Article

User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey

Vuokko Heikinheimo [1,*], Enrico Di Minin [1,2], Henrikki Tenkanen [1], Anna Hausmann [1,2], Joel Erkkonen [3] and Tuuli Toivonen [1]

[1] Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland; enrico.di.minin@helsinki.fi (E.D.M.); henrikki.tenkanen@helsinki.fi (H.T.); anna.hausmann@helsinki.fi (A.H); tuuli.toivonen@helsinki.fi (T.T.)
[2] School of Life Sciences, University of KwaZulu-Natal, 4041 Durban, South Africa
[3] Metsähallitus, Parks & Wildlife Finland, 96101 Rovaniemi, Finland; joel.erkkonen@metsa.fi
[*] Correspondence: vuokko.heikinheimo@helsinki.fi; Tel.: +358-2941-50760

Abstract: Protected area management and marketing require real-time information on visitors' behavior and preferences. Thus far, visitor information has been collected mostly with repeated visitor surveys. A wealth of content-rich geographic data is produced by users of different social media platforms. These data could potentially provide continuous information about people's activities and interactions with the environment at different spatial and temporal scales. In this paper, we compare social media data with traditional survey data in order to map people's activities and preferences using the most popular national park in Finland, Pallas-Yllästunturi National Park, as a case study. We compare systematically collected survey data and the content of geotagged social media data and analyze: (i) where do people go within the park; (ii) what are their activities; (iii) when do people visit the park and if there are temporal patterns in their activities; (iv) who the visitors are; (v) why people visit the national park; and (vi) what complementary information from social media can provide in addition to the results from traditional surveys. The comparison of survey and social media data demonstrated that geotagged social media content provides relevant information about visitors' use of the national park. As social media platforms are a dynamic source of data, they could complement and enrich traditional forms of visitor monitoring by providing more insight on emerging activities, temporal patterns of shared content, and mobility patterns of visitors. Potentially, geotagged social media data could also provide an overview of the spatio-temporal activity patterns in other areas where systematic visitor monitoring is not taking place.

Keywords: user-generated content; geotagged social media data; recreational services; visitor survey

1. Introduction

User-generated content is rapidly being recognized as a complementary source of data for traditional spatial datasets [1]. Among other forms of user-generated geographic information, location-based social media provide constant feeds of content-rich data generated by users of different platforms sharing their experiences and observations online. These data have the potential to enrich existing data collection methods for mapping spatio-temporal activity patterns and location-based experiences of people. The potential of geotagged social media for mapping people's activities and movements has been mainly illustrated in urban environments [2–7]. There are also promising results for utilizing spatio-temporal information of social media posts for visitor monitoring in recreational

areas [8–13]. However, it still needs to be validated whether social media data can be used as a complementary data source for spatial decision making.

Social media, in general, refers to computer-based applications used for networking and sharing digital content. Here, we focus specifically on location-based social media data that contains spatial information (location), temporal information (time), and relevant content (text and photos) generated by users of different social media platforms (e.g., Flickr, Instagram, and Twitter). The data can be often accessed in large quantities through Application Programming Interfaces (APIs) which allow queries and retrievals of publicly shared information from the platforms. Users can link the shared content (a post) to a location by using place names within the text, linking the post to a pre-defined point-of-interest, or by sharing the coordinates of their device. Users often also share the same content via multiple platforms (for example, sharing an Instagram post via Facebook). All sources of social media data are not the same, as platforms differ in purpose, popularity, user profiles, and terms of usage regarding data retrieval and sharing [14]. Social media presents only a selected representation of reality and in order to use these data properly, it is critical to acknowledge inherent biases in social media data including gender, age, socioeconomic status, and motivations of potential data contributors [7,15,16]. Comparisons with ancillary data sources have been suggested as one way of overcoming some of the limitations of social media data [17]. For example, census data in urban areas [5,7] and visitation rates in recreational areas [9] provide valuable reference information for geotagged social media posts in different environments.

In contrast to other forms of user-generated geographic information, social media data is often generated for other purposes than specific planning or mapping efforts. Consequently, location-based social media can be categorized as a source of passive and non-authoritative crowdsourced geographic information [1]. Social media data, among other forms of 'crowdsensing', are also recognized as a source of primary sensor data in remote sensing literature [18,19]. For example, geolocated twitter data and unsupervised clustering have been used for mapping urban land use patterns [20]. Georeferenced social media data is often discussed in parallel with different sources of Volunteered Geographic Information (VGI) referring to the role of citizens as sensors [21,22]. However, due to the passive role of data contributors, social media data is not purely 'volunteered' [1], and even when shared openly, taking advantage of these data in academic research requires special considerations of ethical use [23]. On the other hand, data collection from social media is less intrusive compared to active forms of crowdsourcing such as collaborative mapping [1] or public participatory GIS (PPGIS) campaigns [24], as people do not need to make an extra effort to participate in the generation of data. Therefore, due to its passive nature, social media data can potentially capture a different view on peoples' activities and opinions in space and time in comparison to active data collection campaigns.

Establishing an understanding of people's activities and opinions is needed in the sectors of planning and management, for example, in conservation areas [14,25]. Nature-based tourism to protected areas is increasing globally [26] and plays a crucial role in generating much needed funding to support biodiversity conservation [27] and promoting environmental awareness [28]. An important limitation in assessing the potential role of nature-based tourism to support biodiversity conservation is, for many protected areas, the lack of data on visitor counts, as well as the activities and preferences of visitors in order to direct management and marketing efforts [29–31]. As in many other fields, there is a growing interest towards using new technology, social media in particular, both as a communication channel and a data source in nature conservation [32,33]. Traditionally, information on protected area visitors has been collected using visitor surveys and counting. In countries with the most advanced visitor monitoring systems, such as in Finland, visitor surveys are usually being carried out systematically at a certain time interval. An important limitation is that such surveys are time consuming and expensive to carry out. Therefore, social media could be combined with, or sometimes even replace, traditional surveys in order to bridge some of the information gaps in conservation science and practice [14]. In the context of nature-based tourism, the spatial and temporal attributes of social media data have been used for the quantification of visitation rates [8,9], tourism revenues [10], landscape values [11], and travel patterns [34]. Social media content (text

and/or photos) have also been used for mapping cultural ecosystem services [12] and to assess visitors' preferences for biodiversity [35]. Now, there is a need to test to what extent social media content reflects the reported experiences and activities of visitors in different environments and cultural contexts. Specifically, there is need to assess how social media data can be used to infer visitors' activities in protected areas that have been created for recreational purposes, rather than with biodiversity conservation objectives in mind. Such information is very important to inform protected area management and marketing anywhere in the world, but particularly in areas that lack resources to monitor visitors' experiences.

In this paper, we examine the potential of social media data in providing relevant information about visitation to a national park (Figure 1). As a case study area, we use the most popular national park in Finland, the Pallas-Yllästunturi National Park (hereafter PY). PY provides a suitable test site for social media data, as it has long been studied using standardized visitor surveys of the Finnish national park authority Metsähallitus—Parks & Wildlife Finland. The most recent survey was carried out in the area in 2016 including questions about visitors' social media usage.
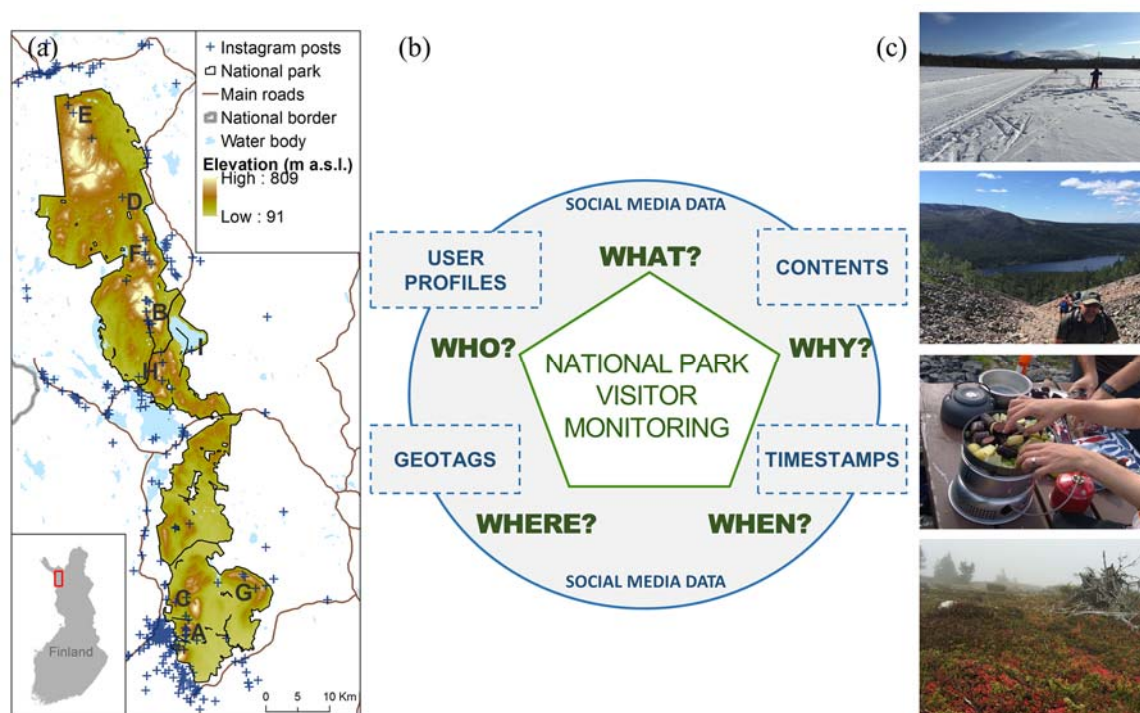

Figure 1. (**a**) Instagram post locations from the case study area of Pallas-Yllästunturi National Park (PY) in Finnish Lapland; (**b**) Framework for using social media data (geotags, timestamps, contents, user profiles) for studying the spatial patterns (where), temporal patterns (when) of visitors/social media users (who) and their activities (what), and motivations (why) (modified from [14]); and (**c**) examples of social media image content from the national park.

The objective of this study was to see how well social media data content corresponds to results derived from traditional national park visitor surveys. Moreover, we wish to understand what complementary information could be derived from social media data regarding visitation patterns and activities in the park. We apply the framework in Figure 1 by exploring if and how the following questions could be answered based on the social media data and visitor survey data: (i) where do people go within the park; (ii) what are their activities; (iii) when do people visit the park and are there are temporal patterns in the activities; (iv) who are the visitors; (v) why do people visit the national park; and (vi) what complementary information can social media provide in addition to the results from the traditional survey. To our knowledge, no previous study has compared social media data to extensive visitor survey data from a similar environment.

2. Materials and Methods

2.1. Study Area and Visitor Survey

There are 39 National Parks in Finland managed by Parks & Wildlife Finland, a unit of Metsähallitus, which is a state-owned enterprise that provides services related to Finland's natural resources (www.metsa.fi/web/en). PY is the most visited National Park in Finland with 538,853 visitors in 2016 (3% increase from 2015) (www.metsa.fi/web/en/visitationnumbers). PY covers an area of 1020 km^2 in the Lapland region close to the Swedish border. The highest peak Taivaskero reaches an elevation of 809 m above sea level and the landscape of the park is a unique combination of Lappish mountains (fells) surrounded by a mosaic of natural peat-bogs and forests. Vegetation in the park ranges from tundra in the fells to herb-rich forests in sheltered gullies. The area has a long land use history by the indigenous Sámi people and traditional practices such as reindeer herding still take place in the park. The national park was established in two phases; the northern part has been a national park since 1938 and the southern part, a former nature reserve, has been included since 2005. In the park, there are year-round services for various recreational activities, such as hiking and cross-country skiing (www.nationalparks.fi/en/pallas-yllastunturinp). Visitor surveys have been conducted in the park since 1998, following the 5-year scheme of Parks & Wildlife Finland. The first two surveys (1998 and 2003) covered only the northern part of the park. After the southern part was added to the national park, the visitor surveys (2010 and 2016) have also been carried out in the entire region.

An on-site visitor survey was conducted in PY during January–October 2016 by Metsähallitus—Parks & Wildlife Finland. The survey form included questions for collecting basic information about the national park visitors—length of stay, activities in the park, visited locations, expenditures, opinions about services, and socio-economic background information including age, gender, and home location [31] (see supplementary material, Figure S1).

In addition to the standard survey questions, respondents were asked to fill in a questionnaire about social media usage (Figure S2). There, respondents were asked whether or not they are a member of any social media platform and if they have shared/intend to share their national park experience in social media. The questionnaire also included more detailed questions about their use of different platforms and motivations for sharing content.

Face-to face interviews were conducted by the park personnel in 23 locations across the park, during 142 days, distributed throughout the popular seasons from winter/spring to autumn. The survey's sampling effort was spatially balanced across the park according to visitation rates of the park's different sections, based on information obtained from continuous visitor counting. Forms were also available in selected wilderness huts so that visitors could answer the survey independently. The survey was available both in Finnish and in English and the target group of the survey was all visitors over the age of 15.

We used Pearson's Chi-square test to compare the likelihood of sharing national park experiences in social media between different types of users. One-way ANOVA was used for testing the statistical difference in age between different social media user groups. Statistical tests were implemented using the R software (Version 3.2.3) [36].

2.2. Social Media Data Collection

Metadata for geotagged social media posts were collected from the Instagram API (www.instagram.com/developer) using the media search endpoint in spring 2016. Data collection was conducted using a custom-made tool written for the Python programming language. All publicly available posts geotagged within a 10-km buffer zone of the Pallas-Yllästunturi National Park from the period of January 2014–May 2016 were requested from the API using the center points of 2 × 2 km grid cells (collection centroids) as input coordinates in the query. All posts geotagged inside or within 100 m radius from the National Park border were taken into account for within-park analysis and were subject to manual classification. In addition, there were 246 posts geotagged to location 'Pallas-Yllästunturin kansallispuisto/Pallas-Yllästunturi National Park' which was attached to coordinates

4 km outside the park borders. These posts were included in park-level statistics, but filtered out when detecting most tagged sub-regions within the park. The main steps of data collection and processing are illustrated in Figure 2.
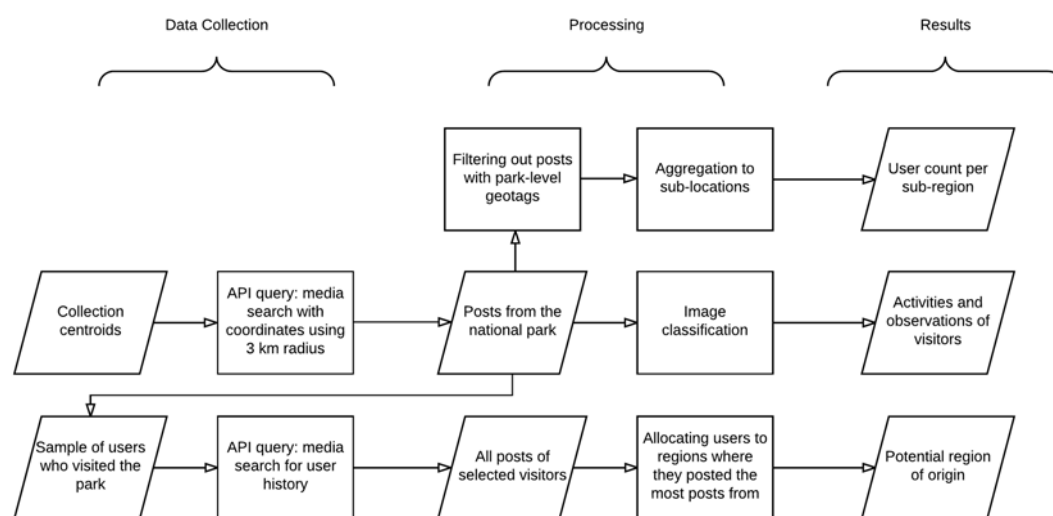

Figure 2. Social media data collection and pre-processing. API, Application Programming Interface.

The location information of the Instagram posts at the time of data collection was attached to pre-defined points-of-interest. In practice, Instagram-users have chosen a pre-defined location from a list when geotagging their photo and thus, the exact coordinates in the dataset are aggregated to these points-of-interest (not the exact coordinates of the user's mobile device).

Instagram was chosen as the source of social media data because of its popularity in the study area and data availability at the time of designing the study. However, due to recent changes in the API policy in June 2016, gathering information from openly shared Instagram posts has become more difficult. According to the visitor survey, 85% of Instagram-users had shared/intended to share their national park experiences online, which further supports the use of Instagram as a data source for visitor monitoring. The number of active Instagram users in Finland has been estimated as 740,000 (13% of population) in 2015, with young adults aged 18–34 (36% of estimated users) as the most active ones (napoleoncat.com/blog/en/instagram-user-demographics-in-selected-european-countries).

2.3. Mapping Most Popular Places within the Park from Social Media Data

Geotagged social media data were aggregated to surveyed sub-regions based on their coordinates. Posts that were tagged with location names referring to the park as a whole were filtered out: Inside the park, 310 photos were geotagged to the park-level with geotags such as 'Pallas-Yllästunturi National Park' referring to the park as a whole, but these posts were technically positioned in one single coordinate location. After filtering out ambiguous geotags, the social media data was aggregated to the same spatial units as in the survey (question 4 in Figure S1) and the resulting rankings were compared using the Spearman rank correlation test.

2.4. Activities and Social Media Content Analysis

The content of pictures posted on Instagram was manually classified according to main subject of the picture (see illustration of classification scheme in supplementary materials, Figure S3). Firstly, we checked if the photo content was relevant for the study area. For example, advertisement and other images clearly not posted by PY visitors were discarded. Secondly, photos were classified according to six main categories defined by the presence or absence of people, activities, landscape, animals, and infrastructure. Thirdly, a more detailed classification was made under each of the main categories. Pictures showing people were further classified according to the number of people present in the photo for detecting the group size. In order to take into account the whole group, the person

taking the photo was added to the count of people in the picture if they were not visible in the photo. Photos marked as 'activities' included pictures showing either people engaged in an activity, the equipment directly used in such activity (e.g., skis, photo cameras), or the outcome (e.g., berries, ski tracks) of performing it. Pictures showing activities were further classified according to types of activities indicated in the visitor's survey (question 9a, Figure S1). This was in order to be able to compare information extracted from social media with results from the survey. Subjective categories included in the visitor survey, such as 'observing nature', were excluded from the manual classification of photos. All other activities observed in the pictures but not included in the visitor survey were aggregated as 'other activities'. These included, for example, reindeer ride, husky ride, sledging, snow scootering, kayaking, kiteboarding, and swimming.

Landscape photos were further classified by indicating the presence or absence of landscape features, such as snow, water, trees, aurora borealis, or other special weather or light conditions. Pictures showing animals were more specifically classified into wild animals and domestic animals. Finally, infrastructure was also further classified according to the type of building (e.g., wilderness huts). It was also noted if the photo was taken indoors.

The classification was performed and double checked for consistency by two people using a form in Microsoft Access 2013. Photo content was accessed online through url-links. Photos which were not publicly available or removed by the user were marked as 'not available'. The classification was finalized after consultation with members of Parks & Wildlife Finland during a joint workshop in October 2016.

We used a Pearson's Chi-square test for assessing the likelihood that the frequency distribution of activities detected from social media (observed activities) were consistent with activities reported in the survey (surveyed activities).

2.5. Detecting Home Location from Social Media Data

Potential home location was estimated for a sample of 291 Instagram users who had visited PY National Park by detecting the country or region from which the user had posted the most photos from. For the 291 users, we collected metadata from all publicly posted content (Figure S4) available from the Instagram API in May 2016. Posts that were geotagged in PY National Park and its surroundings were excluded from the analysis. After excluding posts from the proximity of PY, each user was allocated to the region from which they had posted the most pictures from.

We used a Pearson's Chi-square test in order to assess the likelihood that the frequency distribution of potential home locations of social media users (observed locations) were consistent with visitors' home locations as obtained by the survey (surveyed locations).

3. Results

3.1. Survey

The total number of respondents for the visitor survey was 1927. By season, 57% of visitors filled out the survey during winter (January–May 2016) and 43% during the summer (June–October 2016). Out of all of the respondents, 56% were women and 44% were men. In addition, 63% of all respondents answered the questions about social media usage. Among these, 44% (28% of all respondents) reported that they had shared/intended to share their national park experiences in social media, while 61% (38% of all respondents) reported that they use social media (Table 1).

Table 1. Social media usage among survey respondents (*n* = 1927).

Category	*n*
Social media user	740
Not a member of any social media platform	480
No information	707
Total	**1927**

Facebook was the most popular platform (36% of all survey respondents) followed by Instagram (13%), Twitter (7%), Flickr (1%), and other platforms (7%). Furthermore, 37% of social media users (14% of all survey respondents) reported the use of more than one social media platform in the survey. The majority of Facebook users (62%) reported not using another platform, while most of the Instagram users (96%) were also members of Facebook. Those Facebook users who also used Instagram were more likely to share their national park experiences online compared to those Facebook users who did not use Instagram ($\chi^2 = 28$, *df* = 687, *p*-value < 0.05).

The average age of all the respondents was 54 (min = 16, median = 57, max = 93, sd = 14.94). Among the social media users, the average age was 47 (min = 16, median = 48, max = 93, sd = 14.90), and for visitors who were not members of any social media platform the average age was 60 (min = 25, median = 62, max = 93, sd = 14.67). On average, Instagram-users were younger than visitors who did not use social media at all ($F_{1690} = 422.8$, *p*-value < 0.001) (Figure 3).
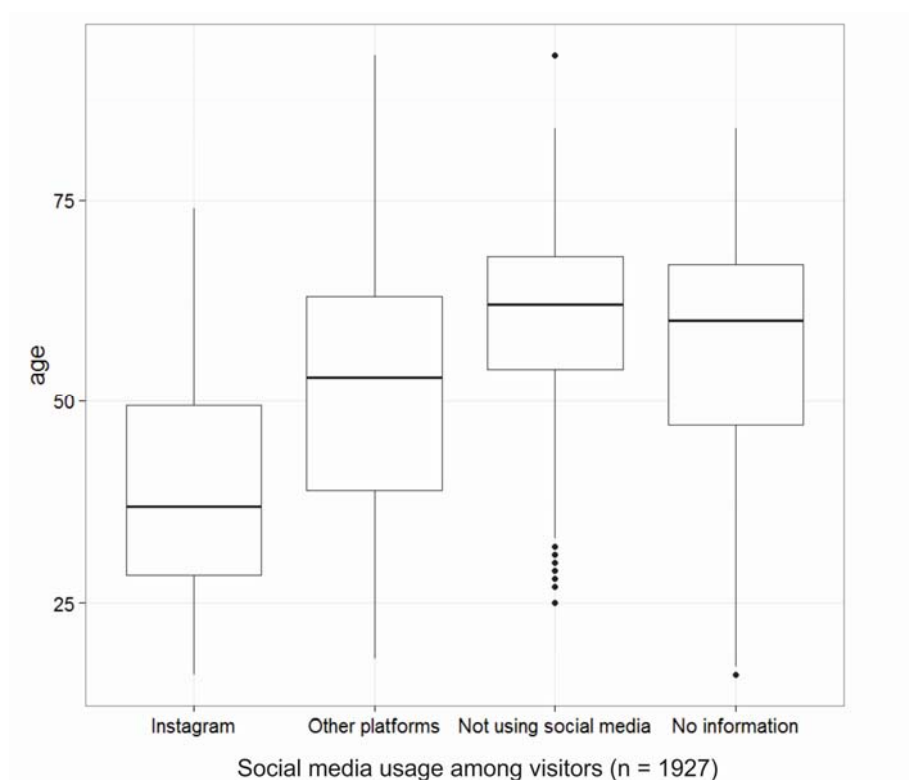

Figure 3. Boxplots showing the age distribution for different groups among the survey respondents (*n* = 1927). Each box indicates the median, the first quartile, the third quartile, and the full range (vertical line and dots) of respondents' age in each group. Groups in the figure are non-overlapping. For example, if a responded used both Instagram and Facebook, they are counted only as Instagram-users in the boxplot.

3.2. Social Media Data

Metadata for 19,939 geotagged photos posted by 7700 users were obtained within a 10 km buffer-zone around the study area. Inside the National Park border, or within 100 m radius from it, there were 4244 photos posted by 2016 users.

Overall, 98% of the photos were available online at the time of image classification. Out of the available photos, 2% were classified as not relevant, 53% contained information about activities, and 44% about people (35% of all available pictures contained both people and activities). Pictures classified as landscape photos were 67% of all available posts. Animals were detected in 6% of the available photos, while 12% contained infrastructure (for example, buildings, fences, duckboards).

3.3. Visitation Patterns in the Park

The comparison between reported visitations to nine different sub-regions (indicated with letters A-I in Figure 1a) in the park and the amount of social media posts from the same areas showed that it is possible to identify the most popular areas in the park from social media data (Table 2). In less-popular sub-regions, the number of social media users was relatively small compared to the two most popular sub-regions. All in all, we found high correlation (r_s = 0.67, *p*-value < 0.05) between the ranking of observed and surveyed sub-regions.

Table 2. Sub-region popularity based on visitor survey and geotagged Instagram posts. Location of sub-regions A-I are presented in Figure 1a. One respondent/user could be allocated to more than one sub-region. For the social media data, not all posts could be allocated to any of the sub-regions.

Rank	Share of Respondents	Sub-Region Rank, Survey		Sub-Region Rank, Social Media	Share of Social Media Users
1	58.5%	A		A	47.3%
2	33.1%	B		B	34.1%
3	22.0%	C		F	3.1%
4	14.3%	D		C	2.1%
5	13.7%	E		I	1.8%
6	13.2%	F		E	1.8%
7	13.2%	G		D	1.6%
8	13.1%	H		H	1.1%
9	9.1%	I		G	0.1%

3.4. Activities of Visitors

Hiking and cross-country skiing were the most popular activities in PY both based on the survey and social media content (Figure 4). Downhill skiing and snowboarding were relatively less popular among the survey respondents compared to social media data. Other surveyed activities such as Nordic walking, bird watching, and picking berries were better captured in the survey than from social media data. Overall, we found no significant difference in the distribution of frequencies between the surveyed activities and activities observed from social media (χ^2 = 304, *df* = 288, *p*-value = 0.2475). At the same time, social media data revealed other activities that were not captured by the survey. These included kiteboarding, sledding, husky ride, reindeer ride, snow scootering, kayaking, and swimming.

3.5. Temporal Patterns of Activities

Social media data reflects well the overall monthly variation in the amount of visitors in the park (Figure 5a), but also reflects the temporal patterns of activities. We inspected the activities by season both from the visitor survey and social media data (Figure 5b,c). In the winter season (January–May), almost all survey respondents had selected cross-country skiing as one of their activities. In the summer and autumn season (June–October) the main activity was hiking (including walking) in the visitor survey (Figure 5b). Social media content revealed similar temporal patterns for the most popular activities; snow sports were most popular in winter, hiking during summer (Figure 5c). In addition to surveyed activities, social media data contained seasonal information of the observed environment, for example, the presence/absence of snow in the landscape (Figure 5d).
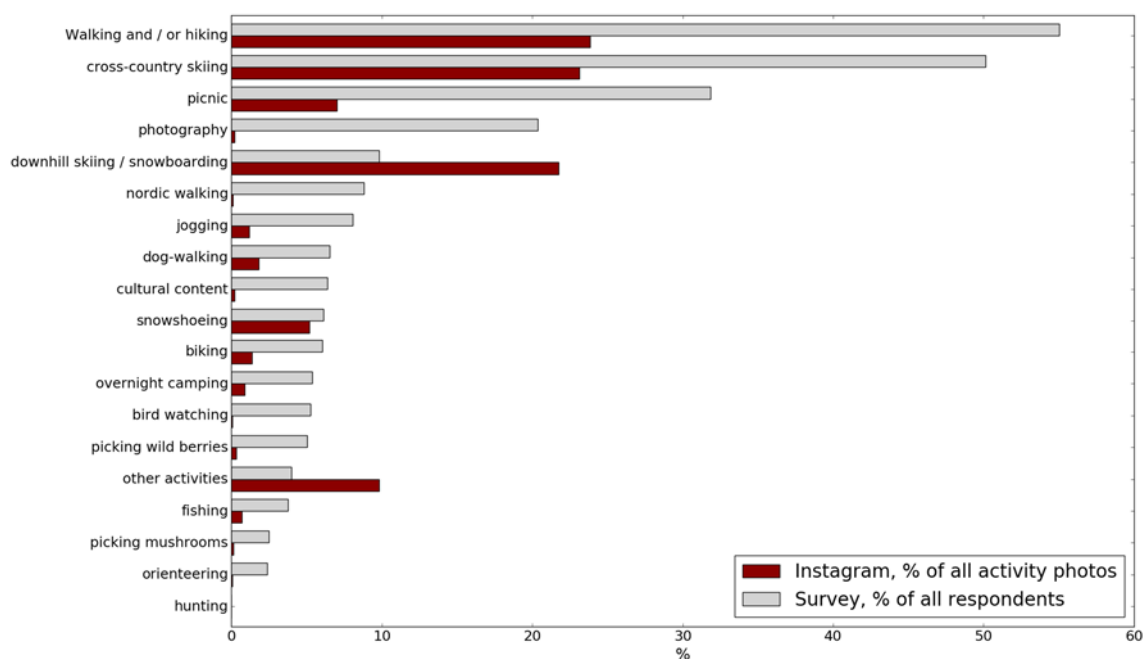
Figure 4. The proportion of respondents/Instagram photos per activity in PY National Park. Activities are sorted by popularity in the survey.
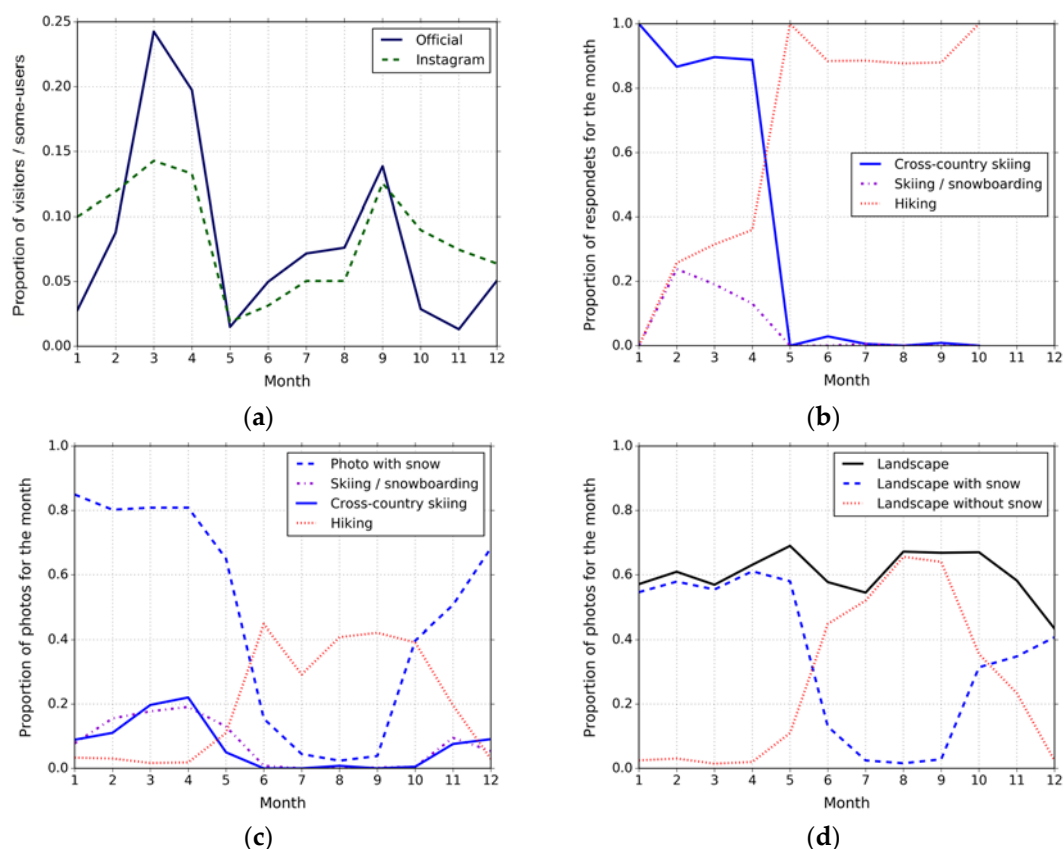

Figure 5. Temporal comparison of official visitor information and social media content: (**a**) Official visitor counts and active daily users in Instagram; (**b**) most popular activities in the visitor survey; (**c**) most popular activities in the social media data; (**d**) landscape photos with and without snow.

3.6. Who Are the Visitors?

Based on the survey, the majority (96%) of visitors had their residence in Finland. International visitors came mainly from Europe (4%) and a few from North-America (<1%). Based on social media data, the most common potential home location was Finland when Instagram users (subset of users for whom user history had been collected from the API) were allocated to the country where they had posted most photos from (Figure 6).

Within Finland, most of the visitors were from the capital region of Helsinki, both based on the visitor survey and social media data (Figure 6a). We found no significant difference between the distribution of frequencies of surveyed home locations and potential home locations observed from social media ($\chi^2 = 190$, $df = 180$, p-value = 0.2903).

The results also helped reveal the typical group sizes of PY visitors (Figure 7). Based on the survey, the median group size was 3, and based on the number of people in Instagram photos median group size was 2. There is potential for using Instagram photo for group-size estimation for small-to medium sized groups (group size 2–10), but not necessarily for detecting people traveling alone or in very large groups.
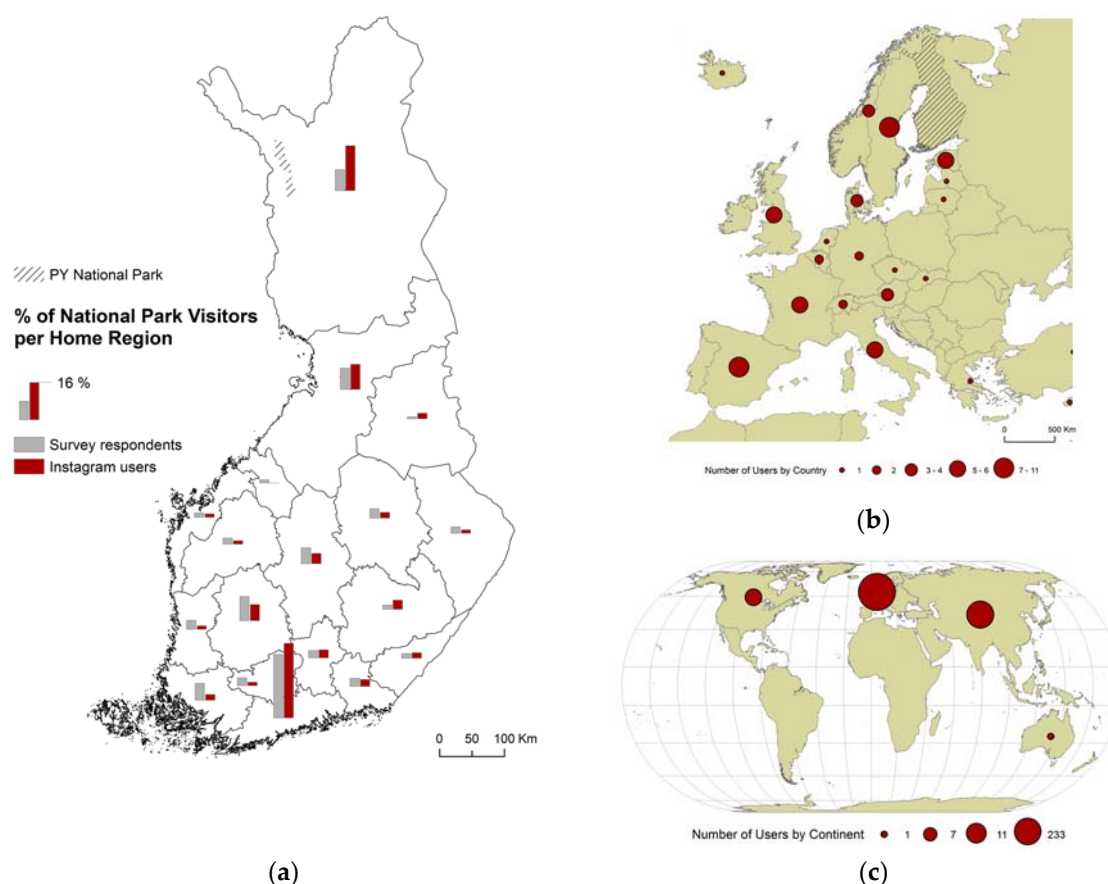

(**a**)

(**b**)

(**c**)

Figure 6. (**a**) Proportion of surveyed home locations (grey) and potential home locations observed from social media (dark red) by region in Finland; (**b**) Number of Instagram users per country and (**c**) per continent from which they had posted most pictures from.
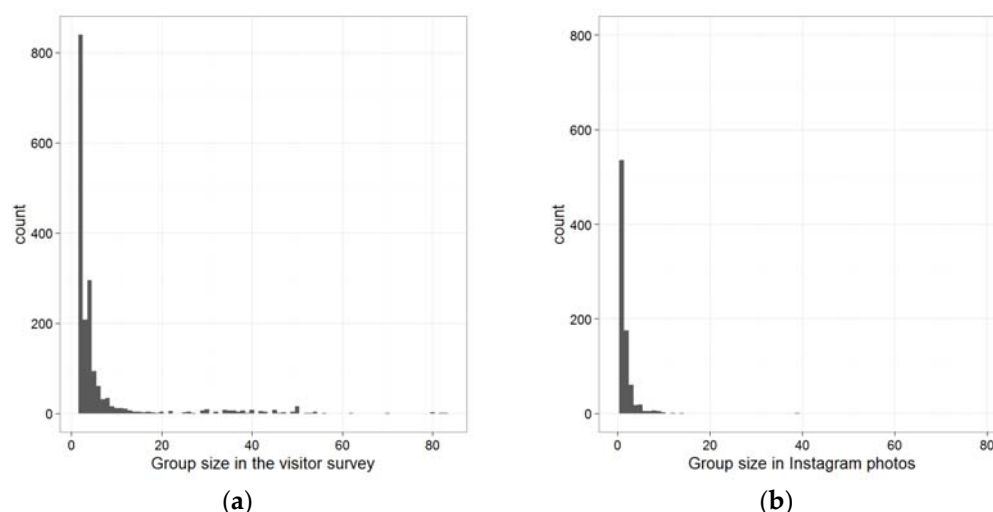
Figure 7. The group size based on (**a**) the survey and (**b**) interpreted from the Instagram images.

4. Discussion

In this study, we compared social media data and systematically collected visitor survey data from the most popular national park in Finland (Pallas-Yllästunturi National Park). This multifaceted comparison suggested that data derived from social media could be used both as an additional and complementary source of information to traditional survey data. In comparison to snap-shot like surveys, social media can provide a source for continuous monitoring of what is happening in the area. It may reveal changes in trends and bring up emerging activities in the park. Such information is crucial for conservation authorities to inform marketing and management. Our comparison also shows that social media data may be able to provide information that is comparable to that collected by traditional surveys.

Social media data was successfully used to detect the most popular sub-regions in the park. However, in areas with lower number of social media posts the results were not as significant. The number of respondents and social media users in less popular locations was relatively small and might have been affected by the survey locations and coordinate accuracy of the social media data. Visitors might be more prone to tag their photos with park-level references (for example, 'Pallas-Yllästunturi National Park') instead of more precise location names within the park in the absence of local knowledge or available pre-defined tags in the social media platform. Other potential reasons for error and bias include poor mobile phone coverage in remote parts of the park (which would influence the amount of social media posts), distance from infrastructure (people might be more prone to post from their accommodation), and activity profile of the sub-region (there were more post in the proximity of ski slopes which are popular destinations among younger age groups).

Social media data could be used to detect some of the most popular activities and their temporal patterns, as well as present new and emerging activities. The most popular activities were the same both in the survey results and in social media. However, less popular activities could be captured only by using either the traditional survey (e.g., Nordic walking and birdwatching) or social media data (e.g., kitesurfing and snow scootering). Considering the dominant age group for Instagram posts, social media content could possibly be used to get a broader and more dynamic picture of emerging activities practiced by younger people in different parts of the national park. Also, social media was able to reveal emerging activities (such as winter biking) not taken into account in repeated surveys.

General patterns of visitors' home locations were also the same both on social media and in the survey results. Most of the visitors were from Finland and Europe. Of course, our approach assumed that users posted more pictures from their home region, which might not be true for users with lower number of posted photos (one might only post a photo from a special trip). However, in the absence of home location information in the user profiles, this method provides an overall picture of the areas in which each user has been most active. Other platforms might have additional profile information

available for estimating the ethnicity or home location of users and the use of such information has been demonstrated in studies from urban environments using Twitter [7].

Understanding the inherent biases in the geographies and user base of data samples from social media have been recognized as key areas of further research in recent literature [15,17]. There are several examples of comparisons of Twitter data against official census data in inhabited areas [5,7], and similar validation efforts can be made using visitor statistics [9,10]. In this paper, we aimed at comparing the spatio-temporal activity patterns presented in social media to surveyed activities in order to make a further validation effort on social media content from our study area. Our results suggest that social media provides comparable information with visitor surveys from areas with adequate amount of data representing the most popular activities. For less popular places and activities, the data sources can be seen as complementary. In PY, despite rigorous sampling efforts across the park, the visitor survey may be somewhat biased towards more elderly visitors. Results from social media are affected by the fact that the sample is self-selected and probably skewed towards younger age groups. As social media data brings in the view of younger visitors and the survey captures more traditional visitors, the two sources of information can actually be considered as complementary to each other.

In conclusion, we believe that social media data can potentially have important implications in informing visitor monitoring and protected area management. As even the best functioning conservation authorities lack both human and financial resources to carry out continuous and repetitive user surveys required to keep protected area management up-to-date, our study highlights that social media data may provide a rapid and cost-efficient alternative to traditional surveys. Continuous monitoring of social media would, for example, allow conservation authorities to better understand spatio-temporal changes in visitor preferences; help assess visitors' profiles and socio-economic backgrounds; understand visitors' sentiments via content analysis; and identify emerging activities, which cannot be captured by pre-defined surveys. The potential of social media is even broader for the practical park management (e.g., to map traffic hotspots or littering in the park) and could be actively used as a source of VGI. In practice, the use of social media data would be facilitated by (1) the development of easy to use tools for the purpose; (2) capacity building of park personnel (as accessing and using social media data requires different expertise than those needed to collect and analyze survey data); and (3) increasing the number of posts and users of social media in the parks through, for example, promoting specific hashtags related to place names, activities, or nature sightings. All in all, more research and practical development is needed before social media data can operationally be used in monitoring visitors of recreational areas. Meanwhile, social media data provides an additional dynamic view of the users and use of parks.

Supplementary Materials: The following are available online at http://www.mdpi.com/2220-9964/6/3/85. Table S1: Place names of the surveyed sub-regions, Figure S1: The survey form for the 2016 visitor survey in Pallas-Yllästunturi National Park conducted by Metsähallitus—Parks & Wildlife Finland; Figure S2: Questions about social media usage in the visitor survey 2016; Figure S3: The classification scheme used to classify the pictures from the national park. One picture could be assigned to more than one category at each level; Figure S4: Geographical distribution of all pictures posted by the Instagram users who have posted from PY national park.

Author Contributions: Vuokko Heikinheimo, Tuuli Toivonen, and Enrico Di Minin conceived and designed the study; Henrikki Tenkanen and Anna Hausmann contributed by social media data collection and development of analysis tools; Joel Erkkonen led the collection of the survey data; V.H. analyzed the data; V.H., T.T., and E.D.M. wrote the paper, with contributions from all authors.

References

1. See, L.; Mooney, P.; Foody, G.; Bastin, L.; Comber, A.; Estima, J.; Fritz, S.; Kerle, N.; Jiang, B.; Laakso, M.; Liu, H.-Y.; Milčinski, G.; Nikšič, M.; Painho, M.; Pődör, A.; Olteanu-Raimond, A.-M.; Rutzinger, M. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 55.
2. Zhang, W.; Derudder, B.; Wang, J.; Shen, W.; Witlox, F. Using Location-Based Social Media to Chart the Patterns of People Moving between Cities: The Case of Weibo-Users in the Yangtze River Delta. *J. Urban Technol.* **2016**, *23*, 91–111.
3. Andrienko, G.; Andrienko, N.; Bosch, H.; Ertl, T.; Fuchs, G.; Jankowski, P.; Thom, D. Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics. *Comput. Sci. Eng.* **2013**, *15*, 72–82.
4. Longley, P.A.; Adnan, M. Geo-temporal Twitter demographics. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 369–389.
5. Steiger, E.; Westerholt, R.; Resch, B.; Zipf, A. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Comput. Environ. Urban Syst.* **2015**, *54*, 255–265.
6. Zheng, Y.-T.; Zha, Z.-J.; Chua, T.-S. Mining Travel Patterns from Geotagged Photos. *ACM Trans. Intell. Syst. Technol.* **2012**, *3*, 1–18.
7. Longley, P.A.; Adnan, M.; Lansley, G. The Geotemporal Demographics of Twitter Usage. *Environ. Plan. A* **2015**, *47*, 465–484.
8. Levin, N.; Kark, S.; Crandall, D. Where have all the people gone? Enhancing global conservation using night lights and social media. *Ecol. Appl.* **2015**, *25*, 2153–2167.
9. Wood, S.A.; Guerry, A.D.; Silver, J.M.; Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* **2013**, *3*, 2976.
10. Sonter, L.J.; Watson, K.B.; Wood, S.A.; Ricketts, T.H. Spatial and Temporal Dynamics and Value of Nature-Based Recreation, Estimated via Social Media. *PLoS ONE* **2016**, *11*, doi:10.1371/journal.pone.0162372.
11. Van Zanten, B.T.; Van Berkel, D.B.; Meentemeyer, R.K.; Smith, J.W.; Tieskens, K.F.; Verburg, P.H. Continental-scale quantification of landscape values using social media data. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12974–12979.
12. Richards, D.R.; Friess, D.A. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecol. Indic.* **2015**, *53*, 187–195.
13. Levin, N.; Lechner, A.M.; Brown, G. An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. *Appl. Geogr.* **2017**, *79*, 115–126.
14. Di Minin, E.; Tenkanen, H.; Toivonen, T. Prospects and challenges for social media data in conservation science. *Front. Environ. Sci.* **2015**, *3*, doi:10.3389/fenvs.2015.00063.
15. Ruths, D.; Pfeffer, J. Social media for large studies of behavior. *Science* **2014**, *346*, 1063–1064.
16. Li, L.; Goodchild, M.F.; Xu, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 61–77.
17. Crampton, J.W.; Graham, M.; Poorthuis, A.; Shelton, T.; Stephens, M.; Wilson, M.W.; Zook, M. Beyond the geotag: Situating "big data" and leveraging the potential of the geoweb. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 130–139.
18. Ganti, R.K.; Ye, F.; Lei, H. Mobile Crowdsensing: Current State and Future Challenges. *IEEE Commun. Mag.* **2011**, *49*, 32–39.
19. Toth, C.; Jóźków, G. Remote sensing platforms and sensors: A survey. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 22–36.
20. Frias-Martinez, V.; Frias-Martinez, E. Spectral clustering for sensing urban land use using Twitter activity. *Eng. Appl. Artif. Intell.* **2014**, *35*, 237–245.
21. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
22. Newman, G.; Wiggins, A.; Crall, A.; Graham, E.; Newman, S.; Crowston, K. The future of citizen science: Emerging technologies and shifting paradigms. *Front. Ecol. Environ.* **2012**, *10*, 298–304.
23. Boyd, D.; Crawford, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Inf. Commun. Soc.* **2012**, *15*, 662–679.
24. Brown, G.; Kyttä, M. Key issues and research priorities for public participation GIS (PPGIS): A synthesis

based on empirical research. *Appl. Geogr.* **2014**, *46*, 122–136.

25. Hausmann, A.; Slotow, R.; Burns, J.K.; Di Minin, E. The ecosystem service of sense of place: Benefits for human well-being and biodiversity conservation. *Environ. Conserv.* **2015**, *43*, 117–127.

26. Balmford, A.; Beresford, J.; Green, J.; Naidoo, R.; Walpole, M.; Manica, A. A Global Perspective on Trends in Nature-Based Tourism. *PLoS Biol.* **2009**, *7*, e1000144.

27. Balmford, A.; Green, J.M.H.; Anderson, M.; Beresford, J.; Huang, C.; Naidoo, R.; Walpole, M.; Manica, A.; Ceballos-Lascurain, H.; Eagles, P.; et al. Walk on the Wild Side: Estimating the Global Magnitude of Visits to Protected Areas. *PLOS Biol.* **2015**, *13*, e1002074.

28. Ardoin, N.M.; Wheaton, M.; Bowers, A.W.; Hunt, C.A.; Durham, W.H. Nature-based tourism's impact on environmental knowledge, attitudes, and behavior: A review and analysis of the literature and potential future research. *J. Sustain. Tour.* **2015**, *23*, 838–858.

29. Knight, A.T.; Cowling, R.M. Embracing Opportunism in the Selection of Priority Conservation Areas. *Conserv. Biol.* **2007**, *21*, 1124–1126.

30. Hausmann, A.; Slotow, R.; Fraser, I.; Di Minin, E. Ecotourism marketing alternative to charismatic megafauna can also support biodiversity conservation. *Anim. Conserv.* **2016**, *20*, 91–100.

31. Kajala, L.; Almik, A.; Dahl, R.; Dikšaitė, L.; Erkkonen, J.; Fredman, P.; Jensen, F.S.; Karoles, K.; Sievänen, T.; Skov-Petersen, H.; et al. Visitor Monitoring in Nature Areas—A Manual Based on Experiences from the Nordic and Baltic Countries. Available online: https://www.naturvardsverket.se/Documents/publikationer/620-1258-4.pdf 534 (accessed on 15 March 2017).

32. Joppa, L.N. Technology for nature conservation: An industry perspective. *Ambio* **2015**, *44*, 522–526.

33. Pearson, E.; Tindle, H.; Ferguson, M.; Ryan, J.; Litchfield, C. Can We Tweet, Post, and Share Our Way to a More Sustainable Society? A Review of the Current Contributions and Future Potential of #Socialmediaforsustainability. *Annu. Rev. Environ. Resour.* **2016**, *41*, 363–397.

34. Orsi, F.; Geneletti, D. Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. *J. Nat. Conserv.* **2013**, *21*, 359–368.

35. Hausmann, A.; Toivonen, T.; Slotow, R.; Tenkanen, H.; Moilanen, A.; Heikinheimo, V.; Di Minin, E. Social Media Data can be used to Understand Tourists' Preferences for Nature-based Experiences in Protected Areas. *Conserv. Lett.* **2017**, doi:10.1111/conl.12343.

36. R Core Development Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.