



## OPEN

## Using social media to quantify nature-based tourism and recreation

## SUBJECT AREAS:

SOCIOECONOMIC  
SCENARIOS

CONSERVATION

ENVIRONMENTAL ECONOMICS

ECOSYSTEM SERVICES

Spencer A. Wood<sup>1,2</sup>, Anne D. Guerry<sup>1,2</sup>, Jessica M. Silver<sup>1,2</sup> & Martin Lacayo<sup>2</sup><sup>1</sup>Natural Capital Project, School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA, <sup>2</sup>Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA, USA.

Scientists have traditionally studied recreation in nature by conducting surveys at entrances to major attractions such as national parks. This method is expensive and provides limited spatial and temporal coverage. A new source of information is available from online social media websites such as flickr. Here, we test whether this source of “big data” can be used to approximate visitation rates. We use the locations of photographs in flickr to estimate visitation rates at 836 recreational sites around the world, and use information from the profiles of the photographers to derive travelers’ origins. We compare these estimates to empirical data at each site and conclude that the crowd-sourced information can indeed serve as a reliable proxy for empirical visitation rates. This new approach offers opportunities to understand which elements of nature attract people to locations around the globe, and whether changes in ecosystems will alter visitation rates.

Received

1 July 2013

Accepted

2 October 2013

Published

17 October 2013

Correspondence and requests for materials should be addressed to S.A.W. (woodsp@stanford.edu)

Recreation and tourism are important components of many national and local economies and they contribute in innumerable ways to quality of life, sense of place, social connection, physical wellbeing, learning, and other intangibles. Information on patterns of recreation and tourism and the factors that influence behavior in these realms is typically collected using site-specific surveys or interviews. The recent emergence of social media creates exciting alternative possibilities to assess how people use and respond to nature and other cues for recreation and tourism. One problem, however, is that while they generate “big data”, it is often unclear how to tease meaning and useful information from social media. Here we assess the relationship between the locations of photographs from the image-sharing website flickr and empirically derived visitation rates at sites around the world. This is the first study to ground-truth the use of data from social media to predict visitation rates, freeing researchers from time- and labor-intensive surveys and revolutionizing the use of social media to understand where people recreate.

A key reason for studying patterns of recreation or tourism is the economic significance of this industry. The total contribution of travel and tourism to the world’s gross domestic product (GDP) in 2011 was approximately \$6 B USD (9% of GDP), with expected growth to \$10 B USD by 2022<sup>1</sup>. On more regional scales, recreation and tourism represent a significant fraction of many local economies. The economy of the Caribbean region, for example, is dominated by tourism, with 15% of GDP from tourism and 17% of the available workforce employed in the tourism sector<sup>2</sup>. In the US, English et al.<sup>3</sup> classify 156–338 nonmetropolitan counties in the US as “tourism dependent”, meaning 10% of income and 15% of jobs in these counties result from tourism. Of course, economic impacts are only one way of measuring the importance of recreation and tourism. These activities are critical contributors to diverse aspects of human wellbeing<sup>4</sup>. For example, outdoor recreation is a spiritual experience for many people<sup>5–8</sup>, and social interactions in nature contribute to building a sense of place<sup>9–12</sup>.

A major and growing portion of the global penchant for travel and recreation is “nature-based”, involving interactions with or appreciation of the natural environment<sup>13</sup>. For these types of activities, characteristics of the environment influence people’s decisions about where, when, and how to recreate. SCUBA divers, for example, select destinations based on the water clarity, water temperature, and diversity of marine life<sup>14,15</sup>. Bird-watchers are drawn to the best places to see target species<sup>16</sup>, which inevitably are places where natural systems support populations of desirable birds<sup>17</sup>.

Several previous studies have succeeded in quantifying the degree to which recreation depends on environmental attributes such as species richness<sup>18</sup>, the diversity of habitats<sup>19,18</sup>, precipitation<sup>20</sup>, and temperature<sup>21</sup>, as well as to other attributes such as infrastructure and cultural attractions<sup>22,23</sup>. However, because people’s motivations for how and where to recreate vary from place to place and depend on many factors including their origin and destination, each study must assemble new empirical data. Existing empirical data are often too coarse to illuminate the relationship between visitation and salient features of a particular location, and represent point



samples as opposed to landscape data. Empirical surveys are also expensive to conduct and are limited in coverage.

To avoid having to conduct costly and time-consuming surveys, researchers have long sought automated methods for collecting empirical data on visitation rates to multiple locations. Shoval and Isaacson<sup>24</sup> recorded the movement patterns of several tourists in Israel who volunteered to carry global positioning system (GPS) locators, an approach since repeated in other locations<sup>25</sup>. Pettersson and Zillinger<sup>26</sup> supplemented survey and GPS data with aerial images to estimate the number of people attending a sporting event in Sweden. While these techniques provide rich datasets, there remains a need for methods of estimating visitation rates across multiple spatio-temporal scales, in natural and built environments, and in both developed and developing countries. We explore a new approach, using the density of existing geolocated photographs posted to the online photo-sharing website flickr as local data that reveal where people recreate.

## Results

Since its launch in 2004, over 71 M flickr users have uploaded over six B photographs to the image-sharing website. Approximately 197 M of these photographs are geotagged, meaning the image has been assigned specific coordinates of latitude and longitude (Fig. 1). Since 2010, 40–50 M geotagged photographs have been uploaded annually. Most geotagged photographs on land are from locations in Europe (40%), North America (39%), and Asia (13%). Sixteen percent of all geotagged photographs are from marine and coastal environments. Most images are taken in the United States, United Kingdom, and France. On a per-area basis, Vatican City, Macau, and Gibraltar have the greatest number of geolocated pictures, while Antarctica and Chad have the lowest density of images.

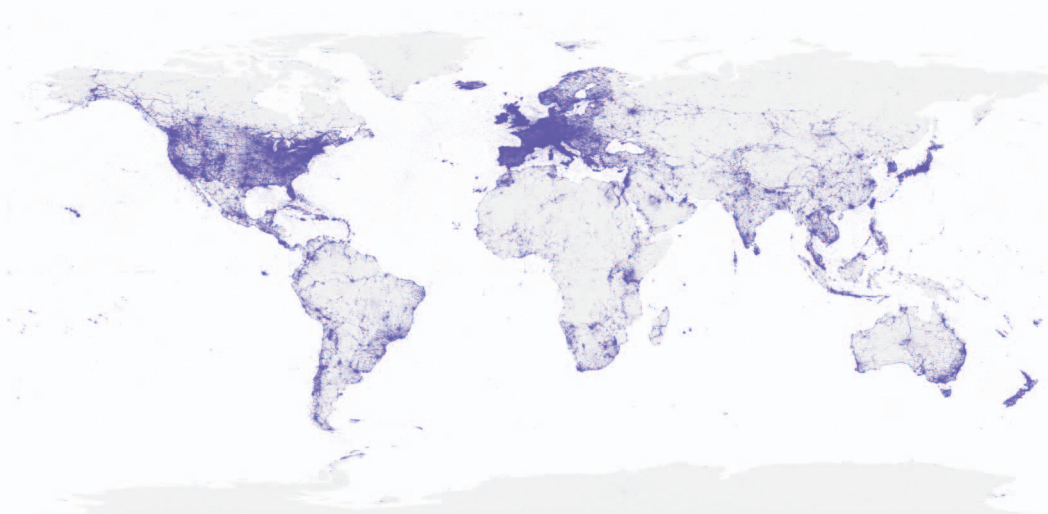
**Visitation rates.** Crowd-sourced information from flickr photographs corresponds well with empirical information about where people go (Fig. 2). Unsurprisingly, only a portion of visitors to any site take geotagged photographs and share them on flickr. This is observable in Figures 2–4 as the distance of each point from the grey 1:1 line representing equal densities of visitors and photographs. Although there is appreciable variation across all sites, there is a reliable statistical relationship between the number of people counted and the flickr-generated estimate of user-days (Fig. 2–4). The relationship

between the empirical estimates of mean annual visitor user-days ( $\overline{EUD}$ ) and those derived from photographs ( $\overline{PUD}$ ) is best described by a power function  $Y = Y_0X^\beta$ , where  $\beta = 0.698$  and  $Y_0 = 13.7$  ( $F_{1,831} = 388.34$ ,  $p < 0.001$ ,  $R^2 = 0.386$ ; Table 1). Absolute visitation rates are less variable within datasets comprising sites from smaller geographies or similar types of use (Fig. 2).

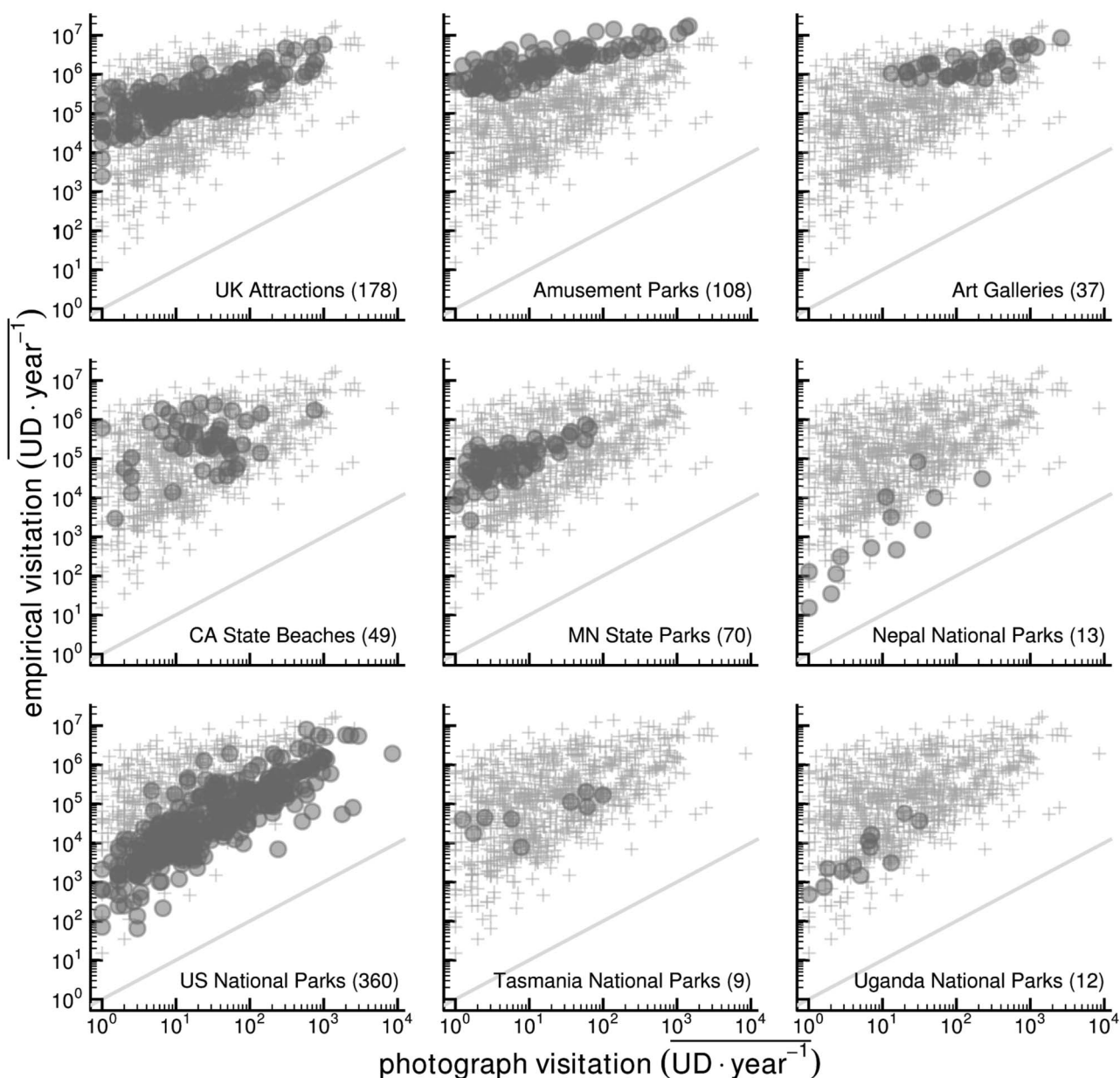
Type of attraction ( $A$ ) and income-level of the country ( $I$ ) do not change the observed relationship between empirical and photograph user-day estimates, but they do explain some of the variation across sites, both in total visitation rates and in the proportion of users who choose to take and upload photographs to flickr. Neither  $A$  nor  $I$  alter the slope of the scaling relationship, according to the results of the ANCOVA. Interactions between the covariate  $\ln \overline{PUD}$  and the two factors  $A$  and  $I$  all non-significant (results not shown). Therefore, we refit the model by removing interactions with  $\ln \overline{PUD}$  and assuming parallel slopes. All results presented here are for the simplified model. Predictors  $A$  and  $I$ , however, do change the intercept,  $Y_0$  (Fig. 3 and 4). Across the 836 sites in this study, cultural attractions have higher visitation rates than natural ones ( $F_{1,831} = 29.31$ ,  $p < 0.001$ ) and sites in high-income nations have more annual visitors than sites in low-income countries ( $F_{1,831} = 9.61$ ,  $p = 0.002$ ). Also,  $A$  and  $I$  interact, such that the higher visitation to cultural attractions occurs primarily at sites in low-income countries ( $F_{1,831} = 101.05$ ,  $p < 0.001$ ) while visitation rates in high-income nations are indistinguishable.

The lowest  $\overline{EUD}$  values were for Khaptad and Dhorpatan National Parks in Nepal ( $\overline{EUD} = 14$  and  $\overline{EUD} = 34$ , respectively). The next least visited site was Hamilton Grange National Memorial, in New York City, with an average of 64 user-days each year from 2008–2010. During this period, the facility was closed while the structure was moved to a new location two blocks away. The two most highly visited sites according to  $\overline{EUD}$  were the Magic Kingdom Theme Park in Florida (over 17 M user-days) and Disneyland in California (over 15 M user-days). The natural attraction with the highest annual visitation was Yellowstone National Park in the United States ( $\overline{EUD} = 5.9$  M user-days). Twenty-five of the 836 study sites had zero photos during the years for which we have  $EUD$  estimates.

**Visitor origins.** Data from photographs uploaded to flickr also serve as a good indicator of the country of origin for travelers (Fig. 5). Originating countries of incoming visitors surveyed at border



**Figure 1** | Locations of the approximately 197 M geotagged photographs uploaded to flickr from 2005–2012. Figure created using the maps package for R.



**Figure 2** | Average user-days per year based on photographs (x-axis) vs. empirical data (y-axis) at 836 sites worldwide. Panels display points from the individual datasets listed in Supplementary Table S1 (○) atop all data (+). Grey line is 1:1.

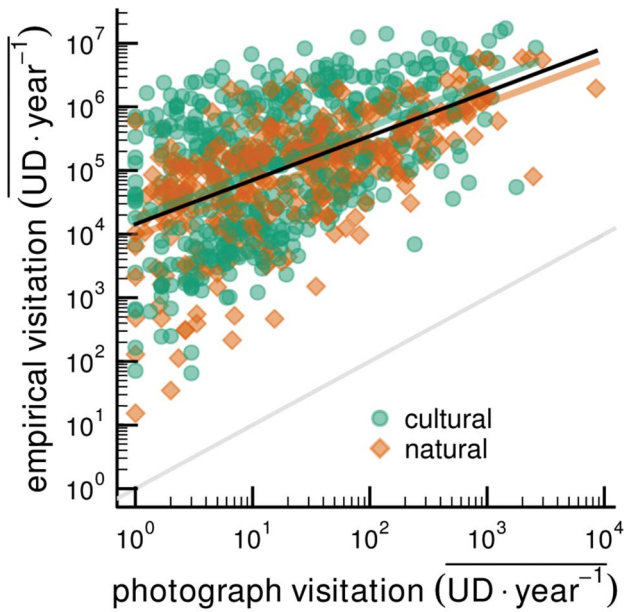
crossings (*EOC*) are related to the home countries reported in the profiles of flickr users who took photos within the same focal country (*POC*), measured as proportions of incoming people ( $\beta = 0.735$ ,  $F_{1,102} = 169.12$ ,  $p < 0.001$ ,  $R^2 = 0.620$ ). In our analysis, the flickr users traveling to five destination countries (*D*) originate from 55 nations from around the world. These nations represent a wide range of population sizes and income levels. A nonsignificant *POC* · *D* interaction term in the ANCOVA shows that the scaling relationship between *EOC* and *POC* does not differ across the five destination countries considered here. Furthermore, the intercept of the relationship between *EOC* and *POC* is also constant across all levels of *D*. Thus, we removed the *D* and *POC* · *D* terms from the model and present the results of a simplified ANCOVA in Table 2. In summary, the home location reported by flickr users who take and upload pictures in a particular country can predict the originating countries of travelers to that nation. Furthermore, the relationship between photo- and empirically-derived visitation rates is equal across the five destination countries examined.

**Visitation over time.** Visitation rates estimated from flickr images match expectations over time at sites selected as examples of cultural (Zuccotti Park and Black Rock Desert) and natural (Vermont) attractions. There is a single spike in *PUD* at Zuccotti Park that corresponds to the duration of the encampment there (Fig. 6a). Increasing numbers of photos begin to appear on flickr when the protests start on September 17, 2011 and drop abruptly when the park is closed on November 15, 2011. In the Black Rock Desert, there is an annual spike in the numbers of photos taken that corresponds to the three-week period surrounding the Burning Man festival each year (Fig. 6b). Similarly, there is a period of higher *PUD* in southern Vermont each October (Fig. 6c) during the prime month for viewing the fall foliage.

## Discussion

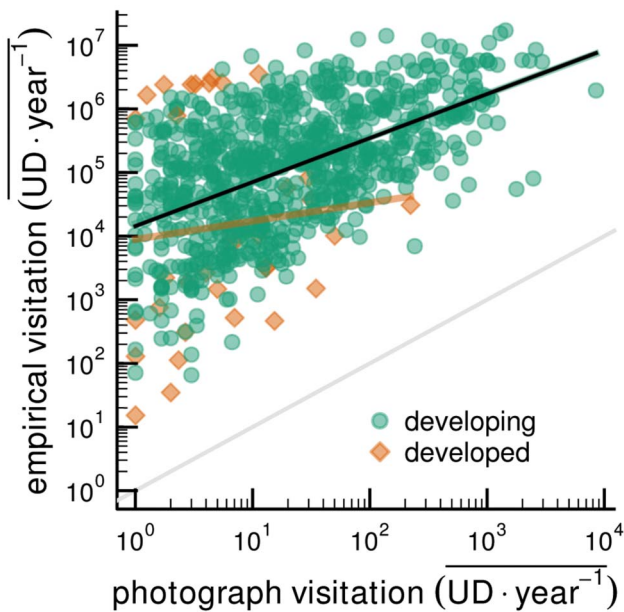
A lack of useful information about where people go during their leisure time has hindered progress toward understanding what draws and repels people to and from various recreation sites around the





**Figure 3** | Average photograph-based and empirical estimates of user-days per year. Sites are colored according to whether they are primarily a cultural (green circles [n = 498]) or natural (orange diamonds [n = 338]) attraction. Black line depicts the overall trend across all sites. Grey line is 1:1.

world. Here, we show that crowd-sourced information can offer new perspectives on this old problem, revolutionizing the way we study people and understand their choices. We hypothesized that pictures could indicate visitors, and furthermore, that photographs uploaded to an image-sharing website could record people’s choices and provide useful data worldwide. Our comparison of visitation data collected from 836 sites in 31 countries with data generated from geotagged photographs uploaded to flickr shows that the



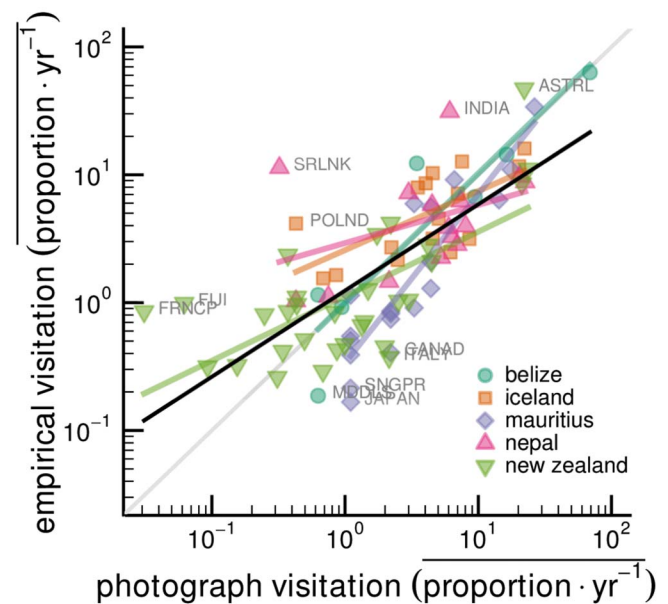
**Figure 4** | Average photograph-based and empirical estimates of user-days per year. Sites are colored according to whether they are located in a country with a low (green circles [n = 35]) or high (orange diamonds [n = 801]) Gross National Income. Black line depicts the overall trend across all sites. Grey line is 1:1.

**Table 1** | Results of an ANCOVA testing the effects of the covariate photo-user-days ( $\ln \overline{PUD}$ ) and the factors attraction type ( $A$ ) and destination income ( $I$ ) on empirically estimated visitation rates ( $\ln \overline{EUD}$ )

Factor	df	SS	MS	F	p
$\ln \overline{PUD}$	1	1322.02	1322.02	388.34	<0.001
$A$	1	99.77	99.77	29.306	<0.001
$I$	1	32.7	32.7	9.605	0.002
$A \cdot I$	1	344	344	101.05	<0.001
residuals	831	2828.95	3.4		

crowd-sourced data are indeed a suitable proxy for the more traditional time- and labor-intensive empirical estimates. This represents a significant advancement, as this new proxy measure of visitation can be applied almost anywhere: in developed and developing countries, data-poor and data-rich locations, urban areas and wilderness. Wherever people are taking and uploading pictures we can use that information to indicate their visit and learn from it.

There is considerable variation among sites in the concordance between empirical and photograph-based visitation rates. However, if used carefully, this relationship could inform future studies of marginal or absolute changes in visitation rates. Many questions can be answered through scenario-based assessments of relative changes in visitation across alternative management regimes. Managers could use geotagged photographs to explore marginal changes in visitation rates with changes in ecosystem health, site access, or tourism infrastructure under alternative future scenarios, irrespective of the variability in the relationship between  $\overline{EUD}$  and  $\overline{PUD}$ . Studies requiring absolute visitation rates should note that the slope of the relationship between empirical and photograph-based visitation rates is consistent across income levels,  $I$ , and attraction types,  $A$ , and it is only the height of the function that varies. In other



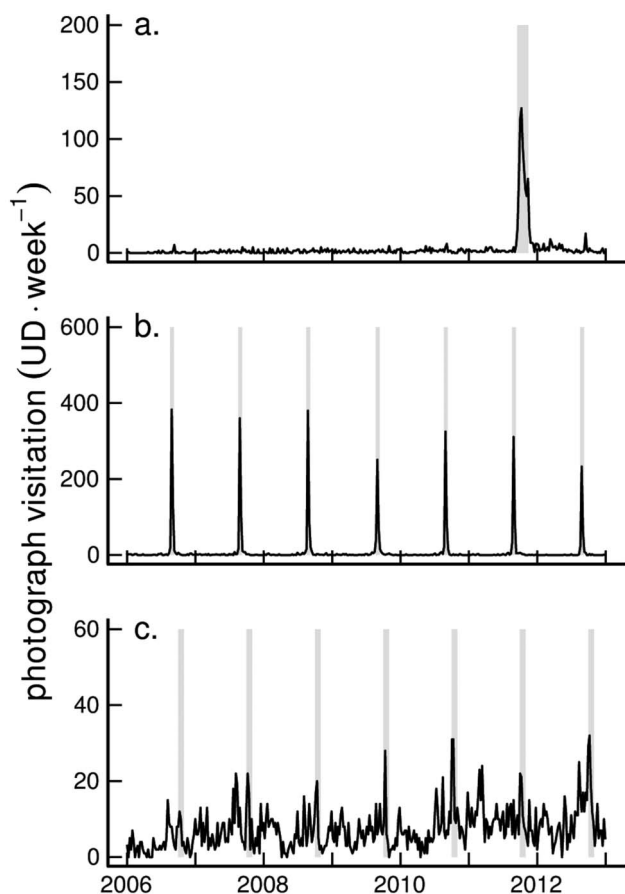
**Figure 5** | The average proportion and originating country of travelers who arrived to five destination countries each year, according to stated home locations of flickr users who took at least one photograph within the country (x-axis) and immigration data (y-axis). Names of outlying originating countries are abbreviated. Datasets are distinguished by colors and symbols and described in Supplementary Table S2. Black line depicts the overall trend across all sites. Grey line is 1:1.



**Table 2** | Results of an ANCOVA testing the relationship between the proportion of visitors from each origin countries, to five destination countries, based on photographs (the covariate  $\ln \overline{POC}$ ) and empirical estimates ( $\ln \overline{EOC}$ )

Factor	df	SS	MS	F	p
$\ln \overline{POC}$	1	56.231	56.231	169.12	<0.001
residuals	102	33.914	0.332		

words, globally, visitation rates derived from field data and images are consistently scaled with a slope of 0.70, but the absolute visitation rate varies with local socioeconomic conditions and attributes of the site. The precision of predictions will hinge on the similarity of the study sites in both geography and the types of attractions. Absolute visitation rates are less variable across sites within nations or from a single destination type, such as a state park or an art gallery (Fig. 2). Wherever some local visitation data are available, the height (and potentially the slope) of the relationship between  $\overline{EUD}$  and  $\overline{PUD}$  could be adjusted to suit the research questions and particular study region. We encourage researchers and practitioners to seek additional socioeconomic factors beyond income and attraction type that explain local variability in the absolute visitation rate. Despite the variability in  $\overline{EUD}$  and  $\overline{PUD}$ , this study presents promising new evidence that visitation rates can be quantified, both in relative and



**Figure 6** | Total photo-user-days each week from 2006–2012 in (a) Zuccotti Park, the site of the occupy protest, (b) Black Rock Desert, the site of the Burning Man festival, and (c) southern Vermont, an area popular for viewing autumn foliage. Grey shading indicates time periods from the start of the protest until people were barred from the park (a), three weeks spanning the annual week-long festival (b), or the month of October each year (c).

absolute terms, using geotagged photographs and a few easily-measured variables of a site.

Home countries given by users on flickr correspond with the home countries of travelers recorded at immigration entry points, making crowd-sourced data not only useful for estimating visitation rates, but also for understanding where visitors originate. Because the time and money that people spend traveling indicates how much they value the destination, these data on the origin and destination of recreators are enormously beneficial for economic methods for valuing recreation sites. One preferred approach for quantifying value is to use a “travel cost model” which uses the cost of travel to estimate peoples’ willingness to pay to recreate at particular sites<sup>27</sup>. Travel cost studies are often criticized for not accounting for people who visit multiple sites on a single trip away from home. Crowd-sourced visitation data can potentially address this issue since users often upload images throughout their journey.

The ability to estimate visitation rates without survey data allows for models that can anticipate changes in visitation in response to changes in ecosystems, relative to other types of change (in built features, social capital, etc.). Random utility models are one example of an economic technique for quantifying the marginal benefits of natural environments and other attributes. Typically, telephone surveys are conducted asking respondents where they live, which recreational sites they visit, and why. These individuals’ choices about which sites to visit reflect their preferences for certain characteristics of sites and the tradeoff between the costs (e.g., travel) and benefits (e.g., presence of wildlife) of the trip. Here, we show that the same data can be gathered using the locations of photographs and spatial data on the presence of features such as swimming beaches, cultural events, or other attractions. Enticing evidence that this approach is suitable for understanding people’s choices is demonstrated by the match of flickr photos to known temporal aggregations of people in Zuccotti Park, Black Rock Desert, and southern Vermont (Fig. 6). We offer these as initial examples and hope to spark further use of this approach to understand what draws and repels people to and from particular places.

Of course, this method is imperfect. There may be biases in who is taking digital photographs and uploading them to social media sites. Different recreational activities may be more or less suited to taking photographs. Surfers, for example, while likely possessing cameras and internet access, may prefer not to take photographs while surfing. Also, the perceived value of a trip may influence whether an individual takes or shares photographs, resulting in a bias against images from visitors who travel shorter distances from home. We observe, for example, that tourists visiting Nepal from neighboring Sri Lanka and India upload fewer photographs to flickr than predicted based on the overall trend (Fig. 5). Similarly, local visitors may be less inspired to take or share photographs of commonly-visited sites. While we find strong correlations between the crowd-sourced information and empirical data at attractions, such as national parks, we do not look at correlations between crowd-sourced information and visitation to more mundane locations, like shopping centers, that might be popular sites for recreation by local people. Further work is needed to explore the utility of this approach at locations that are not major attractions or landmarks. Other social media such as geotagged tweets might serve as more effective proxies for some types of recreation, particularly in urban areas.

New technologies and digital social media have begun making vast amounts of geolocated data available for a wide range of creative purposes, including art, targeted advertising, crime prevention, and scientific research. Some authors are rightfully raising concerns about the appropriate and ethical use of these data, and the potential for apophenia: to see patterns in “big data” where none actually exist<sup>28</sup>. In response to their calls for more critical assessments of digital data, this study vets a novel method for using geotagged photographs from flickr to provide sources of information for understanding where



people go. We conclude that crowd-sourced information can not only break the log-jam of expensive empirical data requirements for predicting and valuing how changes in the landscape alter recreation and tourism, but also can provide revolutionary information for understanding questions about where people recreate, in ways unimaginable before the existence of the internet and social media.

## Methods

While data from social media are fascinating, a critical question that will determine their utility for understanding where people go is: how well do they reflect on-the-ground visitor surveys and records? If we can establish reliable statistical relationships between image- and field-based records then we will have a powerful new tool for tracking how people interact with nature during recreation. To address this question, we first compare photograph- and field-based estimates of visitation rates at recreational sites around the world. Then, we assess how demographics of flickr users, specifically their home country, compare to survey responses by tourists entering through immigration checkpoints into five nations.

**Visitation rates.** We assembled data from nine independent empirical datasets that quantified visitation to 836 sites in 31 countries around the world (Supplementary Table S1). These empirical datasets represent a wide range of attractions, from amusement parks to national parks and from historic battlefields to art galleries. In each study, total empirical annual visitor user-days (*EUD*) were counted at a defined recreation site such as a park or museum. User-days are defined as one person spending a portion of one day within a site. The smallest and largest sites in the dataset are both in the USA, ranging from the 80 m<sup>2</sup> Thaddeus Kosciuszko US National Memorial in Pennsylvania to the Gates of the Arctic National Park and Preserve in Alaska (30,448 km<sup>2</sup>). All nine datasets were available publicly on the world wide web. For a study to be included in our analyses, we required measurements of total annual user-days to at least nine sites for two years between 2005–2011.

To create an alternative measure of visitation rate at the same sites, we used metadata associated with photos uploaded to the online photo-sharing website flickr. We used flickr's public API to download metadata for photographs that were taken within the bounds of each site. Along with the location of pictures, we also stored the photographer's identification number and the date that the image was taken. We used these two additional variables to convert the metadata, often including numerous photographs from the same person on the same day, into total annual user-days over the same time period as the *EUD* data were collected at each site. Thus, for photographs, user-days are defined as the total number of days, across all users, that each person took at least one photograph within each site (*PUD*).

To account for potential variation in the scaling relationship between *EUD* and *PUD*, we recorded two additional attributes for each of the 836 study sites. First, each site was characterized as either a cultural or natural attraction (dummy variable *A*). Cultural sites are those visited primarily for recreation in the built environment and socializing with other people (e.g., amusement parks) or learning about human culture (e.g., art galleries, historic battlefields) irrespective of the natural setting. Natural attractions, on the other hand, are sites where people go primarily to appreciate the flora, fauna, or natural setting (e.g., state beaches, recreation areas, botanical gardens). In reality, the two categories are not exclusive, so we selected the category capturing the primary reason that most people visit each site. Second, we characterized each site as either high- or low-income (dummy variable *I*) based on the nation's per-capita gross national income (GNI) published by the World Bank. We divided the World Bank's four income levels into two categories, low and high, with low-income countries having a per-capita GNI below \$4,035 USD in 2008.

We used analytical methods to assess the correspondence between *EUD* and *PUD* estimates. For all analyses, we aggregated repeated annual measures at each site into a single value of average annual *EUD* and *PUD* across all years that empirical data were collected. We fit a general linear model (GLM) in which  $\ln \overline{PUD}$ , *A* (cultural vs. natural), *I* (low vs. high), and all interactive effects were regressed against  $\ln \overline{EUD}$ . We used an analysis of covariance (ANCOVA) to test two main hypotheses: 1) In  $\overline{EUD}$  is related to  $\ln \overline{PUD}$  and 2) the slope of the relationship between the response  $\ln \overline{EUD}$  and covariate  $\ln \overline{PUD}$  is equal across levels of factors *A* and *I*. We evaluated hypothesis (2) based on the  $\ln \overline{PUD} \cdot A \cdot I$ ,  $\ln \overline{PUD} \cdot A$ , and  $\ln \overline{PUD} \cdot I$  interaction terms in the ANCOVA.

**Visitor origins.** We collected a second empirical dataset of origin countries of tourists arriving to five destination countries (see Supplementary Table S2). Empirical estimates of the proportion of visitors from each country (*EOC*) derive from passenger surveys conducted at entry points such as airports and overland border-crossings. We gathered values from public sources, found online, that reported the *EOC* per origin country for at least one calendar year between 2005–2010. For comparison, we assembled a dataset of the origination locations of users who uploaded photographs to flickr taken within the destination country during the same time that the empirical data were collected. flickr users have the option to list a "current location" in their profile. We assumed each user's current location was the origination of each trip and used this information to calculate the proportion of photographers originating from each country (hereafter, *POC*).

We then explored the relationship between the *EOC* and *POC* values using a GLM. First, we aggregated repeated annual *EOC* and *POC* estimates into a single value of

average annual proportion across all years that empirical data were collected. The GLM regressed  $\ln \overline{POC}$ , along with a categorical variable, *D*, representing the destination country, and an interaction term against  $\ln \overline{EOC}$ . We used an ANCOVA to test whether *ln EOC* is related to  $\ln \overline{POC}$  and whether the slope of this potential relationship varies across the five destination countries of the analysis.

**Visitation over time.** Finally, we present examples of how geotagged images can track the aggregate movement of people responding to cultural and environmental cues. As examples of cultural attractions, we examine visitation rates to Zuccotti Park, in New York City, the epicenter of the Occupy Movement of 2011, and to the Black Rock Desert, Nevada, home of the Burning Man festival. In the first, we expect higher *PUD* values from the start of the protest on September 17, 2011 until people were barred from the park on November 15, 2011. In the second, we expect annual peaks of visitation, measured as *PUD*, surrounding the annual week-long festival. To display the influence of a natural attribute on visitation, we plot seasonal patterns of *PUD* in southern Vermont where "leaf-peepers" are drawn by the colors of the foliage annually in October.

- World Travel and Tourism Council. *Travel and Tourism: Economic Impact* (2012). URL [http://www.wttc.org/site\\_media/uploads/downloads/world2012.pdf](http://www.wttc.org/site_media/uploads/downloads/world2012.pdf) Accessed March 2013.
- Carr, L. M. & Heyman, W. D. Jamaica bound? Marine resources and management at a crossroads in Antigua and Barbuda. *Geogr J* **175**, 17–38 (2009).
- English, D. B., Marcouiller, D. W. & Cordell, H. K. Tourism dependence in rural America: estimates and effects. *Soc Nat Resour* **13**, 185–202 (2000).
- Russell, R. *et al.* Humans and nature: how knowing and experiencing nature affect well-being. *Annu Rev Env Resour* **38**, in press (2013).
- Trainor, S. F. & Norgaard, R. B. Recreation fees in the context of wilderness values. *J Park Recreat Adm* **17**, 100–115 (1999).
- Sweatman, M. M. & Heintzman, P. The perceived impact of outdoor residential camp experience on the spirituality of youth. *World Leis J* **46**, 23–31 (2004).
- Heintzman, P. Nature-based recreation and spirituality: a complex relationship. *Leis Sci* **32**, 72–89 (2009).
- Heintzman, P. The spiritual dimension of campers' park experience: management implications. *Manag Leis* **17**, 291–310 (2012).
- Eisenhauer, B. W., Krannich, R. S. & Blahna, D. J. Attachments to special places on public lands: an analysis of activities, reason for attachments, and community connections. *Soc Nat Resour* **13**, 421–441 (2000).
- Stedman, R. C. Is it really just a social construction?: the contribution of the physical environment to sense of place. *Soc Nat Resour* **16**, 671–685 (2003).
- Ryan, R. L. Exploring the effects of environmental experience on attachment to urban natural areas. *Environ Behav* **37**, 3–42 (2005).
- Willox, A. C. *et al.* From this place and of this place: climate change, sense of place, and health in Nunatsiavut, Canada. *Soc Sci Med* **75**, 538–547 (2012).
- Balmford, A. *et al.* A global perspective on trends in nature-based tourism. *PLoS Biol* **7**, e1000144 (2009).
- Williams, I. D. & Polunin, N. V. Differences between protected and unprotected reefs of the western Caribbean in attributes preferred by dive tourists. *Environ Conserv* **27**, 382–391 (2000).
- Uyarra, M. C., Watkinson, A. R. & Ct, I. M. Managing dive tourism for the sustainable use of coral reefs: validating diver perceptions of attractive site features. *Environ Manage* **43**, 1–16 (2009).
- Naidoo, R. & Adamowicz, W. L. Biodiversity and nature-based tourism at forest reserves in Uganda. *Environ Dev Econ* **10**, 159–178 (2005).
- Puhakka, L., Salo, M. & Sksjrvi, I. E. Bird diversity, birdwatching tourism and conservation in Peru: a geographic analysis. *PLoS One* **6**, e26786 (2011).
- Loureiro, M. L., Macagno, G., Nunes, P. A. & Tol, R. Assessing the impact of biodiversity on tourism flows: an econometric model for tourist behaviour with implications for conservation policy. *J Environ Econ Policy* **1**, 174–194 (2012).
- Neuvonen, M., Pouta, E., Puustinen, J. & Sievnen, T. Visits to national parks: effects of park characteristics and spatial demand. *J Nat Conserv* **18**, 224–229 (2010).
- Loomis, J. B. & Richardson, R. B. An external validity test of intended behavior: comparing revealed preference and intended visitation in response to climate change. *J Environ Plann Man* **49**, 621–630 (2006).
- Richardson, R. & Loomis, J. B. Climate change and recreation benefits in an alpine national park. *J Leis Res* **37**, 307–320 (2005).
- Mills, A. S. & Westover, T. N. Structural differentiation: a determinant of park popularity. *Ann Tourism Res* **14**, 486–498 (1987).
- Hill, G. W. & Courtney, P. R. Demand analysis projections for recreational visits to countryside woodlands in Great Britain. *Forestry* **79**, 185–200 (2006).
- Shoval, N. & Isaacson, M. Application of tracking technologies to the study of pedestrian spatial behavior. *Prof Geogr* **58**, 172–183 (2006).
- McKercher, B., Shoval, N., Ng, E. & Birenboim, A. First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong. *Tourism Geogr* **14**, 147–161 (2012).
- Pettersson, R. & Zillinger, M. Time and space in event behaviour: tracking visitors by GPS. *Tourism Geogr* **13**, 1–20 (2011).
- Ward, F. A. & Loomis, J. B. The travel cost demand model as an environmental policy assessment tool: a review of literature. *Western J Agr Econ* **11**, 164–178 (1986).



28. boyd, d. & Crawford, K. Six provocations for big data. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society* DOI:10.2139/ssrn.1926431 (2011).

## Acknowledgments

The authors are grateful to Peter Kareiva, Katie Arkema, Eric Fischer, Greg Guanell, Steve Polasky, Mary Ruckelshaus, Rich Sharp, CK Kim, Kent Kovacs, Mike Papenfus, Derric Pennington, Jodie Toft, Gregg Verutes, and the Santa Fe Institute. Funding was provided by the Gordon and Betty Moore Foundation, Google, the Tides Foundation, World Wildlife Fund, and the Summit Foundation.

## Author contributions

S.A.W., A.D.G., J.M.S. and M.L. designed the study. S.A.W., J.M.S. and M.L. performed the analyses. S.A.W. and A.D.G. wrote the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Wood, S.A., Guerry, A.D., Silver, J.M. & Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* 3, 2976; DOI:10.1038/srep02976 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>