

Article

Identifying Temporal Patterns of Visitors to National Parks through Geotagged Photographs

Carolina Barros *, Borja Moya-Gómez, and Juan Carlos García-Palomares

Transport, Infrastructure, and Territory–tGIS. Geography Department, Universidad Complutense de Madrid, 28040 Madrid, Spain; bmoyagomez@ucm.es (B.M.-G.); jcgarcia@ghis.ucm.es (J.C.G.-P.)

* Correspondence: dibarros@ucm.es

Received: 24 September 2019; Accepted: 4 December 2019; Published: 6 December 2019

Abstract: Visitor data is essential for decision-making, policy formulation, and monitoring of protected areas. In this context, the data on the temporal distribution of visitors is essential to characterize influx and seasonality, and even to measure the carrying capacity of a site. However, obtaining information from visitors often involves high costs and long production times. Moreover, traditional visitor data has a limited level of detail. New sources of data can provide valuable information regarding the timing of visits. In this study, we tested the use of geotagged data to infer the temporal distribution of visitors to 15 Spanish national parks, and we identified temporal patterns of the visits at three levels: monthly, weekly, and daily. By comparing official monthly visitor counts and geotagged photographs from Flickr, we observed that the number of monthly users who upload photos significantly reflects the number of monthly visitors. Furthermore, the weekly and daily distributions of the Flickr data provided additional information that could contribute to identifying the periods of highest visitor pressure, design measures to manage the concentration of visitors, and improve the overall visitor experience. The results obtained indicate the potential of new data sources for visitor monitoring in protected areas and to open opportunities for future research. Moreover, monitoring tourism in protected areas is crucial to ensure the sustainability of their resources and to protect their biodiversity.

Keywords: geotagged data; national parks; visitor monitoring; temporal patterns

1. Introduction

National parks are often the leading destinations for nature-based tourism because they show the most representative ecosystems within a country: National parks offer plenty of opportunities for recreation, education, and connection with nature. Eagles and McCool [1] have addressed the importance of park tourism as support for local economies and for the conservation of the environment they protect. Furthermore, park tourism has been regarded as an influential factor in raising awareness on the preservation of natural areas [2]. This importance is reflected in the growing number of visitors observed in the 15 parks that form the national network of protected parks in Spain. Over the last ten years, visitors to the national park network have increased by 50% [3,4].

On the other hand, more visitors in national parks can produce negative impacts on the natural environment, such as ecosystem degradation [1,5,6], wildlife distress [7], and the interruption of natural cycles [8]. Thus, park managers must have consistent and accurate information on visitors' characteristics to properly manage and protect these areas.

Visitor data is essential for decision-making, policy formulation, and monitoring of protected areas management [9]. Data that characterizes visitors is crucial for a variety of planning tasks, which include: The identification of trends in demand, generation of forecasts, allocation of infrastructure and services within a park, scheduling of maintenance tasks, staff allocations, and the provision of

resources [1]. In particular, temporal data on park visitation is among the most important to characterize visitor influx and identify the influencing factors of park tourism [10]. Knowledge of the temporal distribution of visitors is used to measure the flow, seasonality, visitor intra-movement, and even the carrying capacity.

Traditional methods of measuring visitors include visitor surveys, direct observation, and on-site counters [10]. However, the collection of such data often entails high costs in terms of time and money. Moreover, traditional visitor data has a limited level of detail [11,12]. New methods have emerged to collect spatial and temporal data of visitors such as GPS. In this sense, GPS loggers have been proven useful in gathering high detail data to characterize visitors and their movement [13–15].

We are currently witnessing a revolution in data production, favored by information and communication technologies, which presents endless opportunities for research [16]. The emergence of new data sources, such as user-generated content from social media, provides new options for assessing tourism and recreation in national parks [12]. Particularly, data from photograph sharing websites seems to be suitable for tourism research, as photography and tourism are strongly linked [17]. From photo-sharing services, Flickr allows free access to their data through an application programming interface (API), which makes it a valuable resource for researchers and park planners.

Free access social media data is an interesting alternative to traditional surveys and GPS devices for data collection because of its reduced acquisition costs, easy access, and high spatial and temporal resolution [18]. Indeed, social media data is a cost-free byproduct of digital interactions from their users [19]. Such data provide a digital “footprint” of their users’ activities that come in a variety of formats, such as photos, texts, audio, and video, which can be used to analyze the spatial and temporal behavior of people.

The potential of data from social media has led to a growing interest in using these data sources to study nature-based tourism in protected areas. Research has mainly focused on geotagged data from photo-sharing websites such as Flickr and Instagram (no longer available for download). Such data had been used to quantify recreation at natural and cultural sites at a global scale [12], model visitation rates in national parks [20], measure parks’ popularity [21], map visitor flows [22,23], identify factors contributing to distribution patterns [11,24], and measure visitor use and spatial patterns [25]. Furthermore, data from web-share services, such as Wikiloc and GPSies, had been used to measure the intensity of use for mountain biking in natural environments [26] and to compare data sources for assessing visitation to parks [27]. Nevertheless, further research on the applicability and validity of social media data as a source of information for visitor monitoring is needed [25].

We contribute to the literature by providing an exploratory analysis of the time-based component of geotagged data with three different time measurements; monthly, weekly, and daily, and how well this new source of data matched official visitor data. Several papers in the literature have demonstrated the usefulness of crowdsourced data for behavioral study purposes. However, few previous papers have used social networks to analyze temporal dynamics in visits to protected areas. When they have done so, they were limited to monthly temporal rhythms, without disaggregating the weekly or daily distribution. Furthermore, this analysis was conducted in a country that has one of the most diverse landscapes in Europe, making it a privileged destination for nature-based tourism. Such diversity provides suitable candidates for testing the use of social networks to study the temporal rhythms of visitors with varied characteristics and locations, thus allowing a more in-depth analysis of the robustness of the source and the methods used. Therefore, this is an initial effort to test the feasibility of using alternative data sources for visitor monitoring in a region with highly diverse kinds of natural spaces.

This research explores the potential of using geotagged data as proxy indicators of the temporal distribution of national park visitors. By comparing official statistics of monthly visitor counts provided by Organismo Autónomo de Parques Nacionales (OAPN) with monthly Flickr users, we assessed the accuracy of geotagged data to infer the monthly distribution of visitors in the 15 parks that form the Spanish network of national parks. Then, we identified and described the temporal patterns of the visits at three levels; monthly, weekly, and daily, through data analysis and clustering techniques. To this end, the article is structured into five sections. After this introduction, the situation

of the Spanish national parks and their main characteristics as a case study are introduced. The third section shows the sources and methods used, highlighting the use of data from photographs of the social network Flickr to study temporal rhythms of visits to natural spaces as the work's main contribution. The fourth section introduces the results we obtained, and the fifth discusses them and presents the research's final conclusions.

2. Study Area

Spain is a pioneering country in Europe in its commitment to the protection of nature [28]. The First Law of National Parks was approved in 1916 and led to the establishment of the first national parks: Picos de Europa and Ordesa. At present, there are 15 national parks in Spain, which are representative of unique landscapes within Europe [4]. Of these parks, two are combined land and sea (Archipiélago de Cabrera and Islas Atlánticas de Galicia), five are mountain areas (Picos de Europa, Ordesa, Aigüestortes, Sierra Nevada, and Sierra de Guadarrama), two are wetlands (Doñana and Tablas de Daimiel), two are Mediterranean forests (Cabañeros and Monfragüe), and one is a subtropical forest of lauraceous species (Garajonay) [29] (Figure 1). This classification of parks based on characteristics and large areas will also be used in analyzing visitors' temporal distribution.



Figure 1. Location and typology of Spanish National Parks.

The 15 national parks cover a total area of 3486 km², which accounts for only 0.76% of the country's entire territory. The newest park, Sierra de Guadarrama, was established in 2013, very near the city of Madrid. The size of national parks ranges from 30 to 858 km² (Table 1). The Sierra Nevada is the largest park at 858.83 km², followed by Picos de Europa and Doñana. At the opposite end of the scale is Islas Atlánticas de Galicia; while its total extent is 84.8 km², only 11.95 km² are dry land. It is followed by the Tablas de Daimiel Park, at 30.3 km², and Garajonay at 39.84 km².

Table 1. Characteristics of Spanish National Parks.

National Parks	Size (km ²)	Distance (Km) to Nearest City * ¹	Year of Declaration	Visitors 2017	Visitors/km ²
Picos de Europa	671.3	74.0	1918	2,047,956	3051
Ordesa y Monte Perdido	157.0	66.0	1918	566,950	3612
Caldera de Taburiente	46.9	11.0	1954	525,961	11,215
Teide	189.9	40.8	1954	4,327,527	22,788
Aigüestortes	141.2	106.6	1955	560,086	3967
Doñana	542.5	53.3	1969	288,589	532
Tablas de Daimiel	30.3	24.6	1972	170,098	5614
Timanfaya	51.1	22.9	1974	1,723,276	33,743
Garajonay	39.8	79.9	1981	836,359	20,993
Archipiélago de Cabrera	100.2	53.8	1991	126,143	1259
Cabañeros	408.6	70.4	1995	112,760	276
Sierra Nevada	858.8	21.8	1999	732,657	853
Islas Atlánticas de Galicia	84.8	15.1	2002	440,661	5196
Monfragüe	184.0	195.6	2007	288,589	1569
Sierra de Guadarrama	339.6	23.9	2013	2,691,890	7927

¹ * city > 100,000 hab. Source: [4].

Management of national parks is shared between the OAPN and the local governments. All parks have an infrastructure for recreation and tourism services, such as visitor centers, information points, parking sites, camping facilities, and nature trails. However, the system for visits and control over them varies greatly. At some small parks with especially vulnerable ecosystems, such as Islas Atlánticas de Galicia, Timanfaya, and Doñana, there is regulated control over the number of visits. At these parks, a permit from the administration is required before the visit. However, access to most parks is free.

As such, the number of visitors to Spanish national parks varies significantly. From 2007 to 2017, the annual number of visits to national parks increased from 10 to 15 million. In 2017, The Teide National park, located in the Canary Islands, was the most-visited park, with 4.3 million visitors, accounting for nearly 30% of total visitors to the national network. In contrast, the Cabañeros national park was the least-visited park, with 100,000 visitors for the same year. Visitor pressure is also highly variable, but in general, parks on the Canary Islands have the greatest pressure, with visitor densities per km² in 2017 at above 20,000 per km² at Timanfaya, Teide, and Garajonay, and more than 10,000 visitors per km² in Caldera de Taburiente.

3. Materials and Methods

The purpose of this study is to validate the use of geotagged photos as indicators of the temporal distributions of visitors to the Spanish national park network and to describe the temporal patterns that these distributions produce. This section introduces the data sources and methodology used, beginning with the collection of data, the processing of raw data and analysis, and presentation.

3.1. Data Collection and Validation

The main data source used in this research is the photo-sharing social media platform Flickr. Flickr is a website for storing and sharing photographs with 75 million users [30]. Geotagged photos within each park were downloaded through the application programming interface (API) provided by Flickr. The API provides free access to photographs as long as it is not for commercial use. The process of collecting geotagged data starts with the collection of all unique photo IDs within a spatial and temporal extent. Using the boundaries of each park, we collected publicly shared geotagged photos that fell inside those boundaries. We downloaded photos taken between January 1, 2010 and December 31, 2016. In a second step, the photo IDs are used to collect additional metadata associated

with each photo. Along with the location of the photos (latitude and longitude), other attributes were collected, such as user ID, the date they were taken, the date they were posted, owner location (nationality), and photo URL address. Both steps were implemented using a Python script. The results were stored in a relational database in a MongoDB application. Then, using XY coordinates, we mapped and analyzed the photographs on an open Geographic Information System (QGIS). A statistical analysis was performed using IBM SPSS software (Armonk, NY, USA).

The resulting dataset has 61,742 records of photos uploaded by 16,190 users, with a total median of 3.8 photos per user. The year 2014 reported the highest number of photos, with over 11,000, and the year 2010 had the lowest number of photos, with nearly 4400 photos. Approximately 0.01% of photos (13) were incorrectly dated; in other words, the date they were taken was not set on the camera, as the year reported on the photos varied from 2033 to 2400. These photos were removed from the analysis.

The distribution of photographs throughout the parks was highly unequal. In the largest and most-visited, such as Teide and Picos de Europa, there were over 10,000 photos and 3000 users (Table 2). On the other hand, the number was reduced in spaces such as Cabañeros, Archipiélago de Cabrera, and Tablas de Daimiel. The results for these three parks should thus be taken very cautiously. The distribution of international users was also highly varied, where they were a majority, especially in parks on the Canary Islands, and very much a minority in parks such as Guadarrama and the Sierra Nevada.

Table 2. Flickr data by National Parks.

National Parks	Photos	Users	Photos/User	Photos/ km ²	Users/km ²	% International Users
Aigüestortes	2118	757	2.8	3.2	1.1	23.6
Archipiélago de Cabrera	647	156	4.1	4.1	1.0	38.2
Cabañeros	583	88	6.6	12.4	1.9	14.3
Caldera de Taburiente	3177	421	7.5	16.7	2.2	68.7
Doñana	2906	547	5.3	20.6	3.9	30.9
Garajonay	1776	473	3.8	3.3	0.9	66.5
Islas Atlánticas de Galicia	2466	1071	2.3	81.4	35.3	20.1
Monfragüe	1991	576	3.5	39.0	11.3	27.2
Ordesa y Monte Perdido	4979	1377	3.6	125.0	34.6	32.3
Picos de Europa	12,357	3147	3.9	123.3	31.4	25.4
Sierra de Guadarrama	6310	1067	5.9	15.4	2.6	5.9
Sierra Nevada	2767	733	3.8	3.2	0.9	30.1
Tablas de Daimiel	979	364	2.7	11.5	4.3	6.1
Teide	15,482	4250	3.6	84.2	23.1	61.4
Timanfaya	3204	1163	2.8	9.4	3.4	73.6
Total	61,742	16,190	3.8	16.1	4.2	39.6

From the user ID and the timestamp of each photo, we calculated the number of unique users per day for each park. We considered this number as a proxy for the number of daily visitors to the parks. Unique users per day were grouped by month to have monthly estimates. To validate our results, we tested whether there was a robust linear relationship between the number of Flickr users and official visitation records [4]. Official records are comprised of the distribution of visitors by month of the year from 2010 to 2016. There are no official statistics available for weekly or daily visitors. Even in some parks (Sierra Nevada, Guadarrama), monthly distributions are estimated indicators.

We compared Flickr's average monthly users with the official average monthly visitors between 2010 and 2016. These values were aggregated by month, and the average monthly users for 2010–2016 were calculated. To test the quality of data collected from Flickr we used a Pearson correlation coefficient, as similar studies have applied a similar approach [11].

3.2. Temporal Analysis

Once Flickr data was downloaded and validated against official data, the temporal distribution of visitors to the 15 parks was analyzed by grouping the data by months to discover monthly distributions throughout the entire period analyzed, by days to discover the distribution over the weeks, and by hours to analyze daily distribution.

For monthly distributions, seasonality was calculated. Seasonality is defined as the tendency of tourist flows to be concentrated in relatively short periods of the year [31]. To examine seasonality during the year in park visitation, we estimated the seasonality ratio (Equation 1) using Lundtorp's proposal [32]. The seasonality ratio is found by dividing the highest value of visitors in a month by average monthly visitors:

$$SRm = \frac{VmMax}{\bar{V}m} \quad (1)$$

where SRm is the Seasonality Ratio, $VmMax$ is the highest number of visitors in a month, and $\bar{V}m$ is the average number of visitors during the months per year. The ratio increases with the degree of seasonal variation. Notice that the lower bound of the Seasonality Ratio equals one, which means the same number of visitors came every month. Its upper bound equals the number of periods compared, which is 12.

Additionally, the distribution of visitors by day of the week and at one-hour intervals was obtained to identify differences between visitor influx at each park in the network with different time horizons. First, we counted users by day of the week for each year and then added these values and calculated the mean value to produce a weekly visitor distribution. Using the same equation as for the Seasonality Ratio, we calculated an indicator to discover the concentration, or lack thereof, of visits on a certain day or days. When this concentration is high over one or several days, this logically occurs on weekends. Thus, a weekly seasonality ration (SRw) was calculated as:

$$SRw = \frac{VwMax}{\bar{V}w} \quad (2)$$

where SRw is the weekly Seasonality Ratio, $VwMax$ is the highest value of visitors in a day, and $\bar{V}d$ is the average number of visitors during days per week. The ratio increases with the degree of weekly variation.

Finally, for daily distribution, we aggregated photos by one-hour intervals and then counted unique users within each interval. Both weekly and daily temporal distributions were useful to visualize and describe seasonal patterns of visitation that were not available from official park data.

To analyze parks based on a seasonality pattern, we used K-means cluster analysis to characterize parks into similar groups. The monthly and weekly seasonality ratios were used as variables for the cluster analysis, along with the proportion of photos taken by international users. One-way ANOVA post hoc test was applied to compare means of the variables for 2, 3, and 4 clusters to select the optimal number of clusters. The comparison showed that 4 was the appropriate number of clusters for the dataset.

4. Results

4.1. Data Validation

Scatter plots show that the majority of parks have a strong linear relation between Flickr users and visitors (Figure 2). Pearson's correlation confirms that all parks have a positive correlation. Four parks have a very high correlation ($r > 0.91$, $p > 0.001$). Five parks have a high correlation ($r > 0.71$), and six parks have a moderate correlation, between 0.7 and 0.54. Out of the last parks, three show moderate correlation, but it was not significant ($p > 0.05$), which could be related to lack of data on monthly visitor numbers for all the years.

Two of these parks are Sierra de Guadarrama and Sierra Nevada. Both parks are very close to two large cities (Madrid and Granada, respectively) (see Table 1), with a high presence of recurring visitors [33]. It is to be expected that recurring visitors take fewer photographs as their visits increase.

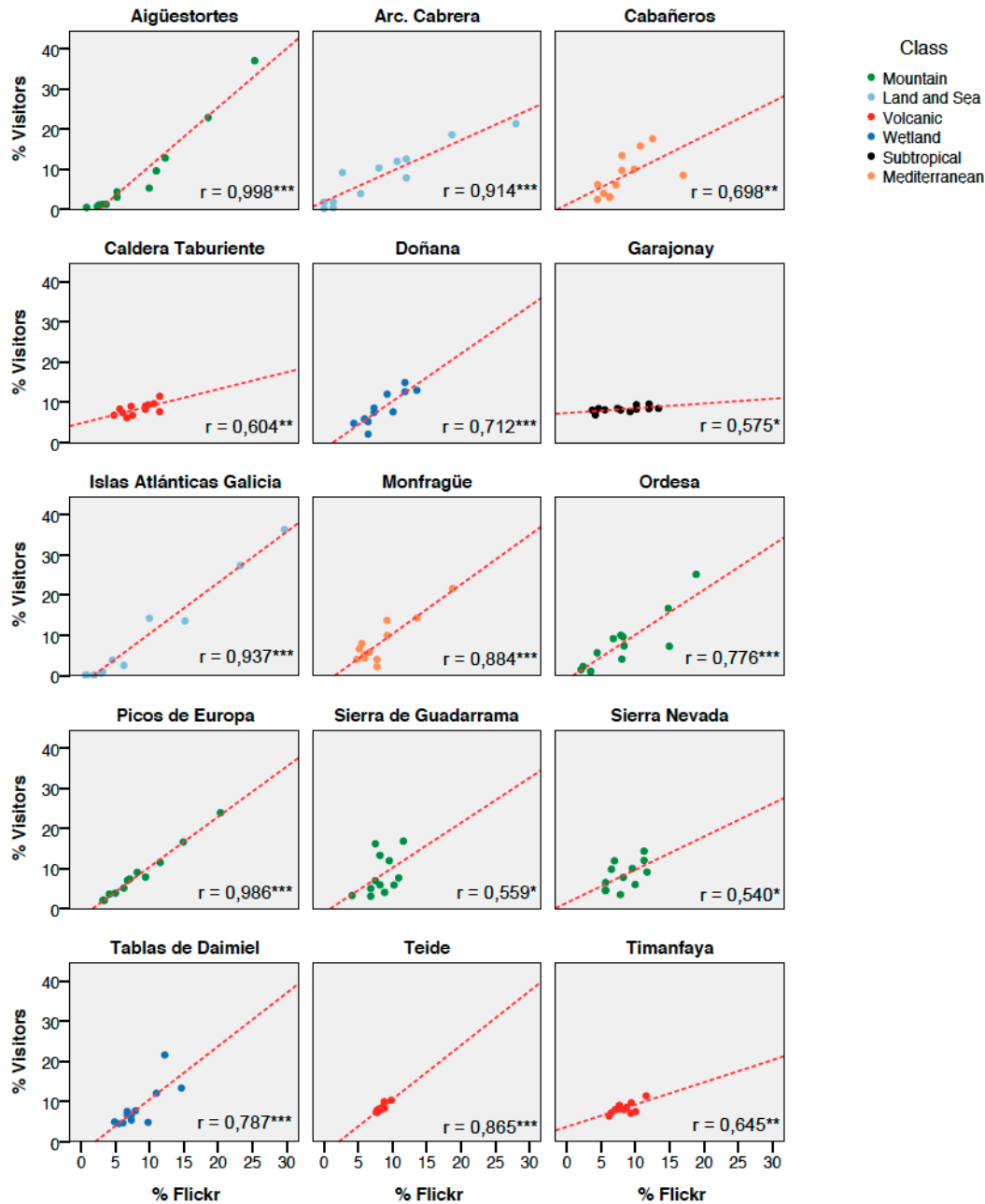


Figure 2. Scatter plots of the monthly distribution of % Visitors and % Flickr users. *** Correlation is significant at the 0.01 level. ** Correlation is significant at the 0.05 level. *Correlation is significant at the 0.10 level.

4.2. Temporal Patterns

Table 3 and Figure 3 show the monthly distribution for the 15 Spanish National Parks derived from Flickr users' estimates. Monthly visitation varies significantly across the parks. However, visitation tends to be highest during the summer months (July and August) for most of them. Visitation peaks often coincide with holidays and vacations for the Spanish population, such as Easter holidays and summer school vacations.

Table 3. Seasonality ratio by month and week.

Park	SRm*	% Max Visitors in One Month	SRw**	% Max Visitors in One Day/Week
Aigüestortes	3	25.0	1.5	21.0
Picos de Europa	2.4	20.0	1.3	19.1
Islas Atlánticas de Galicia	4	33.3	1.3	18.0
Archipiélago de Cabrera	3.4	28.3	1.4	20.6
Monfragüe	2.3	19.2	1.6	23.4
Teide	1.2	10.0	1.1	15.4
Tablas de Daimiel	1.8	15.0	1.7	23.6
Ordesa	2.3	19.2	1.3	19.0
Doñana	1.8	15.0	1.4	19.7
Cabañeros	2.2	18.3	2.1	29.7
Timanfaya	1.4	11.7	1.2	17.6
Taburiente	1.4	11.7	1.1	16.1
Garajonay	1.6	13.3	1.2	17.1
Guadarrama	1.4	11.7	1.9	27.7
Sierra Nevada	1.4	11.7	1.6	22.1

*SRm: Seasonality ratio by month

**SRw: Seasonality ratio by week

Visual inspection of the monthly distributions of the parks (Figure 3) highlights clear differences between parks. For example, most parks (9) show some degree of seasonality, therefore peaks of visitation in one month are easily detected. The summer seasonal peak pattern is much more pronounced in four parks: Islas Atlánticas, Archipiélago de Cabrera, Aigüestortes, and Picos de Europa, whereas other parks (6) show a more regular pattern of visitation throughout the year. The parks located in the Canary Islands are an example. An analysis of the seasonality ratio confirms these differences, with maximum value for these four parks (seasonality ratio by month (SRm) higher than 3 in Islas Atlánticas, Archipiélago de Cabrera, and Aigüestortes). On the opposite end, there is a minimum value of 1.2 for the Teide National Park, and the rest of the Canary Island parks have values lower than 1.5 (Table 3).

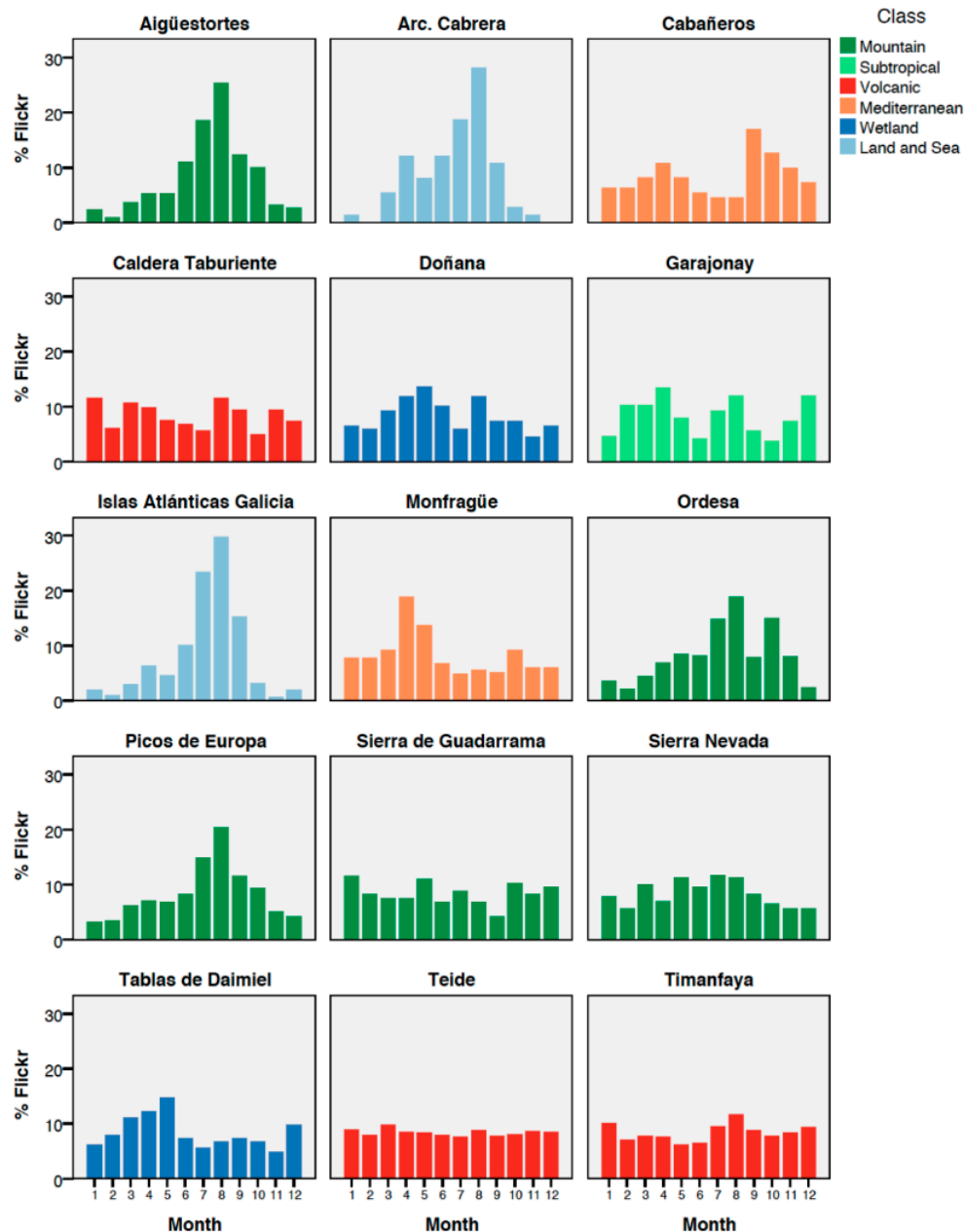


Figure 3. Monthly distribution for Spanish National Parks.

Figure 4 shows the visitor distribution by day of the week for the 15 Spanish National Parks derived from Flickr users from 2010 to 2016. Table 3 shows the results from the seasonality ratio by week (SRw). The results reveal that visitors are concentrated mostly during weekends for the majority of parks. Among the 15 parks, four parks are visited mostly on weekends, gathering 50% or more of total visitors on weekends. Seven parks gathered between 35% and 40% of the visitors on weekends. These parks, whose SRw values are the highest (above 1.5) are parks whose main visitor demand comes from residents of neighboring cities. On the contrary, the parks in the Canary Islands have a fairly constant influx of visitors every day of the week, which could be related to the touristic vocation of the islands and the importance of long-duration foreign tourism (see Table 2).

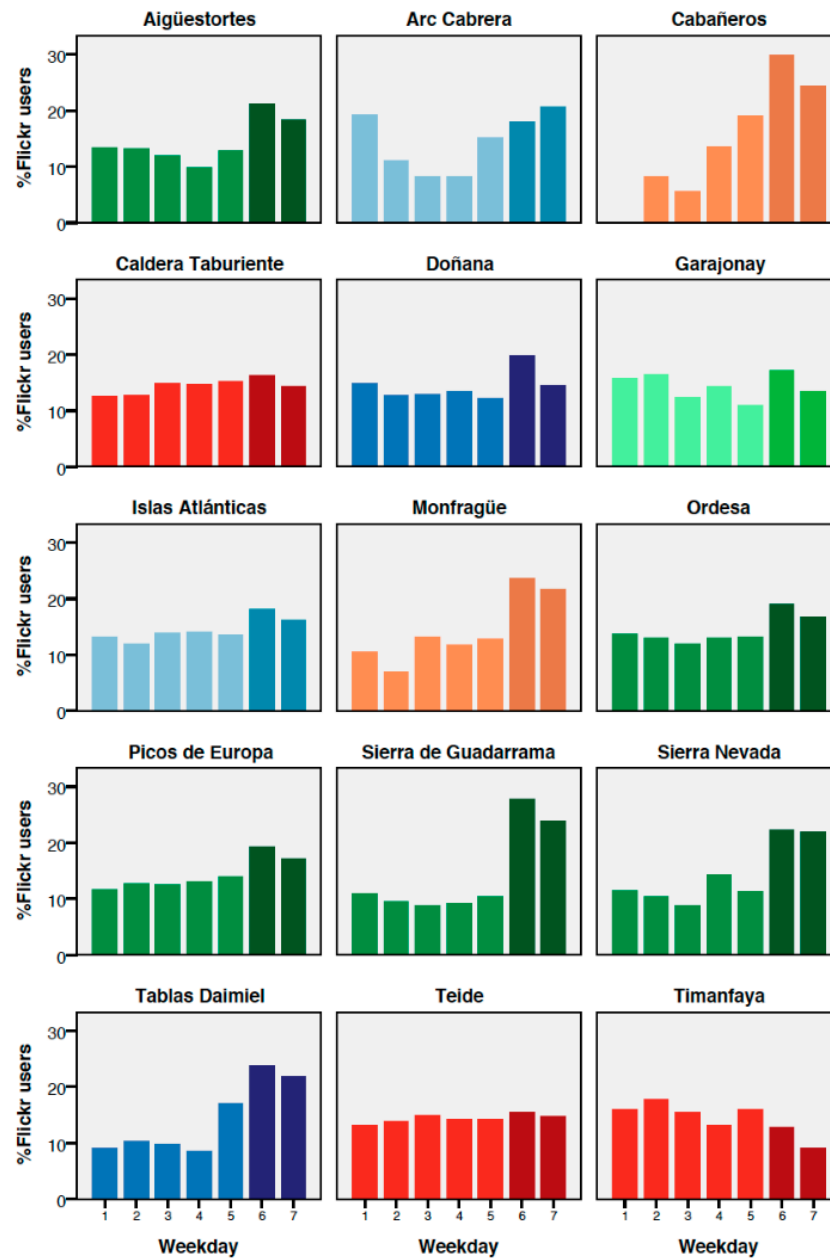


Figure 4. Weekly distribution of visitors to Spanish National Parks.

Figure 5 shows visitor distribution throughout the day (in slots of one-hour intervals, from 7:00 a.m. to 10:00 p.m.) at the 15 National parks. The affluence throughout the day in all the parks bears some common guidelines; for example, the hours with the highest visitor influx in most of the parks are around noon, between 11:00 a.m. to 1:00 p.m., with 12:00 p.m. being the highest attendance time for 10 of the 15 parks.

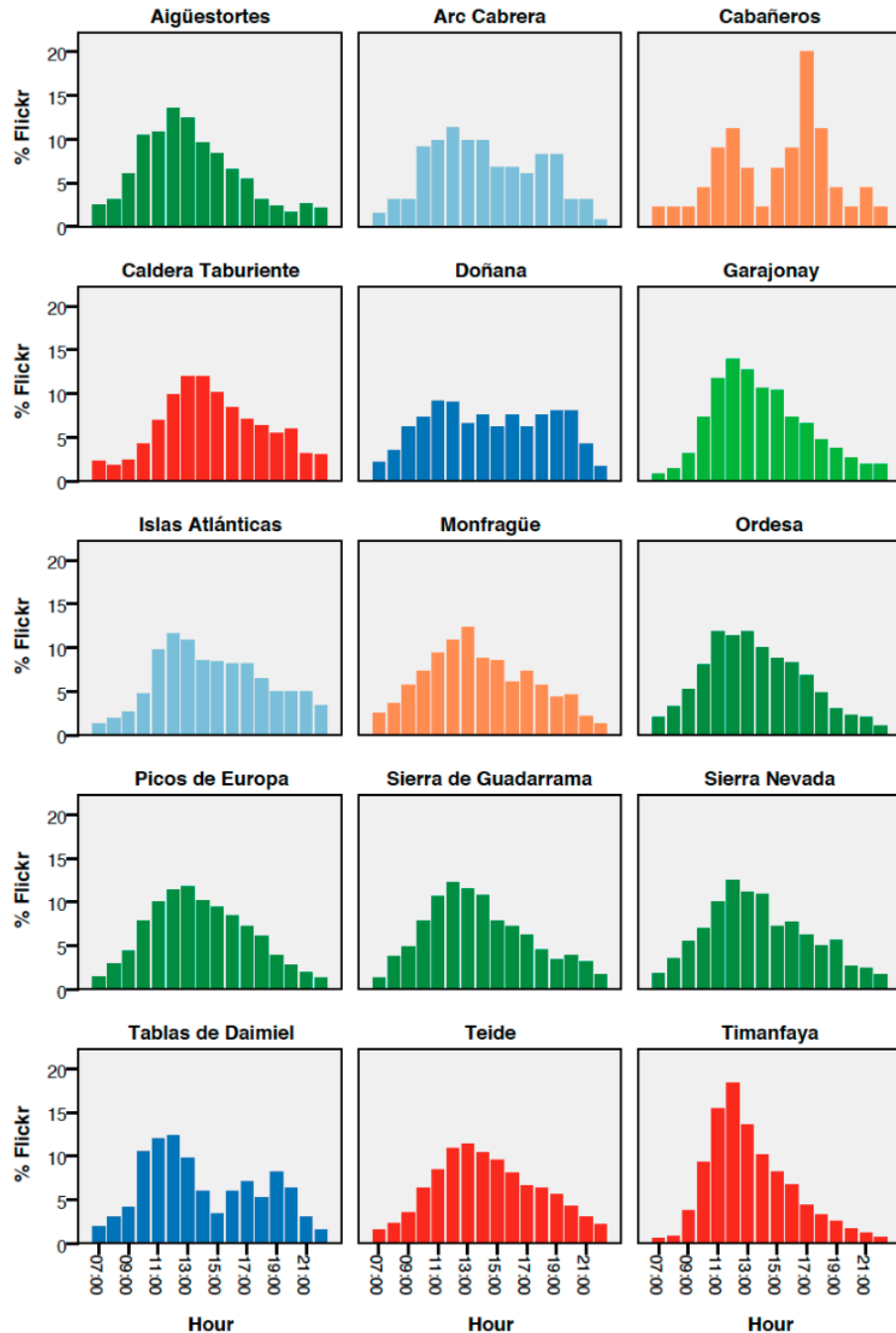


Figure 5. Daily distribution of visitors in Spanish National Parks.

The first photographs in all Parks are normally taken between 10:00 a.m. and 11:00 a.m., when most visits begin. Seven parks concentrated more than 45% of the visits in a three-hour time slot alone. The Timanfaya national park shows the highest proportion of visitors (18%), while the Doñana national park shows the lowest percentage, with 9% at the peak hour. Thus, there is a higher number of photographs from the morning than the afternoon, although these data could also be due to the “return effect”, when users take fewer photographs when retracing their route through the park to return. During night hours, visitor presence decreases to less than 3% in all the parks.

4.3. Typology of Parks Based on the Temporality of Their Visits

A set of four cluster types were identified from the cluster analysis (Table 4 and Figure 6), based on values from seasonality ratio by month (SRm) and by week (SRw) indicators, including one variable to differentiate between park types based on the type of visitors they receive (the percentage of photographs taken by international users). One-way ANOVA post-hoc comparison between clusters shows that the differences of clusters are statistically significant for all the variables (Table 5).

Table 4. Cluster description.

Cluster	N	Variables	Mean	Std. Dev.	Std. Error	Final Center
1	4	SRm	1.40	0.163	0.082	1.40
		SRw	1.16	0.068	0.034	1.16
		% International	0.68	0.054	0.027	0.68
2	5	SRm	2.44	0.321	0.144	2.44
		SRw	1.57	0.310	0.139	1.57
		% International	0.24	0.066	0.029	0.24
3	2	SRm	3.70	0.424	0.300	3.70
		SRw	1.35	0.127	0.090	1.35
		% International	0.29	0.127	0.090	0.29
4	4	SRm	1.60	0.231	0.115	1.60
		SRw	1.63	0.235	0.117	1.63
		% International	0.18	0.142	0.071	0.18



Figure 6. Cluster membership.

Table 5. One-way ANOVA post-hoc comparison between clusters.

	Variable	Sum of Squares	DF	Mean Square	F	Sig.
SRm	Between Groups	8.657	3	2.886	38.153	0.000
	Within Groups	0.832	11	0.076		
	Total	9.489	14			
SRw	Between Groups	0.560	3	0.187	3.537	0.052
	Within Groups	0.581	11	0.053		
	Total	1.141	14			
International	Between Groups	0.598	3	0.199	21.396	0.000
	Within Groups	0.102	11	0.009		
	Total	0.701	14			

Cluster 1 has the lowest values for the seasonality ratio. This cluster consists of parks with a more homogeneous distribution of visitors throughout the year and weeks. Though there is a peak month for visitors, the variation of visitors moves between 5% and 10%, approximately. These parks are characterized by a regular visitor flow during the year, largely international visitors. Such is the case for the parks located in the Canary Islands, which receive a significant influx of international tourism throughout the year.

Cluster 2 shows a high seasonality ratio by month and by week. In this cluster, an average of 20% total visitors occurs in only one month, mostly in August. During the winter season, visitors in Cluster 2 fall to values below 10%, due to the harsh climate. The parks that form the cluster are the mountain and the Mediterranean parks on the Peninsula.

Cluster 3 has slightly lower visitor concentration values throughout the week, but the highest seasonality all year round and international visitor values slightly increase. Two parks form this cluster, and seasonality is concentrated during the summer season. For example, the Islas Atlánticas and Archipiélago de Cabrera parks have their peak season in August.

Cluster 4 shows the highest seasonality ratio by week and a low seasonality by month. The wetland and near-city parks form this cluster, and their seasonality is distributed in different months. For example, Doñana and Tablas de Daimiel have their peak season in the spring months, while the Sierra Nevada has its peak season in July. Most of the visitors are national and highly related to weekend visitors from nearby urban areas (for example, Sierra Nevada and Sierra de Guadarrama have regular visitation throughout the year, which could be due to their proximity to cities like Granada and Madrid, respectively).

This typology is closely related to the classification based on landscape characteristics established by Somoza Medina (2009). As shown in Figure 7, the volcanic parks on the Canary Islands have the most homogeneous visitor distributions.

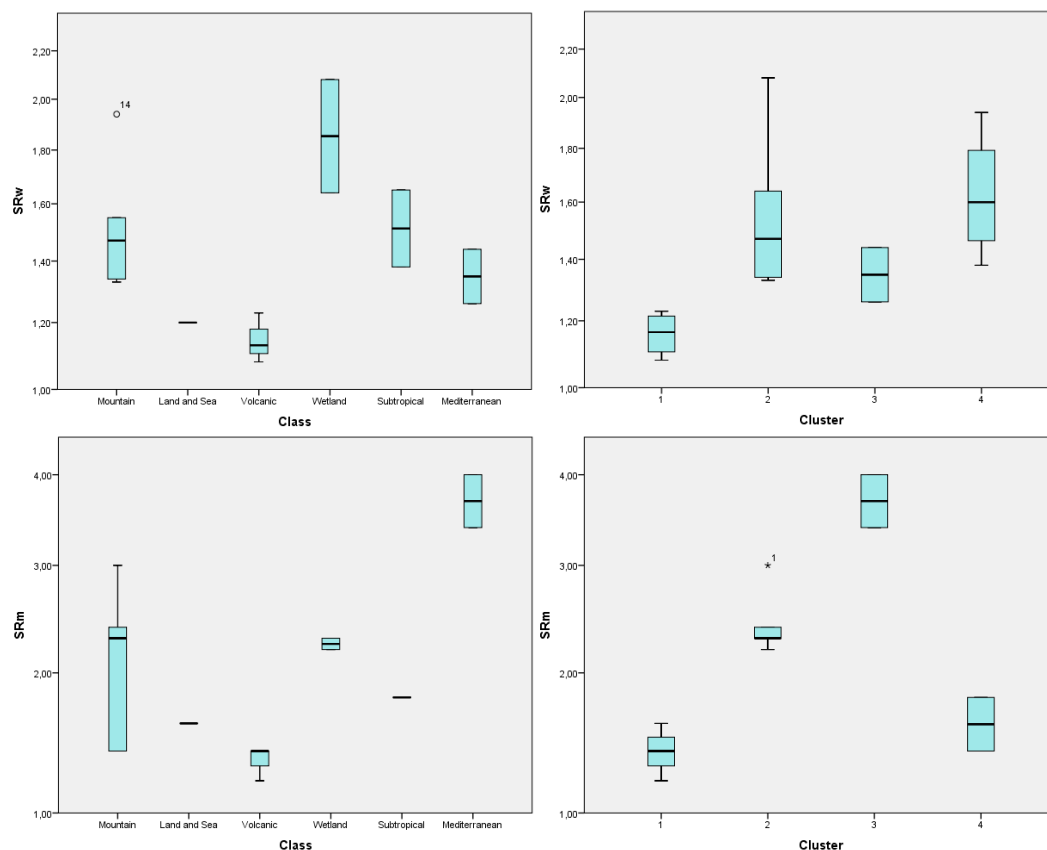


Figure 7. SRm and SRw based on types of national parks.

5. Discussion and Conclusions

This study found strong and significant correlations between geotagged data from Flickr and official visitor data at a monthly level, thus validating the potential to use social media data to infer and describe time patterns for visitors and as a proxy indicator of visitation rates. Moreover, weekly and daily distributions provided interesting insights that could help managers to more precisely identify saturation periods at a higher resolution. Weekly and daily data are helpful for scheduling and transportation tasks and to determine carrying capacity and overcrowding issues. These results are well aligned with previous studies [12,20,21,34] that used similar approaches to test social media validity. All the studies showed a strong relation between geotagged data and visitor statistics.

However, there was variation in the strength of the correlations among parks that could be attributed to: (1) Lack of visitor data for some parks, such as the Cabañeros and the Sierra Nevada parks, which produced moderate, but statistically insignificant, correlation at 0.05 *p*-value; (2) Different methods used for counting visitors, as is the case for the Guadarrama National Park, which has changed its counting method two times since its establishment in 2013 [33]; (3) A lower number of photographs from recurring visitors, which reduces correlations in parks near large metropolitan areas with a high number of domestic and weekend visitors.

Many authors have postulated that natural conditions are determinants for visitation patterns in protected areas [35,36]. In our study area, the geographical location seems to be an important factor for high seasonality. Mountain parks, such as Aigüestortes and Picos de Europa, generated high seasonality ratios. This also occurred with parks near the coast (Archipiélago de Cabrera and Islas Atlánticas de Galicia), which are attractive for beach tourism. In parks with medium seasonality values, the park's natural environment could be the factor for seasonal peaks outside the summer season. For example, in the Doñana national park, the highest visitation rate is in May. This coincides with the breeding season for many bird species, which means visitors have more chances for bird watching.

In contrast, low seasonality in some of Spain's national parks could be produced by external factors. Proximity to large population centers could increase the number of visitors in each month; for example, the Sierra Nevada and Sierra de Guadarrama parks, which are very close to large cities like Granada and Madrid, respectively. Furthermore, parks located in tourist destinations with high demand could be influenced by the regular tourist flows where tourists choose to visit national parks as an additional activity during their vacation, which is the case in the parks located on the Canary Islands.

Cluster analysis was useful for grouping parks with similar seasonal patterns, thus helping to study underlying factors for visitor seasonality in the Spanish National Parks network. Based on seasonality levels, high, medium, and low, K-mean clustering resulted in four clusters. All clusters grouped parks with different ecosystems, sizes, and numbers of visitors.

The high temporal concentration of visitors on a daily basis in some parks could indicate overcrowding. Mapping these distributions, using the geotagged components, could show and verify where crowding occurs. Thus, the carrying capacity of the national parks could be exceeded. In only three hours, the Timanfaya, Garajonay, Aigüestortes, Ordesa, and Sierra de Guadarrama parks concentrate up to 50% of observed activity on Flickr. While it is previously mentioned that this distribution may be conditioned by the "return effect," everything seems to indicate a greater visitor presence during the morning. Further research about where visitors are at peak hours could help park managers to design actions for redistributing visitor flows during the day.

Official visitor data provided by park agencies is essential for park management and nature-based tourism research. However, visitation data lack detail in terms of temporal resolution for most protected areas. Obtaining such data through surveys could be very expensive and time-demanding [20,37]. Hence, social media can be a potential alternative data source for assessing visitor monitoring and obtaining powerful insights into temporal visitation patterns in protected areas. In our study area, none of the parks had weekly or daily distribution of the visitors, so geotagged data from Flickr users helped to illustrate the temporal behavior of visitors during the day and their influx across the week.

Temporal distributions extracted from geotagged photographs can be useful for park managers who need up-to-date information on visitors. For example, visitor distribution at different scales can be used for management tasks such as team scheduling, hiring additional staff, monitoring points of interest, or planning visiting schedules. Moreover, geotagged data can be mapped, which can be useful to alert park managers about where and when crowding issues occur, thus providing useful information for making decisions to deal with them.

This study explored the potential of geotagged data to measure temporal distributions of visitors in a national network of protected areas. The main advantage of social media data collection for visitor monitoring is the capability to collect a vast amount of data with high temporal and spatial resolution at relatively low costs. The attributes associated with geotagged data reveal several characteristics of visitors to national parks. The photos' timestamp was valuable for temporal and seasonality analysis in this and other studies [11,20]. The findings are useful and show that new data sources could be used as complementary data to obtain up-to-date and more detailed data for national park management. For example, in parks where the budget for visitor monitoring is scarce, geotagged data provide significant findings of when visitors come to the park and how long they stay. For parks where visitor monitoring is carried out, geotagged data provide additional data that, when combined with surveys, could give a more accurate view of the temporal distribution of visitors.

Works such as ours that validate data using the digital footprint of visitors to protected areas, whether the figures come from social networks or other sources (mobile phones), enable data collection with a high temporal and spatial resolution that is not always available, particularly in the case of the spatial distribution of visitors within the parks, and this avoids dependence on traditional sources for research.

Nevertheless, geotagged data have some limitations that need to be addressed. First, data from social media are biased by the website's popularity, the users and demographics. Highly productive users also cause bias in geotagged data that could cause overestimation in the analysis. We address this bias by counting users instead of photographs. Additionally, in less popular parks there are less geotagged data, so the significance of the analysis in such parks can be reduced. Furthermore, more studies are needed to explore the full potential of geotagged data, for example, deducing visitor intra-movements, measuring usage intensity, the popularity of parks, and others. Further additional research is necessary to address geotagged data limitations and bias.

Author Contributions: The data analysis and writing of the article was done by C.B. The data collection and text editing was done by B.M.-G. The method application and analysis were done by C.B. and J.G.-P. The results interpretation was made by all authors. All authors have read and approved the final manuscript.

Funding: The authors gratefully acknowledge funding from the Ministerio de Ciencia, Innovación y Universidades and the European Regional Development Fund (ERDF) (Project DynMobility, RTI2018-098402-B-I00).

Acknowledgments:

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Eagles, P.F.J.; McCool, S.F. *Tourism in National Parks and Protected Areas: Planning and Management*; CABI, Ed.; CABI: Wallingford, UK, 2002; ISBN 0851997597, 9780851997599.
2. Ballantyne, R.; Packer, J.; Falk, J. Visitors' learning for environmental sustainability: Testing short- and long-term impacts of wildlife tourism experiences using structural equation modelling. *Tour. Manag.* **2011**, *32*, 1243–1252.
3. Organismo Autónomo de Parques Nacionales. *Memoria de la Red de Parques Nacionales, Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid, SP*; 2015. https://www.miteco.gob.es/es/red-parques-nacionales/divulgacion/memoria-2015_tcm30-378646.pdf

4. MAPAMA. OAPN Boletín de la Red de Parques Nacionales. Available online: <http://www.mapama.gob.es/es/red-parques-nacionales/boletin/visitantes-teide.aspx> (accessed on 14 March 2019).
5. Farrell, T.A.; Marion, J.L. Identifying and assessing ecotourism visitor impacts at eight protected areas in Costa Rica and Belize. *Environ. Conserv.* **2001**, *28*, 215–225.
6. Valentine, P. Review: Nature-based tourism. In *Special Interest Tourism*; Belhaven Press: London, UK, 1992; pp. 105–127.
7. Buultjens, J.; Ratnayake, I.; Gnanapala, A.; Aslam, M. Tourism and its implications for management in Ruhuna National Park (Yala), Sri Lanka. *Tour. Manag.* **2005**, *26*, 733–742.
8. Orams, M.B. Using Interpretation to Manage Nature-based Tourism. *J. Sustain. Tour.* **1996**, *4*, 81–94.
9. Manning, R.E. How much is too much? Carrying capacity of national parks and protected areas. In Proceedings of the Monitoring and management of Visitor Flows in Recreational and Protected Areas, Vienna, Austria, 30 January–2 February 2002; pp. 306–313.
10. Cessford, G.; Muhar, A. Monitoring options for visitor numbers in national parks and natural areas. *J. Nat. Conserv.* **2003**, *11*, 240–250.
11. Tenkanen, H.; Di Minin, E.; Heikinheimo, V.; Hausmann, A.; Herbst, M.; Kajala, L.; Toivonen, T. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Sci. Rep.* **2017**, *7*, 17615.
12. 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.
13. Hallo, J.C.; Beeco, J.A.; Goetcheus, C.; McGee, J.; McGehee, N.G.; Norman, W.C. GPS as a Method for Assessing Spatial and Temporal Use Distributions of Nature-Based Tourists. *J. Travel Res.* **2012**, *51*, 591–606.
14. Orellana, D.; Bregt, A.K.; Ligtenberg, A.; Wachowicz, M. Exploring visitor movement patterns in natural recreational areas. *Tour. Manag.* **2012**, *33*, 672–682.
15. Meijles, E.W.; de Bakker, M.; Groote, P.D.; Barske, R. Analysing hiker movement patterns using GPS data: Implications for park management. *Comput. Environ. Urban Syst.* **2014**, *47*, 44–57.
16. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
17. Prideaux, B.; Coghlan, A. Digital cameras and photo taking behaviour on the Great Barrier Reef—Marketing opportunities for Reef tour operators. *J. Vacat. Mark.* **2010**, *16*, 171–183.
18. Di Minin, E.; Tenkanen, H.; Toivonen, T. Prospects and challenges for social media data in conservation science. *Front. Environ. Sci.* **2015**, *3*, 63.
19. Tsou, M.H. Research challenges and opportunities in mapping social media and Big Data. *Cartogr. Geogr. Inf. Sci.* **2015**, *42*, S70–S74.
20. Sessions, C.; Wood, S.A.; Rabotyagov, S.; Fisher, D.M. Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *J. Environ. Manag.* **2016**, *183*, 703–711.
21. 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.
22. Lee, J.Y.; Tsou, M.H. Mapping spatiotemporal tourist behaviors and hotspots through location-based photo-sharing service (Flickr) data. *Lect. Notes Geoinf. Cartogr.* **2018**, 315–334.
23. Orsi, F.; Geneletti, D. Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. *J. Nat. Conserv.* **2013**, *21*, 359–368.
24. Walden-Schreiner, C.; Leung, Y.F.; Tateosian, L. Digital footprints: Incorporating crowdsourced geographic information for protected area management. *Appl. Geogr.* **2018**, *90*, 44–54.
25. Heikinheimo, V.; Minin, E.D.; Tenkanen, H.; Hausmann, A.; Erkkonen, J.; Toivonen, T. User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 85.
26. Campelo, M.B.; Nogueira Mendes, R.M. Comparing webshare services to assess mountain bike use in protected areas. *J. Outdoor Recreat. Tour.* **2016**, *15*, 82–88.
27. Norman, P.; Pickering, C.M. Using volunteered geographic information to assess park visitation: Comparing three on-line platforms. *Appl. Geogr.* **2017**, *89*, 163–172.
28. Organismo Autónomo de Parques Nacionales. *Plan Director de la Red de Parques Nacionales*; Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid 2011; pp. 690–695.

29. Somoza Medina, J. The national park concept in Spain: Patriotism, education, romanticism and tourism. In *Tourism and National Parks: International Perspectives on Development, Histories, and Change*; Frost, W., Hall, C.M., Eds.; Routledge: Abingdon, UK, 2009; Volume 14, pp. 143–154, ISBN 0415471567.
30. The Internet Archive Yahoo Timeline. Available online: <https://web.archive.org/web/20080713214826/http://yhoo.client.shareholder.com/press/timeline.cfm> (accessed on 3 Mar 2017).
31. Allcock, J.B. Seasonality. In *Tourism Marketing and Management Handbook*; Witt, S.F., Moutinho, L., Eds.; Prentice Hall: London, UK, 1989; pp. 387–392, ISBN 013925885X.
32. Lundtorp, S. Measuring tourism seasonality. In *Seasonality in Tourism*; Lundtorp, S., Baum, T., Eds.; Pergamon Oxford: Oxford, UK, 2001; pp. 23–50, ISBN : 0-08-043674-9.
33. Comisión de Gestión del Parque Nacional Sierra de Guadarrama (PNSG). Coordinadora del *Memoria de Actividades*, Madrid, SP, 2016. Available online: <https://www.parquenacionalsierraguadarrama.es/en/downloads/category/5-docs> (accessed on 1 December 2019).
34. Hamstead, Z.A.; Fisher, D.; Ilieva, R.T.; Wood, S.A.; McPhearson, T.; Kremer, P. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Comput. Environ. Urban Syst.* **2018**, *72*, 38–50.
35. Butler, R.W. Seasonality in tourism: Issues and implications. In *Seasonality in Tourism*; Baum, T., Lundtorp, S., Eds.; Pergamon Oxford: Oxford, UK, 2001; pp. 5–22.
36. Hadwen, W.L.; Arthington, A.H.; Boon, P.I.; Taylor, B.; Fellows, C.S. Do climatic or institutional factors drive seasonal patterns of tourism visitation to protected areas across diverse climate zones in eastern Australia? *Tour. Geogr.* **2011**, *13*, 187–208.
37. Schägner, J.P.; Maes, J.; Brander, L.; Paracchini, M.L.; Hartje, V.; Dubois, G. Monitoring recreation across European nature areas: A geo-database of visitor counts, a review of literature and a call for a visitor counting reporting standard. *J. Outdoor Recreat. Tour.* **2017**, *18*, 44–55.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).