Tourism Management 72 (2019) 249–256

Contents lists available at ScienceDirect

Tourism Management

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

Quantifying nature-based tourism in protected areas in developing countries by using social big data

Yoonjung Kim[a,b], Choong-ki Kim[a,*], Dong Kun Lee[b], Hyun-woo Lee[a], Rogelio II. T. Andrada[c]

[a] Division of Natural Resources Conservation, Korea Environment Institute, Sejong, Republic of Korea
[b] Department of Landscape Architecture and Rural System Engineering, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea
[c] Institute of Renewable Natural Resources, College of Forestry and Natural Resources, University of the Philippines, Los Baños, Philippines

ARTICLE INFO

Keywords:
Nature-based tourism
Social big data
Cultural ecosystem service
Protected area
Spatial visitation pattern
ASEAN heritage park

ABSTRACT

Spatial visitation patterns and its features on nature-based tourism are difficult to assess using only a field-based survey, which is costly and labor intensive. However, understanding of a protected area's visitation status is critical, as it can strongly influence the sustainability of natural resources. Hence, it is important to identify 'where people visit' and 'why people visit,' to evaluate the features attractive to tourists. In this regard, we proposed and applied social big data to investigate nature-based tourism in an ASEAN Heritage Park. Overall, our research was able to effectively illustrate spatial patterns of visitation using 10 years of Flickr geo-tagged photographs. Hotspots of high visitation were identified, while revealing the local spatial impact of distributed attributes. This study offers insights into the applicability of social big data to protected-area management and its potential in reinforcing existing field-based participatory approaches.

1. Introduction

Nature-based tourism is a cultural service defined as 'the non-material benefits people obtain from ecosystems through spiritual enrichment, cognitive development, reflection, recreation, and aesthetic experiences' (Assessment, 2005). In addition, it offers the opportunity to experience the various benefits of ecosystems by facilitating the human-ecosystem relationship (Chan, Satterfield, & Goldstein, 2012; Daniel et al., 2012; Moyle et al., 2017; Torland, Weiler, Moyle, & Wolf, 2015). In fact, the increase in public visitation, reflecting nature-based tourism, may significantly impact the management of protected areas (PA). Major management elements of PA, such as spatial zoning for supporting attributes (e.g. accommodations, restaurants), budget planning, and development of recreation programs, can be strongly influenced by the density and spatial variation of visitation. On the other hand, an increase in economic benefit along with increased visitors support the financial stability of conservation activities that enhance the management capacity (Steven, Castley, & Buckley, 2013). In this regard, the Convention on Biological Diversity (CBD) emphasized the importance of the evaluation of nature-based tourism and attention to sustainable management to prolong the benefits of reducing poverty and promoting environmental protection (CBD, 2016). In this vein, an assessment of spatial visitation patterns and their characteristics is strongly required to obtain essential information on tourism management.

The most representative method for evaluating nature-based tourism is to quantify potential tourism areas through geographical spatial analysis, expenditure analysis, and surveys of people's satisfaction/preference through questionnaire or interview (Nahuelhual, Carmona, Lozada, Jaramillo, & Aguayo, 2013). As a matter of fact, though identification of spatial visitation patterns has been emphasized, evaluation of nature-based tourism generally has relied on questionnaire- or interview-based assessment that considers some of the diverse stakeholders (Chan et al., 2012). Furthermore, the collected field information, in general, is often too limited to be able to fully reflect the spatio-temporal characteristics of visitation or consider multi-destinations of visitors (Hanson, 1980; Heberling & Templeton, 2009; Sessions, Wood, Rabotyagov, & Fisher, 2016; Önder, 2017). In this regard, there is a need to reflect various preferences and impressions on tourist choices in destinations, and reinforce pre-existing quantified approaches toward cultural services (Assessment, 2005). However, it is challenging in reality to consider a large number of visitors on multi-destinations, as the collection of such field data is expensive and labor-intensive (Heikinheimo et al., 2017). Hence, fundamental questions in relation to public visitation such as 'Which locations do people prefer?', and 'What are the characteristics of visitation patterns regarding features inside tourist destinations?' are difficult to identify.

* Corresponding author. Bldg B, 370 Sicheongdaero, National Research Complex, Sejong, 30147, Republic of Korea.
E-mail address: ckkim@kei.re.kr (C.-k. Kim).

https://doi.org/10.1016/j.tourman.2018.12.005
Received 8 February 2018; Received in revised form 5 December 2018; Accepted 6 December 2018

Compared to field data, social big data has gained significant recognition for the exponential growth of its usage. The generation of data, such as geo-tagged photographs and tweets, has been estimated at 2.5 exabytes per day (Bello-Orgaz, Jung, & Camacho, 2016; IBM, 2015). The growth of social big data is particularly meaningful for the assessment of nature-based tourism as a cultural service, because it can provide insights into the preferences, ideas, values, and behavior of the diverse collection of visitors (Willemen, Cottam, Drakou, & Burgess, 2015). Accordingly, previous studies have identified the ability of geo-tagged photographs to effectively represent overall visitation trends in a spatial context (García-Palomares, Gutiérrez, & Mínguez, 2015; Kurashima, Iwata, Irie, & Fujimura, 2010; Salas-Olmedo, Moya-Gómez, García-Palomares, & Gutiérrez, 2018; Wood, Guerry, Silver, & Lacayo, 2013). Specifically, distributions of crowd-sourced images uploaded on Flickr correlated well with observed visitation data at recreational sites (Sessions et al., 2016; Wood et al., 2013). Using social big data, attempts have been made to reveal the management issues of PA, by analyzing the variation in spatial visitation patterns and its relationship with tourism attributes. For instance, Hausmann et al. (2017) identified that preference to visit sub-Saharan protected areas is increased where key species are present (Hausmann et al., 2017). L. Sonter, Watson, Wood, and Ricketts (2016) discovered the influence of changes in landscape attributes on the visitation rate, quantified using geo-referenced images on Flickr (Sonter et al., 2016).

Such an application of big data is in its early stages. To assess the applicability and effectiveness of social big data on PA management, there is a need to verify and examine the use of crowd-sourced information (Heikinheimo et al., 2017). Furthermore, due to the fact that investigating visitation satisfaction using a field survey is generally conducted rather than social big data analysis, it is important to offer insights into how social big data can supplement or reinforce the pre-existing approach. In particular, PAs in developing countries have high conservation value and recognition but poor field data necessary to quantify nature-based tourism. Moreover, compared to unique values, the PAs in developing countries are experiencing a rapid growth in visitation, compared to those in developed countries (Balmford et al., 2009). However, as far as we know, the application of social big data to major PAs in developing countries is rarely conducted. Therefore, though the usage of social big data is necessary for application to every PA, because of the ecological and cultural importance and necessity, the applicability and effectiveness of the social big data approach should be also considered in developing countries.

Hence, this study aims to apply and validate the innovative modeling approach using social big data to offer key information supporting sustainable management at the ASEAN Heritage Park, which is the selected protected area with outstanding biodiversity across the ASEAN region. In particular, we aim to identify spatial visitation patterns by applying 10-year-accumulated geotagged photographs from Flickr (www.flickr.com). Furthermore, to analyze characteristics of preference, the spatial regression relationship between the identified visitation pattern and the distribution of attributes—natural and cultural attractions and tourism-supporting artifacts—was investigated by conducting geographically weighted regression. As a whole, this study offers insight into the application of social big data for tourism management in PA, including the effectiveness and limitation of its usage, and its advantage in reinforcing an existing field survey. It is expected that the results of this study will contribute to providing an understanding of the effective usage of crowd-sourced big data for the sustainable tourism management of PA, where field data may be limited.

2. Method

2.1. Study area

Southeast Asia is geographically known for its unique group of countries that share organism-inhabiting ecosystems that are common to the countries in the region. ASEAN Heritage Parks (AHPs) are hotspots for biodiversity, boasting of a unique group of ecosystems that host more than 300 threatened and endangered species of vertebrates. In particular, AHPs are defined as "protected areas of high conservation importance, preserving in total a complete spectrum of representative ecosystems of the ASEAN region" (ASEAN Centre for Biodiversity, 2017).

The 38 AHPs represent a myriad of habitats that range from mountain peaks and caves to mangrove forests and coral reefs. The coral-reef ecosystems of Indonesia, Malaysia, the Philippines, and Thailand rival that of Australia's Great Barrier Reef. Incidentally, the beaches and mangroves of Southeast Asia are also breeding grounds for turtles and various marine life, such as sharks, reptiles, and birds. Furthermore, there are 5 AHPs that are also designated UNESCO World Heritage Sites, which cements the status of AHPs as the key repositories of not only natural and genetic resources but also cultural resources that represent the uniqueness of the region. This makes them a highly conducive venue for not just biodiversity conservation but also environmental education and ecotourism (ACB (ASEAN Centre for Biodiversity), 2017).

AHPs promote the facilitation of nature-based tourism and its sustainable tourism management. Though each AHP has different characteristics (e.g., accessibility, flora, and fauna species), the status of nature-based tourism should be identified for entire AHPs. All the designated AHPs are comprehensively managed in coordination with ASEAN member states. To initially evaluate the usage of social big data in such AHPs, in this study, the 15 AHPs with the largest number of geotagged photographs were investigated to explore the applicability of social big data (Fig. 1). Among the evaluated AHPs, the AHP with the highest number of geo-tagged photographs was designated as the main study site to offer insights into the utilization of social big data. In this regard, Tarutao National Marine Park, located in Thailand, was evaluated. The location of Tarutao National Marine Park and the investigated 15 AHPs are illustrated in Fig. 1.

2.2. Data collection

We base our analysis on crowdsourced information with high accessibility, which is open to public. Specifically, Flickr (www.flickr.com) was the chosen source of social big data for this study, as photographs were available as open-source data. In particular, the number of mean annual photographs per user (photo-user-day) of each area was quantified by applying a Python model called natcap.invest.recreation from the Natural Capital Project (www.naturalcapitalproject.org). In this study, the coordinates derived from Flickr-geotagged images from 2005 to 2014 were used.

Data regarding the location of attributes inside the Tarutao National Marine Park, was obtained using Open Street Map (OSM; www.openstreetmap.com). Open Street Map is a crowd-sourced spatial information source, in which geographical information is volunteered, enabling users to acquire spatial data that is fairly accurate (Haklay, 2010). Coordinates of major tourist attractions and supporting artifacts can be acquired as a point shape file, though certain PA may have limited official information. In this study, a total of 168 coordinates describing natural attractions (e.g., waterfall, beach), cultural attractions (e.g., historical prison site), and tourism-supporting artifacts (e.g., accommodation facility) were collected.

2.3. Identification of spatial visitation pattern

Spatial visitation patterns were identified by quantifying the proxy 'photo-user-day' (PUD), developed by Wood et al. (2013). PUD is the estimated proxy for the number of visitors. It indicates the annual average number of people in a day who uploaded at least one image (Wood et al., 2013). If PUD is quantified as 10, it indicates that the site has an average of 10 visitors a day over the course of a year. The PUD

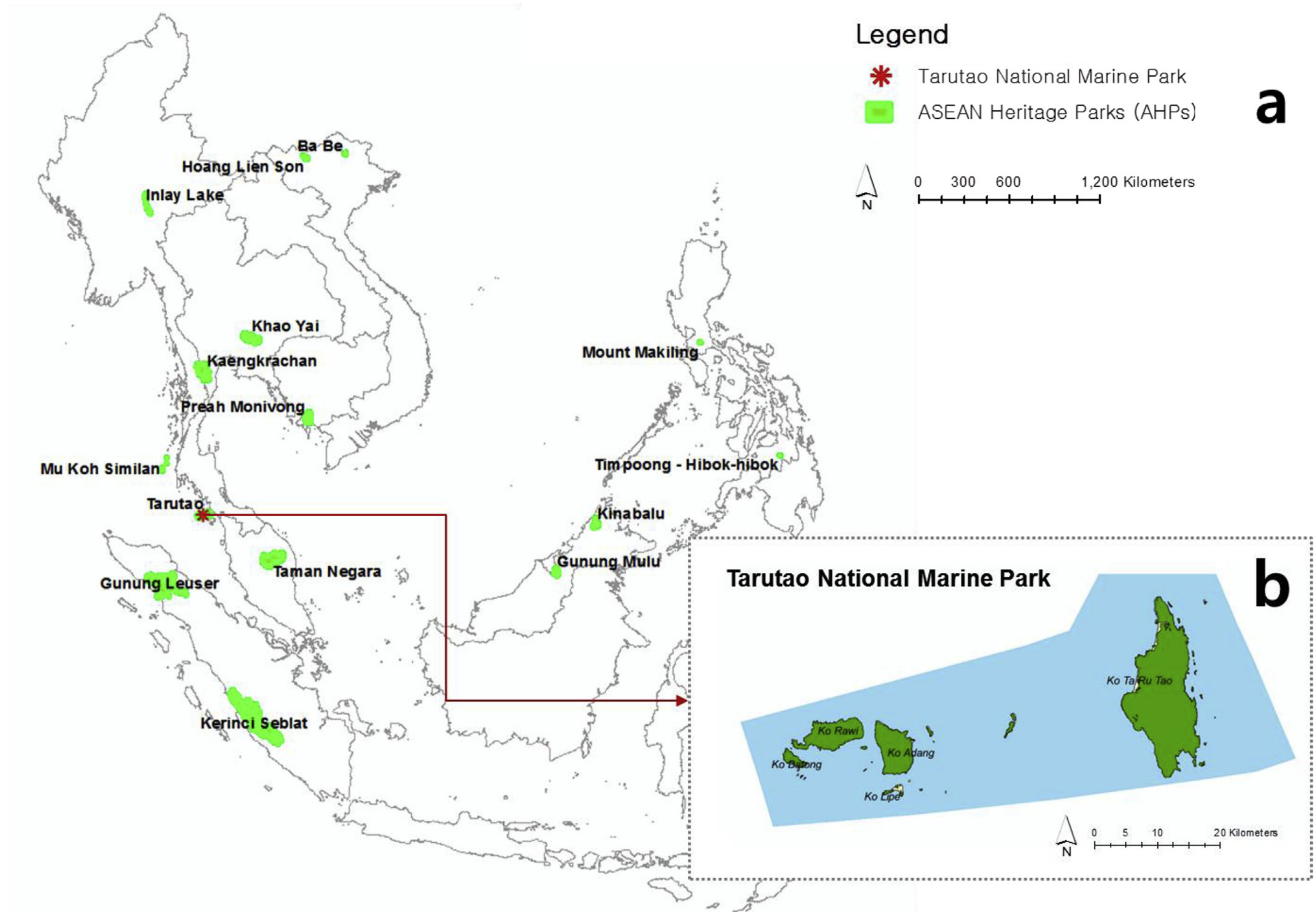


Fig. 1. Map of a) the evaluated 15 ASEAN Heritage Parks (AHPs) and b) Tarutao National Marine Park located in Thailand. Coordinates from geotagged photographs were evaluated for the 15 AHPs using social big data, particularly for Tarutao National Marine Park.

itself should not be used as an indicator reflecting the actual number of visitors. Instead, it should be interpreted as a proxy reflecting the variation in visitation densities for each area, as it cannot represent all past visitors at the considered site.

In this study, PUD was calculated per 1 km × 1 km grid. Each grid was constructed in hexagonal form. The diameter for each hexagon was 1 km. In particular, the PUD value based on the 10-year-accumulated geo-tagged photographs from 2005 to 2014 was quantified. The identified PUD was converted into a map using ArcGIS 10.6, reflecting varying visitation densities. Tarutao National Marine Park, divided into 1 km × 1 km, contained a total of 2076 hexagonal grids. The total analyzed area was 1349 km^2. The number of hexagonal grids containing at least one photograph was total 274. To illustrate the visitation hotspot, every PUD value on study site was divided into top 5%, 10%, 25%, and 50%, revealing the visitation hotspots that have relatively high visitation.

2.4. Validation of PUD

To validate the PUD values, we compared them with the observed tourism revenue at the Tarutao National Marine Park. The collected information on tourism revenue (e.g., admission charges) reflects the activities of the visitors at their multi-destinations. The revenues were aggregated on a monthly basis. However, with regard to the visitations that occurred in 2014, Tarutao National Marine Park did not account for the visitation revenue in July. Thus, 11 datasets, which include monthly revenues in 2014 for all months except July, were compared with each monthly mean PUD value. In particular, we quantified the Pearson correlation coefficient as follows: 1) Top 1% PUD values (visitation hotspots where people frequently visit) were compared with the observed tourism revenue; 2) the overall PUD values per 1 km × 1 km were compared with the observed tourism revenue.

Furthermore, to validate the identified visitation pattern, face-to-face focus group interviews were conducted with five regional experts including regional management officers and scholars in AHPs. The suitability of the identified visitation pattern compared to local knowledge was investigated. The discussion was conducted on November 4, 2016 in Seoul, Republic of Korea. The interview and discussion lasted about two hours.

2.5. Evaluating visitation pattern regarding distribution of attraction and artifact

In this study, geographically weighted regression (GWR) was performed to evaluate the characteristics of preference that reveal why people visit. Regarding the distribution of natural attractions, cultural attractions, and tourism-supporting artifacts, the regression impact of such attributes on visitation pattern was analyzed using spatial statistics. GWR measures the local spatial impact between dependent and independent variables by calculating regression coefficient at each individual location (Fotheringham, Brunsdon, & Charlton, 2003). Compared to ordinary statistical analysis, including ordinary least squares (OLS) linear regression, GWR can draw information reflecting the 'local spatial pattern.' It is a type of local statistics that shows how a relationship varies over space to understand possible hidden causes of spatial patterns (Fotheringham et al., 2003). R^2 and regression coefficients are quantified by performing GWR, which reveals an independent variable's impact on a dependent variable. In this study, attributes—natural attractions, cultural attractions, and tourism-supporting artifacts—were considered independent variables (S1, S2). The dependent variable was PUD, the proxy for visitation. To perform GWR, the kernel (neighborhood) type and bandwidth (distance limit of analysis) method were selected using ArcGIS10.6. We applied the adaptive kernel that reflects the spatial density of the features. The bandwidth is the distance around each observation point, which determine the size of the observation spatial range (Fotheringham et al., 2003). The bandwidth is defined by selecting the number of neighborhood cells. If the determined number of neighborhood cells is smaller, the observation spatial range of the spatial regression analysis is more confined. We performed GWR with a default value of '30' and a smaller value of '10'. To reflect a more confined impact range, we defined the number of neighborhood cells as '10'.

3. Results

3.1. Applicability of social big data in 15 AHPs

Aggregating the crowdsourced information from 2005 to 2014, Tarutao National Marine Park was determined to have the most preferred visiting spot—maximum PUD—among the 15 AHPs (Table 1). Considering the fact that an AHP is a selected protected area with high levels of conservation and recognition, Tarutao National Marine Park was assessed to have a typically high preference in the ASEAN region. The highest number of PUD in the hexagonal unit was 33.3.

The 15 AHP spatial visitation preferences were calculated based on the collected coordinates of geo-referenced images. The maximum PUD for each evaluated protected area and the median value of PUD are illustrated.

3.2. Spatial visitation pattern of Tarutao National Marine Park

The use of social big data for Tarutao National Marine Park identified the number of PUD, revealing where people particularly visit (Fig. 2). Visiting densities in every area of the PA were identified by generating the map showing variations in the number of PUD. Hotspots having the top 5% visitation rates were revealed in the study site (Fig. 2). The results indicated that Ko Lipe island and Ko Tarutao island were the most frequently visited islands. Ko Adang and Ko Rawi islands were also identified as frequently visited places.

We validated the calculated PUD within the observed field-collected data by performing Pearson correlation analysis (Fig. 3). The quantified PUD was significantly correlated with the visitation revenue (r = 0.90, $p < 0.001$). The estimation of visitation for the most visited hotspot (top 1% across Tarutao National Marine Park) also showed a significant correlation with the visitation revenue (r = 0.83, $p < 0.001$). Conducted focus group interviews validated the identified spatial visitation pattern's applicability. The opinions of regional managers and experts on visitation hotspots and related visitation patterns were compared within the identified PUD, corresponding well with the local knowledge on where people mostly visit and representative visitation patterns. In particular, the identified map was the most effective piece of information to investigate overcrowding place and the areas where tourism activities should be facilitated.

Table 1
Maximum and median PUD (Photo-User-Days/yr) per evaluated AHPs (ASEAN Heritage Parks).

Name of ASEAN Heritage Park	Country	Maximum PUD	Median PUD
Tarutao National Marine Park	Thailand	33.3	1.15
Taman Negara	Malaysia	20.6	0.70
Kinabalu	Malaysia	16.1	1.90
Inlay Lake	Myanmar	8.1	2.95
Khao Yai	Thailand	6.7	1.10
Mount Timpoong - Hibok-hibok	Philippines	6.7	0.45
Ao Phangnga - Mu Koh Surin – Mu Koh Similan	Thailand	5.4	1.45
Hoang Lien Son - Sa Pa	Vietnam	3.3	0.60
Preah Monivong	Cambodia	3.3	0.60
Kaengkrachan Forest Complex	Thailand	2.9	0.65
Mount Makiling	Philippines	2.8	0.90
Gunung Leuser	Indonesia	1.6	0.40
Gunung Mulu	Malaysia	1.6	0.60
Ba Be	Vietnam	1.2	0.55
Kerinci Seblat	Indonesia	1	0.45

3.3. Impact of attraction and artifact on visitation pattern

The results of GWR showed the influence of three categorized attributes—natural attractions, cultural attractions, and tourism-supporting artifacts—on identified visitation patterns (Fig. 4). R^2 revealed the validity of the spatial regression analysis regarding each considered attribute and surrounding visitation pattern. The regression coefficient illustrated the direction and size of the impact on visitation pattern for the three categorized attributes.

The highest regression coefficient was observed for the natural attractions ($\beta = 6.5$), while the lowest was detected for the tourism-supporting artifacts ($\beta = -0.3$). Specifically, compared to cultural attractions and tourism-supporting artifacts, natural resources had a wider impact on visitation densities. The strong impact of natural attractions was mainly detected in the northern part of Ko Tarutao, southern part of Ko Adang, Ko Lipe, and Ko Batong. On the other hand, the impact of cultural attractions on visitation patterns was mainly found in the eastern part of Ko Tarutao. As for tourism-supporting artifacts, Ko Lipe and part of Ko Tarutao, Ko Rawi, and Ko Adang were areas showing higher visitation due to the distribution of tourism-supporting artifacts.

4. Discussion

Identification of spatial patterns of public visitation is necessary to evaluate the benefit and appropriateness of tourism activities in PA. The importance of mapping and quantification on the benefits of nature has been emphasized, as it can effectively improve the process of decision making to manage natural resources (Burkhard, Kroll, Nedkov, & Müller, 2012; Daily et al., 2009; Daily & Matson, 2008). The problem is that the field data for PA is often too limited to quantify the preference of nature-based tourism regarding actual visitation patterns. Therefore, in this study, we have proposed and applied an innovative modeling approach that uses social big data to reveal spatial visitation patterns inside ASEAN Heritage Park. Moreover, as there is barely any information on 'how to apply the social big data in PA management,' or 'how to complement existing field surveys on visitation satisfaction with the usage of social big data', we provide insight into the application of social big data for PA tourism management.

4.1. Applicability of social big data to evaluate nature-based tourism

Consideration of diverse visitors and their multi-destinations is a main challenge in tourism management of PA. To overcome the challenge, mapping tools such as public participatory geographic information systems (PPGIS) were developed that link public and expert perceptions of geographical information (Gregory Brown, 2004). However, labor-intensive techniques and time to apply such tools on a large scale are required. On the other hand, extracted GPS-based location information from social media can be effectively applied in large-scale assessment. Using GPS coordinates from Flickr geotagged images, the median and maximum PUD of the 15 AHPs, which were widely distributed across various ASEAN countries, were quantified. Since social big data embraces various spatio-temporal ranges of diverse visitors

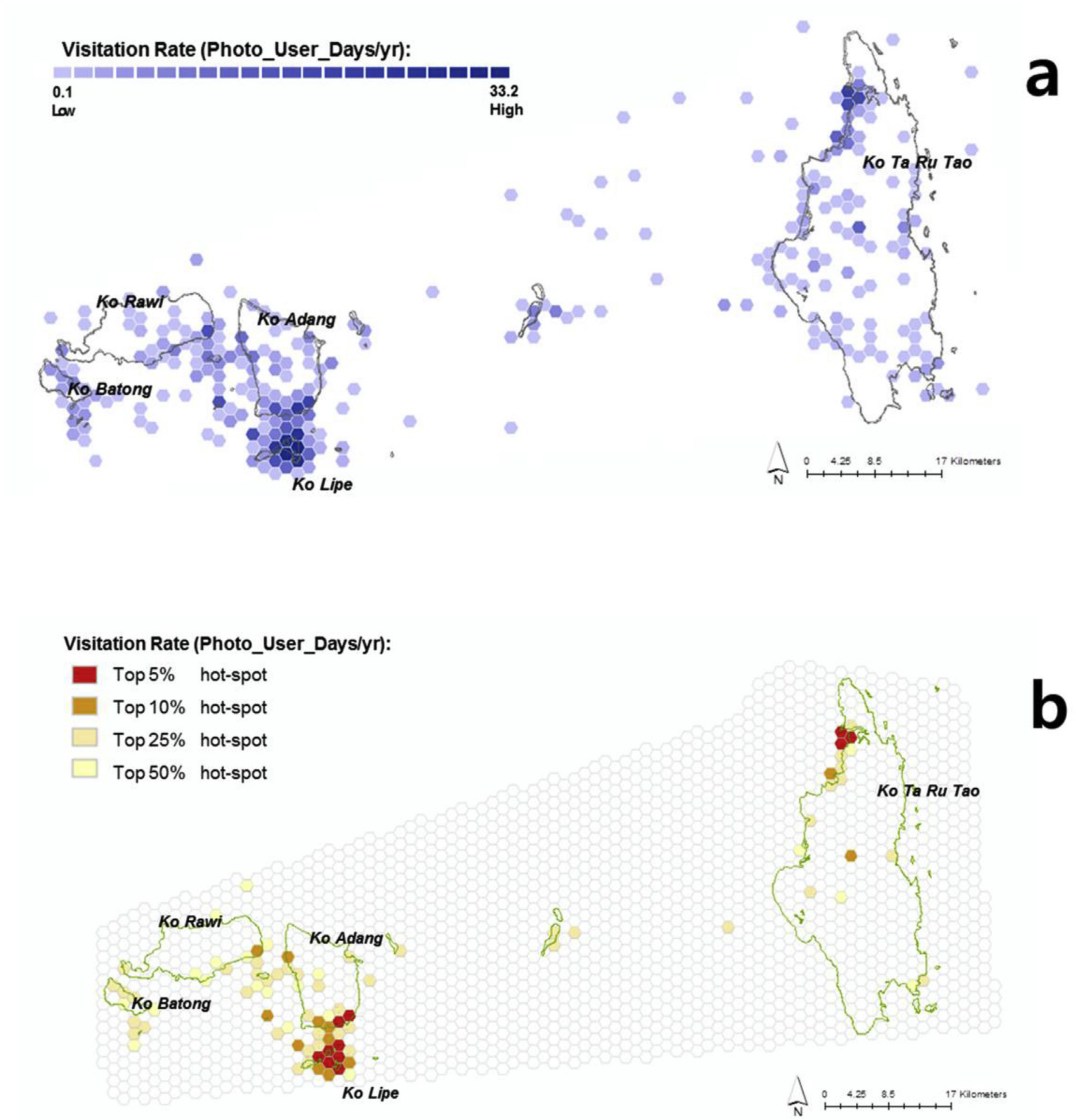


Fig. 2. Evaluated spatial visitation pattern of Tarutao National Marine Park. a) Visitation rate expressed as PUD (Photo-User-Days/yr) of Tarutao National Marine Park; b) Visitation hotspots having top 5%, 10%, 25%, 50% PUD values.

with high cost-efficiency, we assumed that the PUD can be used for the comparison of nature-based tourism in a domestic and global context.

In line with that, the identified PUD in Tarutao National Marine Park was evaluated to be consistent with the observed data. The overall PUD in Tarutao National Marine Park showed a significant relationship with the observed tourism revenue (Pearson's r = 0.90, p = 0.001). While only considering the region with the top 1% in visitation, the Pearson coefficient indicated a high degree of correlation (r = 0.83, p = 0.001). In agreement with these results, S. Wood et al. (2013) and C. Sessions et al. (2016) validated the objectivity of PUD (Sessions et al., 2016; Wood et al., 2013). Regarding 831 tourist destinations in 31 countries, S. Wood et al. (2013) showed that PUD was consistent with observed number of visitors (Pearson's r = 0.6–0.8). C. Sessions et al. (2016) verified the PUD for 38 U.S. National Parks, identifying that PUD is statistically significant with the observed monthly data (Pearson's r = 0.65). In this study, local knowledge of visitation hotspots was also matched with the spatial visitation pattern derived from PUD. Although further verification of social big data should be conducted with sufficient observed data, the results support the applicability of social big data to identify the spatial variance on overall visitation tendencies and hotspots for tourism.

Moreover, the results of GWR between PUD and distribution patterns of attributes showed how the diverse attributes inside PA—natural attractions, cultural attractions, and tourism-supporting artifacts—had influenced public visitation. Visitation patterns from the three categorized attributes were clearly different, which revealed the characteristic of spatial variance of visitation. As location of attributes can be further obtained regarding crowd-sourced information, such as Open Street Map, further analysis of targeted tourism resources can be performed for different management purposes.

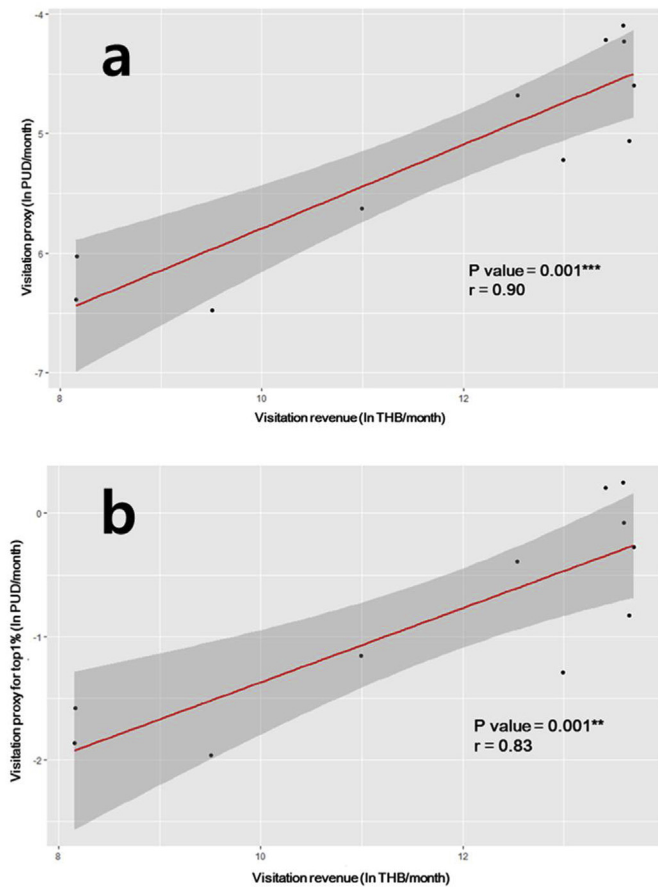


Fig. 3. Correlation between PUD (Photo-User-Days/yr) and tourism revenue. Pearson correlation coefficient for a) verification of PUD (Photo-User-Days/yr) for the entire area, b) verification of PUD (Photo-User-Days/yr) for the top 1% hotspot. THB (Thai Baht) represents the currency of Thailand. Figure created using ggplot2 package for R.

4.2. Effectiveness and usefulness for PA management

Considering that ASEAN countries, as well as other regions, have limited attempts to quantitatively assess cultural services at present (Shoyama, Kamiyama, Morimoto, Ooba, & Okuro, 2017), new types of collected data may contribute to the combination of visitation values of cultural service with spatial management. Factual spatial visiting patterns can be reflected in PA management by supporting policy-makers and park managers. In this study, we identified the degree of visit for each location across the PA. Moreover, we suggested how different types and distributions of attributes in the PA affect visitation patterns. Such information can be used in developing tourism management activities such as the creation of new recreation programs, facilities, or roads to facilitate cultural services in PA. Furthermore, as an unsustainable visitation pattern can turn nature-based tourism into a threat (Daniel et al., 2012; Liddle, 1997; Reed & Merenlender, 2008), social big data's ability to identify the degree of visitation in each location is essential to promote biodiversity conservation (Di Minin, Tenkanen, & Toivonen, 2015). Identified information on visitation hotspots are expected to be used for further analyzing the threatening areas that have exceeded their carrying capacity. Moreover, revealed spatial visitation pattern can contribute to develop focused management plan such as building viewing platform or alternative roadway for sustainable tourism.

Participation of policy-makers and local managers can improve the quantity and quality of data by leveraging modern technology to evaluate specific subjects in a management agenda or certain needs of various visitors (Sessions et al., 2016). This means that the approach to analyzing the daily visitor count can also be extended to correlate the visitation pattern with weather or other subject of management interest (Sessions et al., 2016). In this regard, geo-tagged photographs can be further used to fulfill various objectives of management, such as 'surveying the visitation pattern in the off-season' or 'evaluating a specific location's visitation pattern, where accidents were frequently observed'.

4.3. Coordination with survey-based participatory approach

This study suggested that social big data can be a powerful tool to evaluate nature-based tourism. However, field surveys are required to investigate 'visitation satisfaction' or 'perceived value on nature-based tourism', which is related with public perception and emotion. For instance, the field survey at Tarutao National Marine Park was able to reveal public visitation satisfaction and willingness to pay at distributed attributes (S3). That is, the perceived value should be further investigated regarding people's subjective opinion or choice, since identified spatial visitation patterns cannot solely represent people's various emotions or diverse satisfaction rates for each tourism resource. Thus, in coordinating between the two methodologies, it is important to analyze not only the spatial characteristics of tourism but also the diverse perceived value of visitors.

4.4. Limitation and challenge

There are concerns that big data acquired from social media may have biases. First of all, the amount of social big data can vary depending on the visitor's characteristics and circumstances. For instance, the distance to their home location can impact the frequency of photo-taking activities (Wood et al., 2013). Similarly, the percentage of foreign visitors and nationalities can influence the amount of social big data. Next, the GPS function of a mobile device can highly influence the quality of social big data. Thus, some PAs might have low credibility in the analyzed results due to the malfunction of the GPS signal. Finally, for each country or region, each social-media platform, such as Flickr, Twitter, and YouTube, may show different visitation patterns. Moreover, the number of photographs may differ depending on the user's age or characteristics of attraction.

Therefore, in further analysis, we suggest using multiple big data to reveal the diverse preference across PAs. Furthermore, consideration of the affluent time range is required for data collection. In this study, social big data accumulated for 10 years was used. More attempts should be made to validate the credibility of social big data with sufficient field data. However, even though there are such limitations and obstacles, our findings show that PUD can reveal the spatial visitation patterns of public visitation with high cost-efficiency. To increase the applicability of social big data, the above-mentioned problems need to be considered during further evaluation.

5. Conclusions

This study indicates how the use of geo-referenced images on Flickr can evaluate nature-based tourism in protected areas. Although data from such sources may be limited, the volume of such data possesses value, and it is cost effective. It is suggested that such data can aid the management of protected areas by identifying spatial patterns of visitation, the changes of behavior that can result from inducing changes in signposting, track management and the imposition of restrictions and the provision of alternatives such as viewing platforms to control flows of visitors. As unsustainable tourism can be a risk to the entire ecosystem services in a PA, the potential of social big data to discern such spatial visitation dynamics is a powerful and innovative tool to support tourism management. Hence, regarding the advantages and possibilities of social big data, conducting suggested approach with existing field-based survey can generate robust information to achieve sustainable

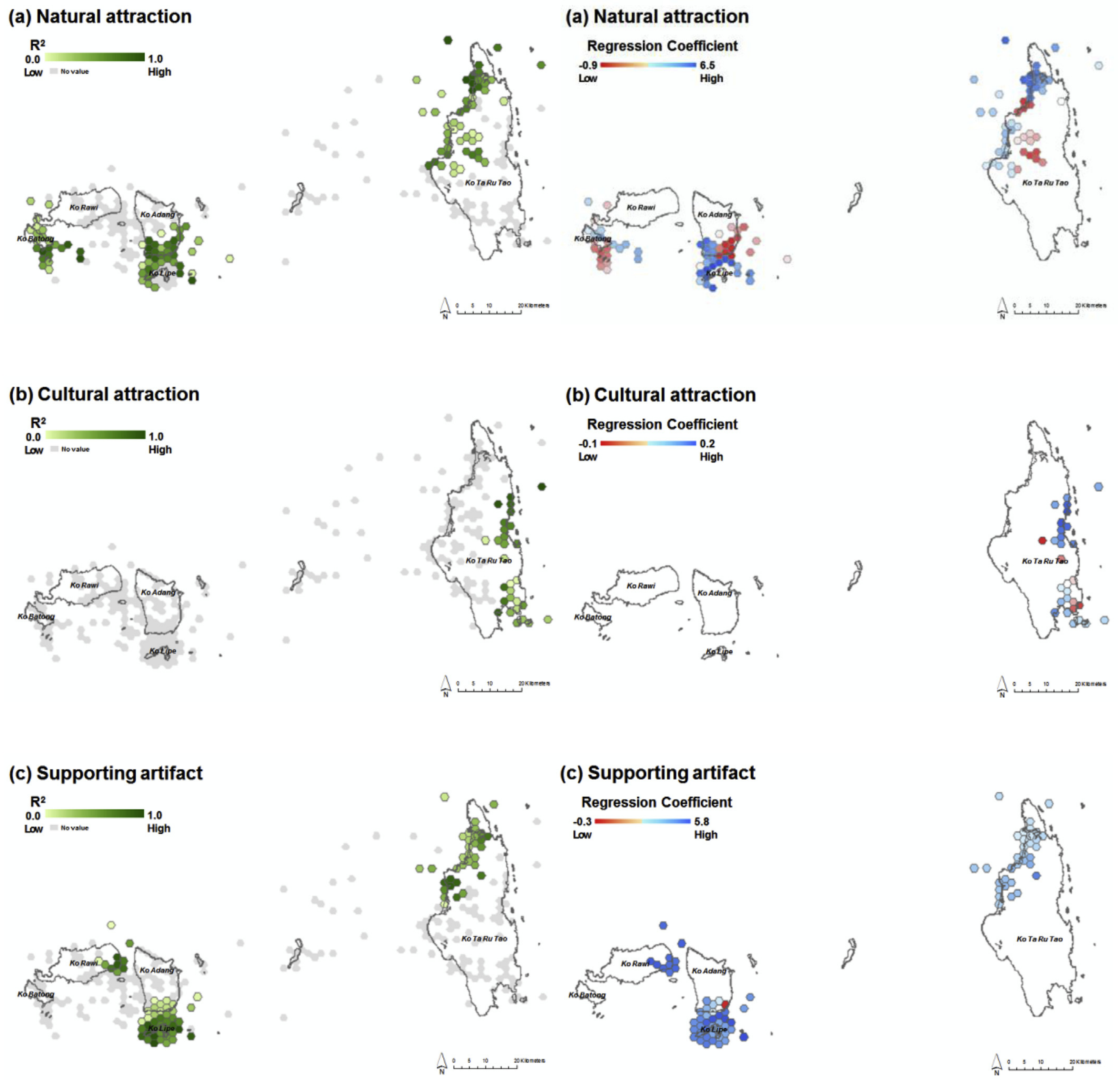


Fig. 4. The results of geographically weighted regression (GWR) illustrate the local spatial impact of attributes on spatial visitation preference. For each GWR model, R^2 and regression coefficient were identified for each area of Tarutao National Marine Park, which showed the impact of the three categorized attributes on public visitation.

tourism.

CRediT authorship contribution statement

Yoonjung Kim: Writing - original draft. **Choong-ki Kim:** Writing - original draft.

Acknowledgements

This research was conducted in regard to "Sustainable management of ASEAN Heritage Parks through valuing and improving eco-tourism", as a Bio-Bridge Initiative under Convention on Biological Diversity. Ministry of Oceans and Fisheries of Korea government (20170325) and Korea Coast Guard (KCG-01-2017-05) also supported the funding of the project. The project was performed by Korea Environment Institute (KEI).

Appendix A. Supplementary data

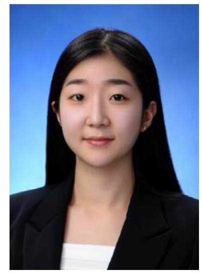
Supplementary data to this article can be found online at https://doi.org/10.1016/j.tourman.2018.12.005.

References

ACB (ASEAN Centre for Biodiversity) (2017). The AHP programme. Retrieved from www.aseanbiodiversity.org.

Assessment, M. E. (2005). *Millennium ecosystem assessment. Ecosystems and human*

wellbeing: A Framework for assessment. Washington, DC: Island Press.
Balmford, A., Beresford, J., Green, J., Naidoo, R., Walpole, M., & Manica, A. (2009). A global perspective on trends in nature-based tourism. *PLoS Biology, 7*(6), e1000144.
Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*, 45–59.
Brown, G. (2004). Mapping spatial attributes in survey research for natural resource management: Methods and applications. *Society and Natural Resources, 18*(1), 17–39.
Burkhard, B., Kroll, F., Nedkov, S., & Müller, F. (2012). Mapping ecosystem service supply, demand and budgets. *Ecological Indicators, 21*, 17–29.
CBD (Convention of Biological Diversity) (2016). CBD/COP/DEC/XIII/3. CBD (convention of biological diversity). Retrieved from https://www.cbd.int/decisions/cop/?m=cop-13.
Chan, K. M. A., Satterfield, T., & Goldstein, J. (2012). Rethinking ecosystem services to better address and navigate cultural values. *Ecological Economics, 74*, 8–18.
Daily, G. C., & Matson, P. A. (2008). Ecosystem services: From theory to implementation. *Proceedings of the National Academy of Sciences, 105*(28), 9455–9456.
Daily, G. C., Polasky, S., Goldstein, J., Kareiva, P. M., Mooney, H. A., Pejchar, L., et al. (2009). Ecosystem services in decision making: Time to deliver. *Frontiers in Ecology and the Environment, 7*(1), 21–28.
Daniel, T. C., Muhar, A., Arnberger, A., Aznar, O., Boyd, J. W., Chan, K. M. A., et al. (2012). Contributions of cultural services to the ecosystem services agenda. *Proceedings of the National Academy of Sciences, 109*(23), 8812–8819.
Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science, 3*, 63.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships.* John Wiley & Sons.
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.
Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682–703.
Hanson, S. (1980). Spatial diversification and multipurpose travel: Implications for choice theory. *Geographical Analysis, 12*(3), 245–257.
Hausmann, A., Toivonen, T. K., Heikinheimo, V. V., Tenkanen, H. T. O., Slotow, R., & Di Minin, E. (2017). Social media reveal that charismatic species are not the main attractor of ecotourists to sub-Saharan protected areas. *Scientific Reports, 7*(1), 763.
Heberling, M. T., & Templeton, J. J. (2009). Estimating the economic value of national parks with count data models using on-site, secondary data: The case of the Great sand dunes national park and preserve. *Environmental Management, 43*(4), 619–627.
Heikinheimo, V., Minin, E. Di, 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.
IBM (2015). Big data and analytics. Retrieved from https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.
Kurashima, T., Iwata, T., Irie, G., & Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 579–588). ACM.
Liddle, M. (1997). *Recreation ecology: The ecological impact of outdoor recreation and ecotourism.* Chapman & Hall Ltd.
Moyle, B. D., Scherrer, P., Weiler, B., Wilson, E., Caldicott, R., & Nielsen, N. (2017). Assessing preferences of potential visitors for nature-based experiences in protected areas. *Tourism Management, 62*, 29–41.
Nahuelhual, L., Carmona, A., Lozada, P., Jaramillo, A., & Aguayo, M. (2013). Mapping recreation and ecotourism as a cultural ecosystem service: An application at the local level in southern Chile. *Applied Geography, 40*, 71–82.
Önder, I. (2017). Classifying multi-destination trips in Austria with big data. *Tourism Management Perspectives, 21*, 54–58. https://doi.org/10.1016/j.tmp.2016.11.002.
Reed, S. E., & Merenlender, A. M. (2008). Quiet, nonconsumptive recreation reduces protected area effectiveness. *Conservation Letters, 1*(3), 146–154.
Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management, 66*, 13–25. https://doi.org/10.1016/j.tourman.2017.11.001.
Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. M. (2016). Measuring recreational visitation at US National Parks with crowd-sourced photographs. *Journal of Environmental Management, 183*, 703–711.
Shoyama, K., Kamiyama, C., Morimoto, J., Ooba, M., & Okuro, T. (2017). *A review of modeling approaches for ecosystem services assessment in the Asian region.* Ecosystem Services.
Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS One, 11*(9), e0162372.
Steven, R., Castley, J. G., & Buckley, R. (2013). Tourism revenue as a conservation tool for threatened birds in protected areas. *PLoS One, 8*(5), e62598.
Torland, M., Weiler, B., Moyle, B. D., & Wolf, I. D. (2015). Are your ducks in a row? External and internal stakeholder perceptions of the benefits of parks in new south Wales, Australia. *Managing Sport and Leisure, 20*(4), 211–237.
Willemen, L., Cottam, A. J., Drakou, E. G., & Burgess, N. D. (2015). Using social media to measure the contribution of Red List species to the nature-based tourism potential of African protected areas. *PLoS One, 10*(6), e0129785.
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*(1), 2976. https://doi.org/10.1038/srep02976.



Yoonjung Kim, PhD candidate (email: kimyj@kei.re.kr) is a researcher at the Korea Environment Institute. Her interests mainly focus on environmental planning and modeling regarding tourism. She has worked on tourism assessment methodologies to calculate the cultural service that nature offers. For the present article, she designed the overall study with the co-authors, collected social big data, performed modeling and field surveys, wrote the first draft, and revised the manuscript.

Choong-ki Kim, PhD (email: ckkim@kei.re.kr) is a research fellow at the Korea Environment Institute. His study areas include ecosystem service, nature-based tourism, and climate change adaptation. His previous work at the Natural Capital Project, Stanford University, involved developing a decision-supporting tool, called Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST), to evaluate the benefits from nature. He applied the InVEST model to various decision contexts to inform decisions for sustainable development. He also worked on the implications of big data and numerical modeling studies. As a corresponding author for this article, he contributed to the design of the overall framework, and also contributed to drafting and revising the manuscript.

Dong Kun Lee, PhD (email: dklee7@snu.ac.kr) is a professor of landscape architecture and rural system engineering at Seoul National University. His research focuses on environmental planning, including various issues such as urban planning, nature-based tourism, ecosystem assessment and protected area management. For this study, he contributed to the study design, spatial regression analysis and application of social big data for tourism management in protected areas.

Hyun-woo Lee, PhD (email: hwlee@kei.re.kr) is a Senior Research Fellow at the Korea Environment Institute. He has done work on national and international environmental policy. His main interests include international environmental cooperation and ecosystem service. For this article, he contributed to the design of the study and provided key knowledge on tourism management issues in protected areas.

Rogelio T. Andrada II, PhD (email: rtandrada@up.edu.ph) is a professor in the College of Forestry and Natural Resources at the University of the Philippines Los Baños. His research is mainly on people's perceptions of tourism and leisure, management of natural resources, and sustainable development. He has also done work on various tourism quantification approaches and their implications. In this study, he supported and co-performed field-based evaluation and provided key local knowledge on the ASEAN Heritage Park for the implications of social big data.