008; Rodriguez, Martinez-Roget, & Gonzalez-Murias, 2018) — can now be addressed simultaneously when appropriate (big) datasets are available. One might argue that this can be achieved by conducting survey based research. An important "add-in" of big data is that the multidimensional characteristics of tourist mobility can now be evaluated across large populations, therefore revealing inherent variations across individuals and segmentation of different types of travelers. Such insights would benefit infrastructure planning and deployment of tourism services in cities.

The multidimensional nature of human mobility has been studied extensively, but mostly over urban residents (Alessandretti, Sapiezynski, Sekara, Lehmann, & Baronchelli, 2018; Gonzalez, Hidalgo, & Barabási, 2008; Jiang, Ferreira, & Gonzalez, 2017; Kang et al., 2010; Pappalardo et al., 2015; Schneider, Belik, Couronné, Smoreda, & González, 2013; Wu et al., 2019; Xu et al., 2016; Xu, Belyi, Bojic, & Ratti, 2018; Xu, Belyi, Santi, & Ratti, 2019; Yuan & Raubal, 2016; Yuan, Raubal, & Liu, 2012). Studies on tourist mobility patterns have been increasing over the years. However, most of them focus on a small number of tourists in a city or a confined geographic area (e.g., parks). The behavioral aspects being researched are rather diverse. For example, by leveraging questionnaire and GPS tracking technology, Mckercher et al. analyze the behavioral patterns of first (233 participants) and repeat visitors (130 participants) to Hong Kong (McKercher, Shoval, Ng, & Birenboim, 2012). They find that first time visitors tended to travel more widely while the activities of repeat visitors were confined to a smaller number of locations. With the GPS trajectories of 68 first-time tourists in Hong Kong, Grinberger et al. introduce a clustering-based method and discover three distinct time-space strategies of these visitors (Grinberger et al., 2014). Some studies have examined the topological structures of tourist movements or trip chains, revealing substantial variations of tourist mobility patterns (Mckercher & Lau, 2008; Yang, Wu, Liu, & Kang, 2017). Despite the fruitful research outcomes, these studies have focused on a small number of visitors (Mckercher & Lau, 2008) or relied on sparse mobility datasets such as Geo-Tagged photos (Yang et al., 2017). There are also studies that leverage GPS and survey-based methods to track the space-time patterns of tourists. Since collections of GPS trajectories and surveys are costly, these studies are usually conducted in a confined geographic area, such as a heritage site (Tchetchik et al., 2009), a summer palace (Xiao-Ting & Bi-Hu, 2012), or natrual & theme parks (Birenboim et al., 2013; Orellana, Bregt, Ligtenberg, & Wachowicz, 2012). There is a lack of research on quantifying tourist mobility patterns at scale, and therefore, a limited understanding of tourist space-time behaviors in cities.

The past decade has witnessed an increasing adoption of mobile phone data for studying human activities and social dynamics (Blondel, Decuyper, & Krings, 2015). The ability to capture whereabouts of large populations make mobile phone data an appealing resource for tourism analysis. One of the early papers in the field analyzes a large-scale mobile roaming dataset to study foreign tourists' space usage in Estonia (Ahas et al., 2007). The study reveals not only the spatial heterogeneity of tourist footprints, but also the seasonal variations, and behavioral differences among visitors of different nationalities. Later in the same study area (Estonia), mobile phone datasets were used to examine the behavioral difference between event tourists and regular visitors (Nilbe, Ahas, & Silm, 2014), to study the "destination loyalty" through mobility of repeat visitors (Kuusik, Tiru, Ahas, & Varblane, 2011; Tiru, Kuusik, Lamp, & Ahas, 2010), or to reveal the spatio-temporal variations of tourist activities (Raun et al., 2016). Despite the rich insights offered by these studies, most of the analyses are performed at an aggregate level. The results mainly describe the collective behavioral patterns of the population or a specific traveler group. The space-time behavior of individual travelers has not been well studied. This results into a limited understanding of the behavioral heterogeneity of individual travelers and the interrelationships among different behavioral dimensions. Many interesting questions await to be answered. For instance, do visitors show uniform or diverging travel patterns in a city? Does a longer duration of stay correspond to more locations visited? Are long-term stayers more diverse than short-term visitors in their spatial and temporal behaviors?

None of the above questions can be addressed without an effective individual level analysis. Although mobile phone data contains rich information on individual travel behavior, such information has not been fully utilized to understand tourist mobility patterns. Existing knowledge on different behavioral dimensions — such as spatial extent of tourist activities, duration of stay in a city, number of locations visited, and movement patterns among these locations — is somewhat scattered and isolated. The interdependency of these behavioral dimensions remains underexplored.

To fill the research gap, this study introduces an analytical framework to provide a multidimensional view of tourist mobility patterns. The efficacy of the framework is demonstrated by analyzing a large-scale mobile phone dataset collected in three cities in South Korea. The major contributions of this research are as follows:

- By analyzing mobile phone trajectories of international travelers who visited any of the three cities in South Korea (Jeonju, Gangneung and Chuncheon) during a period of one year, we introduce nine individual mobility indicators, namely — (1) number of observations days, (2) time span, (3) radius of gyration, (4) diameter of trajectory, (5) total number of activity anchor points, (6) daily number of activity anchor points, (7) total number of origin-destination trips, (8) activity entropy and (9) travel diversity — to quantify important aspects of tourist mobility patterns. These indicators are linked with critical dimensions of human mobility that are covered and investigated in previous literature (Gonzalez et al., 2008; Pappalardo et al., 2016; Rodriguez et al., 2018; Schönfelder & Axhausen, 2003; Xu et al., 2018). They capture different facets of tourist travel behavior, such as their duration of stay in a city (indicator 1 and 2), spatial extent of activities (indicator 3 and 4), locations visited and trips conducted (indicator 5, 6 and 7), as well as diversity of mobility behavior (indicator 8 and 9).
- We examine the statistical properties of these mobility indicators and compare them across cities. The results could reveal inherent variations among individual travelers as well as collective patterns that depict the tourism profile of each city.
- An eigendecomposition approach is introduced to better understand the interdependency of these mobility indicators in each city. The approach takes high-dimensional mobility features of individual travelers as input, and extract a series of principal components to describe the inherent variations across individuals as well as correlations among different mobility indicators.
- Based on the eigendecomposition results, we further employ a dimension reduction technique to describe the key characteristics of each traveler. Since the mobile phone dataset captures the nationality of each traveler, we use such information along with the results of dimension reduction to quantify the behavioral heterogeneity of travelers across different countries and regions.
- Empowered by the eigendecomposition results, we select, in each city, a few traveler groups with distinctive mobility patterns, such as travelers with — (1) a long duration of stay but low mobility diversity; (2) a long duration of stay and high mobility diversity; (3) a short duration of stay and low mobility diversity; (4) a short duration of stay but high mobility diversity. We map the spatial patterns of their activities and examine whether different traveler groups show varying spatial preferences when visiting a city.

We hope the study brings a synergy between classic human mobility analysis (Chen, Ma, Susilo, Liu, & Wang, 2016) and the emerging field of tourism big data (Li et al., 2018). The behavioral insights derived in this study can inspire new hypotheses of tourist travel behavior, and inform location recommendation and deployment of tourism services. The framework can be applied or extended to compatible datasets to understand travel behavior of tourists, residents, and special population groups in cities.

2. Study area and dataset

The mobile phone dataset used in this study was collected by one of the major cellular operators in South Korea. The dataset tracks the location traces of international travelers who visited any of the three cities — Jeonju, Gangneung and Chuncheon — during a period of one year (August 1st, 2017 - July 31st, 2018). Jeonju is a largest city in Jeollabuk-do Province that includes a provincial government building. Jeonju contains a number of touristic attractions. There are, for example, Jeonju hanok village where travelers can experience Korea traditional architectures, clothing and food. As a part of UNESCO Creative Cities Network program, Jeonju has been recognized as one of creative cities for gastronomy. Gangneung is one of top three cities in Gangwon-do Province. Gangneung is a popular tourism destination that contains a variety of natural resources such as sea and mountains. As such, ice hockey games for PyeongChang Olympics as well as a number of winter sports have been held in the city. Along with the coastal views, international baristas have developed coffee shops and community in the city (on the location of Gwaebangsan), which attracts a number of both residents and visitors. Chuncheon is the city where the provincial government building of Gangwon-do Province is located. Chuncheon is connected to Seoul well via a number of effective transportation systems such as several highways, express buses and trains. Since a new subway line between Chuncheon and Seoul has been established in 2012, it shows significant growth of travelers who conduct day trips. As a key travel attraction, Nami Island located in South West area of Chuncheon gains particular attentions from international travelers due to spillover effects of K-pop & drama culture.

In this dataset, each phone user's diary consists of records that document where an individual tended to stay and for how long. The locations were tracked at the level of cellphone towers. Each record tracks the unique ID of the user, the location he or she stayed (lng/lat of cellphone tower), as well as the starting and ending time that define the stay period. Table 1 shows an example of a phone user's records. Each row in the table represents one stay activity and the time periods between consecutive records refer to movements among locations. For example, the first two rows in Table 1 indicate that the visitor stayed at two different locations between [10:05:00–11:25:00] and [11:59:00–14:29:00] respectively, and a trip was conducted in between (i.e., [11:25:00–11:59:00]). The dataset, which was preprocessed by the data provider, tracks each user's whereabouts as a sequence of stays. The data format differs from typical types of mobile phone data, such as call detail records (Gonzalez et al., 2008; Xu et al., 2018) and mobile signaling data (Yan, Wang, Zhang, & Xie, 2018), which track users' geolocations at discrete time points. Note that the densities of cellphone towers in cities would reflect the spatial granularity of the dataset. To better understand their spatial arrangement, for each city, we measure the distance from each cellphone tower to its nearest peer. According to the result, the average nearest distances between cellphone towers are 250, 420 and 443 m, respectively for Jeonju, Gangneung and Chuncheon. Overall, the dataset provides an adequate spatio-temporal resolution for reliable estimations of tourist mobility patterns.

Table 1
Example of an individual's mobile phone records in the dataset.

User ID	Date	Starting Time	Ending Time	Longitude	Latitude
214 * *	2017–11–24	10:05:00	11:25:00	126. * * *	37. * * *
214 * *	2017–11–24	11:59:00	14:29:00	126. * * *	37. * * *
214 * *	2017–11–24	14:48:00	16:37:00	127. * * *	37. * * *
…	…	…	…	…	…
214 * *	2017–11–26	18:49:00	20:29:00	126. * * *	37. * * *
214 * *	2017–11–26	20:36:00	20:55:00	126. * * *	37. * * *

Fig. 1A shows the locations of the three cities in South Korea. According to the dataset, a total of 18,625 phone users visited Jeonju during the study period, compared to 33,219 for Gangneung, and 66,646 for Chuncheon. Fig. 1B to Fig. 1D demonstrate the daily number of visitors to the three cities. We observe a spike in Gangneung that matches with the period of the 2018 Winter Olympics (Feb 9–25, 2018). By that time, Gangneung was the city that hosted the indoor ice events, which serve as a major reason for the dramatic increase of incoming visitors. Since travelers might exhibit unique behavioral patterns during special events, we filter visitors between Jan 20, 2018 and Feb 26, 2018 — the period with notable increase of tourist visits — to minimize the impact of this mega event. This reduces the number of visitors to 15,095 for Gangneung.

Note that in each city we observe a small proportion of visitors with "gap days" (less than 10%). This could refer to travelers who visited a city more than once, or those who turned off their phones on certain days. In this analysis, we do not consider such visitors. This gives us a final dataset with 17,129 visitors for Jeonju, 14,052 for Gangneung, and 65,485 for Chuncheon.

Note that this dataset also documents the nationality of each traveler. In particular, the cellular operator has identified international travelers as inbound tourists who subscribed to their mobile roaming services in South Korea. Fig. 2 shows the nationality segmentation of travelers in the final dataset. In general, a few countries and regions account for a large proportion of tourist visits, but the actual rankings vary among the cities. For instance, the top three origins by travelers are United States, Mainland China, and Japan for both Jeonju (Fig. 2A) and Gangneung (Fig. 2B), but Malaysia, Hong Kong, and Singapore for Chuncheon (Fig. 2C). In Jeonju, the top 16 regions account for 90.5% of the total visits, compared to 93.2% for Gangneung, and 97.4% for Chuncheon.

3. Methodology

3.1. Derive individual mobility indicators from cellphone trajectories

In this study, we derive a collection of individual mobility indicators (IMIs) to characterize tourist behavioral patterns. Some of the indicators can be computed directly from raw mobile phone data, and we refer to them as low-level mobility indicators (LMIs). However, one issue related to mobile phone data is that location recordings depend on the cellphone towers that the mobile devices are connected to. Moreover, the connections could switch between nearby towers due to cellphone load balancing (Csáji et al., 2013). Therefore, cellphone towers might not represent an individual's meaningful activity locations. To address this issue, we introduce an anchor point extraction approach to further process each individual's cellphone trajectory. Several high-level mobility indicators (HMIs) are then derived from the processed cellphone trajectories.

A traveler's cellphone trajectory T in a given city can be represented as a list of tuples $T = \{(l_1, t_1^s, t_1^e), (l_2, t_2^s, t_2^e), \ldots, (l_n, t_n^s, t_n^e)\}$, where l_i denotes the cellphone tower location of the i^{th} record, and t_i^s and t_i^e denote the starting and ending time of the stay. We first compute the following low-level mobility indicators (LMIs):

- N_{day}: number of observation days
- S: time span of trajectory
- R_g: radius of gyration
- D: diameter of trajectory

N_{day} describes a traveler's length of stay in a city, and it is calculated as the total number of observation days in T. Considering that a traveler

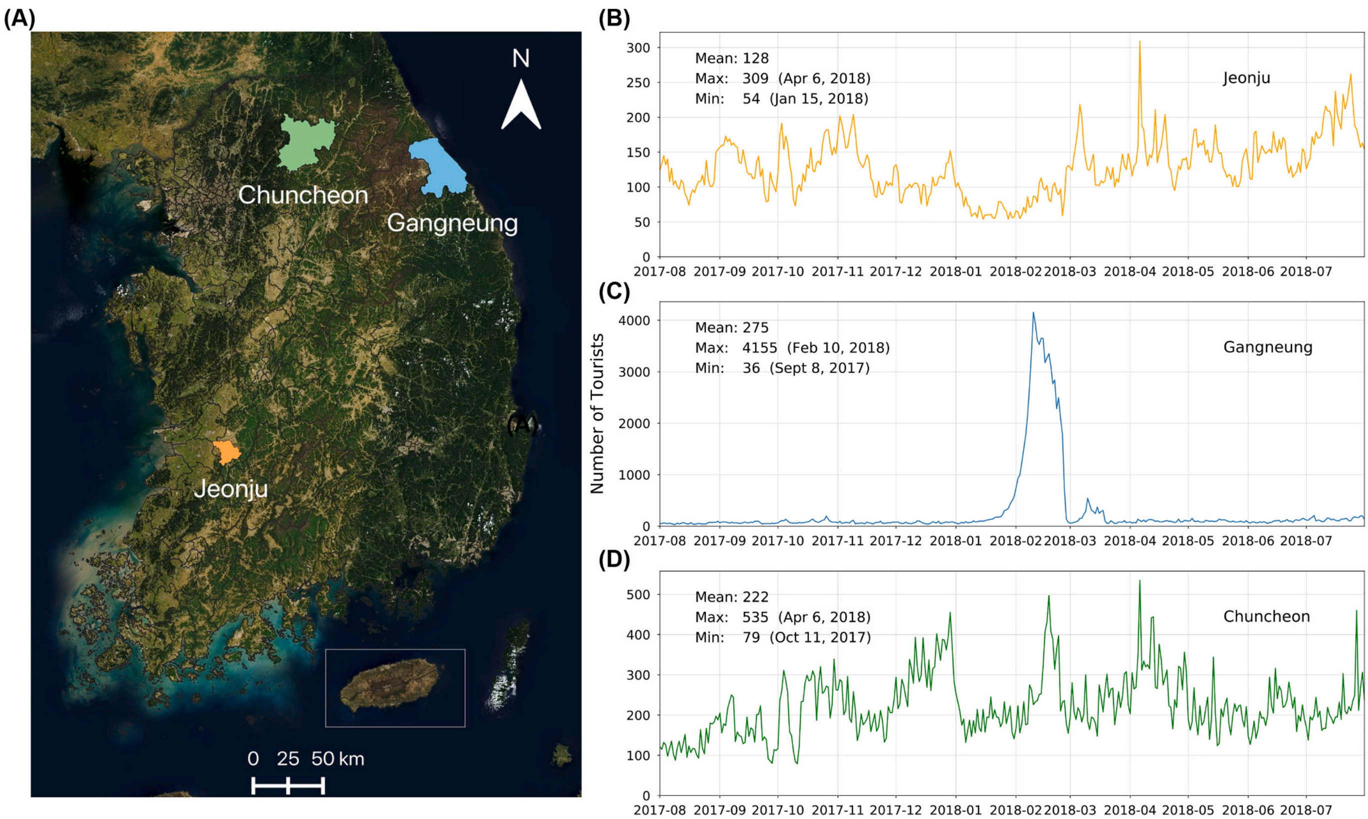


Fig. 1. (A) Locations of the three cities in South Korea; (B–D) daily number of visitors to each city.

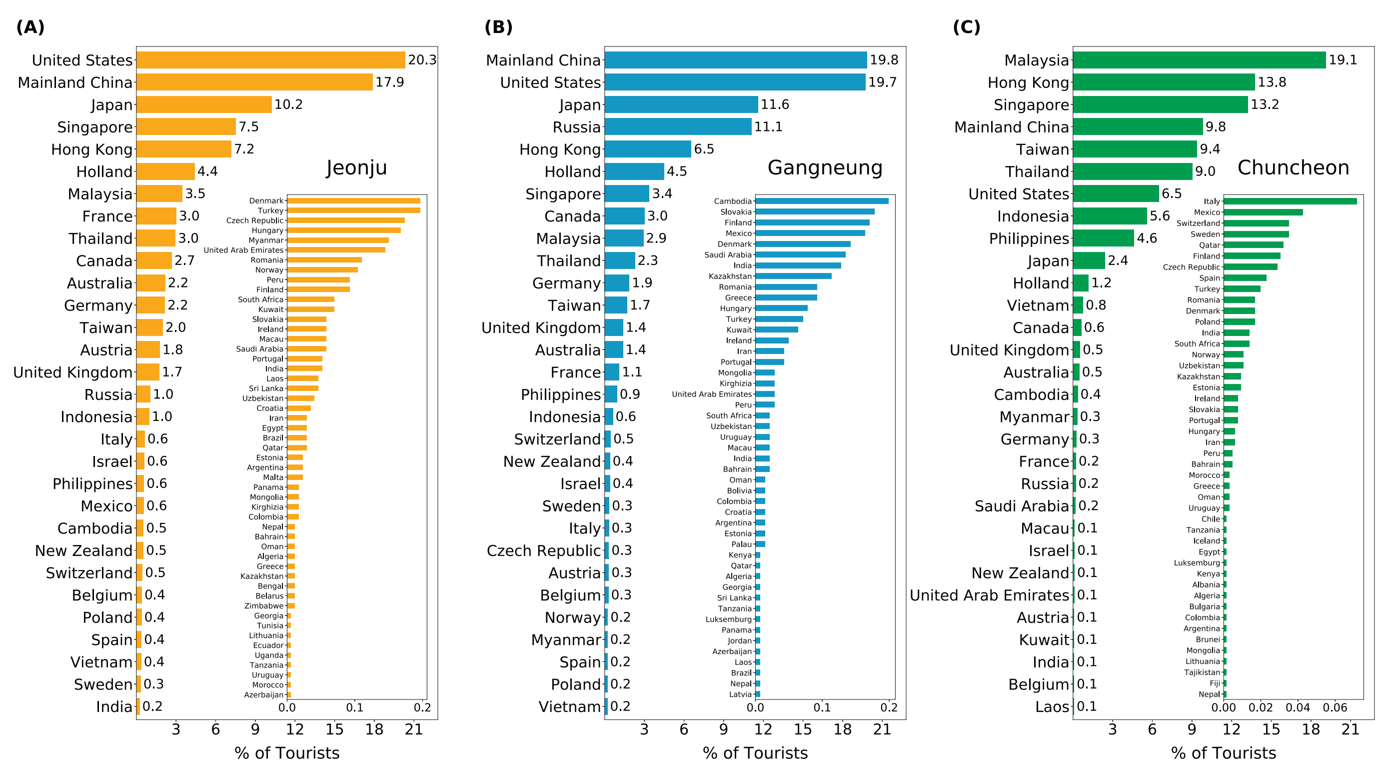


Fig. 2. Percentage of tourists by country or region in the final dataset: (A) Jeonju; (B) Gangneung; (C) Chuncheon. In this dataset, information of tourists from Mainland China, Hong Kong, Macau and Taiwan is provided separately.

could enter or leave a city at any time point of a day, we further compute the time span of the trajectory. Given trajectory T, the time span S is computed as the elapsed time between t_1^s and t_n^e, namely, the time span between the starting time of the first stay activity and the ending time of the last stay activity in the city. In other words, S is a more accurate estimation of a traveler's duration of stay.

Radius of gyration (R_g) has been widely used in existing research to quantify the spatial dispersion of a phone user's activities (Gonzalez et al., 2008; Xu et al., 2018). It is calculated as follows:

$$R_g = \sqrt{\frac{\sum_{i=1}^{n}\left(\overleftarrow{l_i} - \overleftarrow{l_c}\right)^2}{n}} \tag{1}$$

Here $\overleftarrow{l_i}$ denotes the location vector (i.e., x, y coordinates) of l_i, and $\overleftarrow{l_c} = \sum \overleftarrow{l_i}/n$ refers to the center of mass. R_g can be used to measure the spatial dispersion of a phone user's daily activities. A large value of R_g indicates a large activity space, while a small value suggests that the traveler's activities were concentrated within a confined geographic area. As a complement to R_g, we also compute the trajectory's diameter (D), which is measured as the maximum distance between all pairs of cellphone towers in T.

As mentioned above, to compute high-level mobility indicators (HMIs), we first introduce an anchor point extraction approach to further process travelers' cellphone trajectories. The concept of activity anchor point was often used in previous studies to denote a traveler's meaningful activity locations (Dijst, 1999; Schönfelder & Axhausen, 2003; Xu et al., 2016). The extraction of activity anchor points in this study works as follows. Given $T = \{(l_1, t_1^s, t_1^e), (l_2, t_2^s, t_2^e), \ldots, (l_n, t_n^s, t_n^e)\}$, we first compute the total time that the individual stayed at each cellphone tower and sort them in descending order. We then select the cellphone tower with the longest stay duration and group all other towers within a distance of Δd of the selected tower into a cluster. Among the remaining cellphone towers that are not assigned to any cluster, we select the next one with the longest duration of stay and perform the same grouping process. We repeat this procedure until all cellphone towers in T are processed. We refer to these clusters as *activity anchor points*. Regarding the choice of Δd, given that the average nearest distance between cellphone towers in three cities are all below 500 m, we set Δd as 500 m for all three cities.

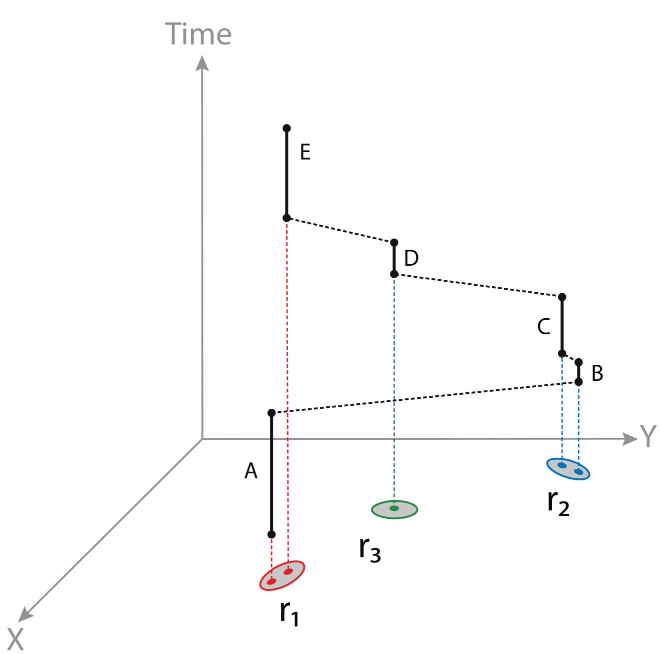


Fig. 3. Extraction of activity anchor points from an individual's cellphone trace T. Each vertical segment corresponds to one mobile phone record in the raw data. A to E denote five distinct cellphone towers traversed by T. r_1, r_2 and r_3 denote the extracted activity anchor points.

Fig. 3 demonstrates the process of anchor point extraction. As shown in the example, an individual's cellphone trace T traverses through five distinct cellphone towers in chronological order: $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$. According to our method, the cellphone tower with the longest duration of stay (A) is selected first. Since tower E is within 500 m of A, it is grouped with A to form the first anchor point r_1. Then, the next tower with the longest duration of stay (C) is selected and grouped with B to form a new anchor point (r_2), leaving D alone to form the third (r_3). Thus, the individual's cellphone trace can be represented as a sequence of activity anchor points: $r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_1$.

In doing so, we are able to transform the cellphone trajectory T into a sequence of stays at the anchor point level $T' = \{(r_1, t_1^s, t_1^e), (r_2, t_2^s, t_2^e), \ldots, (r_n, t_n^s, t_n^e)\}$, where r_i denotes the anchor point associated with the i^{th} record in T. Each anchor point r in T' maps to one of more cellphone towers in close proximity, therefore capturing a more realistic representation of a traveler's activity locations.

We further compute the following HMIs from each traveler's processed cellphone trajectory T':

- A_1: total number of activity anchor points
- A_2: daily number of activity anchor points
- N_{od}: total number of origin-destination trips
- H_1: activity entropy
- H_2: travel diversity

The first high-level mobility indicator, A_1, measures the total number of activity locations visited by a traveler:

$$A_1 = |set(r_1, r_2, \ldots, r_n)| \tag{2}$$

A large value of A_1 indicates that the traveler's activities were distributed across a variety of locations. Note that we also compute the number of activity anchor points visited by the traveler during each observation day and compute the average value, A_2. In other words, A_2 describes how many activity locations on average a traveler visited per day.

Origin-destination (OD) trips describe a traveler's movements between activity locations. In this study, an OD trip is defined based on two sequential stays in T' (e.g., a trip with origin at r_i and destination at r_{i+1}). We introduce N_{od} to measure the total number of intra-urban trips conducted by a traveler.

Activity entropy and travel diversity were used in existing studies to quantify the diversity of an individual's mobility behavior (Pappalardo et al., 2016; Xu et al., 2018). Given a vector $\{p_1, p_2, \ldots, p_{A1}\}$, where p_j denotes the proportion of duration of stay at the j^{th} anchor point in T', the activity entropy is calculated as:

$$H_1 = -\sum_{j=1}^{A_1} p_j log(p_j) \tag{3}$$

Note that $\sum p_j = 1$. A large value of H_1 indicates a high level of activity diversity, while a low value indicates that the traveler spent most of the time at few activity locations.

The travel diversity (H_2) measures how (un)evenly a traveler's trips distributed among different activity locations. Let E denote all the possible origin-destination pairs (without considering direction) in T', the travel diversity is calculated as:

$$H_2 = -\sum_{k \in E} p_k' log(p_k') \tag{4}$$

where p_i' is the probability of observing a trip between the k^{th} origin-destination pair. Note that $\sum p_k' = 1$. Larger values of H_2 indicate higher movement diversity.

3.2. Characterize tourist mobility patterns through eigendecomposition

Eigendecomposition is an important technique for understanding structures in high dimensional data. It has been used in previous studies to uncover rhythms of human activities in cities (Eagle & Pentland, 2009; Reades, Calabrese, & Ratti, 2009; Xu et al., 2019). In this study, we employ the technique to quantify the structural variations of tourist mobility patterns. The approach takes high-dimensional mobility features as input, and extracts a series of principal components (PCs) to describe the inherent data structures. Note that in this research, eigendecomposition is performed separately for each city (Jeonju, Gangneung and Chuncheon).

Given one particular city, the mobility signature of an individual i can be represented as a feature vector $\{x_{i,\,1}, x_{i,\,2}, \ldots, x_{i,\,9}\}$, where $x_{i,\,j}$ maps to the value of j^{th} mobility indicator. Since eigendecomposition is sensitive to the scale of data input, we perform data normalization for each mobility dimension using Min-Max scaling:

$$r_{i,j} = \frac{x_{i,j} - Min_j}{Max_j - Min_j} \tag{5}$$

where $r_{i,\,j}$ denotes the value after normalization, and Min_j and Max_j represent the minimum and maximum value of the j^{th} indicator, respectively. This allows us to transform each mobility indicator into range of [0,1].

The resultant vector is represented as $\Psi_i = \{r_{i,\,1}, r_{i,\,2}, \ldots, r_{i,\,9}\}$, where $r_{i,\,j}$ denotes the j^{th} indicator of the i^{th} individual after normalization.[1] By averaging Ψ_i across all individuals, we obtain the mean feature vector:

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\Psi_i \tag{6}$$

where N denotes the total number of travelers in the city.

A matrix M of size $N \times 9$ is then introduced, with each row being $\Phi_i = \Psi_i - u$ that describes the deviation of the individual's mobility patterns from the mean feature vector:

$$M = \begin{pmatrix} r'_{1,1} & r'_{1,2} & \cdots & r'_{1,9} \\ r'_{2,1} & r'_{2,2} & \cdots & r'_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ r'_{n,1} & r'_{n,2} & \cdots & r'_{n,9} \end{pmatrix} \tag{7}$$

The covariance matrix C of M is then computed as:

$$C = \frac{1}{N}\sum_{i=1}^{N}\Phi_i^T\Phi_i = \frac{1}{N}M^TM \tag{8}$$

We then derive the eigenvectors v_1, v_2, …, v_9 and associated eigenvalues λ_1, λ_2, …, λ_9 from matrix C, with λ_j ranked in descending order. The eigenvectors, which are orthogonal to each other, represent the principal components (PCs) of the matrix M, and the eigenvalues describe the variance explained by each PC. Through eigendecomposition, the mobility signature of individuals can be represented as the linear combination of these PCs (or eigenvectors):

$$\Psi_i = u + A_iV \tag{9}$$

where

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,9} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,9} \end{pmatrix} \tag{10}$$

[1] A vector in this article corresponds to a row vector unless otherwise specified.

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_9 \end{pmatrix} \tag{11}$$

Here, $a_{i,\,j}$ denotes the coefficient of the j^{th} PC for the i^{th} individual. Each row of A contains the loadings of PCs for one individual. Given an individual i, the coefficient of the j^{th} PC indicates to what extent the traveler's mobility behavior deviates from the city average (i.e., the value of the j^{th} indicator in u).

The results of eigendecomposition can reveal structures in high-dimensional mobility features, and suggest whether dimension reduction can facilitate the understanding of tourist mobility patterns. For instance, if the top few PCs are able to explain a significant proportion of total variance in M, it means the mobility signatures of individuals (Ψ_i) can be effectively represented and interpreted through the first few eigenvectors (λ) and the associated coefficients (a).

4. Analysis results

4.1. Statistical properties of the mobility indicators

In this section, we examine the statistical properties of the nine mobility indicators. The results are summarized in Table 2. Regarding the number of observation days (N_{day}), visitors on average spent 2.25 days in Jeonju, compared to 1.86 days in Gangneung, and 1.16 days in Chuncheon. Most of the travelers would stay in a city for less than one week. The time span (S) delineates length of stay at a finer temporal resolution. On average, travelers spent 28.87 h in Jeonju and 20.65 h in Gangneung, while the number drops to 5.68 h for Chuncheon. By comparing the mean, median and 95 percentile of S, we can see that Jeonju and Gangneung were mixed with overnight stayers and same-day visitors. However, Chuncheon were dominated by temporary visitors, and a lot of them stayed in the city for only one or two hours.

By examining the total number of activity anchor points (A_1), we find that an average traveler visited 2.51 activity locations during the stay in Jeonju, compared to 2.36 in Gangneung, and 1.71 in Chuncheon. However, the daily number of activity anchor points (A_2) shows similar distributions across three cities. On the one hand, it suggests a higher diversity of destination choices for individual travelers in Jeonju and Gangneung. On the other hand, it shows that the rate of exploration — reflected by the number of activity locations visited per day by an average traveler — remains constant across cities. From the perspective of travel movements, travelers conducted 3.17 trips on average in Jeonju, compared to 2.44 in Gangneung, and 0.95 in Chuncheon. The median value for Chuncheon is zero, meaning that many travelers came to visit one specific location in the city and then left in the same day (as N_{od} only measures intra-urban trips).

Both D and R_g measure the spatial extent of tourist activities. A large variation is observed in all three cities. The average activity spaces were largest for travelers in Gangneung, followed by Chuncheon and Jeonju. We do not observe a consistent relationship between these values (e.g., mean value of D and R_g) and the total area of cities, which means the size of a city is not a decisive factor that shapes how far tourists traveled. Regarding mobility diversity, travelers show higher activity (H_1) and travel diversities (H_2) in Jeonju and Gangneung than in Chuncheon.

4.2. Structural variations of tourist mobility

The results in Table 2 enable a cross-city comparison of each mobility indicator. However, the interrelationships among different indicators and individual variations in each city remain unknown. For instance, does a longer length of stay mean more trips and activity locations? Which city has the highest level of individual mobility diversity? In this section, we report the results of eigendecomposition to answer these questions.

Table 2
Statistical properties of the nine mobility indicators.

Indicator	Notation	City	Mean	Median	Std	75%	95%
Number of observation days	N_{day}	Jeonju	2.25	2	2.4	2	6
		Gangneung	1.86	1	1.86	2	5
		Chuncheon	1.16	1	0.96	1	2
Time span (hour)	S	Jeonju	28.87	11.7	57.7	25.58	119.37
		Gangneung	20.65	4.23	44.05	20.45	92.04
		Chuncheon	5.68	1.63	22.83	3.67	19.17
Total number of activity anchor points	A_1	Jeonju	2.51	2	2.04	3	6
		Gangneung	2.36	2	2	3	6
		Chuncheon	1.71	1	1.22	2	4
Daily number of activity anchor points	A_2	Jeonju	1.75	1.5	0.9	2	3.5
		Gangneung	1.76	1.09	1.07	2	4
		Chuncheon	1.61	1	0.96	2	4
Total number of trips	N_{od}	Jeonju	3.17	1	6.55	4	13
		Gangneung	2.44	1	5.52	3	11
		Chuncheon	0.95	0	2.58	1	4
Diameter (km)	D	Jeonju	2.14	0.93	2.55	3.95	7.18
		Gangneung	4.81	0.72	7.31	6.37	20.09
		Chuncheon	3.57	0	6.23	6.2	16.83
Radius of gyration (km)	R_g	Jeonju	0.73	0.36	0.9	1.22	2.46
		Gangneung	1.78	0.28	2.8	2.3	7.88
		Chuncheon	1.54	0	2.7	2.62	7.88
Activity entropy	H_1	Jeonju	0.37	0.27	0.4	0.65	1.12
		Gangneung	0.36	0.1	0.45	0.67	1.25
		Chuncheon	0.3	0	0.42	0.64	1.16
Travel diversity	H_2	Jeonju	0.42	0	0.65	0.69	1.79
		Gangneung	0.38	0	0.63	0.69	1.75
		Chuncheon	0.17	0	0.41	0	1.1

Table 3
Mobility variance explained by the principal components ($\times 10^{-2}$).

City	Total Variance	Average Variance	Variance explained by			
			$1^{st}PC$	$2^{nd}PC$	$3^{rd}PC$	$4^{th}PC$
Jeonju	16.6	1.84	10.15 (61%)	3.86 (23%)	1.32 (8%)	0.83 (5%)
Gangneung	16.8	1.87	10.87 (65%)	2.88 (17%)	1.96 (12%)	0.68 (4%)
Chuncheon	11.3	1.26	8.95 (79%)	1.15 (10%)	0.76 (7%)	0.30 (3%)

Table 3 presents the total variance of the mobility signatures for each city and the percentage of variance explained by the top few principal components (PCs). The total variance for Jeonju and Gangneung are similar, meaning that mobility variations among individuals are comparable in the two cities. A lower value for Chuncheon suggests that travelers were less diverse in terms of spatio-temporal activity patterns.

The 1^{st} PCs in all three cities explain a significant proportion of total variance. This indicates a structural variation of mobility behavior among individuals and correlations among the mobility indicators. The first two PCs are able to explain over 80% of the total variance in each city. That means the high-dimensional characteristics of individuals can be effectively represented by the linear combination of the top two PCs.

As shown in Fig. 4, the 1^{st} PCs for Jeonju (Fig. 4A) and Gangneung (Fig. 4E) exhibit similar characteristics and explanatory power. It shows that the spatial extent of activities (D and R_g) and mobility diversity (H_1 and H_2) are important features that distinguish individual travelers. The values of all elements in these PCs are above zero, which indicates a positive correlation among the mobility indicators. More intuitively, it suggests that travelers who stayed longer in the city tended to have larger activity spaces and higher mobility diversities. The 1^{st} PC for Chuncheon explains 79% of the total variance (Fig. 4I). It suggests that the daily number of activity anchor points (A_2), size of activity space (D and R_g), and mobility diversity (H_1 and H_2) are critical dimensions that explain individual variations. It also shows that individual variations cannot be well explained by the length of stay (N_{day} and S) and total number of trips (N_{od}). This is consistent with the findings in Section 4.1.

The 2^{nd} PCs for Jeonju (Fig. 4B) and Gangneung (Fig. 4F) depict another facet of travelers' mobility patterns. The result suggests negative correlations among certain indicators for some travelers. In particular, there exit individuals who stayed in the city for a long time but visited few places per day, which also explain their confined activity spaces and low activity entropy. There could also be travelers who visited many places during a short period of stay. The 2^{nd} PC for Chuncheon has a limited explanatory power (Fig. 4J), and it suggests that some travelers can have high (or low) mobility diversities but small (or large) activity spaces. The results in this section suggest that eigen-decomposition can be effective in revealing the multidimensional characteristics of tourist mobility and structural variations among individuals.

4.3. Mobility patterns across nationalities

Since the top two PCs are able to explain over 80% of the mobility variance, we use the linear combination of these two PCs to express key characteristics of tourist mobility. Performing this dimension reduction gives us a pair of coefficients for each traveler, with each coefficient representing the loading of the corresponding PC. Fig. 5A shows the joint distribution of coefficients for travelers in Jeonju. Areas with darker colors indicate more individuals with the given combination of coefficients. The combination of PC1 and PC2 coefficients can effectively describe the major characteristics of individuals. For instance, User 1 has a high coefficient of PC2 but not necessarily for PC1. Interpreting these values by referencing the two PCs (Fig. 4A and Fig. 4B) suggests that the user had a long duration of stay but low mobility diversity (H_1 and H_2) and a confined activity space (D and R_g). This interpretation is reaffirmed by comparing the user's mobility indicators with the population average (Fig. 5B). User 2 has high coefficients for both PC1 and PC2, indicating that the traveler had a long duration of stay with a large activity space and high mobility diversity. User 3 has negative coefficients for both PC1 and PC2, which correspond to a short duration of stay with low mobility diversity. User 4 has a positive PC1 but negative PC2 coefficient. Although the traveler stayed for only half of a day in Jeonju, he managed to conduct nine trips, linking eight distinct locations that form a large activity space.

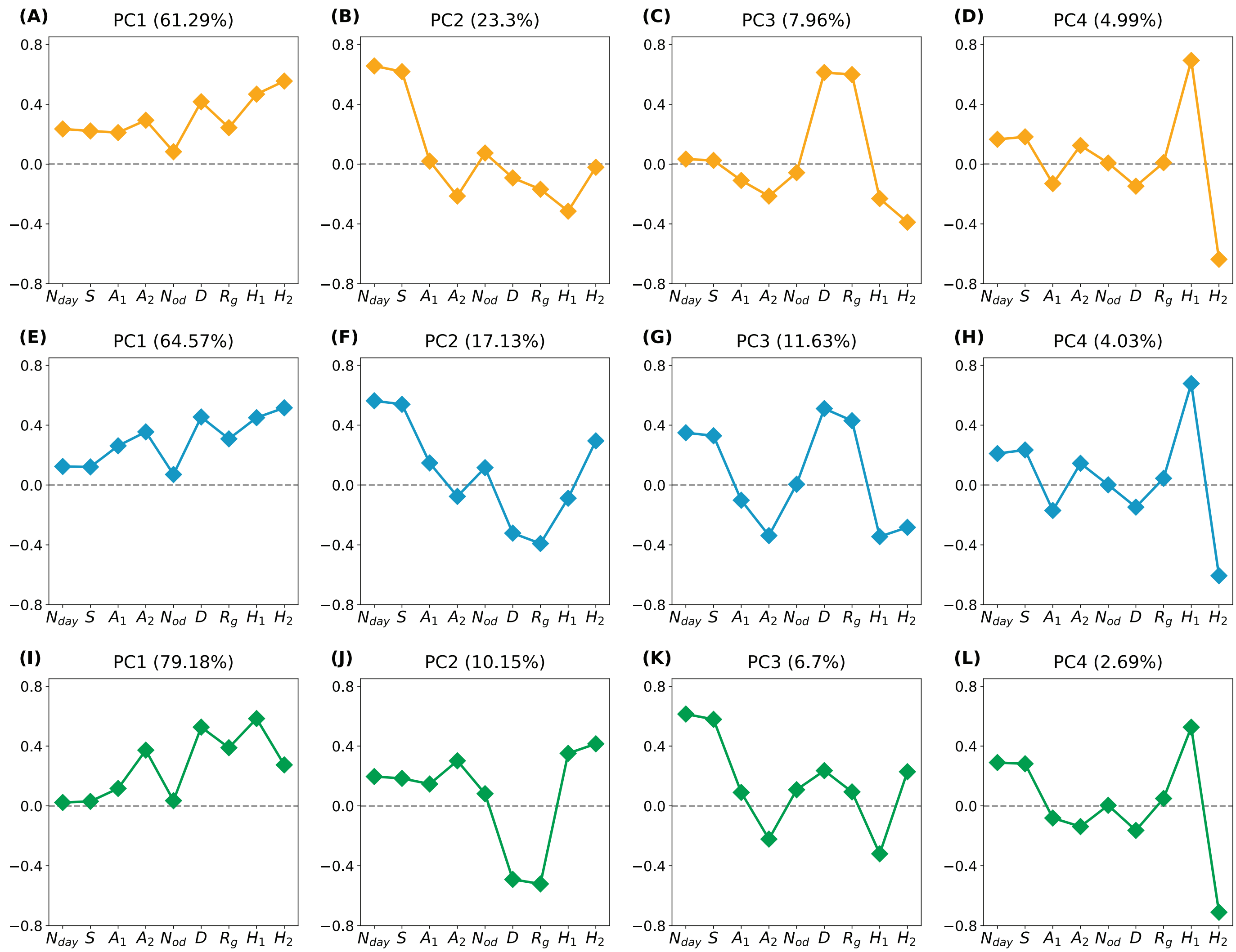
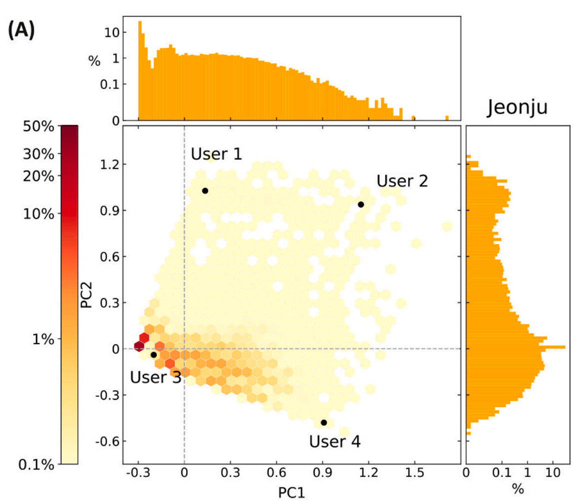


Fig. 4. Results of eigendecomposition: (A-D) Jeonju; (*E*-H) Gangneung; (I-L) Chuncheon.



(B)

Indicators	Mean	1	2	3	4
N_{day}	2.25	14	16	1	1
S	28.87	333.63	363.58	2.07	11.63
A_1	2.51	2	13	1	8
A_2	1.75	1.07	2.69	1	8
N_{od}	3.17	2	45	0	9
D	2.14	1.44	9.76	0.33	5.29
R_g	0.73	0.20	1.70	0.18	1.85
H_1	0.37	0.02	0.80	0	1.76
H_2	0.42	0	2.39	0	2.04

Fig. 5. (A) Joint distribution of PC1 and PC2 coefficients for travelers in Jeonju; (B) mobility indicators of four selected travelers.

Using this technique, we can further compare travelers' mobility patterns across nationalities. As mentioned in Section 2, the top 16 countries or regions account for more than 90% of the travelers in each city. Therefore, we focus on these regions in this analysis. Fig. 6 shows the joint distribution of PC1 and PC2 coefficients for travelers by country or region in Jeonju. Travelers from areas such as United States, Mainland China, Holland, Canada and Germany exhibited a high level of behavioral diversity. In particular, travelers from Mainland China cover

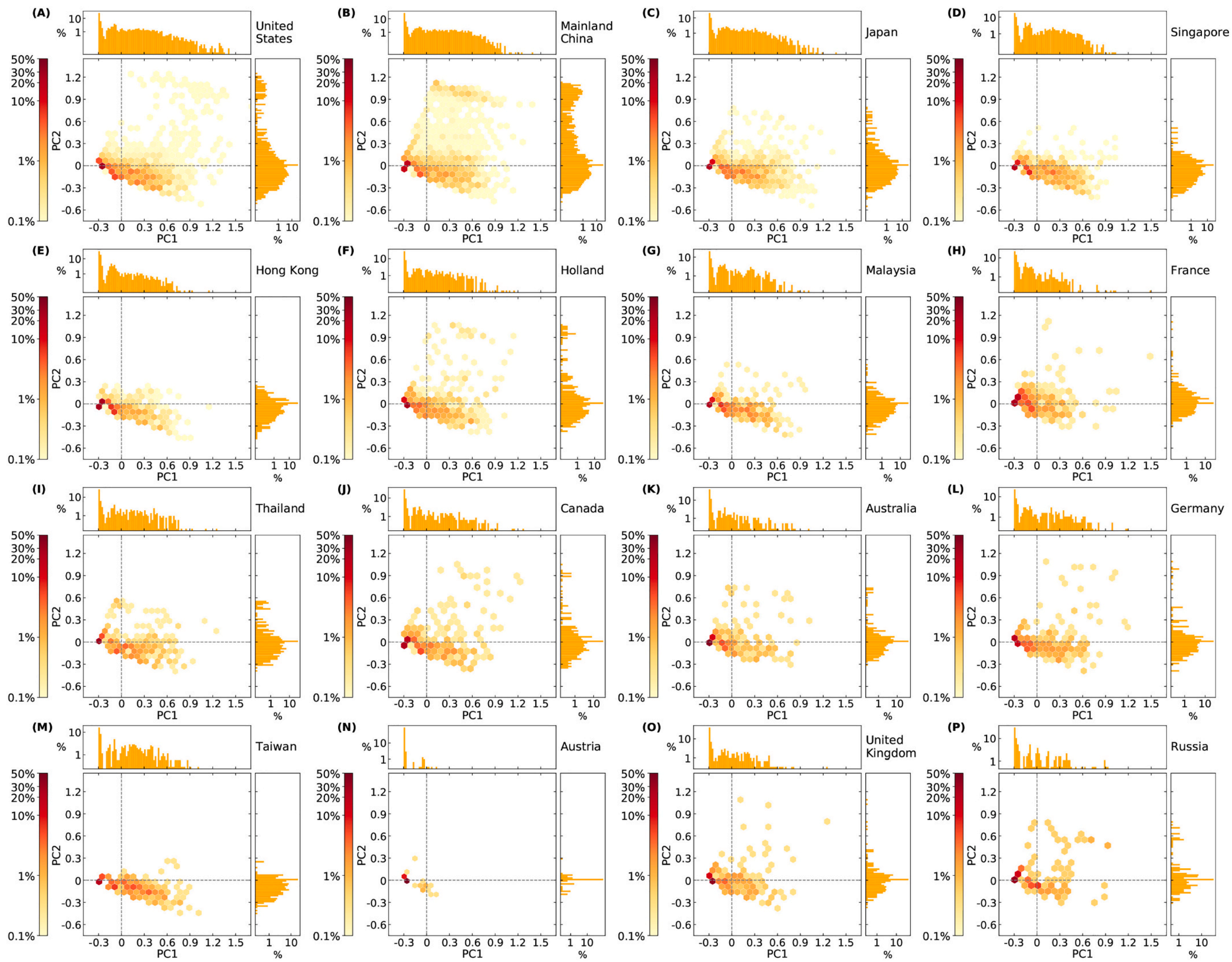


Fig. 6. Joint distribution of PC1 and PC2 coefficients for travelers by country or region in Jeonju.

a large area across the four quadrants (Fig. 6B), which indicates the existence of four distinctive types of visitors shown in Fig. 5. Singapore, Hong Kong, Malaysia and Taiwan show similar distribution patterns. Although travelers from these regions did not stay for long in Jeonju, they showed diverse travel and activity patterns. A low variance is observed among the Austrians (Fig. 6N). Most of the visitors are clustered around (−0.3,0), indicating uniform travel behavioral patterns.

Fig. 7 shows the joint distribution of coefficients for travelers by country or region in Gangneung. Since the top two PCs of Jeonju and Gangneung share similar characteristics (Fig. 4), the distributions here can be interpreted in a similar way. For most of the countries and regions, a significant proportion of travelers is observed in the second and third quadrants, with PC2 coefficients centered around zero. This indicates that many visitors tended to spend a short time visiting few places in Gangneung. However, we do observe many travelers in both the first and fourth quadrants. Its suggest a high diversity of travelers in terms of the locations visited, number of trips conducted, and size of activity space. Meanwhile, the two quadrants reveal a dichotomy of these travelers in their length of stay.

Fig. 8 shows the results for travelers who visited Chuncheon. Compared to the above two cities, Chuncheon has few travelers with a long duration of stay. Most of the nationalities show similar distribution patterns, with tourists widely distributed along the horizontal axis (PC1). Since PC1 explains 79% of the behavioral heterogeneity in Chuncheon (Fig. 4I), the result suggests that most of the nationalities are dominated by travelers with a brief stay, but the mobility patterns among these travelers could vary substantially. We also observe some countries with small interpersonal variations (e.g., Philippines, Vietnam), which indicate similar behavioral patterns among travelers from these countries. Note that for ease of interpretation, we have produced a four-user example for both Gangneung (Fig. A.1) and Chuncheon (Fig. A.2). Readers could refer to the appendix for more details.

Finally, as a high-level summary, we compute the mobility variance among travelers by country or region, similar to what we compute for the overall tourist population (see Table 3). The result is summarized in Fig. 9. In general, no consistent relationship is observed across nationalities. Travelers from certain countries such as United States and Canada show diverse travel patterns in all three cities. For some other areas, such as Mainland China, Holland, Singapore and Hong Kong, travelers are more diverse when exploring certain cities while less diverse when visiting others. This indicates that the mobility characteristics of travelers are jointly affected by their originating countries and the destination cities they tended to visit.

4.4. Spatial patterns of activities for selected traveler groups

The results so far have demonstrated substantial mobility variations among travelers. A follow-up question worth investigating is whether

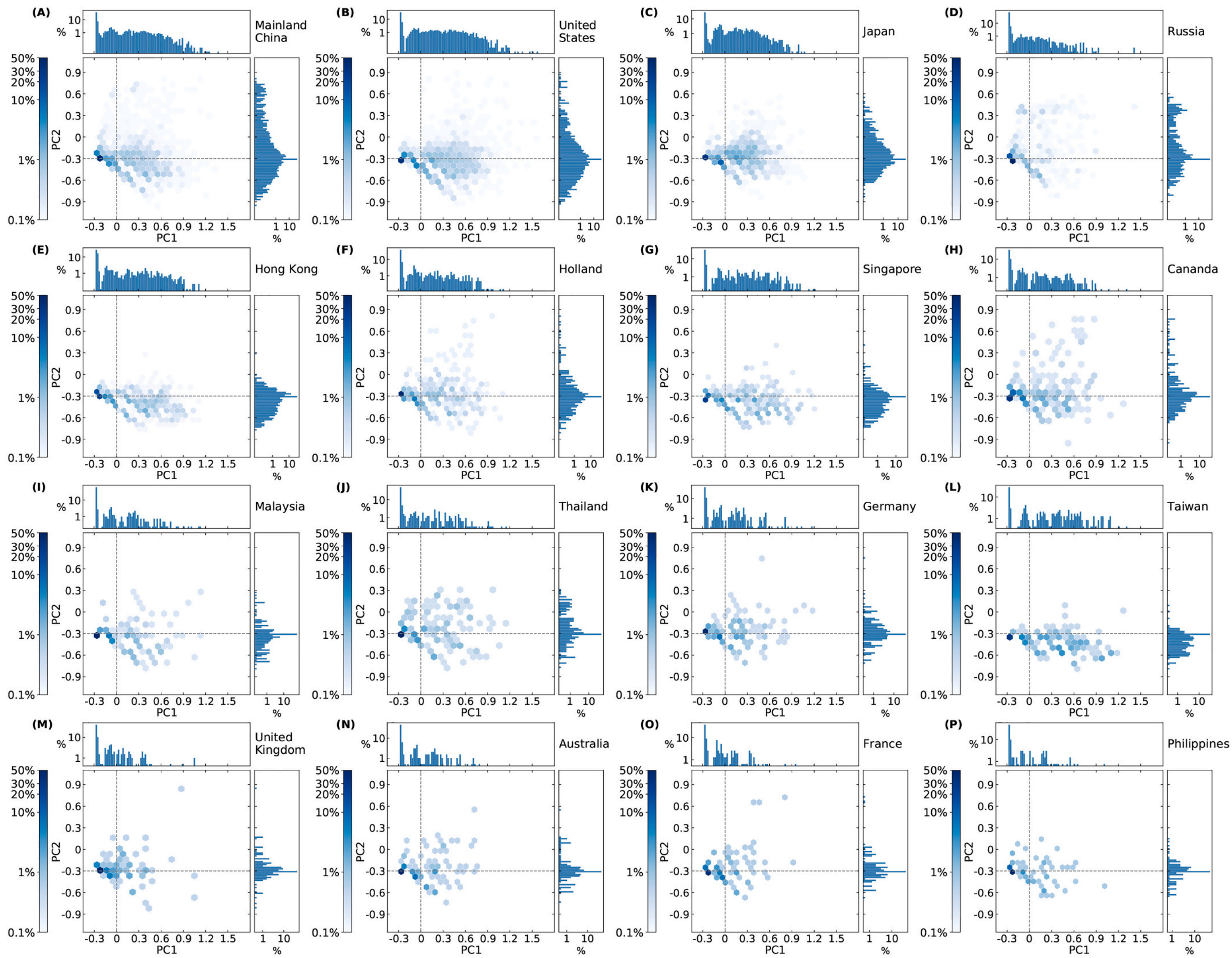


Fig. 7. Joint distribution of PC1 and PC2 coefficients for travelers by country or region in Gangneung.

travelers with different mobility characteristics have similar spatial preferences when visiting a city. In this section, we select, for each city, a few traveler groups with distinctive mobility patterns and explore the spatial patterns of their activities. Fig. 10 illustrates the selection criteria for these traveler groups. For Jeonju, as shown in Fig. 10A, we select four types of travelers that highly resemble the mobility characteristics of the four-user examples illustrated in Fig. 5:

- Type 1: Travelers with a long duration of stay, but low mobility diversity and a confined activity space (192 visitors)
- Type 2: Travelers with a long duration of stay and high mobility diversity (60 visitors)
- Type 3: Travelers with a short duration of stay and low mobility diversity (6647 visitors)
- Type 4: Travelers with a short duration of stay but high mobility diversity (64 visitors)

The same selection criterion is used for Gangneung. This gives us 81 visitors for type 1, 29 visitors for type 2, 7180 visitors for type 3, and 93 visitors for type 4. The interpretations of their mobility characteristics are also similar to those defined for Jeonju. Considering that PC1 for Chuncheon accounts for 79% of the total mobility variance, we select two types of travelers at the two sides of this axis. Type 1 contains 42,614 visitors, who exhibited a confined activity space and low mobility diversity during their travels. Type 2 includes 2523 visitors with a generally large activity space and high mobility diversity. Note that the average time span of these two traveler groups are 2.2 h and 10.3 h, respectively, largely because most of the travelers did not stay for long when visiting Chuncheon. To better understand the spatial preferences of overnight stayers, we incorporate a type 3 by selecting travelers with a time span greater than 24 h (2435 visitors in total).

Given a selected traveler group in a city, to map the spatial distribution of their activities, we revisit their cellphone trajectories at the anchor point level (T') and compute the total number of times each anchor point was visited or used. Note that individual activities are identified at the level of activity anchor points, which could consist of one or several cellphone towers. This introduces an issue when we want to map the density of activities at the level of cellphone towers. To simplify the mapping process, for each activity anchor (of each individual), we identify its representative cellphone tower, defined as the cellphone tower in the anchor point with the highest stay duration. The activity is always allocated to the representative cellphone tower when counting the frequency of activities. Iterating the procedure through all the cellphone trajectories gives us the total number of times each cellphone tower was used by this traveler group.

Another issue is that cellphone towers do not capture the exact locations of tourist activities. In other words, travelers' activities observed at a cellphone tower could occur in the vicinity of that tower due to limited spatial resolution of the data. To tackle this, we adopt uniform hexagons (with a side length of 500 m) as the spatial unit to perform the

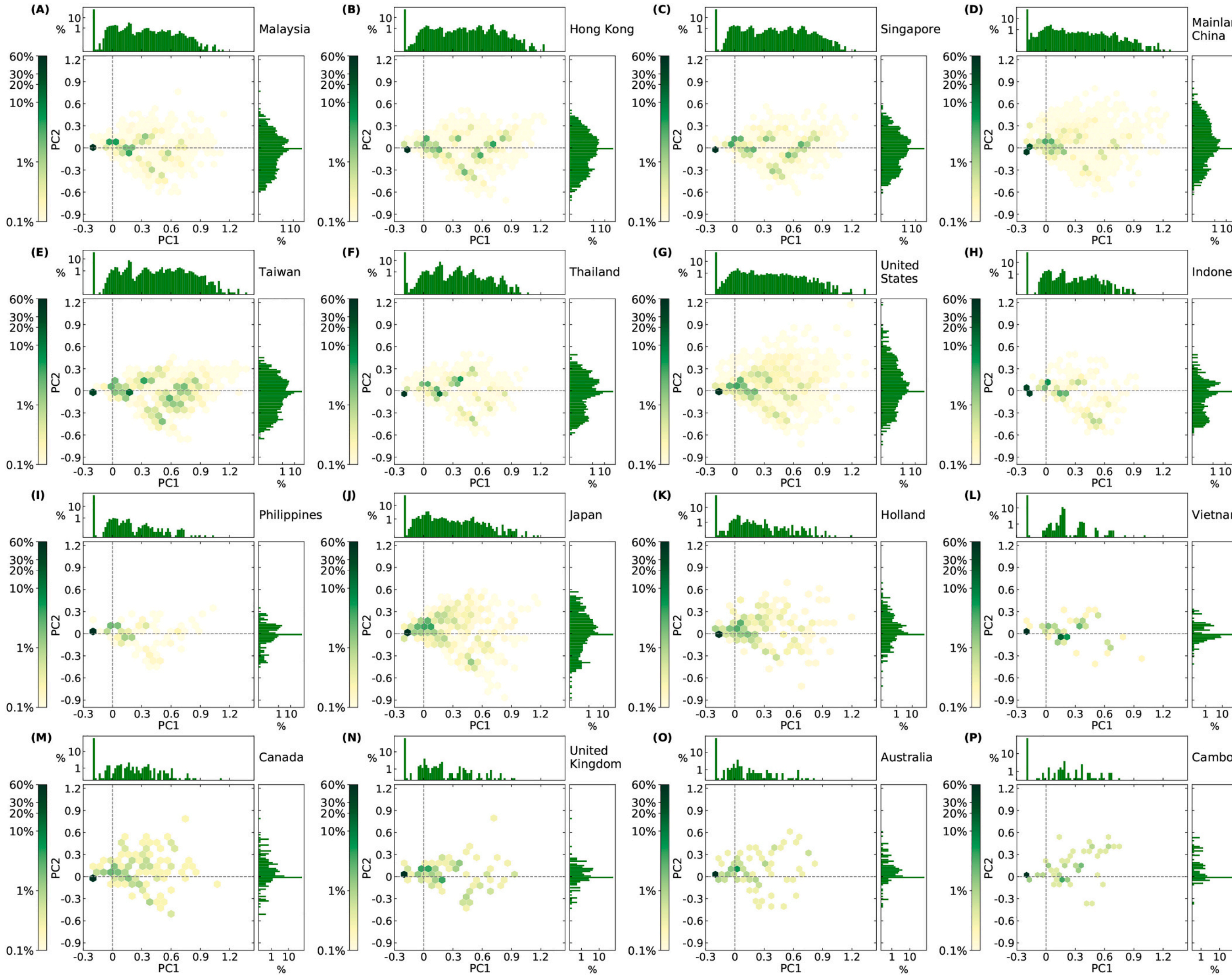


Fig. 8. Joint distribution of PC1 and PC2 coefficients for travelers by country or region in Chuncheon.

spatial mapping. To transform the visitation frequency from cellphone towers to hexagons, we use Thiessen polygons to approximate cellphone towers' service areas. Since a Thiessen polygon could overlap with multiple hexagons, we clip each Thiessen polygon into sub units and assign visitation frequency to each unit based on the ratio of its area and the size of the corresponding Thiessen polygon. We then compute the visitation frequency at the hexagon level by aggregating the numbers of all the concerning sub units. Readers could refer to Fig. B.3 in the appendix for an example of the calculation.

Fig. 11 shows the spatial patterns of activities for the four traveler groups in Jeonju. Hexagons with darker colors denote areas with more tourist visits. Type 1 visitors have long stay durations but low mobility diversity. As shown in Fig. 11A, their activities were clustered around two universities in Jeonju (Jeonju University & Chonbuk National University). The activities of type 2 visitors show a more uniform distribution in the central part of Jeonju (Fig. 11B). Besides activities around Chonbuk University, a substantial amount of visits occurred around Jeonju Hanok Village, a major attraction that is famous for its traditional buildings. Type 3 visitors had a brief stay in Jeonju and they also have a low level of mobility diversity. They serve as the most representative traveler group to the city. As shown in Fig. 11C, most of their activities took place around Jeonju Hanok Village and nearby historical sites such as the Royal Portrait Museum. Some visits were also observed at the Jeonju Express & Intercity Bus Terminal, which may indicate their usage of this transportation hub. Type 4 visitors show similar spatial preferences compared to type 3, and more activities were observed beyond the Hanok Village. The area around Jeonju station — a station of a major intercity railway line (Jeolla Line) in South Korea — also received many visits. It is likely that many visitors in this group have used this transportation service. The results in Fig. 11 demonstrate substantial variations of location choices among different traveler groups.

Fig. 12 shows the results for Gangneung. Type 1 and type 2 visitors are similar in duration of stay, but differ in the level of mobility diversity. Interestingly, the activities of these two groups show similar spatial distributions. Hot spots were observed around the Gangneung Olympic Park, the Central and Seongnam Market, and the Gyeongpodae Beach (Fig. 12A and Fig. 12B). Although individuals in type 2 tended to visit more places than type 1 visitors, at the collective level, they show very similar spatial preferences. Type 3 visitors have a short stay duration and low mobility diversity. Although they did not stay for long in the city, their collective activity patterns were scattered (Fig. 12C). Several areas of interests were observed along the coast (e.g., Tyumonsin Ko, Gyeongpodae Beach, Anmok Beach, Gwaebangsan, and Jeongdongjin Beach). Type 4 visitors have a high level of mobility diversity, meaning that individuals in this group visited more places than type 3 on average. Surprisingly, their activities, at the collective level, were less scattered (Fig. 12D). Their activities were concentrated in Jeongdongjin

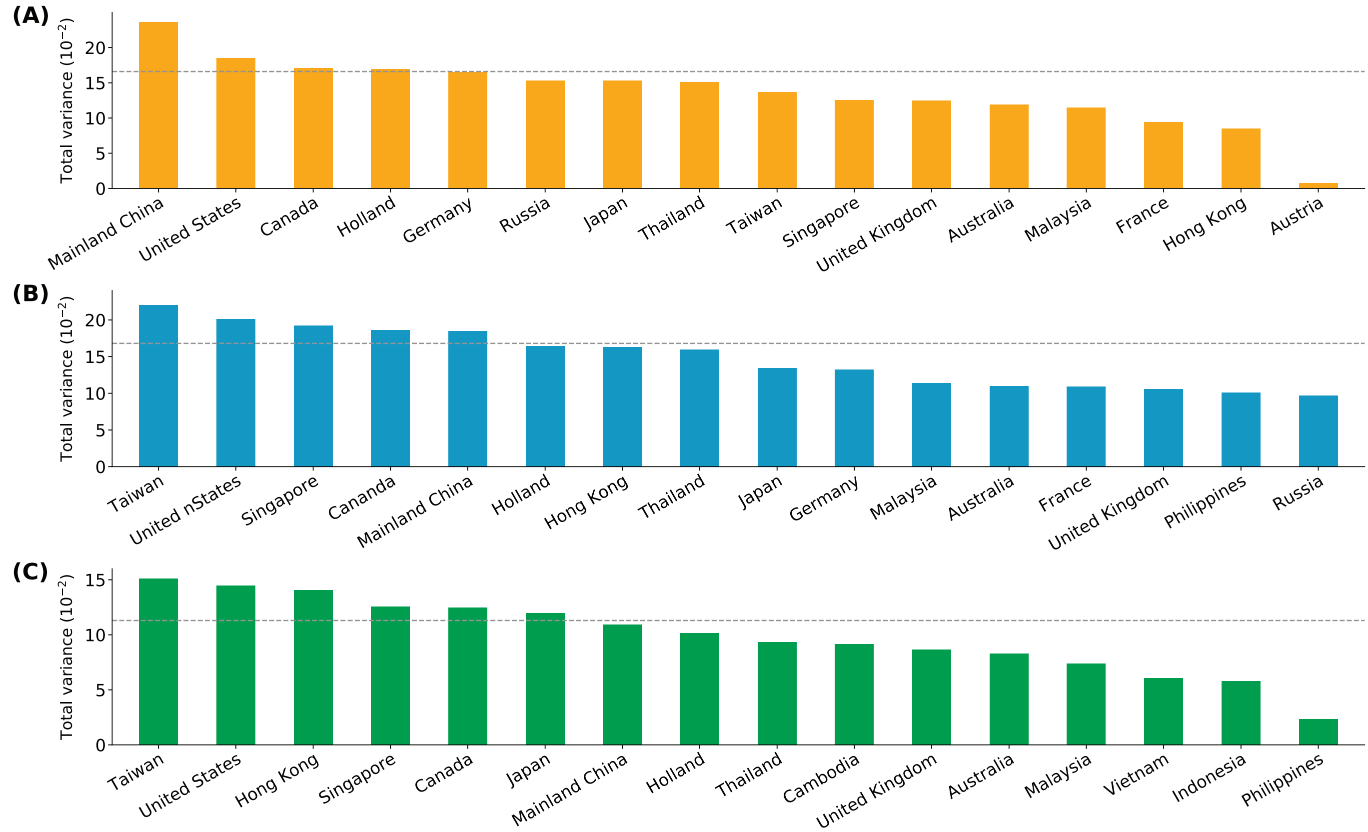


Fig. 9. Mobility variance among travelers by country or region:(A) Jeonju; (B) Gangneung; (C) Chuncheon. Dashed lines denote population average of each city.

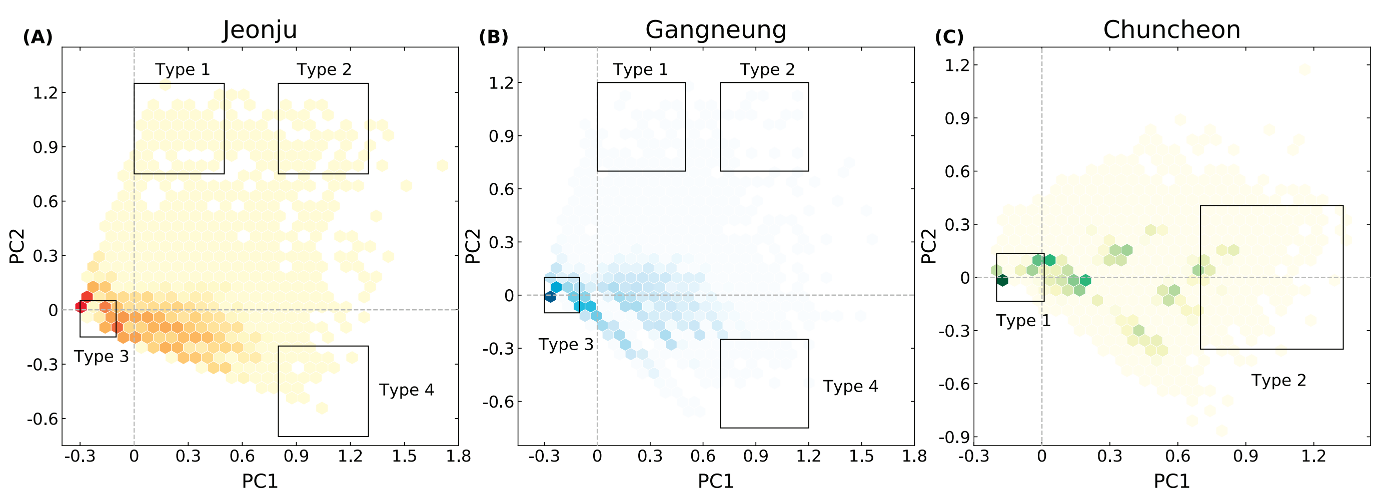


Fig. 10. The selection criterion of traveler groups with distinctive mobility characteristics for each city. For Chuncheon, in addition to type 1 and type 2 travelers, we incorporate a type 3 travelers with a time span greater than 24 h to better represent overnight visitors.

Beach, Tyumonsin Ko and downtown area. The results in Fig. 12C and Fig. 12D reveal an interesting relationship between individual and collective tourist mobility patterns. On the one hand, travelers with a high diversity of location choices at the individual level do not necessarily produce diverse spatial patterns at an aggregate level (e.g., many tourists visited the same set of places during their visits in a city). On the other hand, travelers who visit few places in a city can have very different location choices, which produce more scattered spatial patterns as observed in Fig. 12C.

Three traveler groups are selected for Chuncheon (Fig. 13). Type 1 visitors on average spent only 2.2 h in the city. Strikingly, nearly 70% of their activities were observed in Nami Island (Fig. 13A), a famous attraction 63 km away from Seoul. Nami Island, according to many tuorism guides, is a popular day-trip attraction from Seoul. A convenient way to reach the island is through the Gapyeong station, a railway station of the Gyeongchun Line (the station to the northwest side of Nami Island shown in Fig. 13A). Other areas of interest for type 1 visitors include Gangchon Railbike, Elysian Gangchon Ski, and the local food court in downtown. Fig. 13B shows the spatial patterns of activities for type 2, visitors with a short stay duration but high mobility diversity. Besides Nami Island, activities were also observed at many other places such as Gangchon Rail Park, Deungseon Waterfall Jeong-Yangsa

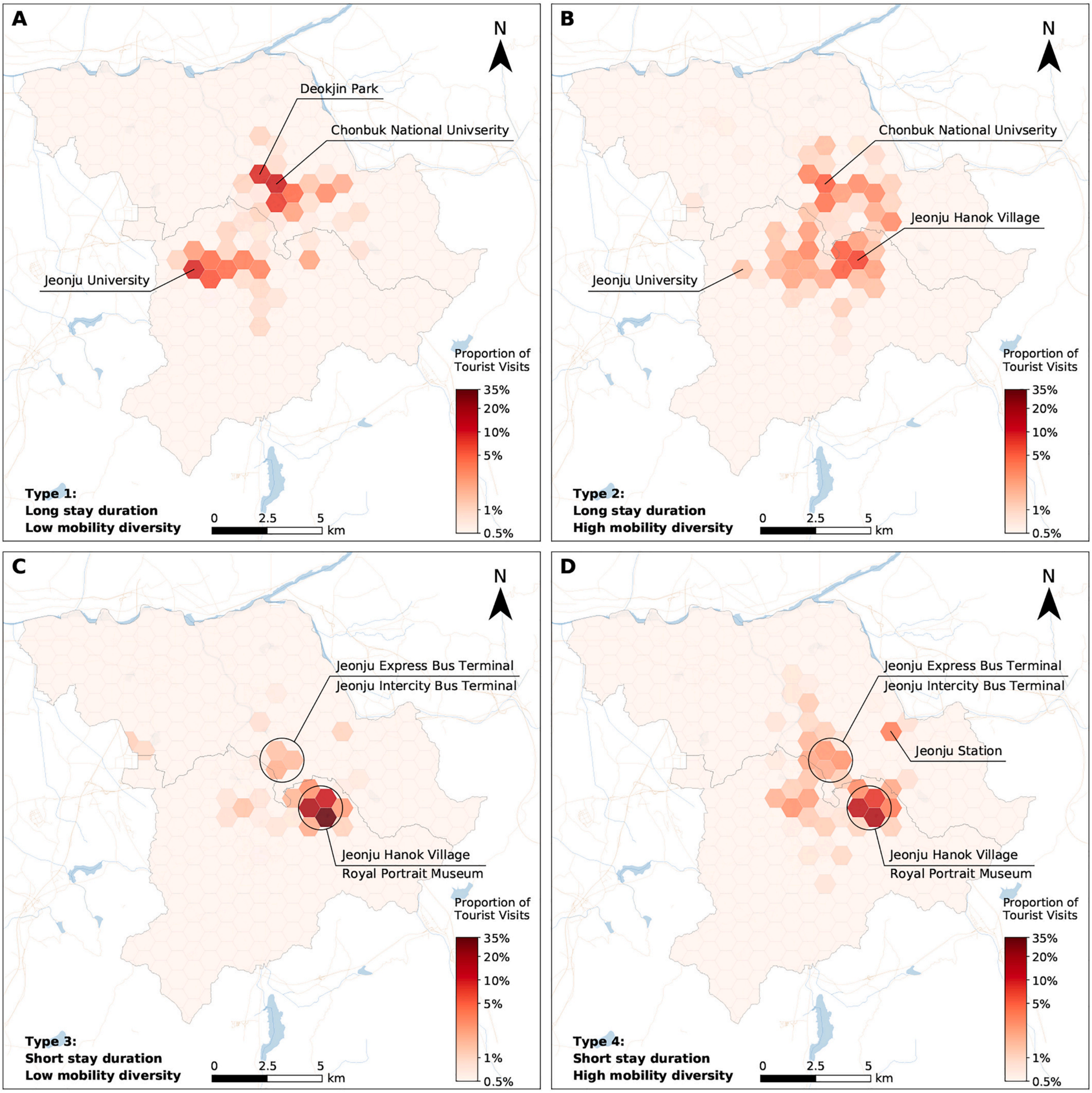


Fig. 11. Spatial patterns of activities for the four traveler groups in Jeonju.

Temple, and Gangchon Station. Chuncheon Dakgalbi Street, Soyang-gang Skywalk and Gubongsan Observatory are also popular places for this travler group. Type 3 refers to a group of travlers with a time span greater than 24 h. As shown in Fig. 13C, Nami island is no longer a top attraction for these overnight stayers. The center of activities have switched to the downtown area. Again, the results in Fig. 13 indicates that tourists with different mobility characteristics tended to explore different sets of places in a city. Interestingly, most of these places lie in the vicinity of the major railway and stations (Gyeongchun Line).

5. Discussion and conclusion

Tourist travel behavior is complex in nature. Such complexities are embodied in different behavioral dimensions, such as their duration of stay in a city, locations visited, and organization of trips and movements. By analyzing mobile phone trajectories of international travelers to three different cities in South Korea (Jeonju, Gangneung and Chuncheon), we find that tourists explored cities in different ways. Such behavioral heterogeneity reveals the existence of different types of travelers that are of interests to urban planners and tourism stakeholders.

In this research, we introduce nine mobility indicators to enable a comprehensive depiction of individual tourist mobility patterns. These indicators capture a traveler's mobility behavior from either a spatial, temporal, or spatio-temporal perspective. By examining the statistical properties of these indicators, we find that tourist travel behavior differs across cities. Jeonju and Gangneung were mixed with overnight stayers and same-day visitors, while Chuncheon were dominated by temporary

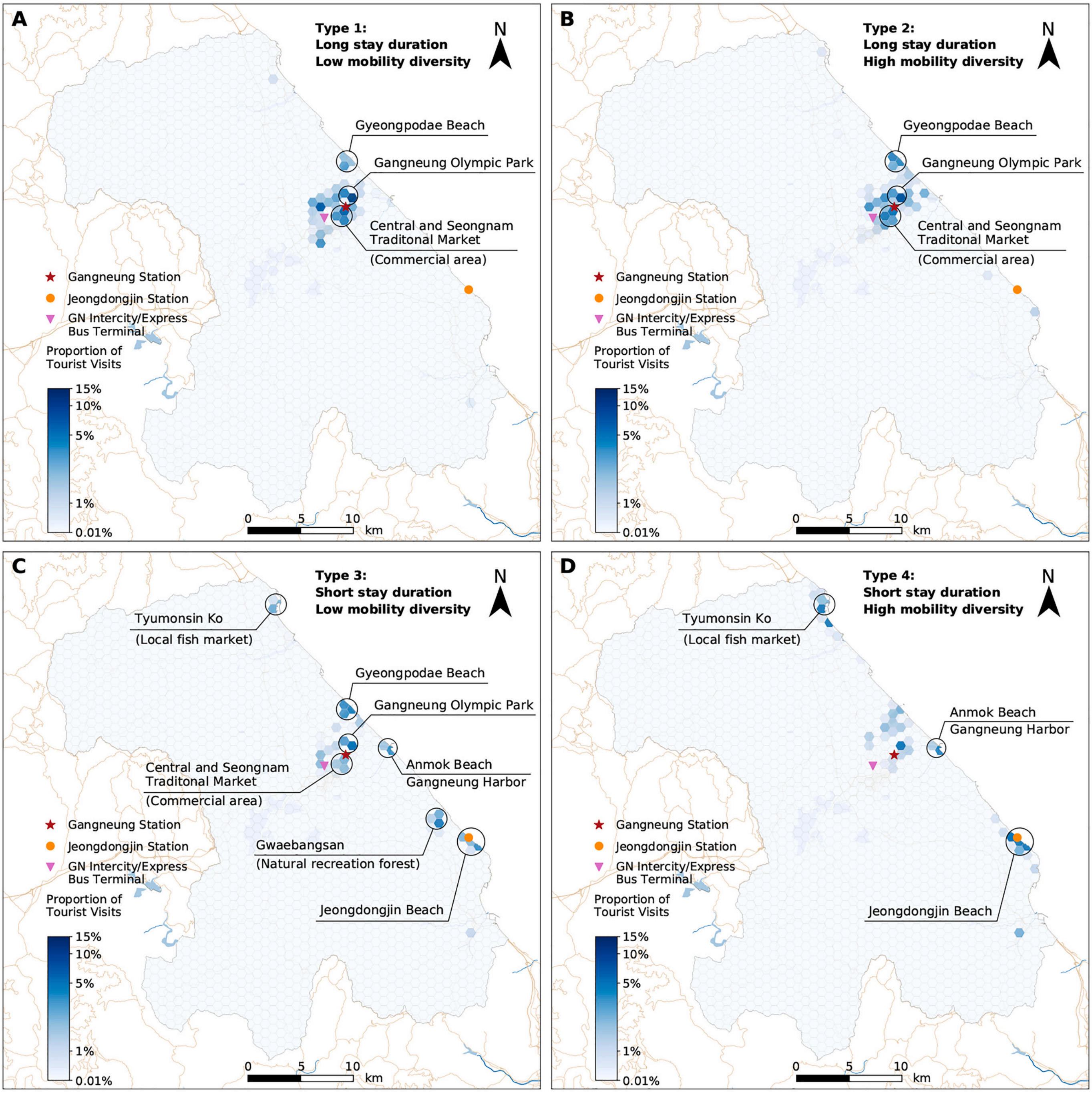


Fig. 12. Spatial patterns of activities for the four traveler groups in Gangneung.

visitors, and many of them stayed in the city for only one or two hours. Since travelers stayed longer in Jeonju and Gangneung, they tended to visit more places than travelers in Chuncheon. However, longer stays and more locations do not necessarily translate into larger activity spaces or more trips. For instance, although travelers in Jeonju and Gangneung visited a similar number of locations during their journeys, there is substantial difference in the spatial extent of their activities. Another interesting finding is that although a higher diversity of destination choices is observed for travelers in Jeonju and Gangeneung, the rate of exploration — as reflected by the number of activity locations visited per day by an average traveler — remains constant across three cities. The statistical results depict a comprehensive picture of tourist profiles in each city, and also an intertwined relationship among different mobility indicators.

An eigendecomposition approach is then used to better understand the interdependency of these mobility indicators in each city. The approach takes high-dimensional mobility features of travelers as input, and extract a series of principal components (PCs) to describe the inherent variations across individuals. For all three cities, the first one or two PCs are able to describe a significant proportion (over 80%) of mobility variance across individuals. In particular, the spatial extent of activities (D and R_g) and mobility diversity (H_1 and H_2) are important features that distinguish individual travelers in both Jeonju and Gangneung. For Chuncheon, the daily number of activity anchor points (A_2), size of activity space (D and R_g), and mobility diversity (H_1 and H_2) are critical dimensions that explain individual variations. This reveals a strong interdependency of certain mobility indicators that can be leveraged to further segment travelers with specific behavioral

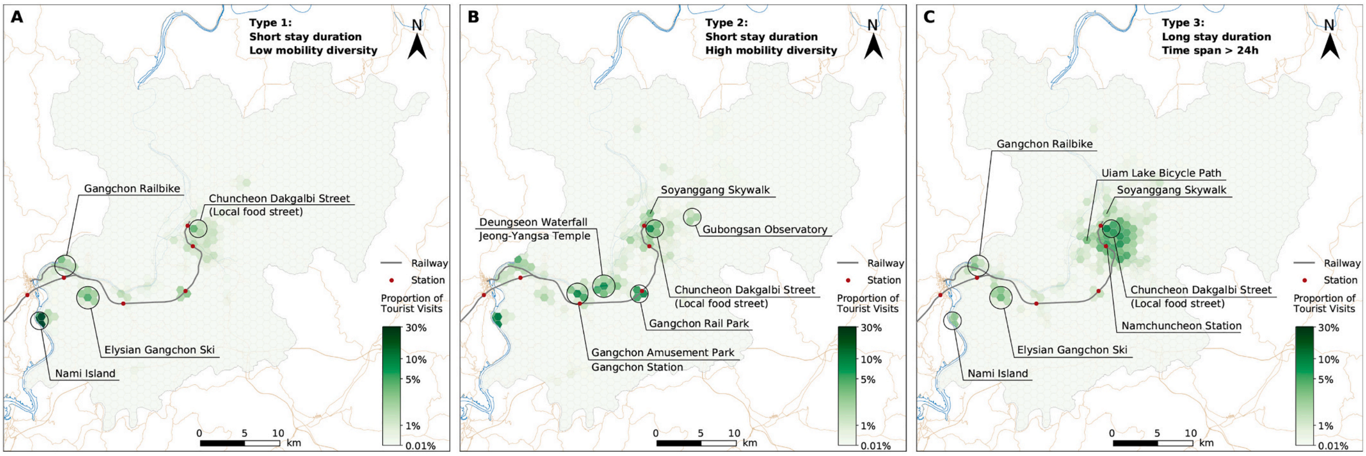


Fig. 13. Spatial patterns of activities for the three traveler groups in Chuncheon.

characteristics.

Since the top two PCs are able to explain over 80% of the mobility variance, we use the linear combination of these two PCs to express key characteristics of individual travel behavior. By leveraging this dimension reduction technique along with the nationality of travelers provided in the dataset, we demonstrate the efficacy of the eigendecomposition approach in capturing travelers' mobility patterns from different originating countries. We are able to identify nationalities of travelers with a remarkable level of mobility diversity, or those with variations mainly observed within few dimensions (e.g., travelers from certain countries who did not stay for long, but showed diverse travel and activity patterns), or those who showed uniform behavioral patterns (e.g., Austria travelers to Jeonju).

Finally, we select a few traveler groups with distinctive mobility patterns in each city and examine the spatial patterns of their activities. A few important findings are worth noting. First, substantial differences are observed in the spatial preferences of different traveler groups. In particular, the duration of stay in a city is a crucial factor that shapes tourists' location choices. Tourists with a short or long duration of stay tended to visit different sets of places, a finding that is universal across the three cities. Second, a high level of mobility diversity at the individual level does not necessarily mean diverse location choices at the group level. This can be illustrated by the spatial activity patterns of type 3 and type 4 travelers in Gangneung (Fig. 12). On the one hand, travelers could visit few places individually, but form a great variety of location choices collectively (Fig. 12C). On the other hand, travelers with a high mobility diversity could produce less diverse location choices at the group level (Fig. 12D). The uniform spatial preferences among these travelers can be affected by latent factors such as similarities in sociodemographics and group tours.

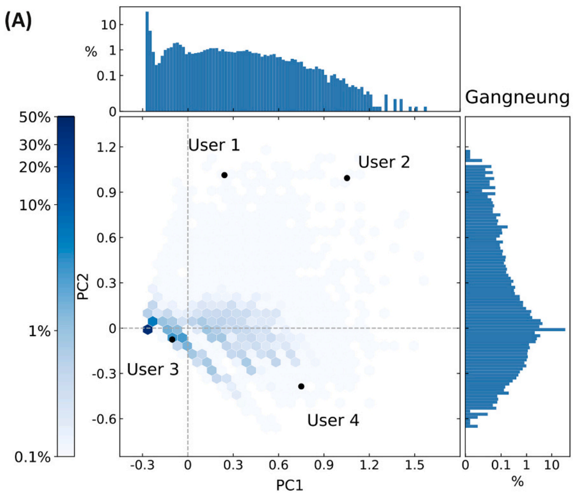
Our findings have several important implications for planning and policy recommendations. First, the majority of visitors to the three cities have a short duration of stay. The places they are able to reach and explore are limited by the travel time budget. Yet, their spatial preferences are quite diverse from an aggregate perspective (e.g., Fig. 12C). Therefore, future transport development could consider improving the connectivity of popular places for short-term visitors. Such improvements (e.g., reduction in travel time between places) would enable visitors to have more flexible travel plans, for example, to visit more attractions within a limited amount of time or spend more time at preferred places. The deployment of location recommendation services should also consider the behavioral diversity of short-term visitors. Instead of recommending few iconic places, highly customized travel plans will be attractive to some visitors if their favorite places can be connected within a reasonable amount of time. Second, tourists with a long duration of stay tended to visit different sets of places compared to short-term visitors. Many of these places are located in the central parts of cities where economic and cultural activities proliferate. Improving the vitality and quality of these places could possibly enrich travelers' experiences, which can be helpful to building destination image and attracting repeat visitors. A follow-up question, though, is how cities could optimize their strategies in catering different types of travelers (e.g., same-day visitors and overnight stayers), who contribute to cities' tourism revenue in different ways (Ashworth & Page, 2011; Rodriguez et al., 2018). In this sense, cities could leverage behavioral insights derived from big data (e.g., mobile phone trajectories) to explore these questions and better define their positions. These considerations are more meaningful today given the major disruptions caused by COVID-19 (Gössling, Scott, & Hall, 2020) and the economic downturn (Fernandes, 2020).

The current work can be improved or extended from the following perspectives. Although we have demonstrated the varying spatial preferences of different traveler groups, more nuanced differences in their travel behavior remain underexplored. It would be meaningful to leverage other approaches — such as trajectory clustering and sequential pattern mining — to identify travelers who explore cities with different space-time strategies. Since the mobile phone dataset captures dwell time of tourists at various locations, a more detailed analysis on time use could yield additional insights into their spatio-temporal behaviors (Xu, Li, Xue, Park, & Li, 2020). Such insights would inspire more realistic travel recommendations that consider travelers' preferences with respect to time allocation. It would also be interesting to perform analysis over different periods of the year to identify potential seasonable variations in tourist mobility patterns. To sum up, this work provides a multidimensional view of tourist travel behavior by linking classic human mobility analysis and the emerging field of tourism big data. The framework can be applied to other types of mobility observations (e.g., geocoded social media, WiFi, GPS) to understand travel behavior of different population groups in cities.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments on earlier versions of the manuscript. This research was jointly supported by the Research Grant Council of Hong Kong (NO. 25610118), the National Natural Science Foundation of China (NO. 41801372; NO. 41671387), and the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A3A2098438). The authors would like to express their great appreciation to Korea Tourism Organization (particularly to Team of Tourism Big Data) for their full support in initiating the project and collecting the big data set as well as sharing their insights on this project.

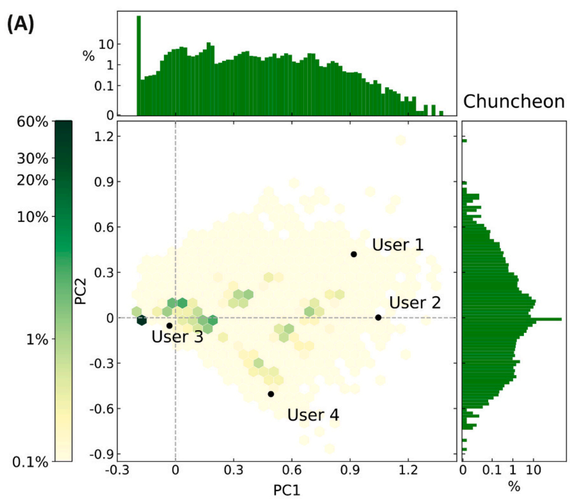
Appendix A. Four-user example for Gangneung and Chuncheon



(B)

Indicators	Mean	1	2	3	4
N_{day}	1.86	15	14	1	1
S	20.65	340.87	313.37	2.62	7.7
A_1	2.36	4	14	2	4
A_2	1.76	1.27	3.36	2	4
N_{od}	2.44	8	55	1	4
D	4.81	4.32	18.42	2.69	28.07
R_g	1.78	0.64	2.38	1.75	11.03
H_1	0.36	0.14	1.09	0.28	1.22
H_2	0.38	1.04	2.63	0	1.04

Fig. A.1. (A) Joint distribution of PC1 and PC2 coefficients for travelers in Gangneung; (B) mobility indicators of four selected travelers.



(B)

Indicators	Mean	1	2	3	4
N_{day}	1.16	11	1	1	1
S	5.68	239.72	7.85	8.87	2.53
A_1	1.71	14	5	2	2
A_2	1.61	2.91	5	2	2
N_{od}	0.95	28	5	1	1
D	3.57	28.05	23.61	3.43	20.88
R_g	1.54	2.91	10.46	2.38	12.9
H_1	0.3	1.07	1.56	0.09	0.4
H_2	0.17	3.04	1.61	0	0

Fig. A.2. (A) Joint distribution of PC1 and PC2 coefficients for travelers in Chuncheon; (B) mobility indicators of four selected travelers.

Appendix B. Calculate visitation frequency at the hexagon level

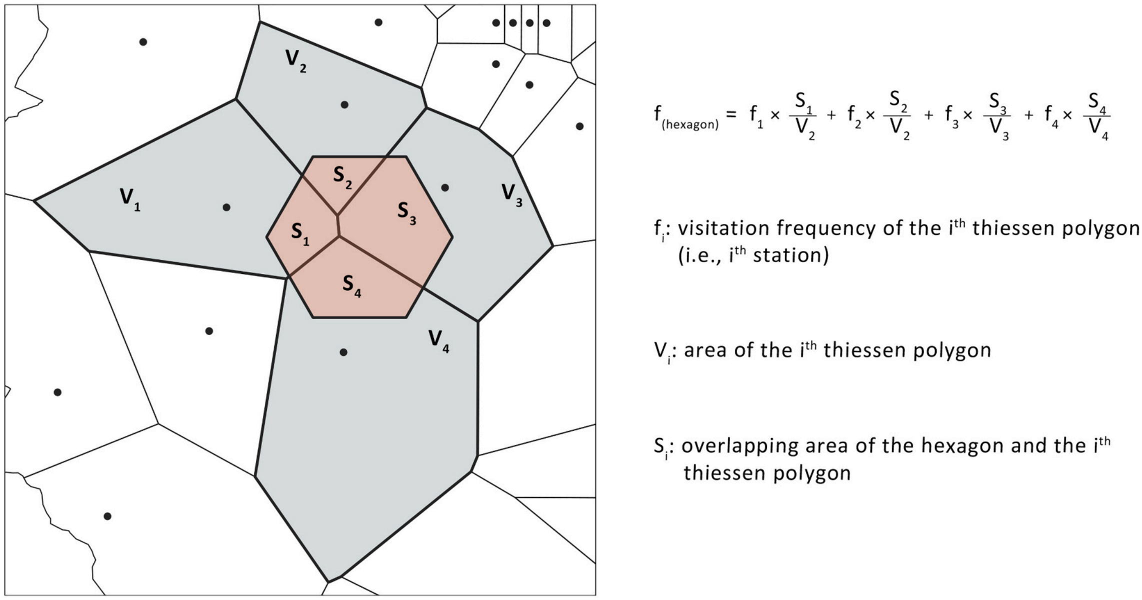


Fig. B.3. Convert the visitation frequency of tourists from cellphone towers to hexagons. Thiessen polygons are used to approximate the service areas of the cellphone towers.

References

Ahas, R., Aasa, A., Mark, Ü., Pae, T., & Kull, A. (2007). Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management, 28*(3), 898–910.

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (2018). Evidence for a conserved quantity in human mobility. *Nature Human Behaviour, 2*(7), 485–491.

Ashworth, G., & Page, S. J. (2011). Urban tourism research: Recent progress and current paradoxes. *Tourism Management, 32*(1), 1–15.

Birenboim, A., Anton-Clavé, S., Russo, A. P., & Shoval, N. (2013). Temporal activity patterns of theme park visitors. *Tourism Geographies, 15*(4), 601–619.

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science, 4*(1), 10.

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies, 68*, 285–299.

Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J.-C., Huens, E., Van Dooren, P., … Blondel, V. D. (2013). Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications, 392*(6), 1459–1473.

De Cantis, S., Ferrante, M., Kahani, A., & Shoval, N. (2016). Cruise passengers' behavior at the destination: Investigation using gps technology. *Tourism Management, 52*, 133–150.

Dijst, M. (1999). Two-earner families and their action spaces: A case study of two dutch communities. *GeoJournal, 48*(3), 195–206.

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology, 63*(7), 1057–1066.

Fernandes, N. (2020). *Economic effects of coronavirus outbreak (covid-19) on the world economy*. Available at SSRN 3557504.

Gonzalez, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779.

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: A rapid assessment of covid-19. *Journal of Sustainable Tourism*, 1–20.

Grinberger, A. Y., Shoval, N., & McKercher, B. (2014). Typologies of tourists' time–space consumption: A new approach using gps data and gis tools. *Tourism Geographies, 16*(1), 105–123.

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017). Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data, 3*(2), 208–219.

Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., & Ma, X. (2010). Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *, 2010. 18th international conference on geoinformatics* (pp. 1–7). IEEE.

Kim, Y., Kim, C.-k., Lee, D. K., Lee, H.-w., & Andrada, R. I. T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management, 72*, 249–256.

Kuusik, A., Tiru, M., Ahas, R., & Varblane, U. (2011). Innovation in destination marketing: The use of passive mobile positioning for the segmentation of repeat visitors in Estonia. *Baltic Journal of Management, 6*(3), 378–399.

Lau, G., & McKercher, B. (2006). Understanding tourist movement patterns in a destination: A gis approach. *Tourism and Hospitality Research, 7*(1), 39–49.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.

Mckercher, B., & Lau, G. (2008). Movement patterns of tourists within a destination. *Tourism Geographies, 10*(3), 355–374.

McKercher, B., Shoval, N., Ng, E., & Birenboim, A. (2012). First and repeat visitor behaviour: Gps tracking and gis analysis in Hong Kong. *Tourism Geographies, 14*(1), 147–161.

Mou, N., Yuan, R., Yang, T., Zhang, H., Tang, J. J., & Makkonen, T. (2020). Exploring spatio-temporal changes of city inbound tourism flow: The case of shanghai, China. *Tourism Management, 76*, 103955.

Nilbe, K., Ahas, R., & Silm, S. (2014). Evaluating the travel distances of events visitors and regular visitors using mobile positioning data: The case of Estonia. *Journal of Urban Technology, 21*(2), 91–107.

Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management, 33*(3), 672–682.

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications, 6*, 8166.

Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., & Giannotti, F. (2016). An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics, 2*(1–2), 75–92.

Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management, 57*, 202–212.

Reades, J., Calabrese, F., & Ratti, C. (2009). Eigenplaces: Analysing cities using the space–time structure of the mobile phone network. *Environment and Planning. B, Planning & Design, 36*(5), 824–836.

Rodriguez, X. A., Martinez-Roget, F., & Gonzalez-Murias, P. (2018). Length of stay: Evidence from Santiago de Compostela. *Annals of Tourism Research, 68*, 9–19.

Saluveer, E., Raun, J., Tiru, M., Altin, L., Kroon, J., Snitsarenko, T., … Silm, S. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research, 81*, 102895.

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society Interface, 10*(84), 20130246.

Schönfelder, S., & Axhausen, K. W. (2003). Activity spaces: Measures of social exclusion? *Transport Policy, 10*(4), 273–286.

Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies, 18*(5), 587–606.

Tchetchik, A., Fleischer, A., & Shoval, N. (2009). Segmentation of visitors to a heritage site using high-resolution time-space data. *Journal of Travel Research, 48*(2), 216–229.

Tiru, M., Kuusik, A., Lamp, M.-L., & Ahas, R. (2010). Lbs in marketing and tourism management: Measuring destination loyalty with mobile positioning data. *Journal of Location Based Services, 4*(2), 120–140.

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46*, 222–232.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports, 3*, 2976.

Wu, L., Yang, L., Huang, Z., Wang, Y., Chai, Y., Peng, X., & Liu, Y. (2019). Inferring demographics from human trajectories and geographical context. *Computers, Environment and Urban Systems, 77*, 101368.

Xiao-Ting, H., & Bi-Hu, W. (2012). Intra-attraction tourist spatial-temporal behaviour patterns. *Tourism Geographies, 14*(4), 625–645.

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems, 72*, 51–67.

Xu, Y., Belyi, A., Santi, P., & Ratti, C. (2019). Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface, 16*(160), 20190536.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., & Ratti, C. (2019). Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems, 75*, 184–203.

Xu, Y., Li, J., Xue, J., Park, S., & Li, Q. (2020). Tourism geography through the lens of time use: A computational framework using fine-grained mobile phone data. *Annals of the American Association of Geographers*, 1–25.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation, 42*(4), 625–646.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Lu, F., Chen, J., … Li, Q. (2016). Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers, 106*(2), 489–502.

Yan, L., Wang, D., Zhang, S., & Xie, D. (2018). Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: A case study in Shanghai. *Transportation*, 1–16.

Yang, L., Wu, L., Liu, Y., & Kang, C. (2017). Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from flickr. *ISPRS International Journal of Geo-Information, 6*(0), 345.

Yuan, Y., & Raubal, M. (2016). Analyzing the distribution of human activity space from mobile phone usage: An individual and urban-oriented study. *International Journal of Geographical Information Science, 30*(8), 1594–1621.

Yuan, Y., Raubal, M., & Liu, Y. (2012). Correlating mobile phone usage and travel behavior–a case study of Harbin, China. *Computers, Environment and Urban Systems, 36*(2), 118–130.

Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems, 54*, 144–153.