Using social media to quantify spatial and temporal dynamics of nature-based recreational activities

Authors: Francesca Mancini[a], George M. Coghill[b], David Lusseau[a]

[a]Institute of Biological and Environmental Sciences, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen AB24 2TZ, UK.

[b]Department of Computing Science, University of Aberdeen, Meston Building, Meston Walk, Aberdeen AB24 3UE, UK.

Corresponding author: Francesca Mancini

Room 418, Zoology Building, Tillydrone Avenue, Aberdeen AB24 2TZ, UK.

Telephone: 01224 274106. Email: r03fm14@abdn.ac.uk

Big data offer a great opportunity for nature-based recreation (NbR) mapping and evaluation. However, we need to assess when and how it is appropriate to use this resource. We used Scotland as a case study to validate the use of data from Flickr to quantify NbR on a national scale and at different spatial and temporal resolutions.

We used Flickr photos to estimate visitation in the Cairngorms National Park (CNP) and compared this dataset to a time series of visitor numbers obtained from the CNP authority. Then, we compared the spatial distribution of photographs of wildlife taken in Scotland to a dataset obtained from a survey. Finally, we identified hotspots of wildlife watching in Scotland and investigated how they changed between 2005 and 2015.

Both short- and long-term trends of NbR were validated and the spatial information present in geotagged pictures of wildlife is accurate at a resolution as fine as 10x10 Km. Caution is necessary if using this data at a finer resolution. Our findings have implications for planning and management of NbR as pictures uploaded on Flickr can be used as a proxy for NbR at spatial and temporal scales that are relevant for ecosystem management.

Keywords: digital conservation, Moran eigenvector spatial filtering, big data, Flickr, time series analysis, wildlife watching

1. Introduction

Recreation is a key cultural ecosystem service provided by nature. Tourism is often a primary income for local communities (Curtin 2003; Silva 2013), it can dominate national economies and play a key role in nations' macroeconomics (O'Connor et al., 2009). Nature-based tourism involves interactions with the natural environment and it represents a big component of global recreation (Balmford *et al.* 2009). This type of recreation is an important issue because of its economic contribution to conservation (Gossling 1999; Brightsmith *et al.* 2008), the health benefits it brings to humans (Russell *et al.* 2013) and its role in alleviating poverty (Ferraro & Hanauer 2014). Nature-based tourism, such as wildlife watching, was initially welcomed by conservation and environmental organisations as an eco-friendly alternative to other consumptive activities, such as hunting and fishing (Tisdell & Wilson 2002). However, there is growing evidence that these activities, if not managed properly, can have negative effects on the environment (McClung *et al*. 2004; Reed & Merenlender 2008; Pirotta & Lusseau 2015). Quantifying temporal and spatial patterns of wildlife watching can help management by identifying areas that are under high pressure from these activities and areas that could be sustainably developed to redistribute this pressure.

The widespread use of the Internet and the popularity of smartphones and social media websites offer the opportunity to use the data generated by their billions of users (Worthington *et al.* 2012; Leighton *et al.* 2016). To our knowledge, five studies in the field of ecosystem services and nature-based tourism have compared data from social media to visitor statistics obtained through more traditional methods (e.g. surveys or censuses) (Wood *et al.* 2013; Levin *et al.* 2015, 2017; Sessions *et al.* 2016; Hausmann *et al.* 2017). All of these studies have demonstrated that there is a correlation between these two types of data. (Wood *et al.* 2013) and (Levin *et al.* 2015) tested the use of social media to quantify nature-based tourism in protected areas and selected tourist sites across the globe; (Levin *et al.* 2017) and (Sessions *et al.* 2016) tested the same on a national scale, respectively in Australia and the United States; while (Hausmann *et al.* 2017) validated the use of the content of

55 photos posted online to infer tourists' preferences inside the Kruger National Park. These studies
56 have focused only on protected areas and selected visitor attractions and none of them addressed
57 the question of whether the spatial patterns of geotagged photos from social media correspond to
58 those found in data collected using conventional methods. In addition, only two of them tested the
59 validity of the temporal component of social media data as a proxy for short- and long-term
60 temporal patterns of nature-based recreation (Wood *et al.* 2013; Sessions *et al.* 2016). Therefore, we
61 still don't know if this proxy is valid to quantify cultural ecosystem services outside protected areas,
62 at which spatial resolution it is accurate and whether we can use it to infer short- and long-term
63 temporal patterns of nature-based recreation.

64 However, following promising results from the first study (Wood *et al.* 2013), the use of data from
65 social media to quantify recreational ecosystem services and tourists' preferences has become
66 increasingly popular (Keeler *et al.* 2015; Richards & Friess 2015; Gliozzo *et al.* 2016; van Zanten *et al.*
67 2016; Martínez Pastur *et al.* 2016; Tenerelli *et al.* 2016; Sonter *et al.* 2016; Tieskens *et al.* 2017;
68 Yoshimura & Hiura 2017). These studies used geotagged photos hosted on different social media
69 websites at spatial and temporal scales and resolutions that have not been tested before. This study
70 aimed to fill in these knowledge gaps, by: 1) testing whether the spatial patterns of geotagged
71 photos posted on the photo-sharing website Flickr are a good proxy for spatial trends in nature-
72 based recreation at a national scale and outside protected areas; 2) identifying the spatial resolution
73 at which this proxy is appropriate at this scale; 3) determining whether the temporal information
74 contained in these data can be used to infer short- and long-term patterns in nature-based
75 recreation.

76 We addressed these questions using unique data available in Scotland. Wildlife watching contributes
77 roughly £127 million per year to the Scottish economy (Bryden *et al.* 2010), generating 2763 full-
78 time equivalent (FTE) jobs (Blake *et al.* 2010). The main groups of charismatic wildlife that attract
79 tourists to Scotland are birds, seals, whales and dolphins (Curtin 2013). We tested whether temporal

patterns of pictures taken at a nature-based recreation site (the Cairngorms National Park – CNP) and posted on Flickr were reproducing patterns found in time series of visitor numbers obtained from the CNP authority. We then tested the validity of the spatial distribution of geotagged pictures of wildlife posted on Flickr as a proxy for spatial patterns of wildlife watching activities in Scotland using a dataset collected through a Scotland-wide tourism survey (Land Use Consultants 2016). Finally, we use this validation to assess the geography and long-term temporal patterns of wildlife tourism in Scotland.

2. Materials and methods

2.1. Data collection

In order to accomplish the aims of this study, we downloaded data from the Flickr Application Programming Interface (API) to create 3 datasets. Two of these datasets were then compared to visitor statistics obtained from two sources: time series of visitor numbers provided by the CNP authority and a spatial dataset of wildlife watching from the Scottish marine recreation and tourism survey (Land Use Consultants 2016). In the next three section we describe these datasets before providing details on the analysis.

2.1.1. Flickr and CNP

Time series from the CNP authority - The CNP visitation dataset contained the number of visitor days to the park (a person spending at least a portion of a day at the CNP) per month from 2009 to 2014. These visitor days were estimated by the STEAM model (STEAM visitor days – SVD) (Global Tourism Solutions Ltd. 2006), which uses data on accommodation occupancy rates, visitor surveys, number of visitors to paid attractions and other data supplied by the CNP authority.

Flickr - Data from Flickr were collected through Flickr application programming interface (API https://www.flickr.com/services/api/) and R (R Core Team 2015), using the packages RCurl (Lang & the CRAN team 2015; version 1.95.4.7), XML (Lang & the CRAN team 2015b; version 3.98.1.3) and

104 httr (Wickham 2016; version 1.1.0) to communicate with the API, request and download the data.
105 Dates and geographic coordinates associated with the pictures were used to select only those taken
106 in the CNP between 2009 and 2014 (R code available at https://github.com/FrancescaMancini/Flickr-
107 API). We downloaded different metadata associated with the photos: picture and photographer ID,
108 the date when the photo was taken, the geographic coordinates of where the picture was taken and
109 the tags. In order to avoid bias coming from having a small number of very active users, we used the
110 combination of photographer-ID and date to delete multiple photos from the same user on the same
111 day. Therefore, the number of data points in the dataset represents the number of Flickr visitor days
112 (FVD).

113 *2.1.2. Flickr and wildlife watching*

114 Flickr - The second Flickr dataset contained information about wildlife watching Scotland-wide. We
115 used keywords (bird, seal, dolphin and whale) to query Flickr API and select all the geotagged photos
116 of the main groups of wildlife sought by tourists in Scotland and we downloaded the same metadata
117 as for the CNP photos. For each of the different wildlife groups, we again deleted multiple photos
118 from the same user on the same day. Using the tags associated with the photos, we eliminated all
119 the pictures that were not relevant, such as pictures of statues or paintings and pictures taken in
120 zoos. The text that was associated with each picture was examined and a list of keywords for non-
121 relevant photos was compiled, then following the same method from (van Zanten *et al.* 2016), we
122 used that list of keywords to filter out irrelevant pictures (see R code available at
123 https://github.com/FrancescaMancini/Flickr-API). We recognise that this method is not perfect and
124 it might have left some non-relevant photos behind, but we suggest that, considering the high
125 number of pictures we processed, this was the best way to ensure that the mistakes introduced by
126 the filtering procedures were not biasing the results.

127 *2.1.3. Flickr and the Scotland tourism survey*

Scottish tourism survey - The dataset from the Scottish marine recreation and tourism survey (Land Use Consultants 2016) (available at http://live-marinedatascotland.getnucivic.com/dataset/scottish-marine-recreation-and-tourism-survey-2015) contained spatial information on trips to coastal areas in Scotland made by the survey respondents between October 2014 and October 2015 to conduct wildlife watching activities. The online survey asked respondents to draw polygons on a map of Scotland, corresponding to the areas they had used for wildlife watching activities in that period. All the polygons were then stacked on top of each other and counted, resulting in a raster file with a resolution of 1 Km.

Flickr - The third Flickr dataset was a subset of the Scotland-wide one described in section 2.1.2, from which we selected only the geotagged pictures taken between October 2014 and October 2015. A buffer of 2 Km inside the coastline was created in ArcGIS (ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute) to select only pictures taken in a coastal environment. In ArcGIS, we created 3 grids (5 Km, 10 Km and 20 Km) and counted the number of FVD and of visitor days from the survey (SuVD) in each cell.

2.2. Analysis

2.2.1. Temporal validation

In order to test whether the temporal patterns shown by the Flickr and the CNP authority time series were similar, we used Wavelets Analysis (WA) (Cazelles *et al.* 2008). WA decomposes the variance of the time series in its oscillating components, thus detecting significant periodicities. The advantage of this method compared to other spectral decompositions is that WA does not assume stationarity of the time series, but it allows the main frequency component to change through time by estimating the signal's spectral characteristics as a function of time (the wavelet power spectra). We used the Morlet mother wavelet to perform this decomposition. This continuous and complex function allows the extraction of both time-dependent amplitude and phase of the time series. WA also allows the analysis of patterns of covariation between two time series. We compared the time

153 series of SVD and FVD using the wavelet coherence, which identifies the linear correlation between
154 two signals. In order to assess statistical significance of the association between the two time series
155 we used a random noise resampling scheme, where the null hypothesis tested is that the association
156 between the two signals is not different from that expected by chance alone (Ménard *et al.* 2007).
157 We also computed the phase difference to test whether the two time series were synchronised or
158 out of phase. This analysis was conducted in R using the package WaveletComp version 1.0 (R code
159 available at https://github.com/FrancescaMancini/Flickr-Statistical-Analysis) (Roesch & Schmidbauer
160 2014).

161 *2.2.2. Spatial validation*

162 We compared the spatial distribution of wildlife tourists obtained from Flickr and the one obtained
163 from the survey at three different spatial scales: 5 Km, 10 Km and 20 Km. We fitted Generalised
164 Linear Models (GLM) to the three datasets using the number of FVD in each cell as the response
165 variable and the number of SuVD in each cell as explanatory variable. The data contained a high
166 number of zeros, so we fitted a GLM with a binomial error distribution to the presence/absence of
167 FVD in a cell and a GLM with a negative binomial error distribution to the number of FVD present.
168 This allowed us to test two hypotheses: 1) the probability of finding at least one picture by one user
169 on Flickr is higher for areas with higher SuVD 2) the number of users posting pictures on Flickr is
170 higher for areas with higher SuVD. Since densely populated areas tend to have a higher average
171 number of Flickr users (Fig. A1), we used population abundance for each grid cell as model weights.
172 The data for population size was a 1Km resolution raster of estimates of population count available
173 at http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count (Centre for International
174 Earth Science Information Network - CIESIN - Columbia University, 2016. Gridded Population of the
175 World, Version 4 (GPWv4): Population Count. doi:10.7927/H4X63JVC). The residuals of the GLMs
176 were spatially correlated and directional variograms, estimated using R package gstat version 1.1.0
177 (R code available at https://github.com/FrancescaMancini/Flickr-Statistical-Analysis) (Pebesma

178 2004), showed that the spatial autocorrelation was anisotropic (Fig. A2). We therefore used spatial
179 eigenvector mapping (SEVM) to derive explanatory variables for the GLMs (Griffith & Peres-Neto
180 2006). This method decomposes a matrix of relationships between data points into eigenvectors
181 that capture spatial effects. The eigenvectors can then be included as explanatory variables in the
182 GLM to remove the effect of spatial autocorrelation on the analysis. First we used a Delaunay
183 triangulation of the grid cells centres to define neighbours, to which we assigned row-standardised
184 spatial weights. A set of Moran eigenvectors (ME) were then calculated from these weights and
185 those that best reduced the spatial autocorrelation of residuals were selected and included as spatial
186 covariates in the GLMs. We used AIC to select only the ME that improved the explanatory power of
187 the model. This analysis was conducted in R using the package spdep version 0.5.92 (R code available
188 at https://github.com/FrancescaMancini/Flickr-Statistical-Analysis) (Bivand *et al.* 2013; Bivand &
189 Piras 2015).

190 *2.2.3. Wildlife watching hotspots and long-term temporal patterns*

191 We investigated spatio-temporal patterns in wildlife watching hotspots by producing density maps
192 of the geotagged pictures posted on Flickr. We used a two-dimensional kernel density estimator
193 from the R package ggplot2 (Wickham 2009), where the bandwidth is calculated using Scott's rule of
194 thumb (Scott 1992) (R code available at https://github.com/FrancescaMancini/Flickr-Statistical-
195 Analysis).

196 **3. Results**

197 In total, we downloaded metadata on 29,336 pictures (4699 unique FVD) taken in the CNP between
198 2009 and 2014 and uploaded on Flickr.

199 The query to Flickr API returned 92,229 results (36,998 FVD) for pictures with the word "bird" in
200 tags, title or description taken in Scotland between 2005 and 2015. From the search with the word

"seal" we obtained 7212 photos (2571 FVD) and from the search with the word "dolphin" and the word "whale" we obtained 5994 pictures (1634 FVD).

3.1. Temporal validation

The power spectra of the Flickr and CNP survey time series were very similar, with significant 12-month cycles throughout the 5-year period (Fig. A3a-b), showing that visitation has a strong seasonal component in both measures. This similarity was supported by the strong coherence between the Flickr and the empirical time series around the same 12-month oscillations, which was also constant through time (Fig. 1a). The phase difference for this 12-month cycle was constant around 0 (Fig.1b), indicating that the two time series were synchronised. Cross-correlation was also significant at a period of 6 and 3 months, but this was not consistent throughout the time period (Fig.1a).

3.2. Spatial validation

The survey dataset indicated that the areas more intensely used for wildlife watching are around the west coast of Scotland, the Moray Firth, the Firth of Forth and the Tay estuary (Fig. A4). Spatial distribution of the pictures from Flickr also identified the last three as areas of high visitation (Fig. A1). The Flickr dataset also contained pictures taken on the west coast of Scotland, but not in the same density as shown by the survey dataset. This area of the country is not highly populated, so there might be a certain bias in the number of Flickr users uploading pictures. When the Flickr data was normalised by population size, the west coast appeared as an area of high visitation (Fig. A5).

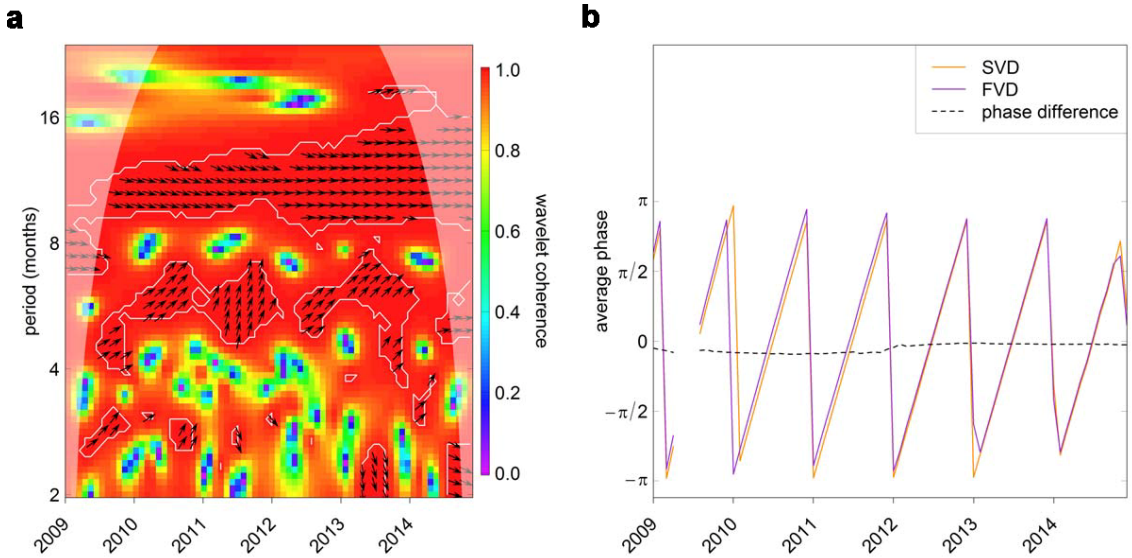

Fig.1. Results of wavelet analysis. a) Wavelet coherence between the two time series. Colour code from purple (low values) to red (high values). The arrows indicate synchrony of the two time series: arrows pointing to the right means the oscillations are synchronised. Arrows are only plotted within white contour lines that indicate significance. The shaded area near the edges in the graphs is the cone of influence, and indicates the range of the graph where the results are not reliable because of edge effects. b) Phases of the oscillations of the two time series (orange and purple lines) computed in the 8-16 periodic band where there is significant correlation. The dotted line is the phase difference.

The probability of finding a Flickr picture taken in a certain cell was higher in cells with higher SuVD (Fig. 2). This was true for each of the spatial scales tested (20km: estimate = 0.37, *SE* = 0.04, *Z* = 9.5, *p-value* < 0.001; 10km: estimate = 0.36, SE = 0.02, Z = 14.7, *p-value* < 0.001; 5km: estimate = 0.28, *SE* = 0.001, *Z* = 21.1, *p-value* < 0.001). The higher the number of visitors captured by the survey, the higher the probability of finding photos on Flickr taken in that area.

We found a correspondence between the number of FVD and the number of SuVD at 10km and 20km resolution but not at 5km (Fig. 3; 20km: estimate = 0.1, *SE* = 0.01, *Z* = 8.1, *p-value* < 0.001; 10km: estimate = 0.02, *SE* = 0.001, *Z* = 3.4, *p-value* < 0.001; 5km: estimate = 0.01, *SE* = 0.006, *Z* = 1.7, *p-value* > 0.05). The number of Flickr users taking a picture in a cell was higher in cells with higher number of visitors captured by the survey, but only at a 10 and 20 Km resolution.

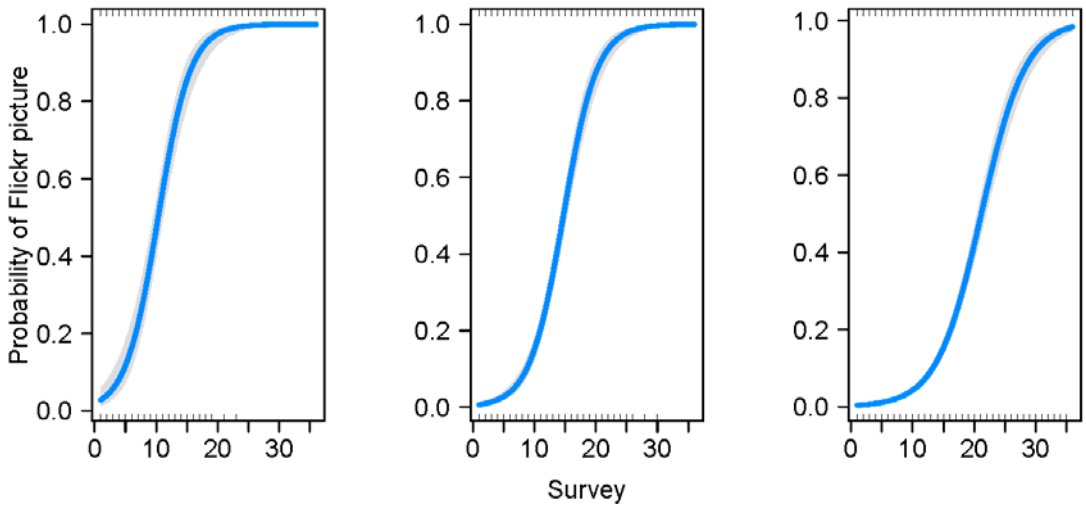

Fig. 2. Results of binomial GLMs. Left: results at the 20 Km resolution; centre: results at the 10Km resolution; right: results at the 5Km resolution. Predictions from the models (blue line) are plotted on the response scale with confidence intervals (shaded areas around the prediction curve). Tick marks on the x-axis represent data.

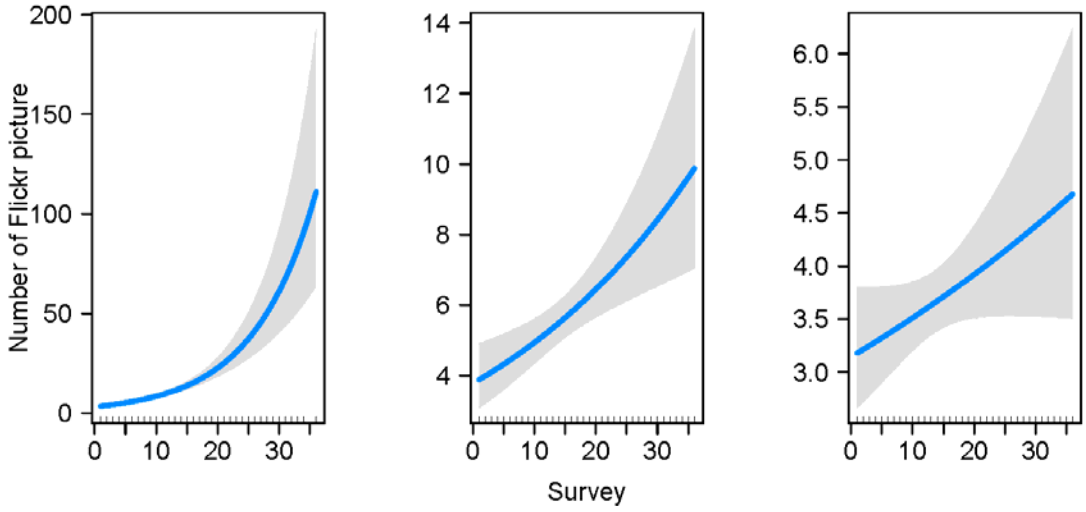

Fig. 3. Results of negative binomial GLMs. Left: results at the 20 Km resolution; centre: results at the 10Km resolution; right: results at the 5Km resolution. Predictions from the models (blue line) are plotted on the response scale with confidence intervals (shaded areas around the prediction curve). Tick marks on the x-axis represent data.

3.3. Wildlife watching hotspots and long-term temporal patterns

The density maps revealed spatio-temporal patterns of wildlife watching hotspots in Scotland. Birdwatching (Fig. 4) seems to be concentrated around Edinburgh and Glasgow, however when FVD in Edinburgh and Glasgow were excluded from the dataset, hotspots in the Moray Firth, Orkney, Shetland, the Isle of Mull and the North-West coast started to appear (Fig. A6). A seasonal plot of the same pictures shows that this high density around urban areas spreads out towards the West coast and the islands during spring and summer (Fig. A7).

Seal watching seems to be concentrated initially around the west coast, the Firth of Forth and Shetland (Fig. 5). It is worth noticing the appearance of another hotspot from 2008 corresponding to Newborough in Aberdeenshire, becoming very important after 2011.

Dolphin and whale watching maps showed a consistent hotspot at Chanonry Point in the Moray Firth (Fig. 6), with the appearance of a second hotspot from 2013 in Aberdeen.

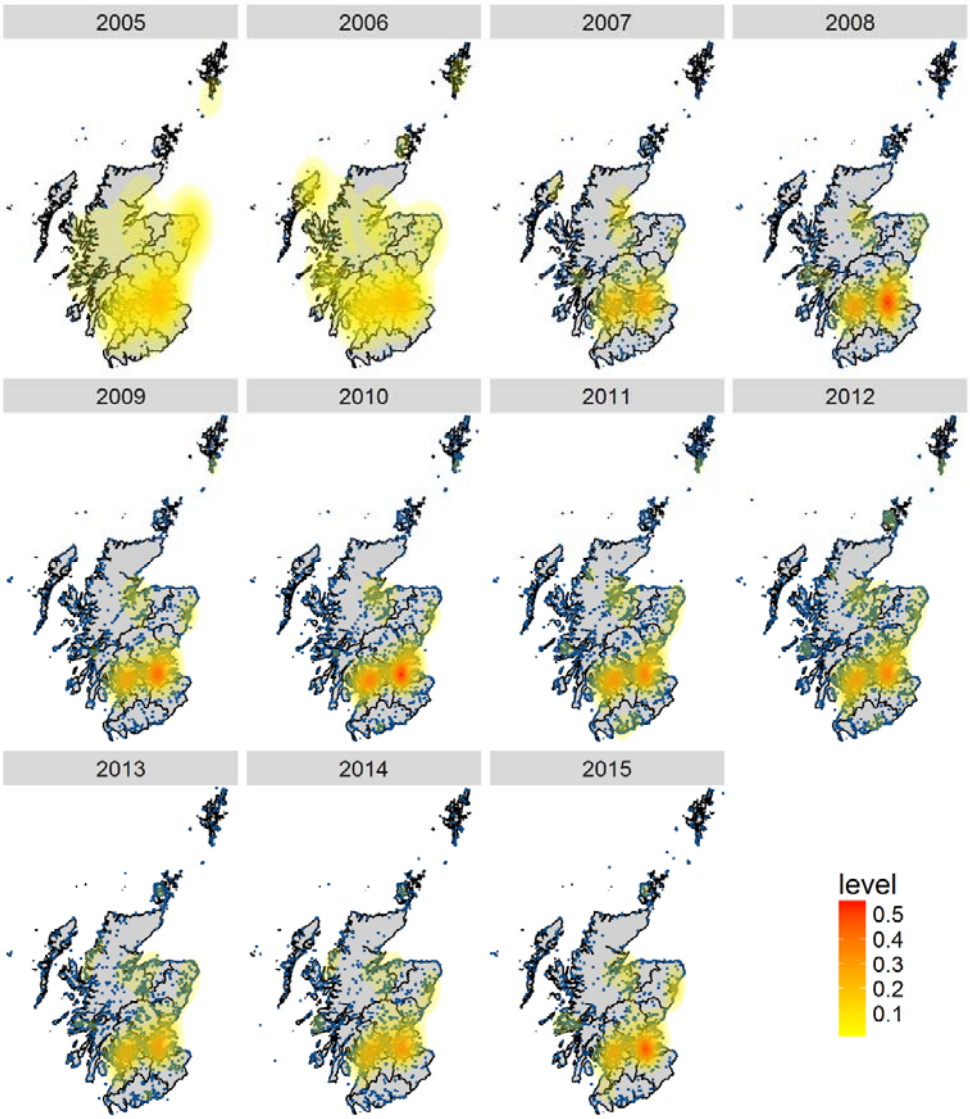

Fig.4. Bird watching density maps. Each panel represents the density of FVD in a different year, from 2005 to 2015. The blue dots on the maps are the data. Different colours represent different density levels, from low (yellow) to high (red).

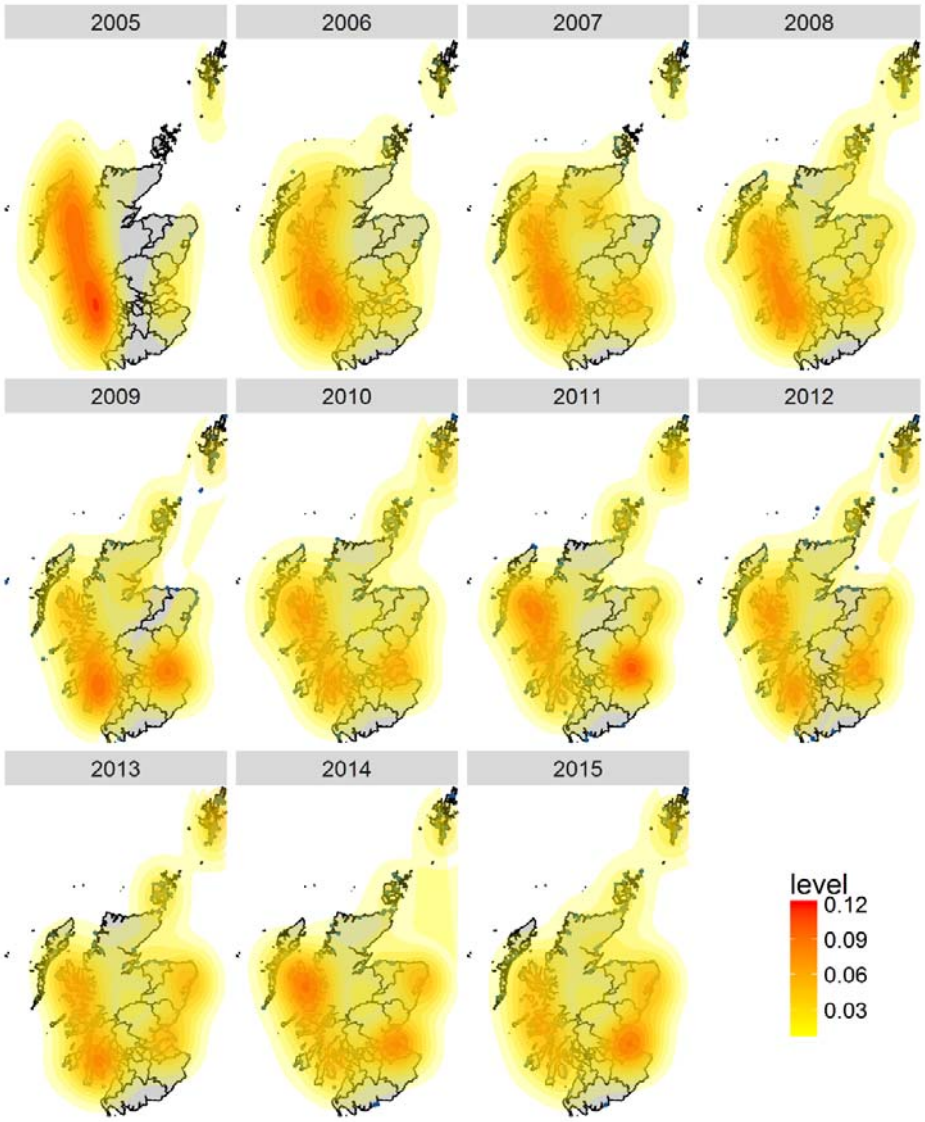

Fig.5. Seal watching density maps. Each panel represents the density of FVD in a different year, from 2005 to 2015. The blue dots on the maps are the data. Different colours represent different density levels, from low (yellow) to high (red).

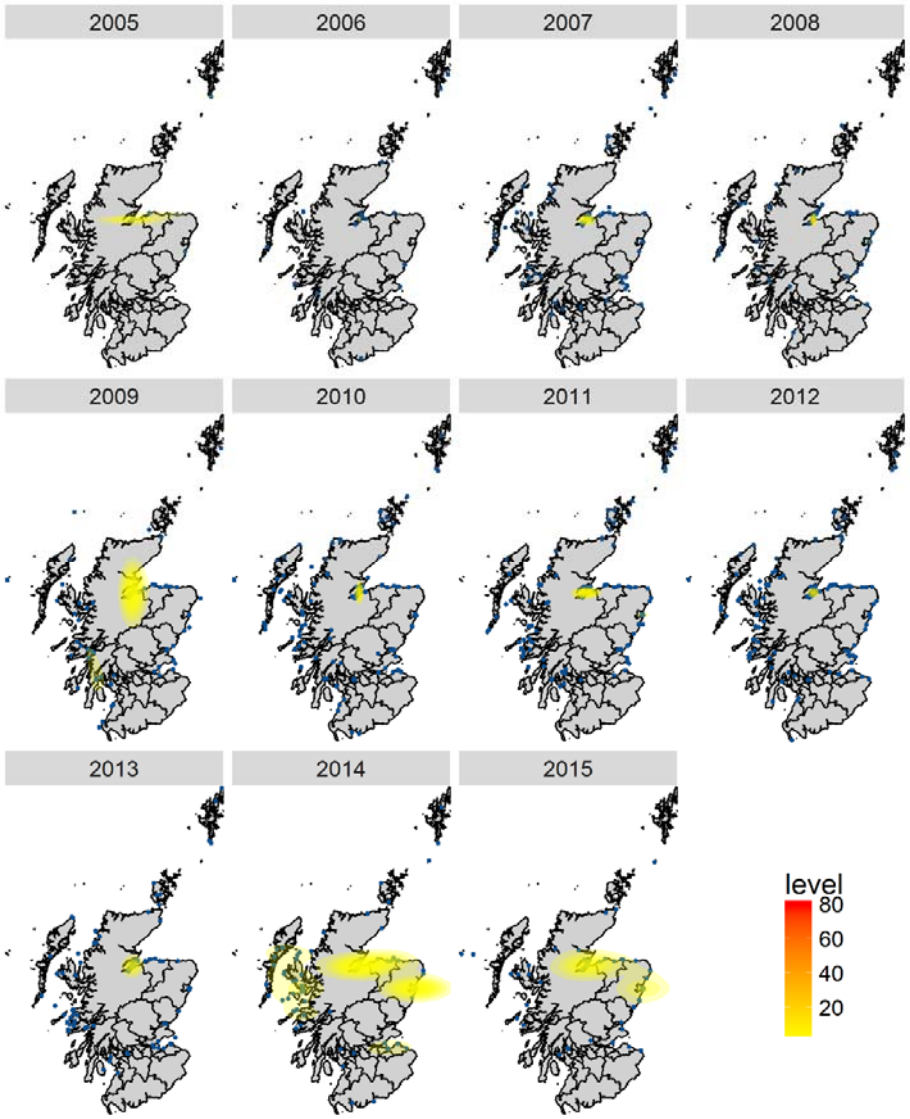

Fig.6. Dolphin and whale watching density maps. Each panel represents the density of FVD in a different year, from 2005 to 2015. The blue dots on the maps are the data. Different colours represent different density levels, from low (yellow) to high (red).

4. Discussion

In this study, we were able to compare spatial and temporal patterns in pictures from Flickr to those obtained through traditional methods. The results are in line with previous work (Wood *et al.* 2013; Levin *et al.* 2015, 2017; Sessions *et al.* 2016; Hausmann *et al.* 2017), suggesting that geotagged pictures uploaded on Flickr can be a good proxy for nature-based recreational activities. Novel

279 results from our study reveal that this data source can reliably quantify both short- (Fig.1) and long-
280 term temporal patterns of nature-based recreation (Fig.4-6; see Appendix B for further analysis and
281 validation of temporal trends), as well as the spatial distribution of wildlife watching activities on a
282 national scale (Fig.2 and Fig.3). This study also showed that the spatial resolution at which we can
283 use this proxy can be as fine as 5 x 5 Km (Fig.2), but, at a resolution finer than 10 x 10 Km, the
284 number of Flickr users taking pictures of wildlife may not be a reliable measure for volume of
285 recreation (Fig.3). This could be due to lack of precision in entering geographic coordinates of the
286 location where the photo was taken. If GPS signal is not available or the photo is taken with a device
287 that does not record geographic coordinates, then users can add the spatial reference during the
288 uploading of the photo on the website by indicating on a map the location where the picture was
289 taken. This process might not be accurate enough to provide reliable information about where
290 people go at a very fine spatial scale. This result calls for caution when using this proxy to make
291 inferences about spatial patterns of visitation rates on a regional scale and fine resolution.

292 Using both spatial and temporal information from geotagged photos uploaded on Flickr at the same
293 time provided insights into long-term spatio-temporal patterns of wildlife watching activities that
294 further validated the use of this proxy and gave information that is relevant for the management of
295 these activities. The majority of pictures of birds were taken around Edinburgh and Glasgow.
296 However, after excluding these urban areas from the dataset the density maps revealed other
297 hotspots of bird watching such as the Moray Firth, Orkney, Shetland, the Isle of Mull and the North-
298 West coast (Fig. A6). The density maps also detected a change in the bird watching hotspots with
299 seasons (Fig.A7), consistent with a movement from the area around Edinburgh and Glasgow to more
300 remote areas on the west coast, the Moray Firth and the islands. We were therefore able to capture
301 the movement of people towards tourism destination (Blake *et al.* 2010; Land Use Consultants
302 2016). The seal hotspot maps (Fig. 5) revealed high activity in the Firth of Forth, Tay estuary and the
303 West coast where special areas of conservation (SACs) and haul out sites are present for both grey
304 and harbour seals (Morris *et al.* 2014). This map also showed the appearance of a seal watching

305 point in Newborough after 2008. This site now holds 26% of grey seals (*Halichoerus grypus*) and 1%
306 of harbour seals (*Phoca vitulina*) in the East Coast of Scotland Seal Management Area and has
307 recently been proposed as a new designated haul-out site to protect the seals (Marine Scotland
308 2015). The popularity of this site kept increasing from 2008 through to 2015 and this information
309 could be important to determine if protection measures need to be put in place given the increasing
310 pressure from volume of recreation. The whale and dolphin watching density maps (Fig. 6) showed a
311 very consistent hotspot on the East coast, corresponding to Chanonry Point in the Moray Firth,
312 which is a very popular dolphin watching destination
313 (https://www.tripadvisor.co.uk/Attraction_Review-g783108-d2366132-Reviews-Chanonry_Point-
314 Fortrose_Ross_and_Cromarty_Scottish_Highlands_Scotland.html). The same maps also revealed the
315 emergence of a new dolphin watching hotspot in Aberdeen. The hotspot only appears in 2013, after
316 the launch by the Royal Society for the Protection of Birds (RSPB) of dolphin watching events from
317 Aberdeen harbour as part of the "Dates with nature" projects
318 (http://www.rspb.org.uk/discoverandenjoynature/seenature/datewithnature/details.aspx?id=34036
319 6). This result indicates that such organised events can attract tourists and create a hotspot, offering
320 an opportunity for managers to influence tourism demand. Further analysis could explore whether
321 the birth of a new hotspot creates new demand or whether it could shift tourists' attention from
322 overexploited destinations to unused ones.

323 Data from social media still presents some limitations that need to be acknowledged, some of which
324 also apply to traditional sampling methods (Wood et al., 2013; Li et al., 2013). Also, there is a bias
325 resulting from densely populated areas having more Flickr users than sparsely populated ones (Levin
326 et al., 2015; Fig.4 and Fig. A.1). Different species of wildlife may be more or less suited to be
327 photographed: for instance, there are far more pictures of birds than dolphins uploaded on Flickr,
328 partly due to the fact that encounters with dolphins are less common than encounters with birds,
329 and partly to the fact that taking a picture of a dolphin is more difficult than taking a picture of a bird
330 and it might require specialised equipment. Therefore, it might not be possible to compare volume

331 of tourism dedicated to different species. Furthermore, the perceived value of a trip may influence
332 whether an individual takes or shares photographs, producing a bias against images from visitors
333 who visit areas closer to their home (Wood *et al.* 2013). Half of the respondents to the marine
334 recreation and tourism survey lived within one mile of the coast (Land Use Consultants 2016) and
335 they might have reported using an area where they would not normally take pictures because of its
336 proximity to home. This could explain some of the differences between the two datasets.

337 In conclusion, despite limitations, the number of geotagged pictures uploaded on Flickr can be used
338 as a proxy for wildlife watching activities at spatial and temporal scales that are relevant for
339 ecosystem management in global regions where this social media is prevalently used (Wood *et al.*
340 2013). This opens new avenues to study tourist behaviour and decisions. The fact that we can use
341 this data at a scale as fine as 10 Km means that we can now make more precise inference on
342 tourists' preferences on larger areas. This information has also implications for wildlife tourism
343 management and conservation of targeted species. First, we can now easily and cheaply quantify
344 wildlife tourism in areas that are not monitored, allowing us to assess whether some areas,
345 therefore wildlife populations, are receiving increasing pressure from tourism, while other areas are
346 underutilised. Secondly, organised events such as the RSPB "Date with Nature" can attract tourists
347 and create a recreational hotspot, offering an opportunity for managers to influence demand. Lastly,
348 data from social media can give us a dynamic view of demand for nature-based recreation, which
349 can be used to anticipate changes in visitation in response to changes in ecosystems and socio-
350 economic development. In a fast-changing world, the ability to predict changes is going to be
351 increasingly important to plan for a sustainable use of the natural environment.

352 **Acknowledgements**

353 This work was funded by the University of Aberdeen and Scottish Natural Heritage (SNH) and their
354 support is gratefully acknowledged. We thank MASTS (the Marine Alliance for Science and
355 Technology for Scotland) for their role in funding this work and B. Leyshon and F. Manson (SNH) for

fruitful discussion. All the data used in this study are publicly available and were collected following terms and conditions of the data providers. The data are anonymised and we did not collect any personal information.

Data accessibility

All the data and R scripts are available at https://github.com/FrancescaMancini/Flickr-API and https://github.com/FrancescaMancini/Flickr-Statistical-Analysis.

References

1. Balmford, A., Beresford, J., Green, J., Naidoo, R., Walpole, M. & Manica, A. (2009). A global perspective on trends in nature-based tourism. *PLoS biology*, **7**, e1000144.
2. Bivand, R., Hauke, J. & Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, **45**, 150–179.
3. Bivand, R. & Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, **63**, 1–36.
4. Blake, A., Curtin, S., Richards, S., Vaughan, R., Edwards, J. & Fletcher, J. (2010). *The Economic Impact of Wildlife Tourism in Scotland. Report by the International Centre for Tourism and Hospitality Research Bournemouth University for Scottish Government Social Research*.
5. Brightsmith, D.J., Stronza, A. & Holle, K. (2008). Ecotourism, conservation biology, and volunteer tourism: A mutually beneficial triumvirate. *Biological Conservation*, **141**, 2832–2842.
6. Bryden, D.M., Westbrook, S.R., Burns, B., Taylor, W.A. & Anderson, S. (2010). *Assessing the economic impacts of nature based tourism in Scotland. Scottish Natural Heritage Commissioned Report No. 398.*
7. Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S. & Stenseth, N.C. (2008). Wavelet analysis of ecological time series. *Oecologia*, 287–304.

8. Curtin, S. (2013). Lessons from Scotland: British wildlife tourism demand, product development and destination management. *Journal of Destination Marketing and Management*, **2**, 196–211.
9. Curtin, S. (2003). Whale-Watching in Kaikoura: Sustainable Destination Development? *Journal of Ecotourism*, **2**, 173–195.
10. ESRI. (2013). ArcGIS Desktop: Release 10.2. Redlands CA: Environmental Systems Research Institute.
11. Ferraro, P.J. & Hanauer, M.M. (2014). Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 4332–4337.
12. Gliozzo, G., Pettorelli, N. & Haklay, M. (Muki). (2016). Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. *Ecology and Society*, **21**, art6.
13. Global Tourism Solutions Ltd. (2006). Overview of STEAM, GTS. URL http://www.globaltourismsolutions.co.uk/steam-model
14. Gossling, S. (1999). Ecotourism: a means to safeguard biodiversity and ecosystem functions? *Ecological Economics*, **29**, 303–320.
15. Griffith, D.A. & Peres-Neto, P.R. (2006). Spatial Modeling in Ecology: the Flexibility of Eigenfunction Spatial Analyses. *Ecology*, **87**, 2603–2613.
16. Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V. & Di Minin, E. (2017). Social Media Data can be used to Understand Tourists´ Preferences for Nature-based Experiences in Protected Areas. *Conservation Letters*, **0**, 1–10.
17. Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T. & Downing, J.A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, **13**, 76–81.
18. Land Use Consultants. (2016). Scottish Marine Recreation and Tourism Survey 2015.

19. Lang, D.T. & the CRAN team. (2015a). RCurl: General Network (HTTP/FTP/...) Client Interface for R. R package version 1.95-4.7. URL https://cran.r-project.org/package=RCurl.
20. Lang, D.T. & the CRAN team. (2015b). XML: Tools for Parsing and Generating XML Within R and S-Plus. R package version 3.98-1.3. URL https://cran.r-project.org/package=XML.
21. Leighton, G.R.M., Hugo, P.S., Roulin, A. & Amar, A. (2016). Just Google it: Assessing the use of Google Images to describe geographical variation in visible traits of organisms. *Methods in Ecology and Evolution*, **7**, 1060–1070.
22. Levin, N., Kark, S. & Crandall, D. (2015). Where have all the people gone? Enhancing global conservation using night lights and social media. *Ecological Applications*, **25**, 2153–2167.
23. Levin, N., Mark, A. & Brown, G. (2017). An evaluation of crowdsourced information for assessing the importance of protected areas. *Applied Geography*, **79**, 115–126.
24. Marine Scotland. (2015). *Consultation on Possible Designation of a Seal Haul-out Site. Proposal in respect of the Ythan Estuary*.
25. Martínez Pastur, G., Peri, P.L., Lencinas, M. V., García-Llorente, M. & Martín-López, B. (2016). Spatial patterns of cultural ecosystem services provision in Southern Patagonia. *Landscape Ecology*, **31**, 383–399.
26. Ménard, F., Marsac, F., Bellier, E. & Cazelles, B. (2007). Climatic oscillations and tuna catch rates in the Indian Ocean: A wavelet approach to time series analysis. *Fisheries Oceanography*, **16**, 95–104.
27. Morris, C., Duck, C., Lonergan, M., Baxter, J., Middlemas, S. & Walker, I. (2014). *Method used to identify key seal haul-out sites in Scotland for designation under the Marine (Scotland) Act Section 117*.
28. Pebesma, E.J. (2004). Multivariable geostatistcs in S: the gstat package. *Computers & Geosciences*, **30**, 683–691.
29. Pirotta, E. & Lusseau, D. (2015). Managing the wildlife tourism commons. *Ecological Applications*, **25**, 729–741.

30. R Core Team. (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.r-project.org/.
31. Reed, S.E. & Merenlender, A.M. (2008). Quiet, Nonconsumptive Recreation Reduces Protected Area Effectiveness. *Conservation Letters*, **1**, 146–154.
32. Richards, D.R. & Friess, D.A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*, **53**, 187–195.
33. Roesch, A. & Schmidbauer, H. (2014). WaveletComp: Computational Wavelet Analysis. R package version 1.0. URL https://cran.r-project.org/package=WaveletComp.
34. Russell, R., Guerry, A.D., Balvanera, P., Gould, R.K., Basurto, X., Chan, K.M. a., Klain, S., Levine, J. & Tam, J. (2013). *Humans and Nature: How Knowing and Experiencing Nature Affect Well-Being*.
35. Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization* (D.W. Scott, Ed.). John Wiley & Sons, Inc., Hoboken, NJ, USA.
36. Sessions, C., Wood, S.A., Rabotyagov, S. & Fisher, D.M. (2016). Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management*, **183**, 703–711.
37. Silva, L. (2013). How ecotourism works at the community-level: the case of whale-watching in the Azores. *Current Issues in Tourism*, **18**, 196–211.
38. Sonter, L.J., Watson, K.B., Wood, S.A., Ricketts, T.H., Danielson, P. & Xian, G. (2016). Spatial and Temporal Dynamics and Value of Nature-Based Recreation, Estimated via Social Media (J. Yang, Ed.). *PLOS ONE*, **11**, e0162372.
39. Tenerelli, P., Demšar, U. & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, **64**, 237–248.
40. Tieskens, K.F., Schulp, C.J.E., Levers, C., Lieskovs, J., Kuemmerle, T., Plieninger, T. & Verburg,

P.H. (2017). Characterizing European cultural landscapes: Accounting for structure, management intensity and value of agricultural and forest landscapes. *Land Use Policy*, **62**, 29–39.

41. Tisdell, C. & Wilson, C. (2002). Ecotourism for the survival of sea turtles and other wildlife. *Biodiversity and Conservation*, **11**, 1521–1538.
42. Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer Internatinal Publishing.
43. Wickham, H. (2016). httr: Tools for Working with URLs and HTTP. R package version 1.1.0. URL https://cran.r-project.org/package=httr.
44. Wood, S. a, Guerry, A.D., Silver, J.M. & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports*, **3**, 2976.
45. Worthington, J.P., Silvertown, J., Cook, L., Cameron, R., Dodd, M., Greenwood, R.M., McConway, K. & Skelton, P. (2012). Evolution MegaLab: a case study in citizen science methods. *Methods in Ecology and Evolution*, **3**, 303–309.
46. Yoshimura, N. & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services*, **24**, 68–78.
47. van Zanten, B.T., Van Berkel, D.B., Meentemeyer, R.K., Smith, J.W., Tieskens, K.F. & Verburg, P.H. (2016). Continental-scale quantification of landscape values using social media data. *Proceeding of the National Academy of Science*, **113**, 12974–12979.