

Quantifying Nationality Bias in Social Media Data on Different Platforms for Visitor Monitoring in Nikko National Park, Japan

Masahiro Kajikawa^{Aa}, Takafumi Miyasaka^{Aa*}, Yutaka Kubota^A, Akihiro Oba^B, Katori Miyasaka^C

^A Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan; ito10121@gmail.com; miyataka@nagoya-u.jp, bremen_de@hotmail.com; kokoroyutaka0204@gmail.com

^B Research and Development Initiative, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan; perry@tamacc.chuo-u.ac.jp

^C College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan; miyasaka.katori@nihon-u.ac.jp

KEYWORDS

Country of origin
Nature-based tourism
Representativeness
Sampling bias
Social networking service
Spatial visitor distribution
Twitter

ABSTRACT

Geotagged social media data have been used widely for visitor monitoring in protected areas. The data might, however, over or underestimate visitors from specific countries due to nationality bias, i.e., differences between nationality of actual visitors versus those visitors who post on social media. This study aimed to quantify nationality bias in social media data for visitor monitoring. We conducted a questionnaire survey in Nikko National Park, Japan. Questions covered the nationality of visitors and their usage of social media, and other visitor attributes and behavior. Foreign visitors had significantly different attributes and behaviors compared to Japanese. Non-Japanese Asian visitors were overrepresented in Instagram and Facebook data. In comparison, the X platform was more representative of all visitors. Nationality bias in different platforms needs more attention and further study in different areas.

Introduction

Nature-based tourism in protected areas has grown worldwide in recent years. It has been recognized that nature-based tourism can improve human health and well-being, generate revenue for nature conservation, contribute to local economies, and foster local and indigenous identity (Winter et al., 2020). Visitors can enjoy a wide range of activities by taking advantage of the unique natural features of each protected area. Meanwhile, excessive use of natural areas can result in their degradation or loss and a subsequent decrease in visitor satisfaction (Luque-Gil et al., 2018). It is necessary to

carry out proper visitor monitoring and management to achieve quality experiences for them without damaging nature resources.

Data on visitors and their behavior in protected areas are crucial for monitoring and management, but these data are often not easy to collect. Recreational activities in protected areas vary greatly depending on the attributes of visitors, such as age, gender, and size of party (Batra, 2009). Field surveys are commonly used to investigate visitors' attributes, as well as their purpose and behavior, but data collected in this way have limited spatiotemporal res-

^a The two authors, Masahiro Kajikawa and Takafumi Miyasaka, contributed equally to this work

* Corresponding author: Takafumi Miyasaka; e-mails: miyataka@nagoya-u.jp, bremen_de@hotmail.com

doi: 10.5937/gp27-46222

Received: August 30, 2023 | Revised: November 22, 2023 | Accepted: November 27, 2023

olution due to insufficient budgets and human resources (Cessford & Muhar, 2003; Chua et al., 2016).

A possible solution is to use data that can be obtained from social media. Geotagged photos and messages posted on social media by visitors have been considered as particularly useful sources of information because they are generated on site. An increasing number of attempts have been made to collect visitor information with high spatiotemporal resolution from geotagged social media posts (Sotiriadis, 2017). Previous studies using geotagged data on social media have analyzed the spatial distribution of visitors (Kim et al., 2019), balance between demand and supply of nature-based cultural services (Yoshimura & Hira, 2017), the activities and interests of visitors (Heikinheimo et al., 2017), and the satisfaction of visitors with their nature experiences (Lu & Stepchenkova, 2012).

A common issue pointed out in previous studies that used social media is data bias in posters' attributes. Examples of possible data bias include an emphasis on younger people, who are more likely to use social media (van Zanten et al., 2016), gender differences in posting frequency (Keeler et al., 2015), and differences in posting frequency based on the distance from home (Wood et al., 2013). However, even though we can recognize the existence of bias in social media data, the bias may be difficult to quantify because posters often do not disclose their personal information due to privacy concerns. It is therefore challenging to determine who and what the social media data represent and to interpret this data in a reliable way (Fischer, 2012).

The nationality of visitors is an essential sociodemographic characteristic for any social media analysis, as it is often one of the parameters examined for more in-depth analysis (Heikinheimo et al., 2022). For example, this information is crucial for separating locals from visitors in tourism research (García-Palomares et al., 2015). Visitor nationality also merits attention in that nationality differences may result in bias when exploring visitor behaviors on social media if a specific nationality group is more likely to share information than other groups (Sinclair et al., 2020).

The present study focused on nationality bias, i.e., differences between the nationality of actual visitors versus the nationality of those visitors who post on social media. Nationality bias can arise due to significant differences in the use of social media from country to country (including the proportion of the population using social media)

(Blank, 2017). Differences between countries in culture and values can greatly influence visitor behavior. For instance, domestic and foreign travelers can take different travel routes to destinations and prefer different types of activities (Chua et al., 2016), and people from Asian and non-Asian countries often behave differently on their trips (Vu et al., 2015). Thus, social media data may over or underestimate the behavior of visitors from specific countries or regions because of nationality bias.

However, nationality bias is not fully understood. Few studies have explicitly considered it to help understand visitor behavior in protected areas. Sinclair et al. (2020) estimated the residences of visitors to several national parks in Germany using their geotagged photos on Flickr and compared the findings with onsite survey data. They found that foreigners accounted for a significantly higher percentage as a share of Flickr posters than of actual visitors. This previous finding is informative, despite being limited to specific geographical areas and specific cultures. More case studies are needed, particularly in areas where the majority of visitors are from Asian countries or other countries outside of Europe and the Americas. People from Asian and non-Asian countries differ in their use of social media (Brian, 2023) and behavior on their trips (Vu et al., 2015). Furthermore, given that the number of travelers from Asia has increased in the international tourist trade (World Tourism Organization, 2019), for practical reasons it is important to understand nationality bias in Asian countries, which attract Asian tourists with relative ease due to geographic proximity.

This study examined one protected area in Japan that attracts many visitors from Japan and overseas (the latter especially from Asia). Our objective was to quantify possible nationality bias in social media data, in order to improve a general understanding of this issue and the feasibility of using social media for visitor monitoring and management in protected areas. Japan has made inbound tourism a pillar of its economic growth strategy, and in particular has featured national parks as important attractions. Given that the numbers of foreign tourists were on a rising trend in national parks as well as the country as a whole (Japan Tourism Agency, 2020a), it is also crucial in this national context to understand nationality bias in social media data.

Although Twitter was renamed X at the end of July 2023, this paper uses the name as it was when our survey was conducted.

Data and Methods

Study area

Our study site was the Oku-Nikko area of Nikko National Park in Tochigi Prefecture (Figure 1). It was designated a national park in 1934 and extends over three prefectures: Tochigi, Gunma, and Fukushima Prefectures. The park's total area is 114,908 ha, of which private lands occupy 27.1%. It is one of the most popular national parks in Japan, ranking fourth in visitor numbers among all national parks in 2019 (Ministry of the Environment, 2023). Factors for its popularity include easy access from Tokyo, attractive shrines and temples at this World Cultural Heritage site, and outstanding nature to enjoy.

The Oku-Nikko area is located in the southwestern part of Nikko National Park, and its average elevation is about 1,400 m above mean sea level. This part of the national park contains well-known wetlands, waterfalls, lakes, and rich fauna and flora, making it a popular area for hikers. It is especially full of tourists during the fall when the autumn foliage is in full color. Since the area is a principal destination for nature-based tourism within the park, we selected it as our study site.

The study area can be classified into three zones in terms of their attractiveness and visitor characteristics. The southeastern zone is crowded with mass tourists

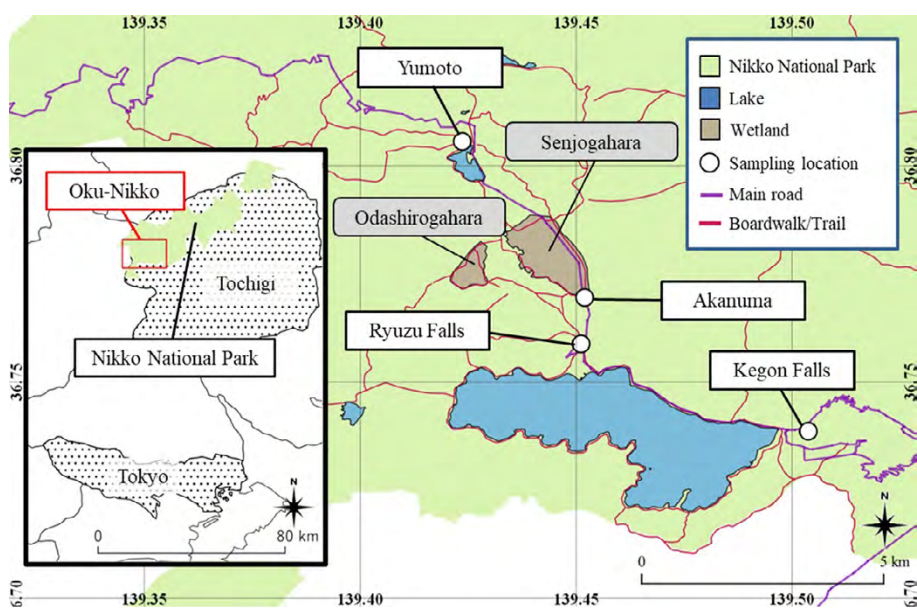


Figure 1. Sampling locations in the Oku-Nikko area of Nikko National Park, Japan

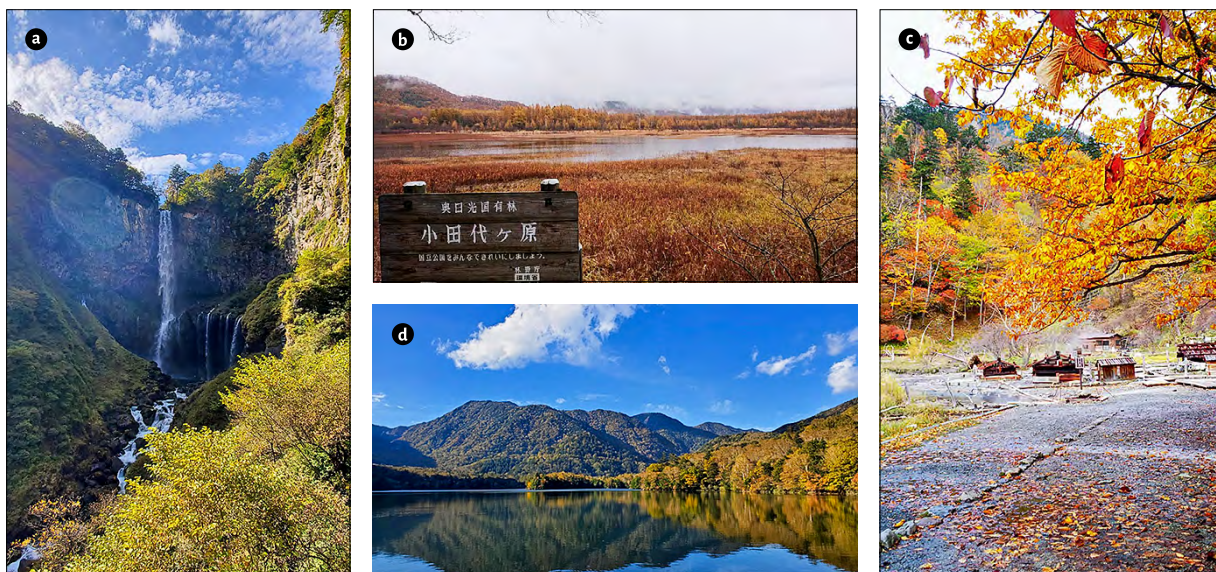


Figure 2. Typical scenery in the Oku-Nikko area of Nikko National Park, Japan: (a) Kegon Falls, (b) Odashirogahara, (c) Yumoto, and (d) Lake Yunoko

mainly attracted to see Kegon Falls (Figure 2a), a nationally famous waterfall, and benefits from relatively easy access from the World Cultural Heritage site “Shrines and Temples of Nikko,” which is east of the Oku-Nikko area. The central zone is characterized by marshes listed under the Ramsar Convention as “Oku-Nikko-shitsugen,” including Senjogahara and Odashirogahara (Figure 2b), and visitors enjoy hiking there. The northern zone centers on a hot springs resort, named Yumoto (Figure 2c), next to Lake Yunoko and surrounded by mountains (Figure 2d), which attracts overnight visitors who enjoy hiking and climbing.

Questionnaire survey

We conducted a questionnaire survey of visitors at the following four locations in the Oku-Nikko area from October 25 (Fri) through 28 (Mon) 2019: Kegon Falls in the southeastern zone (hereafter Kegon Falls), a primary starting point for hikes around the wetlands in the central zone (Akanuma), the hot spring resort in the northern zone (Yumoto), and a famous waterfall, named Ryuzu Falls, located in between the central and southeastern zones (Ryuzu Falls). These four locations attract a diverse range of visitors. As soon as we finished one interview, we immediately approached the next visitor (or group) that passed and confirmed the willingness of respondents to participate in the survey (i.e., informed consent was secured).

Questions asked of visitors covered their nationality and usage of social media, including what social media platforms they used, whether or not they posted on social media to share their travel experiences, and whether or not they used geotags when doing so. We also investigated other basic visitor attributes and behavior, including gender, age, size of party, means of transportation, length of stay, and places visited within the Oku-Nikko area. The survey was available in both Japanese and English, and targeted all visitors over the age of 13.

Our survey period was right before the COVID-19 pandemic started. Since the number of international visitors to Japan dropped during the pandemic, we used 2019 data to adequately detect nationality bias.

Resampling of collected data

When conducting surveys at multiple locations with different levels of popularity, sampling should be done according to the spatial distribution of the number of actual visitors at those locations, in order to avoid over or underestimating characteristics of some locations. In our survey, Kegon Falls was crowded with mass tourists and appeared to have the largest number of visitors among the four sampling locations. However, the number of respondents at Kegon Falls was actually lower than at other locations (Table 1), which indicates that the characteristics of visitors to Kegon Falls would be underestimated in our dataset. The dataset should be resampled with reference to

other visitor count data, but detailed reference visitor data such as the number of visitors in different segments of a natural area are generally not available.

We estimated the relative ratio of actual visitor numbers at the four sampling locations using two parameters: the number of visitors who actually posted geotagged tweets (on Twitter) around each of the locations in October 2019, and the percentage of respondents who answered that they usually posted their travel experiences with geotags (on Twitter). The reason why we used Twitter is as follows. First of all, the number of posts on social media generally has strong correlations with actual visitor numbers in national parks (Wood et al., 2013). Among major platforms, Instagram and Facebook had a large number of users in the world and a great number of posts about nature (Hausmann et al., 2018). These platforms, however, had limited application in research, probably because of restrictions on data access (Ghermandi & Sinclair, 2019). Twitter and Flickr were less restrictive, and many studies had applied data from these platforms to visitor monitoring in protected areas (Ghermandi & Sinclair, 2019). Flickr, however, was rarely used in Japan, whereas the number of Twitter users was larger than that of even Instagram or Facebook in Japan (GlobalWebIndex, 2020). We therefore considered Twitter as the best social media platform for the resampling.

We collected all publicly available geotagged posts on Twitter within the bounding box area defined by 36.698186 N, 36.834750 N, 139.318670 E, and 139.528918 E, which is identical to the map area shown in Figure 1, from October 1 to 31, 2019, using the Twitter API. As a result, we obtained 422 geotagged posts after excluding irrelevant posts, such as weather and earthquake news. We counted how many tweets were posted within a 200 m radius of each sampling location using QGIS version 3.8.0. If the same person posted more than once at the same location on the same day, we counted the posts as one post.

We estimated the relative ratio as shown in Eq. (1) and calculated the size of resamples at each sampling location as shown in Eq. (2) and Table 1,

$$RR_i = \frac{T_i \cdot \left(\frac{S_i}{s_i} \right)}{T_{Kegon} \cdot \left(\frac{S_{Kegon}}{s_{Kegon}} \right)} \quad (1)$$

$$RS_i = S_{Kegon} \cdot RR_i \quad (2)$$

where i is codes for sampling locations (i.e., Kegon Falls, Akanuma, Yumoto, or Ryuzu Falls), RR is the ratio of visitor numbers at location i to those at Kegon Falls, S is the number of samples collected (i.e., the number of total respondents) at location i , s is the number of respondents who answered that they usually posted their travel experi-

Table 1. Result of resampling the dataset collected in the Oku-Nikko area in October 2019

Sampling location	Total sample (S)	Tweeting sample (s)	Tweeter (T)	Relative ratio (RR)	Resample (RS)
Kegon Falls	243	19	55	1	243
Akanuma	371	27	31	0.61	147
Yumoto	231	14	10	0.23	57
Ryuzu Falls	340	28	25	0.43	105

Total sample (S) is the number of total respondents at each of the sampling locations; Tweeting sample (s) is the number of respondents at each of the sampling locations who answered that they usually posted their travel experiences with geotags on Twitter; Tweeter (T) is the number of visitors who posted geotagged tweets around each of the sampling locations; Relative ratio (RR) is the estimated relative ratio of actual visitor numbers among the sampling locations; and Resample (RS) is the number of resamples at each of the sampling locations (cf. Eq (1) and (2)).

ences with geotags on Twitter at location i , T is the number of visitors who posted geotagged tweets around location i , and RS is the number of resamples at location i . Given that Kegon Falls appeared to have the largest number of visitors among the four locations, it was treated as a reference when calculating the relative ratio.

We randomly extracted respondents of the estimated number (RS) at every location and built a new dataset. This resampling was repeated three times to test whether the attribute characteristics of extracted respondents varied at every resampling. The statistical analysis described in the next section was applied to the three resampled datasets for this test. We found that the characteristics of visitor attributes did not generally differ between the three datasets (see Results for details), so we randomly chose one of the datasets for further analysis.

Statistical analysis

First, we compared visitor attributes (except nationality) and behaviors between all visitors and visitors posting with geotags on social media, and between all visitors and

visitors of each nationality group, in order to characterize these different groups of visitors. Then, we analyzed whether the ratio of visitors posting with geotags on social media to all visitors differed by nationality, to test for nationality bias in social media data. To simplify analysis and interpretation of results, we considered social media posters to be only those visitors who used Twitter, Instagram, or Facebook platforms, which accounted for 98.0% of all the respondents who shared their experiences with geotags on social media. Our nationality classifications were Japan, Asia (excluding Japan), and non-Asia (88.0% from Europe and North America, the rest from Australia, New Zealand, and Chile), which is similar to the classification used by Vu et al. (2015) for visitors to Hong Kong to examine the behavior of tourists with different profiles. Asia in our classification included countries defined by the United Nations as East Asia, South Asia, Central Asia, West Asia, and Southeast Asia. We performed the analysis of means (ANOM) or Fisher’s exact test for all comparisons at a 5% significance level. All statistical analyses were conducted using JMP version 15.

Results

Visitor attributes

We found no significant differences in gender, group size, and length of stay between all visitors and visitors posting with geotags on social media regardless of the platforms (Table 2). On the other hand, the age of visitors posting on any platform of social media tended to be younger than that of all visitors, and fewer visitors posting on Instagram or Facebook used cars for traveling and they tended to be first-time visitors compared to all visitors.

The visitor attributes of all visitors were generally similar to those of Japanese visitors, although some statistical differences were detected, probably because of the large sample size. Gender did not differ due to nationality, but other attributes differed between all visitors and

visitors from Asia or outside Asia: age of foreign visitors was younger; the group size of Asian visitors tended to be larger, and that of visitors from the outside Asia tended to be smaller; Asian visitors more often stayed overnight; and most of the foreign visitors did not use cars and were first time visitors.

These results were obtained from all three resampled datasets with only slight differences (Table 3). We assumed that the characteristics found in only one dataset (i.e., length of stay for Facebook and frequency of visit for Twitter) were produced by a fluctuation due to resampling, and excluded them from consideration. The result in Table 2 is based on dataset C in Table 3, which we randomly chose.

Table 2. Differences in visitor attributes between all visitors and visitors posting with geotags on social media, and between all visitors and visitors of each nationality group in Oku-Nikko, Japan

Attribute	All visitors (n = 552)	Social media platform used						Nationality group					
		Twitter (n = 46)		Instagram (n = 80)		Facebook (n = 62)		Japan (n = 483)		Asia (n = 44)		Non-Asia (n = 25)	
Gender													
Male	61.6%	69.6%		55.4%		56.1%		63.3%		51.3%		45.8%	
Female	38.4%	30.4%		44.6%		43.9%		36.7%		48.7%		54.2%	
Age group													
10s	1.5%	6.5%		5.0%		4.8%		1.5%		2.3%		0.0%	
20s	18.8%	43.5%		41.3%		29.0%		17.0%		25.0%		44.0%	
30s	12.9%	17.4%	*	25.0%	*	25.8%	*	10.1%		36.4%	*	24.0%	*
40s	14.9%	13.0%		15.0%		17.7%		14.3%		20.5%		16.0%	
50s	20.8%	13.0%		10.0%		14.5%		22.6%		9.1%		8.0%	
≥60s	31.2%	6.5%		3.8%		8.1%		34.6%		6.8%		8.0%	
Group size													
1	14.5%	10.9%		12.5%		8.1%		12.8%		15.9%		44.0%	
2	59.6%	60.9%		57.5%		59.7%		62.1%		40.9%		44.0%	
3	10.5%	8.7%		6.3%		6.5%		11.4%		6.8%	*	0.0%	*
4	7.6%	8.7%		11.3%		14.5%		6.8%		13.6%		12.0%	
≥5	7.8%	10.9%		12.5%		11.3%		6.8%		22.7%		0.0%	
Transportation													
Car	72.6%	71.7%		55.0%	*	53.2%	*	81.4%	*	15.9%	*	4.0%	*
Other	27.4%	28.3%		45.0%		46.8%		18.6%		84.1%		96.0%	
Length of stay													
Day visitor	55.4%	58.7%		47.5%		45.2%		57.6%	*	38.6%	*	44.0%	
Overnight visitor	44.6%	41.3%		52.5%		54.8%		42.4%		61.4%		56.0%	
Frequency of visit													
First-time visitor	26.5%	34.8%		45.0%	*	45.2%	*	19.1%	*	72.7%	*	88.0%	*
Repeat visitor	73.6%	65.2%		55.0%		54.8%		81.0%		27.3%		12.0%	

* indicates a significant difference from the mean of all visitors at a 5% significance level. Among the visitor attributes surveyed, only gender had some missing values, and its sample size differed: n = 74 for Instagram, n = 57 for Facebook, n = 477 for Japan, n = 39 for Asia, and n = 24 for non-Asia.

Table 3. Differences in analysis results between three resampled datasets

Attribute	Social media platform used			Nationality group		
	Twitter	Instagram	Facebook	Japan	Asia	Non-Asia
Gender						
Age group	A, B, and C	A, B, and C	A, B, and C		A, B, and C	A, B, and C
Group size					A, B, and C	A, B, and C
Transportation		A, B, and C	A, B, and C	A, B, and C	A, B, and C	A, B, and C
Length of stay			A	B and C	B and C	
Frequency of visit	B	A, B, and C	A, B, and C	A, B, and C	A, B, and C	A, B, and C

A, B, and C indicate different datasets. The appearance of the character(s) in each cell indicates that a significant difference from the mean of all visitors was detected at a 5% significance level when using the corresponding dataset(s) (see Table 2 for the result using dataset C).

Visitor behavior

We found that visitors posting with geotags on Instagram or Facebook tended to visit historical buildings and recreational facilities more often (Table 4). As we found with visitor attributes (described above), the behavior of all visitors was generally similar to that of Japanese visitors. Asian people visited historical buildings and recreational facilities more often and wetlands less often. Visitors from outside Asia also visited historical buildings more often. Although they appeared to visit wetlands less often,

no statistical difference was found, a factor probably influenced by the small sample size.

Nationality bias

Rates of posting with geotags on Twitter in all nationality groups were not significantly different from the average of all visitors (Figure 3). On the other hand, Asian visitors posted more with geotags on Instagram and Facebook, whereas Japanese visitors posted less on these platforms than all visitors.

Table 4. Differences in visitor behaviors between all visitors and visitors posting with geotags on social media, and between all visitors and visitors of each nationality group in Oku-Nikko, Japan

Place visited	All visitors (n = 552)	Social media platform used			Nationality group		
		Twitter (n = 46)	Instagram (n = 80)	Facebook (n = 62)	Japan (n = 483)	Asia (n = 44)	Non-Asia (n = 25)
Waterfalls	80.6%	87.0%	87.5%	80.7%	78.9% *	93.2%	92.0%
Wetlands	40.6%	43.5%	46.3%	45.2%	42.7% *	31.8% *	16.0%
Lakes	36.8%	28.3%	38.8%	40.3%	35.2%	47.7%	48.0%
Historical buildings	21.9%	30.4%	42.5% *	45.2% *	17.8% *	54.6% *	44.0% *
Recreational facilities	19.6%	26.1%	31.3% *	37.1% *	16.8% *	50.0% *	20.0%
Other nature spots	4.0%	4.4%	2.5%	6.5%	3.9%	6.8%	0.0%

* indicates a significant difference from the mean of all visitors at a 5% significance level

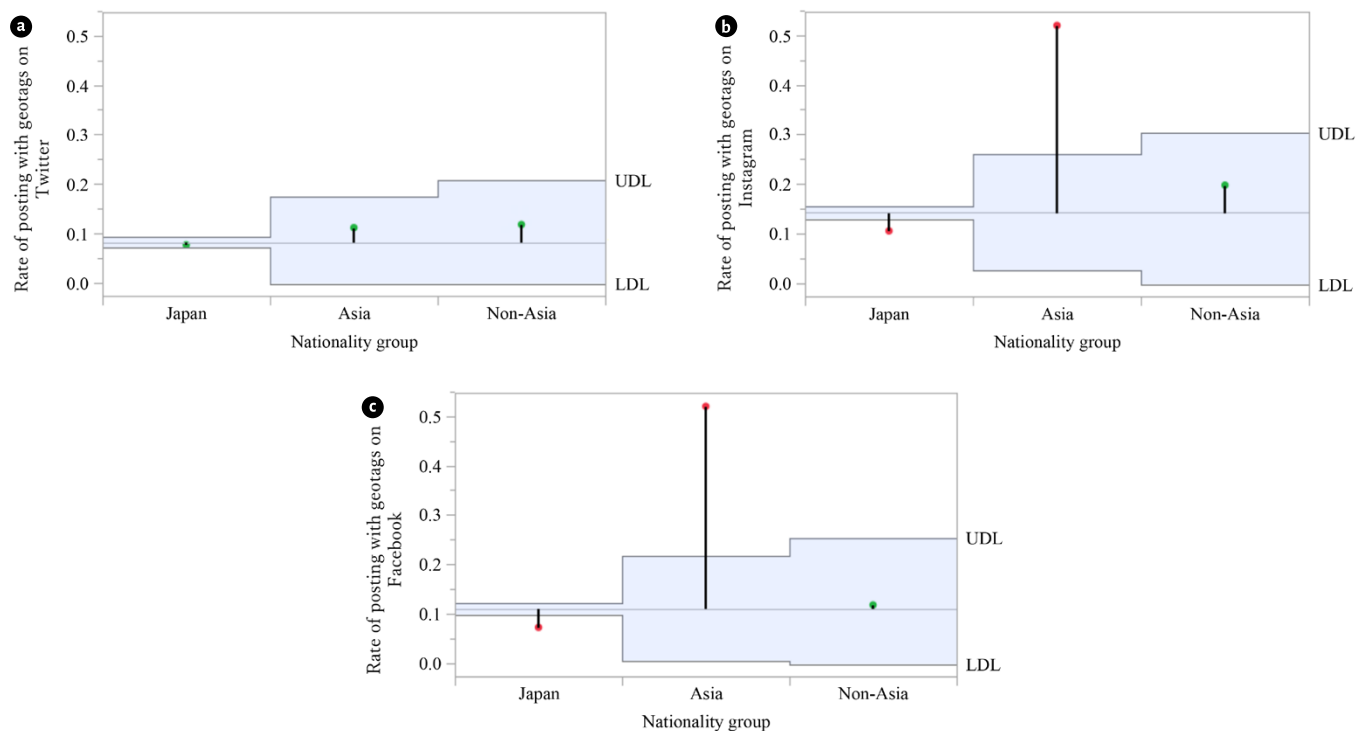


Figure 3. Differences in the rates of posting with geotags on Twitter (a), Instagram (b) and Facebook (c) between all visitors and visitors of each nationality group. UDL, upper decision limit; LDL, lower decision limit. The line in the middle of the blue area shows the average of all visitors. The percentages out of the range from LDL to UDL, shown in red, indicate that the values are significantly different from the average value of all visitors at a 5% significance level

Discussion

Visitor attributes

Many foreign visitors cannot drive a car by themselves because they do not have a driver's license in Japan and tend to use public transportation. The means of transportation therefore largely differed between domestic and foreign visitors. Most Japanese visitors in our survey were repeat visitors, which is consistent with a previous visitor survey conducted in the Oku-Nikko area (Miyasaka et al., 2018), whereas foreign visitors were more likely to be first time visitors. It is obvious that international trips are not easy to repeat in terms of time and cost compared to domestic trips. Liao et al. (2019) also reported that people who frequently travel abroad tended to visit places they have never visit before. Foreign visitors and all visitors or Japanese visitors had opposite patterns for these two attributes, i.e., means of transportation and frequency of visit, and these large differences probably made differences in these attributes of Instagram and Facebook posters, a relatively large number of whom were foreigners. On the other hand, even though visitors from Asia or outside Asia had statistically significant different patterns of group size and length of stay compared to all visitors, the degree of differences would not be large enough to influence those attributes of social media posters. The result that visitors posting on social media were younger compared to all visitors regardless of the platforms would be just a general trend (Heikinheimo et al., 2017; Hausmann et al., 2018), rather than the influence of foreign visitors, who were also younger than all visitors. Regarding group size, Asian visitors were more likely to come in a group of four or more, while visitors from outside Asia were more likely to visit alone. A study of foreigners visiting Shirakawa-go, a popular World Cultural Heritage site in Japan, also showed that many Asian tourists were in groups but that the majority of non-Asian tourists were alone (Ichikawa et al., 2016). The attribute patterns of foreign visitors found in this study are generally consistent with the results of an official nationwide survey of foreign tourist behavior in Japan (Japan Tourism Agency, 2020b), which showed that most foreign visitors to Japan used public transportation, the first time visit was the largest category of visitation frequency of foreigners, and non-Asian visitors tended to be alone compared to Asian visitors.

Visitor behavior

Visitors posting on Instagram and Facebook and foreign visitors, particularly Asian visitors, were more likely to visit historical buildings and recreational facilities. These places were located in the southeastern zone of the study area. Since this zone is relatively easily accessible from the World Cultural Heritage site of the park, many visitors to the zone just stop in and tend to be first time vis-

itors. Many foreign visitors appear to fall in the category of such visitors. Furthermore, wetlands, including a popular but low accessibility wetland, were less visited by foreign visitors, unlike other natural places such as waterfalls and lakes, probably because major waterfalls and lakes are all more accessible, being located near a main road. On the other hand, Japanese visitors tended to focus on natural landscapes including wetlands, implying that repeat visitors might be more attracted by (deeper) nature than artificial places. For example, Miyasaka et al. (2018) reported that about 90% of hikers in the Oku-Nikko area were repeat visitors. These results are consistent with previous findings that foreign tourists were spatially concentrated in major sightseeing spots compared to domestic travelers (Ghermandi et al., 2020; Muñoz et al., 2019).

Nationality bias

Instagram and Facebook users worldwide have been higher in number than Twitter users, but the reverse is true in Japan, where Twitter was the most popular platform in 2019 (GlobalWebIndex, 2020). This difference could explain the nationality bias found only in Instagram and Facebook posters in the area. Although we used data gathered in 2019, Twitter (now X) remains as the most popular platform in Japan (Kemp, 2023), while the number of inbound tourists in Japan has recovered in 2023 to the same level as before the COVID-19 pandemic (Japan National Tourism Organization, 2023). Thus, we believe that our findings are still relevant.

Nationality bias can be expected to vary depending on the status of use of social media, which can differ from country to country. In German national parks, for example, Sinclair et al. (2020) reported that visitor counts derived from Flickr and onsite survey data were highly correlated, but that those from another social media platform, VKontakte, and the onsite survey data were not significantly correlated. This was due to the characteristics of VKontakte being particularly used by Russian speakers, which meant low representativeness of all visitors to German national parks. In addition to such a difference in the usage of social media, nationality bias can also be affected by different visitation rates of foreigners within and between countries. For instance, the capital city had more foreign visitors than other areas in Serbia (Kalinić and Vujičić, 2019), and the dependency of the tourism market on international travelers in the United States was lower than in the EU (Stankov et al., 2018). Considering these differences between countries and subregions in social media usage and inbound tourism conditions, further case studies are needed to elucidate nationality bias and to improve the feasibility of using social media for visitor monitoring and management.

Although there has been little research to date on nationality bias in social media data, Sinclair et al. (2020) found that Flickr data overrepresented international visitors to German national parks, who primarily came from the Netherlands, the United Kingdom, and the United States. The present study demonstrated that data on Facebook and Instagram, but not on Twitter, overestimated Asian visitor numbers and their behavior in a Japanese national park. Ours is the first case study to identify nationality bias in Asia comparing the three major social media platforms. Our results suggest that care is required when using popular worldwide platforms for visitor monitoring, as those platforms do not capture visitors from all geographic regions equally (Heikinheimo et al., 2022).

Foreign visitors might be more likely than Japanese visitors to share their travel experiences using geotags, and this might strengthen the nationality bias we found. The Ministry of Internal Affairs and Communications (2020) reported that people in Japan were more concerned than in other countries, such as China, Germany, and the United States, about providing personal data (including location information) to companies via web services. Also, in general, international visitors tended to see their new experience in a foreign country as more special than domestic visitors (Ghermandi & Sinclair, 2019). This tendency might drive their desire to record details of their travels to aid memory and to share with friends and family. Geotags facilitate that recording of details. Such psychological fac-

tors might boost posting rates by foreign visitors using geotags and reinforce nationality bias.

Limitation and implication

Although Rusdi et al. (2019) showed that there was no significant difference between Asian and non-Asian visitors to Indonesia in the rates of posting on social media while traveling, our study detected nationality bias only in Asian visitors. Our results might be influenced by our small sample size of non-Asian visitors. This is a limitation of our study and is due to the resampling of our original dataset. If we had checked the number of Twitter posts in four sampling locations (Tweeter (*T*) in Table 1) before conducting the questionnaire survey, we could have collected samples that more closely approximated the actual distribution of visitors (estimated by Relative ratio (*RR*) in Table 1), and could have minimized the reduction of sample size by resampling (Total sample (*S*) minus Resample (*RS*) in Table 1). This indicates, however, that checking the number of social media posts at multiple sampling locations beforehand can help collect balanced samples even where actual visitor distribution is unknown. Furthermore, our study suggests that nationality bias can be confirmed using a simple questionnaire survey on visitor nationality and social media usage, combined with analysis of the distribution of social media posts at the sampling locations. This is a theoretical and practical implication for consideration for future research and park management.

Conclusion

The present study quantified possible nationality bias in social media data for visitor monitoring in the Oku-Nikko area of Nikko National Park, Japan. Our result showed that visitors posting with geotags on Instagram and Facebook and visitors from Asia (excluding Japan) and outside Asia (particularly Europe, America, and Oceania) generally had similar patterns of visitor attributes, except group size. The foreign visitors, particularly Asian visitors, also had similar patterns of places visited compared to visitors posting with geotags on the two platforms. These attribute and behavior patterns were different from the patterns of all visitors, most of whom were Japanese (87.5%). In contrast, visitors posting with geotags on Twitter did not significantly differ in visitor attributes or behaviors from all visitors, except in terms of age. In addition, the rates of posting with geotags on Instagram and Facebook were much higher for

Asian visitors and lower for Japanese visitors, compared to all visitors, whereas the rates of posting with geotags on Twitter did not differ among nationality groups. These results indicated that foreign visitors had different characteristics compared to all visitors or Japanese visitors and that in particular, Asian visitors and their behavior would be overestimated based on Instagram and Facebook data.

Our results suggest that in Japan, compared to Instagram and Facebook, Twitter would be more representative of all visitors in areas where most of the visitors are domestic (Japanese). Nationality bias, however, can vary in different areas due to varied social media usage, including platform preferences, and foreigners' visits. More case studies are anticipated in the future, and it is our hope that this study's approach to quantify nationality bias can contribute to them.

References

- Batra, A. (2009). Senior pleasure tourists: examination of their demography, travel experience, and travel behavior upon visiting the Bangkok metropolis. *International Journal of Hospitality & Tourism Administration*, 10(3), 197–212. doi: 10.1080/15256480903088105.
- Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35(6), 679–697. doi: 10.1177/0894439316671698.
- Brian, D. (2023). Social network usage & growth statistics: how many people use social media in 2023? <https://backlinko.com/social-media-users> (21.11.2023).
- Cessford, G., & Muhar, A. (2003). Monitoring options for visitor numbers in national parks and natural areas. *Journal for Nature Conservation*, 11(4), 240–250. doi: 10.1078/1617-1381-00055.
- Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, 57, 295–310. doi: 10.1016/j.tourman.2016.06.013.
- Fischer, F. (2012). VGI as big data: a new but delicate geographic data-source. *GeoInformatics*, 15(3), 46–47.
- García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408–417. doi: 10.1016/j.apgeog.2015.08.002
- Ghermandi, A., Camacho-Valdez, V., & Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism Management*, 77, 104002. doi: 10.1016/j.tourman.2019.104002
- Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: a systematic map. *Global Environmental Change*, 55, 36–47. doi: 10.1016/j.gloenvcha.2019.02.003.
- GlobalWebIndex (2020). Japan: Key Digital Behaviors and Trends over Time and Across Demographics. pp. 9.
- Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, 11(1), e12343. doi: 10.1111/conl.12343.
- Heikinheimo, V., Järv, O., Tenkanen, H., Hiippala, T., & Toivonen, T. (2022). Detecting country of residence from social media data: a comparison of methods. *International Journal of Geographical Information Science*, 36(10), 1931–1952. doi: 10.1080/13658816.2022.2044484.
- Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: a comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, 6(3), 85. doi: 10.3390/ijgi6030085.
- Ichikawa, Y., Hata, T., & Matsui, K. (2016). 日本人・外国人ツーリストの観光特性とイメージにみる白川郷の世界遺産観光 [Characteristics of heritage tourism from the viewpoint of Japanese and foreign tourist's behavior and images in the Shirakawago]. *Tsukuba studies in human geography*, 36, 11–28. (in Japanese)
- Japan National Tourism Organization (2023). Japan Tourism Statistics. <https://statistics.jnto.go.jp/en/graph/> (21.11.2023).
- Japan Tourism Agency (2020a). White paper on tourism in Japan, 2020 (summary). <https://www.mlit.go.jp/kankochou/en/siryou/content/001375676.pdf> (21.11.2023).
- Japan Tourism Agency (2020b). International visitor survey: calendar year 2019. <https://www.mlit.go.jp/kankochou/en/siryou/toukei/content/001335770.xls> (21.11.2023).
- Keeler, B. L., Wood, S. A., Polasky, S., Kling, C., Filstrup, C. T., & Downing, J. A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, 13(2), 76–81. doi: 10.1890/140124.
- Kalinić, C. & Vujičić, M. (2019) A subnational assessment of hotel social media metrics – the case of Serbia. *Geographica Pannonica*, 23(2), 87–101. doi: 10.5937/gp23-19968.
- Kemp, S. (2023). Digital 2023: Japan. <https://datareportal.com/reports/digital-2023-japan> (21.11.2023).
- 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. doi: 10.1016/j.tourman.2018.12.005.
- Liao, Y., Yeh, S., & Jeuken, G. S. (2019). From individual to collective behaviours: exploring population heterogeneity of human mobility based on social media data. *EPJ Data Science*, 8(1), 34. doi: 10.1140/epjds/s13688-019-0212-x.
- Luque-Gil, A. M., Gómez-Moreno, M. L., & Peláez-Fernández, M. A. (2018). Starting to enjoy nature in Mediterranean mountains: crowding perception and satisfaction. *Tourism Management Perspectives*, 25, 93–103. doi: 10.1016/j.tmp.2017.11.006.
- Lu, W., & Stepchenkova, S. (2012). Ecotourism experiences reported online: classification of satisfaction attributes. *Tourism Management*, 33(3), 702–712. doi: 10.1016/j.tourman.2011.08.003.
- Ministry of Internal Affairs and Communications (2020). White paper 2020: information and communications in Japan. <https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2020/2020-index.html> (21.11.2023).

- Ministry of the Environment (2023). 国立公園利用者数 (公園、年次別) [Number of national park visitors by park and year]. https://www.env.go.jp/park/doc/data/natural/naturalpark_04.pdf (21.11.2023).
- Miyasaka, T., Oba, A., Akasaka, M., & Tsuchiya, T. (2018). Sampling limitations in using tourists' mobile phones for GPS-based visitor monitoring. *Journal of Leisure Research*, 49(3–5), 298–310. doi: 10.1080/00222216.2018.1542526.
- Muñoz, L., Hausner, V., Brown, G., Runge, C., & Fauchald, P. (2019). Identifying spatial overlap in the values of locals, domestic and international tourists to protected areas. *Tourism Management*, 71, 259–71. doi: 10.1016/j.tourman.2018.07.015.
- Rusdi, J. F., Salam, S., Abu, N. A., Sunaryo, B., Taufiq, R., & Muchlis, L. S. (2019). Dataset smartphone usage of international tourist behavior. *Data in Brief*, 27. doi: 10.1016/j.dib.2019.104610.
- Sinclair, M., Mayer, M., Woltering, M., & Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management*, 263, 110418. doi: 10.1016/j.jenvman.2020.110418
- Sotiriadis, M. D. (2017). Sharing tourism experiences in social media: a literature review and a set of suggested business strategies. *International Journal of Contemporary Hospitality Management*, 29(1), 179–225. doi: 10.1108/IJCHM-05-2016-0300.
- Stankov, U., Jovanović, T., Pavluković, V., Kalinić, Č., Draškulić-Kovačević, N., & Cimbaljević, M. (2018). A regional survey of current practices on destination marketing organizations' Facebook Pages: the case of EU and U.S. *Geographica Pannonica*, 22(2), 81–96. doi: 10.5937/22-16673.
- Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences*, 113(46), 12974–9. doi: 10.1073/pnas.1614158113.
- 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, 222–32. doi: 10.1016/j.tourman.2014.07.003.
- Winter, P. L., Selin, S., Cerveny, L., & Bricker, K. (2020). Outdoor recreation, nature-based tourism, and sustainability. *Sustainability*, 12(1), 81. doi: 10.3390/su12010081.
- 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. doi: 10.1038/srep02976.
- World Tourism Organization – UNWTO. (2019). International Tourism Highlights, 2019 Edition. pp. 24. doi: 10.18111/9789284421152.
- Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services*, 24, 68–78. doi: 10.1016/j.ecoser.2017.02.009.