Environmental Research Communications

PAPER • OPEN ACCESS

Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis

To cite this article: Claire A Runge *et al* 2020 *Environ. Res. Commun.* **2** 075001

IOP Publishing *Environ. Res. Commun.* **2** (2020) 075001 https://doi.org/10.1088/2515-7620/ab9c33

Environmental Research Communications

RECEIVED
25 February 2020

REVISED
18 May 2020

ACCEPTED FOR PUBLICATION
12 June 2020

PUBLISHED
2 July 2020

PAPER

Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis

Claire A Runge[1], Vera H Hausner[1], Remi M Daigle[2] and Christopher A Monz[3]

[1] Arctic Sustainability Lab, Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway
[2] Département de biologie, 'Université Laval, Québec, Canada
[3] Department of Environment and Society and the Ecology Center, Utah State University, Logan, UT, United States of America

E-mail: claire.runge@uqconnect.edu.au

Keywords: cultural ecosystem services, social media, Arctic, ecotourism, Flickr, passive crowdsourcing, artificial intelligence

Supplementary material for this article is available online

Abstract

In the Arctic, as in many parts of the world, interactions with the natural world are an important part of people's experience and are often recorded in photographs. Emerging methods for automated content analysis of social media data offers opportunities to discover information on cultural ecosystem services from photographs across large samples of people and countries. We analysed over 800 000 Flickr photographs using Google's Cloud Vision algorithm to identify the components of the natural environment most photographed and to map how and where different people interact with nature across eight Arctic countries. Almost all (91.1%) of users took one or more photographs of biotic nature, and such photos account for over half (53.2%) of Arctic photos on Flickr. We find that although the vast majority of Arctic human-nature interactions occur outside protected areas, people are slightly more likely to photograph nature inside protected areas after accounting for the low accessibility of Arctic protected areas. Wildlife photographers travel further from roads than people who take fewer photographs of wildlife, and people venture much further from roads inside protected areas. A large diversity of nature was reflected in the photographs, from mammals, birds, fish, fungi, plants and invertebrates, signalling an untapped potential to connect and engage people in the appreciation and conservation of the natural world. Our findings suggest that, despite limitations, automated content analysis can be a rapid and readily accessed source of data on how and where people interact with nature, and a large-scale method for assessing cultural ecosystem services across countries and cultures.

Introduction

Cultural ecosystem services, the non-material benefits that people derive from interacting with nature, are recognised as important to people across a diversity of societies and cultures [1, 2]. Despite their value to a large number of people, cultural ecosystem services remain challenging to quantify, differ among users, and can change over time [3, 4]. As a result, such benefits have often been neglected in assessments and decision-making around management and use of natural areas [5].

Geotagged photographs are a useful source of information on the places visited by people and the relative magnitude of visitation and, when combined with content analysis, can be a rapid and readily accessed source of data on how people interact with nature (i.e., cultural ecosystem services, CES) across an otherwise unmapped area [6, 7]. Social media has been used to examine human-nature interactions, but has mostly been done on a small spatial scale, often limited to a single protected area. Recent developments in the use of automated image recognition for content analysis of CES have opened the way for rapid assessment of large amounts of visual data [7–10]. Known also as 'automated content analysis', this new technology combines advances in artificial intelligence (i.e., 'machine learning') with visual data available from social media and opens up the possibility of examining patterns of landscape use and nature-interactions across a diversity of people [11] and countries.

These new sources of data allow new questions to be asked about human interactions with nature, such as how culture and socioeconomic status influence the types of nature that people interact with and the types of places they are likely to visit to do so.

In remote parts of the world, such as the Arctic, there are few assessments of cultural ecosystem services and hardly any that examine how nature attracts the increasing number of visitors to this vast region [12]. Automated content analysis of geotagged photographs can fill some of these gaps by providing data on the species and ecosystems that people appreciate through photographs shared on social media. In the Arctic, general tourism and nature-based tourism have been synonymous historically [13, 14]. Visitors are attracted to the pristine scenery and symbolic qualities of the natural attractions in the Arctic [15]. Recent increases in tourism have been attributed to various factors including global climate change, which impels tourists to experience Arctic environments, wildlife, and indigenous cultures before they are gone [16, 17]. Like many other remote areas, the reasons for people visiting the Arctic are poorly mapped, and it is unclear to what extent CES plays a role in attracting a growing number of visitors in the Arctic [12].

In this paper we examined a comprehensive suite of Flickr photographs from across the Arctic to uncover what types of nature experiences are important to people inside and outside protected areas in eight countries. Flickr is especially popular among those sharing nature photos [10, 18], and is therefore appropriate for identifying appreciation of the unique ecosystems, wildlife and the scenic qualities of Arctic landscapes. We separate between biotic and abiotic cultural services similar to the Common International Classification of Ecosystem Services [19], with a particular emphasis on what Richards and Friess [20] call nature appreciation, i.e. 'photographs that primarily depicted animals or plants' and which are also of high relevance to protected areas management. We use an automated content analysis tool, Google's Cloud Vision, to identify the objects and concepts within a set of over 800 000 photographs from the Arctic. We identify the types of nature most photographed and map hotspots of different categories of human-nature interactions across the Arctic including abiotic and biotic nature. Finally, we examine whether people with high nature orientation (i.e. those who take many photos of wildlife) interact with nature in different places compared to other visitors.

Methods

We extracted photos taken in the Arctic between 2004 and 2017 that were posted to Flickr and ran the photos through an automated algorithm, Google's Cloud Vision, to identify the contents. We then manually categorised the component of the natural environment (ecosystem or geosystem service [3]) associated with each automatically selected label. We examined the proportions of photographs depicting nature, the number of Flickr users photographing nature, and the prevalence of the different components of the natural environment depicted inside and outside protected areas.

Unless otherwise stated, all analysis was conducted in R version 3.4.2 [21] using the 'tidyverse' [22], 'sf' [23], 'mgcv' [24], and 'raster' [25] packages. All spatial data was projected to EPSG 102017 (North Pole Azimuthal Lambert equal area) for analysis. The R code associated with the project is available at doi:10.18710/DUANRP.

Extraction of data from Flickr

We first extracted geotagged and publicly shared photo metadata for over 2 million photos from Flickr (www.flickr.com) for the region north of latitude 60°N. Photo metadata included location and date that each photo was taken, user id (key coded by Flickr), image URL, Flickr- and user- generated image tags, and user-generated image title. Data was extracted from the Flickr API (https://www.flickr.com/services/api/) on 4 December 2017. Due to an issue with the data download we re-extracted photos for Iceland (bounded by −27° to −12° longitude and 62° to 68° latitude) on 11 January 2018. We used the R package 'flickRgeo' [26] which provides an R wrapper for the Flickr API.

We define our study region, 'Arctic', as the region within the Arctic Council boundaries [27] north of latitude 60°N. We excluded photos from the extracted dataset that were taken outside this study region. We also excluded photos that were missing urls or geotag coordinates, had null coordinates (0,0) and photos taken prior to January 1 2004, or after December 31 2017. We excluded photos by users who have uploaded only 1 or 2 photos within the study region as they are likely to represent people who are just trialling Flickr by uploading a random photo rather than a photo representing a genuine ecosystem service. These 'test users' account for approximately 36% of users in the Flickr dataset but just 0.95% of photos (appendix S2). We further excluded geotagged points where the photo image was unavailable at the time the content analysis was performed (this can occur when users take down photos or make them private), leaving a total of 805 684 geotagged photos with metadata from 13 596 unique users. We then identified the country each photo was taken in and whether or not the photo was taken within a protected area using the geotagged coordinates. Protected area borders were drawn from CAFF [28] and supplemented with data from Protected Planet [29].

Content analysis of photos

We performed content analysis on all 805 684 photos for the Arctic with the Google Cloud Vision API (V1; https://cloud.google.com/vision/) during April and May 2018. Google Cloud Vision uses machine learning to detect broad sets of categories within an image. Google documentation states the Cloud Vision API can 'identify objects, locations, activities, animal species, products, and more' (https://cloud.google.com/vision/docs/labels), though no further information on the algorithm is publicly available. We extracted the url from the metadata for each Flickr photo then passed this url to the Cloud Vision API Label Detection function with the R package 'RoogleVision' [30]. The API returns a list of labels for each photo and a score associated with each.

The Google Cloud Vision algorithm assigned 6660 unique labels to the photos in our dataset. These included descriptors of physical objects (e.g. 'snow', 'tree'), activities (e.g. 'hiking', 'dog sledding') and concepts (e.g. 'wilderness', 'calm'). We limited our analysis to any labels with score of 0.6 or higher up to a maximum of 20 labels for each photo (median number of labels per photo = 17).

Ecosystem service mapping

We manually categorised any label associated with 4 or more photos (5056 labels) as one of 2 broad cultural services as distinguished by CICES v5.1. [19] (abiotic nature, biotic nature) or non- service (pet, non-nature). We further categorised biotic cultural services associated with nature appreciation into sub-categories 'wildlife', 'bird' or 'plant'. The category assigned to each label and the frequency of each label is listed in appendix S1 which is available online at stacks.iop.org/ERC/2/075001/mmedia. A photograph can represent multiple categories. For example, a photograph of grazing reindeer might be categorised as both 'abiotic' (label = snow) and 'biotic' (label = herd). Photographs may represent additional ecosystem services not discussed in this manuscript (e.g. recreation).

We generated spatial maps for each ecosystem service at 10 km resolution by summing across time, for each 10 km cell i, the number of photo-unit-days (equation (1)) i.e. the number of unique Flickr users n that took at least one photo in a cell i in a given day t, summed across all (5114) days.

$$pud_{cell\ i} = \sum_{t\ 1}^{t\ max} n\ users \tag{1}$$

Comparing Cloud Vision to manual content analysis

We performed manual content analysis on a random sample of photos to evaluate how well Google Cloud Vision performs at identifying ecosystem services from photos. We examined approximately 300 randomly selected photos from each of ten regions: Alaska, Canada, Iceland, Greenland, Norway, Sweden, Finland, Svalbard, Russia, and marine areas. Content analysis validation was performed by four independent people, and 10% of the sample (315 photos) was allocated to all four coders to test intercoder reliability. 182 photos were no longer available on Flickr at the time of validation and content analysis was performed on a total of 2645 unique photos. Excluding unavailable photos, Person A coded 876 photos, Person B 1341 photos, Person C 865 photos, and Person D 862 photos. Coders were instructed to classify the components of the photo according to types of CES (two general categories, abiotic nature, biotic nature, and more specific wildlife, bird, plant) and whether components not related to cultural services were present using the classification scheme in appendix S1. Photos could be assigned multiple words where multiple components are present in the photograph. For instance, a person standing in front of a waterfall would be classified as 'abiotic nature' (the waterfall) and 'non-nature' (the person). Coders were instructed to classify humans as 'non-nature', and domestic animals as 'pets' not 'wildlife'. We calculated the iota coefficient for the interrater agreement across all ecosystem service groups and the four manual coders, and the percent agreement and Fleiss's kappa across the four manual coders for each ecosystem service group individually. We calculated the percent agreement and Cohen's kappa between the manual coding and Cloud Vision treating all four manual coders as a single coder and examining each ecosystem service group individually. Intercoder reliability was calculated using R package 'irr' [31].

Comparing nature photos inside and outside protected areas

Many protected areas in the Arctic are situated in remote or inaccessible areas. In order to determine whether photographs of nature are more likely to be taken inside protected areas while accounting for this bias in accessibility, we first divided the landscape into 10 km diameter hexagonal grid cells. We then calculated the footprint in each cell in each year for a given season by allocating a cell a value of 1 if a photograph of biotic nature had been taken within that season in that cell in a given year, and 0 if not. We excluded cells that fell within the Russian Federation and marine areas due to sparse coverage of these regions by Flickr data. We modelled the footprint inside and outside protected areas, controlling for accessibility. The models took the form of binomial generalized additive models with logit link of footprint as a response variable. Model covariates included the country the photograph was taken in, whether any part of the cell overlapped a protected area, and five accessibility metrics modelled as linear responses (log of distance to airports, log distance to ports, log distance to populated areas, log distance to road, the square root of the length of road within a grid cell). In order to account for spatial autocorrelation we included a fitted thin-plate spline on the variables latitude and longitude of the cell centroid. Intercept, slope and confidence intervals were estimated by a restricted maximum likelihood (REML) estimator and methods for large datasets ('bam' function in R's mgcv package). We chose these accessibility variables after examining candidate variables for correlation. The location of airports, ports and populated areas, and country boundaries were extracted from Natural Earth (www.naturalearthdata.com) using the R package 'rnaturalearth' [32]. Roads were extracted from Global Roads Inventory Project [33]. As the presence of snow and ice limits access to outdoor areas in winter, we ran one model for summer and one for winter. We defined the months of May to October as 'summer' and November through April (of the following year) as 'winter' (e.g. 'winter 2016' includes the months November and December in 2016 and January through April in 2017). The summer model had 91 482 cells with no photos, and 5822 with photos. The winter model had 94 560 cells with no photos, and 2744 with photos.

Landscape use by wildlife photographers

We performed a linear mixed effects analysis of the relationship between landscape use and a user's propensity to take photos of wildlife. We included the importance of wildlife to each user (the square root of the proportion of each user's photos that were of wildlife, defined as non-domestic animal or bird) whether the photo represents wildlife or not, and whether the photo was taken in or near (within 1km) a protected area or not as fixed effects. We included user ID as a random intercept. The response variable was log of the distance from a road (in m) that each photo was taken. We included in the model all the photos taken in the Arctic in summer (432 105 photos; 9545 users). P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Visual inspection of residual plots showed residuals were strongly clustered at zero but evenly distributed either side of zero, with fat-tailed distribution on the Q–Q plot and some heteroscedasticity was present (appendix S6), suggesting that a Gaussian log distribution may not be the best fit to the data. However, given the large sample size and acceptable confidence intervals we find this model to be appropriate for our purposes [34].

Results

Diversity of nature depicted

Almost all (91.1%) of users took one or more photograph of biotic nature (table 1). Over half of Arctic photos (53.2%, 428 286 photos) depicted some form of biotic nature. Flickr may disproportionally attract people interested in nature photography and it is likely that data from the platform is not representative of the full range of photographs that might be taken by people [35, 36]. That bias is useful here as we are particularly interested in examining human-nature interactions. Though vertebrates were present in only a small proportion of photos (7.4% and 3.2% respectively), interactions with vertebrates show a disproportionate importance to Arctic users, with 41.0% of people photographing wildlife and 23.1% photographing birds (table 1). Pets are also photographed by many people. 13.5% of people took a photo of a domestic animal, predominantly dogs, though these account for just 1.5% of the photos in the dataset (table 1). Photographs depict a wide variety of taxa, ecosystem types, and abiotic features (figure 1, appendix S1).

Spatial distribution of nature photographs

The vast majority of nature photographs were taken on land, and in Iceland (figure 2). Norway and Alaska also have large numbers of nature photographs. Most photographs of Arctic nature were taken outside protected areas (figure 3). Across the Arctic, 21.9% of photos taken on land depicting biotic nature were taken inside terrestrial protected areas and 20.9% taken on water were taken inside marine protected areas (figure 3(a)). These proportions are higher than expected given the area held in protected areas (15.5% of Arctic land and 3.4% of marine areas).

Flickr users were more likely to photograph nature inside than outside protected areas areas in summer, but no more or less likely to photograph nature inside than outside protected areas in winter after accounting for the difference in accessibility of protected and unprotected areas. The protected area term of the summer model was significant (coef 0.822, std error 0.045, z-value $= 18.237$, $p < 2 \times 10^{-16}$). The summer model explained 49.9% of the deviance (R^2 0.465, AIC 22134). Removing the protected area term increased the AIC and significantly reduced the deviance explained (ΔAIC $= 329$, Δdf $= -0.957\,68$, Δdeviance -329.41, Pr(>Chi) $< 2 \times 10^{-16}$). The winter model explained 54.9% of the deviance (adjusted $R^2 = 0.451$, AIC 11337). The protected area term of the winter model was significant ($p = 0.012\,40$) in the full model, but removing the protected area term did not

Table 1. Summary of Arctic photographs. A photograph can represent multiple categories of ecosystem services. For example, a photograph of grazing reindeer might be categorised as both 'abiotic' (label = snow) and 'biotic' (label = herd). Number of users = 13 596; number of photos = 805 684, number of labels = 12 061 189. Photographs may represent additional ecosystem services not presented here (e.g. recreation).

Ecosystem/service represented	Number of photos depicting	Percent of photos depicting	Number of users photo-graphing	Percent of users photo-graphing	Highest scoring labels (percent of all photos)[a]	Most frequently used labels (percent of all photos)[b]
Non-nature	750 501	93.1	13 538	99.6	sky (29.74), winter (12.77), vehicle (4.02), sunset (1.96), wind wave (1.81), vacation (1.68), window (1.64), sunlight (1.58), water transportation (1.45), urban area (1.07).	sky (55.76), cloud (24.93), winter (14.15), horizon (12.21), freezing (10.12), atmosphere (9.44), reflection (8.97), phenomenon (7.54), morning (7.48), arctic (7.31).
Abiotic nature	506 875	62.9	12 512	92.0	water (18.29), snow (7.94), terrain (4.89), wave (3.68), water resources (3.53), waterfall (3.31), waterway (2.74), rock (2.29), valley (1.63), soil (1.53).	water (31.31), mountain (27.87), highland (21.35), sea (17.28), fell (15.63), hill (15.48), snow (14.05), loch (12.99), rock (11.81), mountain range (11.57).
Biotic nature	428 286	53.2	12 384	91.1	tree (11.52), wilderness (11.15), wildlife (3.29), tundra (3.22), wood (3.07), nature (2.89), landscape (2.35), woody plant (1.57), rural area (1.34), vegetation (1.2).	tree (20.65), landscape (16.57), wilderness (12.27), grass (12.11), nature (8.89), tundra (8.26), plant (7.63), ecor-egion (6.59), rural area (6.51), grassland (6.14).
Wildlife	59 285	7.36	5577	41.0	wildlife (3.77), organism (0.78), snout (0.76), fauna (0.69), vertebrate (0.56), mammal (0.19), herd (0.14), whales dolphins and porpoises (0.1), terrestrial animal (0.07), marine mammal (0.06).	fauna (4.27), wildlife (3.81), organism (1.57), snout (1.37), mammal (1.23), vertebrate (0.58), terrestrial animal (0.53), marine mammal (0.47), deer (0.36), herd (0.32).
Bird	26 010	3.23	3143	23.1	wing (0.59), waterfowl (0.51), water bird (0.51), seabird (0.5), bird (0.31), shorebird (0.19), eagle (0.07), spar-row (0.06), perching bird (0.05), wren (0.04).	bird (2.8), beak (2.29), seabird (1.18), water bird (1.1), charadriiformes (0.68), ducks geese and swans (0.64), wing (0.6), duck (0.6), shorebird (0.54), water-fowl (0.53).
Domestic animal	11 736	1.46	1829	13.5	dog like mammal (0.57), Siberian husky (0.17), small to medium sized cats (0.15), wolfdog (0.08), street dog (0.06), puppy (0.06), retriever (0.04), tabby cat (0.03), Greenland dog (0.02), dog breed group (0.02).	dog like mammal (1.18), dog (1.04), dog breed group (0.64), dog breed (0.63), Siberian husky (0.25), cat like mammal (0.19), Greenland dog (0.19), small to med-ium sized cats (0.19), cat (0.19), seppala Siberian sled-dog (0.13).

[a] For each photo, we first selected the label that was assigned the highest score by Cloud Vision of the labels associated with that ecosystem service (if any). We then summed the frequency of those labels across all photos. Only the 10 most frequent of those labels are presented here. A full list can be found in appendix S1.

[b] The number of times each label occurs, across all photos, counting any instance a label is associated with a photograph. Only the 10 most frequent labels are presented here. A full list can be found in appendix S1.

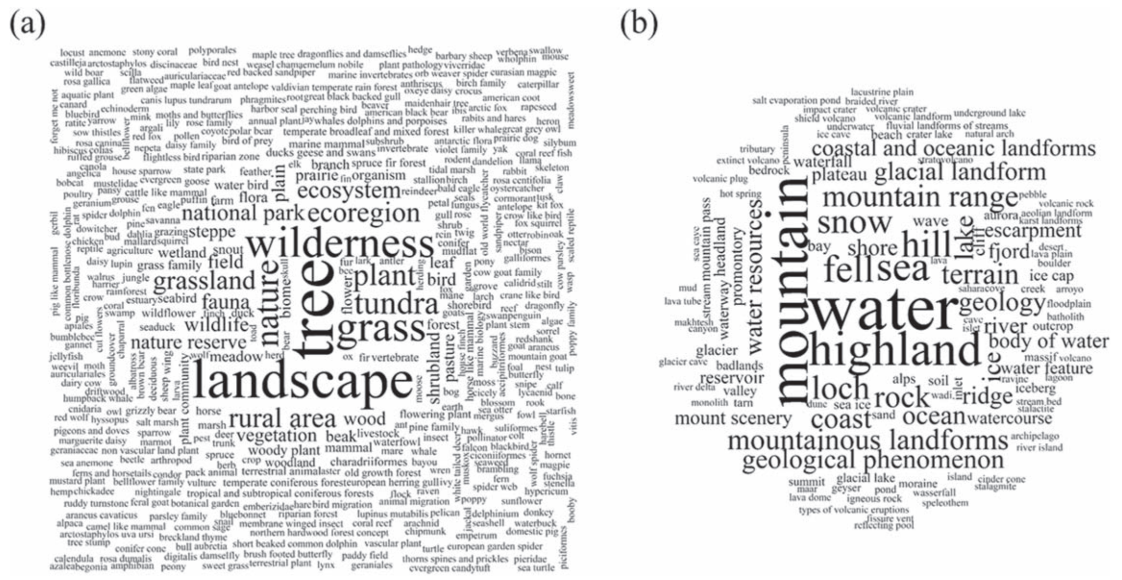

Figure 1. Labels that were assigned to Flickr photographs in the Arctic by Google's Cloud Vision algorithm and found to represent (a) biotic ecosystem services, (b) abiotic ecosystem services. The number of photos that were labelled with each word is indicated by its relative size. A wide variety of taxa, ecosystems, and abiotic land formations have been photographed.

change the deviance explained (ΔAIC5, Δdf $= -0.994\,56$, Δdeviance $= -10.706$, Pr($>$Chi) $= 0.019\,93$). Visual examination of model residuals did not reveal any residual spatial autocorrelation in any of the models. Model coefficients, plots of model residuals and partial plots of variables are included in appendix S3.

Validation of content analysis

Intercoder reliability between manual coders was acceptable (appendix S4; photos $= 315$, coders $= 4$, variables $= 7$, iota $= 0.612$). Percentage agreement was high for most ecosystem service categories, though coders differed widely in their classification of photos as containing plants (ranging from 2%–51% of photos in the validation dataset). Person A reported they classified a photo as containing plant only where the plant was the focus of the photo. At the other extreme, Person D reported they classified a photo as containing a plant anytime a vegetation component was visible, including bare winter branches or snow-covered trees, distant forests and vegetation traces on mountains. This highlights the difficulties of validating a 'black box' machine learning algorithm. When the plant category was removed, intercoder reliability increased (variables $= 6$, iota $= 0.709$).

Intercoder reliability between manual coding and the Cloud Vision coding of ecosystem services was good, with percent agreement ranging from 69.9% to 98.7% (appendix S4; iota $= 0.511$, subjects $= 3944$, coders $= 2$, variables $= 6$, excluding non-nature). Manual coders reported substantially lower proportions of non-nature components than the Cloud Vision algorithm. The reason for this is unclear, but may be an artefact of the wide variety of labels assigned to photos by the Cloud Vision algorithm. For instance, a photo of a seabird with an ocean background (appendix S5) was labelled by Cloud Vision with words that we classified as non-ecosystem service related—(black and white, monochrome photography, monochrome) in addition to the words indicating ecosystem services (bird, seabird, fauna, wave, water). Cloud Vision coded a lower proportion of photos as containing biotic nature components than did the manual coders, which may reflect the difficulty of the algorithm in detecting distant biota or biotic traces (see appendix S5). Cloud Vision also reported higher proportions of plants in photos than the manual coders (36.0% versus 22.5%) though this difference is likely in part a reflection of the wide variability in the ways plants were coded.

Landscape use by wildlife photographers

Users took photographs further from roads inside protected areas (figure 4), and users who took a higher proportion of wildlife photographs travelled further from roads. Photographs representing wildlife were taken slightly further from roads than those that didn't depict wildlife.

The full model (log likelihood $-905\,032$) was significantly more likely than those without terms for importance of wildlife to each user (log likelihood $-905\,045$, Chi-sq 27.443, df $= 1$, Pr($>$Chisq) $= 1.618 \times 10^{-7}$) and whether photo was taken inside a protected area (log likelihood $-913\,847$, Chi-sq 17 630, df $= 1$, Pr($>$Chisq) $< 2.2 \times 10^{16}$). The full model was also significantly more likely than an intercept-only model (log

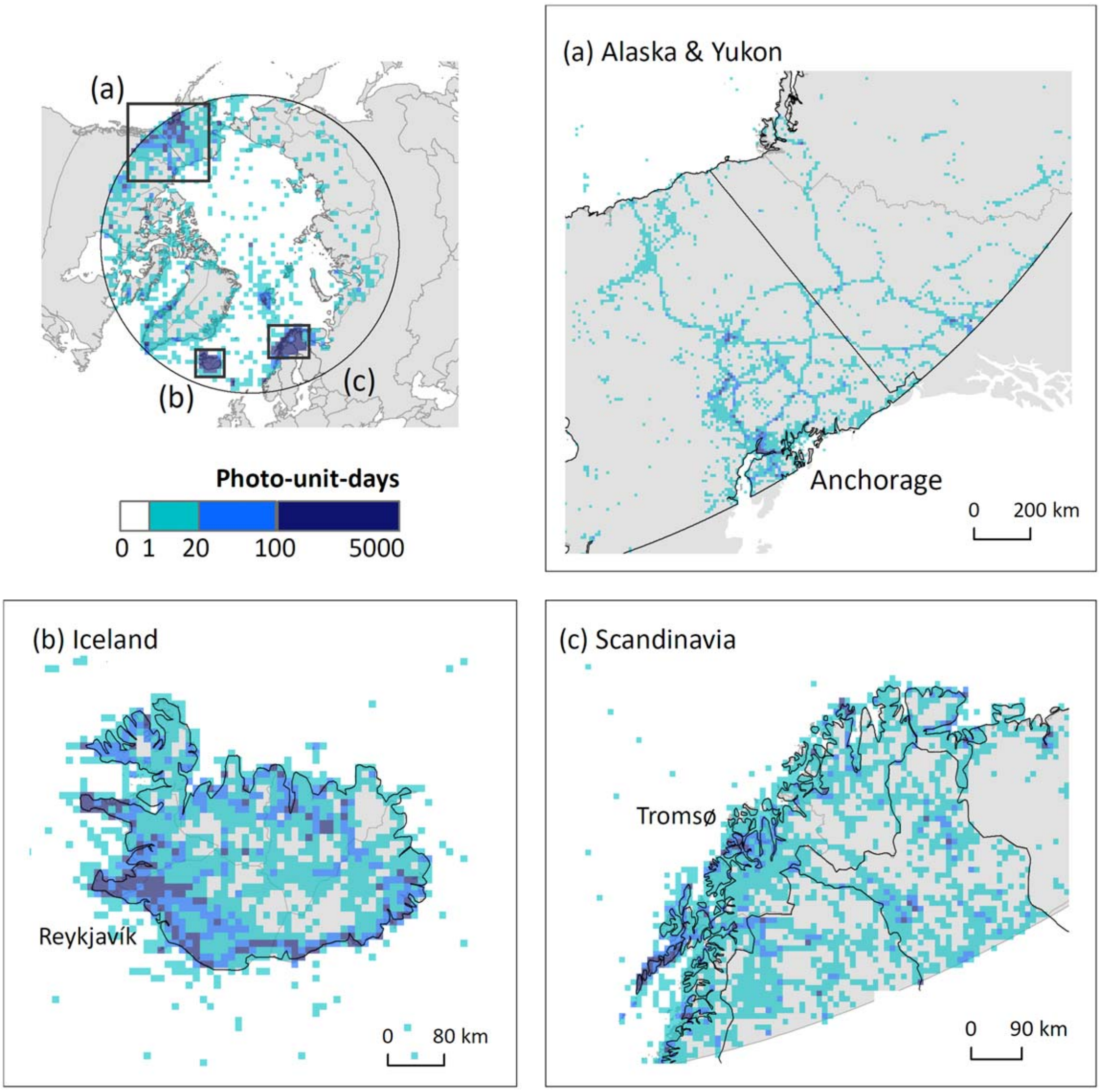

Figure 2. Spatial locations where people photograph biotic nature across the Arctic. In Alaska and Yukon (a) interactions with nature are strongly mediated by road access as they are in (b) Iceland, where most photos are taken a day's drive from the international airport and along scenic coastal routes (c). In Northern Scandinavia, the popular tourist destinations of Lofoten Islands, Tromsø, and North Cape (the most northerly point in mainland Europe) hold highly photographed components of nature.

likelihood −913 871, Chi-sq 17 679, df = 3, Pr(>Chisq) $< 2.2 \times 10^{16}$). The full model was marginally significantly more likely than the model without the term whether a photo represented wildlife (log likelihood −905 035, Chi-sq 6.0485, df = 1, Pr(>Chisq) = 0.013 92). Figure 4 shows the marginal effects of the full model. Model diagnostics are included in appendix S6.

Discussion

The attraction of Arctic nature as an untamed and wild frontier region has a long history, but the high appreciation of nature and the unique wildlife of these cold climates are increasingly actualised by last chance tourism to experience the rapidly disappearing species and ecosystems before they are gone [16, 37]. The opportunity to view and interact with wildlife is a primary motivation for Arctic tourists with 53% of tourists saying it was the main reason for visiting Svalbard [38]. We found over 40% of people took a photo of wildlife, although such photos make up only 7% of all photos (table 1). Many Arctic visitors have expressed disappointment that they did not see more wildlife [39]. This disparity between the desire to interact with wildlife and the opportunity to do so suggests an untapped market for ecotourism. Nature documentaries may contribute to unrealistic expectations that Arctic wildlife is plentiful and easy to see [40]. In many parts of the Arctic wildlife are now elusive and rare. Though this is in part an inherent quality of Arctic ecosystems, wildlife rarity has been exacerbated by historical over-exploitation and the rapidly emerging impacts of climate change

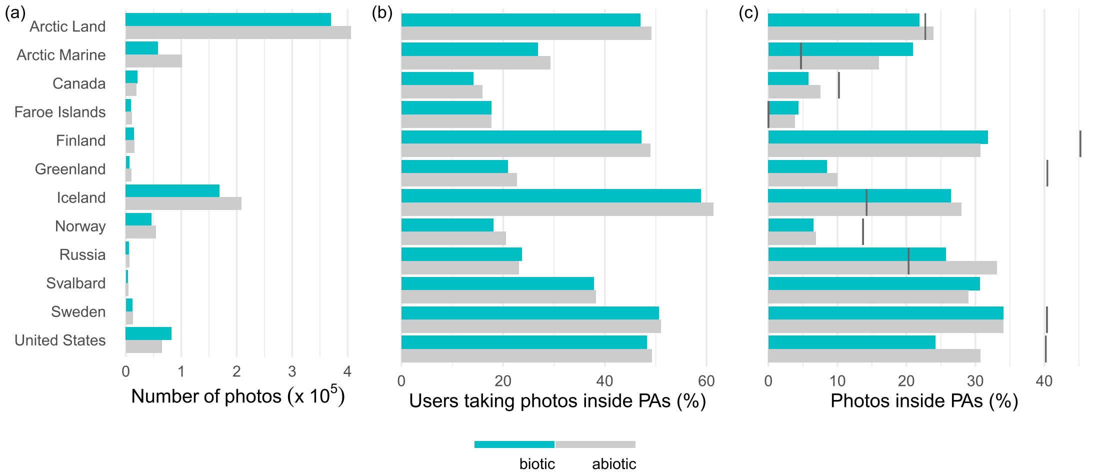

Figure 3. Regional numbers of Flickr photographs representing abiotic and biotic nature across the Arctic for the years 2004 to 2017. (a) Total number of photos representing abiotic and biotic ecosystem services (ES) in each region, with terrestrial Iceland the most photographed location, (b) Proportion (%) of users that took at least one photograph of each ES in each region that did so within a protected area during the study period. Very few Flickr users photographed within North American protected areas, though protected areas are important sites for nature interactions in Iceland and Svalbard. (c) Proportion (%) of photographs representing each ES that were taken within a protected area. Grey horizontal lines indicate the proportion of each region that is protected. In most of the Arctic, protected areas are less photographed than would be expected given the proportion of land they cover. In all three plots bars for each country represent terrestrial areas only. Marine areas and marine protected areas are represented by the bar labelled 'Arctic Marine'.

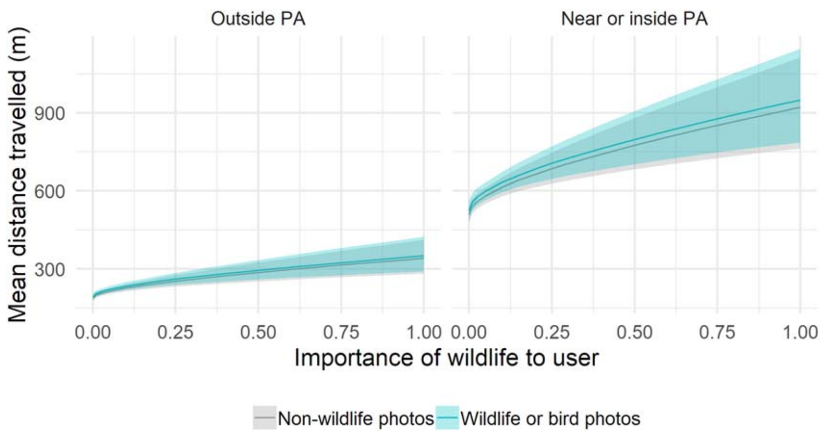

Figure 4. Marginal effects of user characteristics and protected areas on distance from a road at which a photograph is taken. Importance of wildlife to user is measured as the proportion of each user's photos that are of non-domestic wildlife or birds.

[41]. The loss of Arctic biodiversity, both common and iconic, is thus of concern not just to conservationists and Arctic people that rely on the ecosystem services they provide [42], but to people across the world.

Almost all users took one or more photographs of biotic nature and that over half of Arctic photos depicted some form of biotic nature, confirming the importance of cultural ecosystem services to Arctic people and visitors. Rates of nature photography vary around the world, with typically 50%–60% of users photographing nature, and 10%–20% of photographs represent an animal or plant. Case studies in the Pyrenees in Spain found 62% of tourists took a nature photo [43], and a study in Singapore found 20% of photos were of nature (animal or plant) [9] whereas 50.3% of visitors to Mulde Basin in Germany took pictures of nature [8]. Our results are similar to Retka *et al* [44] which found 47.4% appreciating landscapes in a marine protected area in Brazil, while 11.9% took pictures of plants and animals.

Our results indicate that areas outside protected areas are important sites for contact with nature, with almost 80% of nature photographs taken outside Arctic protected areas. These places provide opportunities for nature connection, cultural ecosystem service provision [6] and ecotourism [45], both organised and opportunistic [46–48]. Places outside protected areas tend to be closer to where people live, and provide more regular and more equitable opportunities for people to interact with and experience nature [49]. Accessibility, whether by road, sea or air, is an important determinant of where people, including tourists, go, and of protected area use [50]. Many Arctic protected areas are sited in locations that are difficult to get to, particularly in winter.

For instance, Greenland is home to the world's largest terrestrial protected area, *Kalaallit Nunaanni nuna eqqissisimatitaq* (Northeast Greenland National Park; IUCN Cat II) which covers over a quarter of the landmass of Greenland. Sea ice and lack of roads or nearby airports severely restricts access to this region, and it is rarely visited by tourists. After controlling for the inaccessibility of protected areas, we find that Flickr users are slightly more likely to photograph nature inside than outside protected areas. Moreover, our models show that CES drive patterns of land use, as wildlife photographers are slightly more likely to travel further into wilderness areas, especially in protected areas (figure 4).

Social media data can provide both broad scale (regional, global) and local scale (park and protected area) patterns of cultural ecosystem services and can therefore be an important tool for managing the increasing number of visitors to the Arctic. Our results also allow for more fine-scale identification of areas where the footprint on Arctic ecosystems and biodiversity are greatest, what types of nature different people are interacting with, and indicate target groups and sites where management might be needed. To aid decision-making we have made the datasets freely available for download at doi:10.18710/DUANRP. Additional CES such as recreation, hunting, gathering and social were able to be identified from the photographs by the method we used, and are included in the datasets we have made available. The maps we have made available allow hotspots of different types of CES activity to be identified.

Limitations of social media data and automated content analysis for cultural ecosystem service mapping

Flickr data is limited in arctic areas of the Russian Federation and it is likely that our dataset substantially underestimates cultural ecosystem services in this region. Large numbers of tourists, mostly domestic, visit the Russian Arctic and nature is as an important a part of tourism in Russia as elsewhere [51].

Our qualitative tests of Cloud Vision in early 2018 indicated that while the algorithm can accurately detect many images of human activities, ecosystems, species and taxa, it often fails to detect or misidentifies wildlife that is distant, against a complex background, and wildlife traces such as animal tracks or breaching or breathing whales (appendix S5). Where identification of wildlife to the species level using passively sourced photographs is the goal, it may be more may be more appropriate to develop and train bespoke image classification algorithms using machine learning [52, 53].

Public research access to large datasets of photographs is currently in a state of flux. Following the consolidation of the main social media platforms for geotagged photographs, Flickr is now one of the few sources of free and publicly available geotagged visual data [10]. Flickr changed ownership in 2018 but as of early 2019 data was still accessible through the APIs. However, users are now limited to uploads of 1000 photos on a free account (previously unlimited), which may change the types of photos that users upload and frequency at which they refresh images. In the year between content analysis and manual validation 6.4% of photos in our dataset became unavailable for public viewing. It is unclear whether this is due to users leaving the service, replacing photos on their account with newer photos, or removing public access to their photographs. These dynamics present challenges for the ethical use, analysis, bias correction, interpretation (particularly of trends across time), and archiving of social media data [10, 54].

Conclusion

The recent development of readily accessible and cost-effective AI, combined with the availability of social media data sets, opens exciting opportunities for the analysis and quantification of cultural ecosystems services across large areas and large number of people [8, 9]. Despite the limitations challenging the use of social media data in remote areas, our approach and similar methodologies can undoubtedly support the ongoing efforts to understand cultural ecosystem services and to integrate this knowledge into governance decision-making and environmental accounting [55]. Our analyses do not represent a full assessment of cultural ecosystem services in the Arctic, as we chose to focus more specifically on nature appreciation, but they do emphasize the importance of nature experiences to many. Flickr is popular among those sharing nature photographs, and is perhaps less relevant for what Richards and Tunçer [9] refer to as social recreation, and for identifying activities such as trekking, fishing and hunting. Further research should combine different social media to assess a broader set of cultural services relating to both biotic and abiotic nature.

Acknowledgments

This study was funded by FRAM—High North Research Centre for Climate and the Environment through the Flagship MIKON (Project RConnected) and the Arctic Belmont Forum 'Arctic Observing and Research for Sustainability' (Project CONNECT). The Norwegian collaboration was financed by Norwegian Research Council grant 247474.

Data availability

The R code used in this analysis, raster maps of ecosystem services, and tables listing the labels and ecosystem services identified in each photograph are publicly and freely available for download at doi:10.18710/DUANRP. Photographs underpinning this analysis are available at www.flickr.com.

ORCID iDs

Claire A Runge https://orcid.org/0000-0003-3913-8560

References

[1] Milcu A I, Hanspach J, Abson D and Fischer J 2013 Cultural ecosystem services: a literature review and prospects for future research *Ecol. Soc.* **18** 44
[2] IPBES 2019 *Ch 2.3 Status and Trends—Nature's Contributions to People in Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (Bonn, Germany: Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services)
[3] Small N, Munday M and Durance I 2017 The challenge of valuing ecosystem services that have no material benefits *Glob. Environ. Change* **44** 57–67
[4] Daniel T C *et al* 2012 Contributions of cultural services to the ecosystem services agenda *Proc. Natl Acad. Sci.* **109** 8812–9
[5] Chan K M A, Satterfield T and Goldstein J 2012 Rethinking ecosystem services to better address and navigate cultural values *Ecol. Econ.* **74** 8–18
[6] Oteros-Rozas E, Martín-López B, Fagerholm N, Bieling C and Plieninger T 2017 Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites *Ecol. Indic.* **94** 74–86
[7] Ghermandi A and Sinclair M 2019 Passive crowdsourcing of social media in environmental research: a systematic map *Glob. Environ. Change* **55** 36–47
[8] Lee H, Seo B, Koellner T and Lautenbach S 2019 Mapping cultural ecosystem services 2.0—potential and shortcomings from unlabeled crowd sourced images *Ecol. Indic.* **96** 505–15
[9] Richards D R and Tunçer B 2018 Using image recognition to automate assessment of cultural ecosystem services from social media photographs *Ecosyst. Serv.* **31** 318–25
[10] Toivonen T, Heikinheimo V, Fink C, Hausmann A, Hiippala T, Järv O, Tenkanen H and Di Minin E 2019 Social media data for conservation science: a methodological overview *Biol. Conserv.* **233** 298–315
[11] Song X P, Richards D R and Tan P Y 2020 Using social media user attributes to understand human–environment interactions at urban parks *Sci. Rep.* **10** 1–11
[12] Malinauskaite L, Cook D, Davíðsdóttir B, Ögmundardóttir H and Roman J 2019 Ecosystem services in the Arctic: a thematic review *Ecosyst. Serv.* **36** 100898
[13] Fredman P and Tyrväinen L 2010 Frontiers in nature-based tourism *Scand. J. Hosp. Tour.* **10** 177–89
[14] Hall M and Boyd S 2005 Nature-based tourism in peripheral areas: introduction *Nature-based Tourism in Peripheral Areas: Development or Disaster?* ed M Hall and S Boyd (Clevedon: Channel View Publications) pp 3–20
[15] Ferdamalstofa—Icelandic Tourist Board 2016 *International Visitors in Iceland Visitor Survey Summer 2016* https://www.ferdamalastofa.is/en/recearch-and-statistics/visitor-surveys
[16] Lemelin H, Dawson J, Stewart E J, Maher P and Lueck M 2010 Last-chance tourism: the boom, doom, and gloom of visiting vanishing destinations *Curr. Issues Tour.* **13** 477–93
[17] Müller D K, Lundmark L and Lemelin R H 2013 Introduction: new issues in polar tourism *New Issues in Polar Tourism: Communities, Environments, Politics* ed D K Müller, L Lundmark and R H Lemelin (Dordrecht: Springer Netherlands) pp 1–17
[18] Muñoz L, Hausner V H, Runge C A, Brown G and Daigle R M 2020 Using crowdsourced spatial data from Flickr versus PPGIS for understanding nature's contribution to people in Southern Norway *People Nat.* **2** 437–49
[19] Haines-Young R and Potschin M 2018 *Common International Classification of Ecosystem Services (CICES) V5.1 and Guidance on the Application of the Revised Structure* 53 https://cices.eu
[20] Richards D R and Friess D A 2015 A rapid indicator of cultural ecosystem service usage at a fine spatial scale: content analysis of social media photographs *Ecol. Indic.* **53** 187–95
[21] R Core Team 2017 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria https://www.R-project.org/
[22] Wickam H 2017 tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1 https://CRAN.R-project.org/package=tidyverse
[23] Pebesma E 2018 Simple features for R: standardized support for spatial vector data *R J.* (https://doi.org/10.32614/RJ-2018-009)
[24] Wood S N 2011 Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models *J. R. Stat. Soc.* B **73** 3–36
[25] Hijmans R J 2017 raster: Geographic Data Analysis and Modeling. R package version 3.5-2
[26] Daigle R and Dunnington D 2018 *flickRgeotag v0.0.1* (https://doi.org/10.5281/zenodo.1314717)
[27] Arctic Assessment and Monitoring Programme (AMAP) 2017 Boundary for Arctic Assessment and Monitoring Programme (AMAP) working group of the Arctic Council http://geo.abds.is/geonetwork/srv/eng/catalog.search#/metadata/9BB31CE7-9AE2-453D-9234-0BAEB5C8A33C. Accessed 27/10/2017.
[28] Conservation of Arctic Flora and Fauna (CAFF) 2017 Arctic Protected Areas—2017 http://geo.abds.is/geonetwork/srv/eng/catalog.search#/metadata/2e56ee1f-50a9-4983-88f4-edaa8588950d.Accessed 27/10/2017.
[29] UNEP-WCMC and IUCN 2018 Protected Planet: The World Database of Protected Areas www.protectedplanet.net.Accessed 10/5/2018.
[30] Teschner F 2018 RoogleVision: Access to Google's Cloud Vision API for Image Recognition, OCR and Labeling. R package version 0.0.1. https://github.com/cloudyr/RoogleVision

[31] Gamer M, Lemon J, Fellows I and Singh P 2019 *irr: Various Coefficients of Interrater Reliability and Agreement R package version 0.84.1.* https://CRAN.R-project.org/package=irr
[32] South A 2017 *rnaturalearth: World Map Data from Natural Earth. v0.1.0* https://CRAN.R-project.org/package=rnaturalearth.
[33] Meijer J R, Huijbregts M A J, Schotten K C G J and Schipper A M 2018 Global patterns of current and future road infrastructure *Environ. Res. Lett.* **13** 064006
[34] Lumley T, Diehr P, Emerson S and Chen L 2002 The importance of the normality assumption in large public health data sets *Annu. Rev. Public Health* **23** 151–69
[35] Tenkanen H, Di Minin E, Heikinheimo V, Hausmann A, Herbst M, Kajala L and Toivonen T 2017 Instagram, flickr, or twitter: assessing the usability of social media data for visitor monitoring in protected areas *Sci. Rep.* **7** 17615
[36] van Zanten B T, Berkel D B V, Meentemeyer R K, Smith J W, Tieskens K F and Verburg P H 2016 Continental-scale quantification of landscape values using social media data *Proc. Natl Acad. Sci.* **113** 12974–9
[37] Sisneros-Kidd A M, Monz C, Hausner V, Schmidt J and Clark D 2019 Nature-based tourism, resource dependence, and resilience of Arctic communities: framing complex issues in a changing environment *J. Sustain. Tour.* **27** 1259–76
[38] Visit Svalbard 2018 *Visitor Survey (Gjesteunders)kelse) Svalbard 2018* https://en.visitsvalbard.com/visitor-information/Visit-Svalbard-Insights/visitor-survey
[39] Johnston M E 2006 Impacts of global environmental change on tourism in the polar regions *Tourism and Global Environmental Change* ed Stefan Gössling and Michael Hall C (London: Routledge) (https://doi.org/10.4324/9780203011911)
[40] Wearing S L, McDonald M, Ankor J and Schweinsberg S 2015 The nature of aesthetics: how consumer culture has changed our national parks *Tourism Rev. Int.* **19** 225–33
[41] Conservation of Arctic Flora and Fauna (CAFF) 2013 *Arctic Biodiversity Assessment 2013: Synthesis* Arctic Council https://www.caff.is/assessment-series/arctic-biodiversity-assessment/232-arctic-biodiversity-assessment-2013-synthesis.
[42] Fauchald P, Hausner V, Schmidt J and Clark D 2017 Transitions of social-ecological subsistence systems in the Arctic *Int. J. Commons* **11** 275–329
[43] Donaire J A, Camprubí R and Galí N 2014 Tourist clusters from Flickr travel photography *Tour. Manag. Perspect.* **11** 26–33
[44] Retka J, Jepson P, Ladle R J, Malhado A C M, Vieira F A S, Normande I C, Souza C N, Bragagnolo C and Correia R A 2019 Assessing cultural ecosystem services of a large marine protected area through social media photographs *Ocean Coast. Manag.* **176** 40–8
[45] Weaver D B and Lawton L J 2007 Twenty years on: the state of contemporary ecotourism research *Tour. Manag.* **28** 1168–79
[46] Cox D T C, Hudson H L, Shanahan D F, Fuller R A and Gaston K J 2017 The rarity of direct experiences of nature in an urban population *Landsc. Urban Plan.* **160** 79–84
[47] Lin B B, Fuller R A, Bush R, Gaston K J and Shanahan D F 2014 Opportunity or orientation? Who uses urban parks and why *PLoS One* **9** e87422
[48] Plieninger T, Dijks S, Oteros-Rozas E and Bieling C 2013 Assessing, mapping, and quantifying cultural ecosystem services at community level *Land Use Policy* **33** 118–29
[49] Martinez-Harms M J, Bryan B A, Wood S A, Fisher D M, Law E, Rhodes J R, Dobbs C, Biggs D and Wilson K A 2018 Inequality in access to cultural ecosystem services from protected areas in the Chilean biodiversity hotspot *Sci. Total Environ.* **636** 1128–38
[50] Levin N, Lechner A M and Brown G 2017 An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas *Appl. Geogr.* **79** 115–26
[51] Stepchenkova S and Morrison A M 2006 The destination image of Russia: from the online induced perspective *Tour. Manag.* **27** 943–56
[52] Norouzzadeh M S, Nguyen A, Kosmala M, Swanson A, Palmer M S, Packer C and Clune J 2018 Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning *Proc. Natl Acad. Sci.* **115** E5716–25
[53] Tabak M A *et al* 2019 Machine learning to classify animal species in camera trap images: applications in ecology *Methods Ecol. Evol.* **10** 585–90
[54] Lomborg S and Bechmann A 2014 Using APIs for data collection on social media *Inf. Soc.* **30** 256–65
[55] Guerry A D *et al* 2015 Natural capital and ecosystem services informing decisions: from promise to practice *Proc. Natl Acad. Sci.* **112** 7348–55