Open access • Journal Article • DOI:10.1890/15-0113.1

Where have all the people gone? Enhancing global conservation using night lights and social media — Source link

Noam Levin, Noam Levin, Salit Kark, David J. Crandall

Institutions: University of Queensland, Hebrew University of Jerusalem, Indiana University

Published on: 01 Dec 2015 - Ecological Applications (Ecological Society of America)

Topics: Environmental pollution

Ecological Applications, 25(8), 2015, pp. 2153–2167

Where have all the people gone? Enhancing global conservation using night lights and social media

NOAM LEVIN,[1,2,5] SALIT KARK,[3] AND DAVID CRANDALL[4]

[1]*Remote Sensing Lab, Department of Geography, Hebrew University of Jerusalem, Mount Scopus Campus, Jerusalem 91905 Israel*
[2]*School of Geography, Planning and Environmental Management, ARC Center of Excellence for Environmental Decisions, University of Queensland, Brisbane, Queensland 4072 Australia*
[3]*Biodiversity Research Group, ARC Center of Excellence for Environmental Decisions, School of Biological Sciences, University of Queensland, Brisbane, Queensland 4072 Australia*
[4]*School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405 USA*

Abstract. Conservation prioritization at large scales is complex, combining biological, environmental, and social factors. While conservation scientists now more often aim to incorporate human-related factors, a critical yet unquantified challenge remains: to identify which areas people use for recreation outside urban centers. To address this gap in applied ecology and conservation, we developed a novel approach for quantifying human presence beyond populated areas by combining social media "big data" and remote sensing tools. We used data from the Flickr photo-sharing website as a surrogate for identifying spatial variation in visitation globally, and complemented this estimate with spatially explicit information on stable night lights between 2004 and 2012, used as a proxy for identifying urban and industrial centers. Natural and seminatural areas attracting visitors were defined as areas both highly photographed and non-lit. The number of Flickr photographers within protected areas was found to be a reliable surrogate for estimating visitor numbers as confirmed by local authority censuses ($r = 0.8$). Half of all visitors' photos taken in protected areas originated from under 1% of all protected areas on Earth (250 of ~27 000). The most photographed protected areas globally included Yosemite and Yellowstone National Parks (USA), and the Lake and Peak Districts (UK). Factors explaining the spatial variation in protected areas Flickr photo coverage included their type (e.g., UNESCO World Heritage sites have higher visitation) and accessibility to roads and trails. Using this approach, we identified photography hotspots, which draw many visitors and are also unlit (i.e., are located outside urban centers), but currently remain largely unprotected, such as Brazil's Pantanal and Bolivia's Salar de Uyuni. The integrated big data approach developed here demonstrates the benefits of combining remote sensing sources and novel geo-tagged and crowd-sourced information from social media in future efforts to identify spatial conservation gaps and pressures in real time, and their spatial and temporal variation globally.

Key words: conservation; defense meteorological satellite program (DMSP); Flickr; geo-tagged data; light pollution; night lights; photography hotspots; protected areas; visitation.

INTRODUCTION

As human populations and mobility grow (Brockmann et al. 2006, Seto et al. 2012), our presence and impact in natural areas beyond urban centers is rapidly increasing worldwide. To help prioritize global conservation efforts effectively (Margules and Pressey 2000, Myers et al. 2000, Sanderson et al. 2002, Mittermeier et al. 2005, Brooks et al. 2006) and to better identify areas of potential conflict between humans and biodiversity (Balmford et al. 2001), there is a need to estimate not only where people live and work but also where humans are found in the more remote and natural areas, which are often the targets of protection efforts. However, spatial patterns of human recreational activity remain largely undetected outside populated areas and relatively few well-monitored protected areas (Eagles et al. 2002). This lack of accurate data about human presence in more natural and remote areas, which contain most of the conservation targets, handicaps conservation, management, policy, and investment decisions (Balmford et al. 2015). While government-based population censuses mostly estimate where people live and work (albeit infrequently and expensively; Balk et al. 2006), attempts to determine human impact outside of urban centers at large scales have predominantly had to rely on a range of indirect surrogates for human activity (Sanderson et al. 2002; Balmford et al. 2015) or remain at the local scale. Remote sensing provides a useful tool for mapping land-cover changes (e.g., deforestation; Hansen et al. 2013), land use (e.g., agricultural areas; Bastiaanssen et al. 2000), plant diversity (Levin et al. 2007, Asner and

Manuscript received 19 January 2015; revised 23 March 2015; accepted 24 March 2015. Corresponding Editor: R. L. Knight.
[5] E-mail: noamlevin@mail.huji.ac.il

Martin 2008), certain resource extraction activities (e.g., oil drilling operations leading to gas flaring; Elvidge et al. 2009), and has been used to estimate the isolation of protected areas based on the heterogeneity of vegetation cover (Seiferling et al. 2012). However, direct mapping of human mobility cannot be provided by remote sensing.

The proliferation in the use of GPS, smartphones, Web 2.0 (enabling people to collaborate and share information online), and social media provide new sources for spatial (Goodchild 2007) and other "big data." Big data are characterized by the three Vs: volume (large data sets), velocity (close to real-time data collection), and variety (gathered from many sources mostly without any quality assurance; Goodchild 2013). While social media data have many limitations and the use of social media tools is not uniformly distributed across the world (Li et al. 2013), it does provide new opportunities to collect information which was impossible to gather in the recent past at large scales, such as where people are distributed. Social media data can be very useful for multiple purposes (e.g., Song et al. 2010, Wood et al. 2013), including conservation planning and management.

Here, we develop and present an approach which allows us to quantify human presence beyond populated areas by combining (1) geo-tagged, online, user-generated information from the Flickr photo-sharing website (Crandall et al. 2009) and (2) global data of satellite-based night-light brightness, providing an objective spaceborne indicator of human activity strongly related to urbanization levels and human population density (Elvidge et al. 1997, Levin and Duke 2012, Bennie et al. 2014). In this framework, we tested the following hypotheses: the number of Flickr photographers per unit area is correlated with the number of visitors of a protected area; Flickr metrics (including the mean number of photos and photographers per area) are correlated with night lights at the country level (as they are both driven by population size and by economic activity); and Flickr metrics and night lights will have different spatial patterns, especially outside populated areas.

This global analysis of night lights and social media data allows us, for the first time, to differentiate human activity in densely populated urban areas, which have both high and temporally stable night-light brightness, from human visitation in remote unpopulated regions, which have low night-light brightness. While some human threats to protected areas (such as poaching, deforestation, agriculture, and invasive species) may not always be associated with night lights or with the number of Flickr photographs, our aim here is to investigate what insights can be gained from examining night lights and social media data generated from protected areas using a powerful big data social media tool. We examine and map hotspots of recreational visitation as reflected by the number of Flickr photographs and Flickr photographers per area globally for all countries, protected areas, and biodiversity hotspots at the global scale (Mittermeier et al. 2005), as well as for Earth's 661 Last of the Wild areas (Sanderson et al. 2002). We examine the major factors hypothesized to be related to spatial variation in the number of Flickr photographs. These include human population density, gross domestic product, type and size of protected area, distance to roads, road type, and several other factors. We identify areas that are both highly visited and protected and areas that received large numbers of visitors in recent years but remain unprotected.

Methods

In order to quantify human presence beyond populated areas, we combined two global-scale data sets, including night-light intensity and Flickr photos.

Data sources

Night-light brightness mapping.—We used version 4 of the global DMSP-OLS (Defense Meteorological Satellite Program Operational Linescan System) nighttime stable lights product time series (at 0.0083° resolution globally; *available online*).[6] The stable light product of DMSP-OLS contains the lights from cities, towns, and other sites with persistent lighting, including gas flares; ephemeral events such as fires are excluded from this product, as well as background noise. While the new VIIRS Day/Night Band (DNB) nighttime light imagery has better data quality than that of the DMSP-OLS (e.g., improved spatial resolution and radiometric quality; Miller et al. 2013), a stable lights product is not yet available, and night-light imagery from the VIIRS is only available from 2012 onward. Values in the DMSP-OLS data set were quantized on a 6-bit scale between 0 and 63, denoted here as L for nighttime light brightness. We calculated the mean night-light brightness values of the DMSP-OLS stable lights product between 2004 and 2012, so as to correspond with the time frame of our Flickr data. While this data set can be used to examine trends in night-light brightness (see Bennie et al. 2014, Gaston et al. 2015), our focus was on spatial patterns of night lights and not on their temporal aspects. We used two thresholds to define lit areas: (1) areas corresponding with major metropolitan areas ($L > 15$), and (2) areas corresponding with all stable light sources, including glow from metropolitan areas into nearby rural areas ($L > 0$).

Flickr data analysis.—Flickr is a popular photo-sharing website that currently includes more than 8 billion photos and 87 million registered users, with over 3.5 million photos uploaded daily (Jeffries 2013). In addition to the photo files themselves, Flickr allows

[6] http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

storing of metadata about the uploaded images, including where on Earth the photo was taken; some users specify this manually, while others use GPS-enabled devices (e.g., smartphones) that record geographic coordinates automatically. We collected data from the public Flickr API covering the entire globe, including all geotagged photos uploaded up to May 2013 (following the intensive computing retrieval approach presented by Crandall et al. [2009]). In total, we compiled ~187 million geo-tagged images taken by over one million different users. The vast majority of photos were taken since 2004 using smartphones or digital cameras.

Supporting data sets.—We used Conservation International's Bio\diversity Hotspots GIS layer (*available online*).[7] The Last of the Wild data came from a GIS layer made available by the Wildlife Conservation (*available online*).[8] Country population data was current as of 2012, for most countries (*available online*).[9] Gross domestic product data (as of 2005) came from the World Bank (*available online*).[10] We used Internet penetration rates (proportion of population with regular access to cyberspace, as of 2013/2014; Chinn and Fairlie, 2006) to examine its relationship with the number of Flickr photos available from each country, using data from Internet World Stats (*available online*).[11]

Spatial analysis

We used grid cells of 0.01° × 0.01° globally and calculated the mean stable light brightness per cell between 2004 and 2012. For each cell, we calculated the number of Flickr photos taken and the number of unique photographers (users). We aimed to examine factors driving the spatial patterns in the amount of Flickr data available across different locations worldwide as well as to examine regions of high conservation importance. Therefore, we calculated Flickr and night-light statistics for the following spatial units: all countries ($n = 248$), Earth's terrestrial biodiversity hotspots as defined by Conservation International ($n = 35$; Mittermeier et al. 2005), Earth's 661 Last of the Wild areas (Sanderson et al. 2002), all coastal areas globally (within 5 km of a coastline), and Earth's protected areas (World Database on Protected Areas; for details see Cantú-Salazar et al. 2013, Watson et al. 2014). To adjust for the spatial resolution of the DMSP-OLS night-lights (±3 km), we only considered protected areas larger than 25 km^2. Of the total 173 382 protected areas globally, 15% (26 693) were larger than 25 km^2, covering 98% of the total area protected, and these were included in our analysis.

We calculated the total, mean, and maximum number of Flickr photos and photographers within the different spatial units (countries, protected areas, Earth's biodiversity hotspots, Earth's Last of the Wild regions), as well as the proportion of the area of each of these spatial units containing Flickr data. We used Spearman rank correlations (r_S) and quantile regression (Koenker and Bassett 1978, Koenker and Machado 1999) as calculated in XLSTAT 2014 (Addinsoft, Paris, France) to quantify and test the significance of the relationship between Flickr photography and road density within protected areas. Road data (including dirt tracks and trails) was obtained from the OpenStreetMap (OSM) project (Haklay 2010) provided by Geofabrik (*available online*).[12] OSM started in London in 2004 and has since expanded globally, providing a free digital map of the world. Gröchenig et al. (2014) examined the state of OSM mapping, reporting that mapping progress was the most advanced in Europe and North America (notable countries include the United States, the Netherlands, and Japan). Within the United States, it was found that OSM data was often superior (i.e., more complete) than the U.S. Census TIGER/Line geodata for certain features, e.g., pedestrian paths and trails and cycle paths (Zielstra et al. 2013). OSM data was rasterized to a resolution of 0.001° and reprojected to an equal area projection. To examine the correspondence between Flickr photography and deforestation, we used forest-cover change data from Hansen et al. (2013), calculating the percentage of each protected area that was deforested between 2000 and 2012.

In order to identify currently unprotected sites attracting visitors, we calculated peak areas of photography outside urban and protected areas (after resampling our data to a spatial resolution of 0.05°, i.e., ~5 km). By averaging the percentage of lit areas of a protected area ($L > 15$) and the percentage of a protected area that had Flickr photographers, we calculated an index for identifying the most influenced protected areas by the combined effects of light pollution (which may result both from lit areas within protected areas and from skyglow from nearby cities; Hölker et al. 2010, Gaston et al. 2015) and visitors (as quantified by photos taken within protected areas), with 100% meaning that the entire protected area was both highly lit and photographed.

To investigate how the number of Flickr photos corresponds with actual visitation data collected by managers, we examined 12 countries across most continents, representing both OECD as well as third world countries (Argentina, Australia, Brazil, Canada,

[7] http://www.cepf.net/resources/hotspots/Pages/default.aspx

[8] http://sedac.ciesin.columbia.edu/data/collection/wildareas-v2

[9] http://unstats.un.org/unsd/demographic/products/socind/Dec.%202012/1a.xls

[10] http://data.worldbank.org/indicator/NY.GDP.MKTP.CD/countries?page=1

[11] http://www.internetworldstats.com/list2.htm

[12] http://www.geofabrik.de/data/download.html

Chile, Ecuador, Kenya, Nepal, New Zealand, South Africa, United Kingdom, United States) for which relatively high-quality visitor count data (mean number of annual visits) within protected areas (over 25 km^2, $n = 436$) were available (Appendix: Table A1).

Results

Global distribution of Flickr photos

Eleven percent of all total geo-tagged Flickr photos (20.6 million) were taken within protected areas worldwide (Appendix: Table A2). Nearly 5% of all geo-tagged Flickr photos taken between 2004 and 2013 (9.2 million) originated from non-lit coastal areas outside urban centers and 6.4% of all geo-tagged Flickr photos (11.9 million) were taken in non-lit protected areas (i.e., in remote protected areas where there is no light pollution; Appendix: Table A2). Within protected areas, 16.3% of all Flickr photos were taken in non-lit areas (3.3 of 20.6 million photos; Appendix: Table A2). The distribution of highly photographed protected areas was strikingly nonuniform and followed a power law distribution with an exponent of 1.7. Half of all Flickr photos taken in protected areas globally (10.3 million photos) originated from only 250 of the ~27 000 protected areas (<1%) worldwide. The total number of Flickr photos and photographers within protected areas were strongly correlated ($r = 0.95$), with ~6.5 photos per photographer. The mean annual number of visitors (per km^2) reported by local authorities and the mean number of Flickr photographers (normalized by area) per protected area were strongly correlated ($r = 0.80$, $n = 436$ protected areas; Fig. 1). At the country level ($n = 193$ countries), the percentage lit area ($L > 0$) was strongly correlated with the Flickr metrics of percentage of area with Flickr photographers and with the mean number of Flickr photographers ($r_S = 0.85$; Table 1). These correlations (with the lit area percentage) were weaker when examined for protected areas ($r_S = 0.63$ and 0.61, for percentage of area with Flickr photographers and with the mean number of Flickr photographers, respectively).

Population density and a country's gross domestic product (GDP) were both strongly associated with Internet penetration rates (proportion of population with regular access to cyberspace) and with the number of Flickr photos uploaded from each country (Table 1, Figs. 2 and 3). To compensate for between-country differences in Flickr photo numbers, we also identified for each of 40 selected countries the three most visited protected areas (Fig. 4; Appendix: Tables A3–A7), as well as the three most visited protected areas located in remote (i.e., non-lit) areas. These 40 countries cover nearly two-thirds (63%) of Earth's terrestrial area (including the world's 10 largest countries), 76% of the global human population (including the world's 10 most highly populated countries), and account for 85% of the global GDP, 74% of the world's lit area ($L > 0$), and 80% of the areas covered by Flickr photos.

Earth's most visited protected areas

At the global scale, the most photographed protected areas outside urban centers included Yosemite National Park (USA), the Lake District (UK), the Peak District (UK), the Grand Canyon (USA), and Yellowstone National Park (USA; see Fig. 5 for the spatial distributions of photographers globally and within Yellowstone, Grand Canyon, and Kruger [South Africa] national parks). Based on the total number of photos taken, we found that Yosemite National Park was ranked first in the United States, with a total of 175 169 photos (detailed results in Appendix: Table A3). Based on the mean number of photos per protected area (an estimate per unit area), a measure less affected by area size, the ranking of the most visited protected areas within countries changed. In Australia, for example, the most photographed protected area was the Great Barrier Reef (61 616 photos in total), however when taking into account the size of protected areas, by using the mean number of photos per grid cell, the most photographed protected area in Australia was Queensland's Noosa National Park (mean of 120 photos per grid cell; detailed results in Appendix: Table A4).

When we examined the maximum number of photos per grid cell within each of the protected areas, the rankings changed again, this time highlighting protected areas with highly visited focal attractions. In France, Mont Saint-Michel had over 30 000 photos in an area smaller than 1 km^2, while in Australia, Uluru-Kata Tjuta (formerly known as Ayers Rock) was ranked first with over 6000 geo-tagged photos taken within 1 km^2 (detailed results in Appendix: Table A5). Based on night-light data, we ranked the most photographed remote (i.e., non-lit) protected areas within countries. For example, in South Africa, Kruger National Park was ranked first, with 21 780 photos (detailed results in Appendix: Table A6). We found that based on the number of photos per unit area, on average, more people visit World Heritage Sites and Ramsar Sites worldwide compared with all other types of protected areas combined (Appendix: Table A8). This was also true for the number of Flickr photographers (Appendix: Table A8).

We found that within protected areas, visitors were distributed unevenly across the park and were mostly found in sections that were more accessible by roads and trails (see Fig. 5 for the uneven distribution of photos within protected areas). The tendency of people to visit (i.e., photograph) more accessible areas was quantified for the United States as a case study by examining all roads and tracks in protected areas within the contiguous United States ($n = 2699$). The percentage of area that had photos within protected areas was negatively correlated with distance from roads ($r_S = -0.338$, $P < 0.0001$; pseudo $R^2 = 0.59$ using a 0.9 quantile regression;

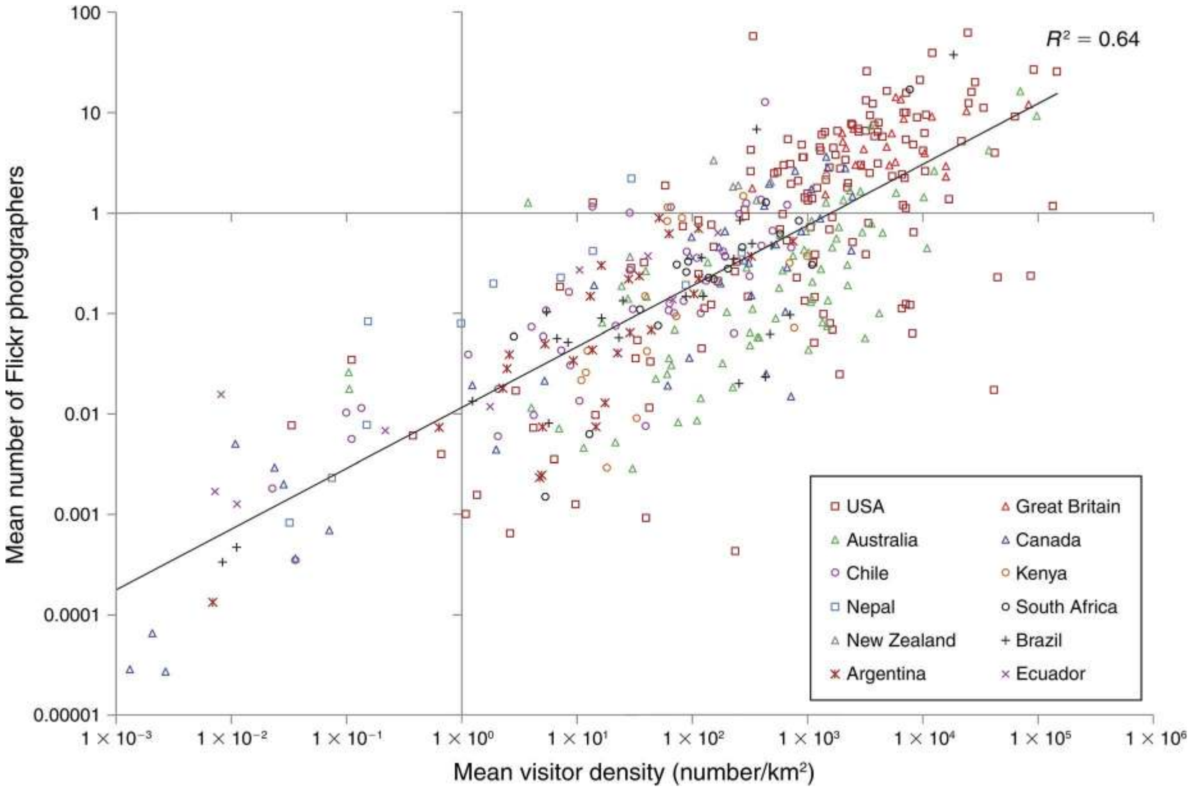

FIG. 1. The relationship between mean annual number of visitors (per km^2) per protected area and mean number of Flickr photographers (normalized by area) in those areas, for protected areas in 12 selected countries for which high-quality visitation data were available. The power relationship between these two variables was $y = 0.0115x^{0.6054}$, i.e., ~1600 annual visitors/km^2 for every Flickr photographer for a grid cell of 0.01°, and ~72 000 annual visitors/km^2 for 10 Flickr photographers for a grid cell of 0.01°. Note the log–log axis scales.

Fig. 6) and was positively correlated with the percentage of area covered by roads ($r_S = 0.354$, $P < 0.0001$). Nonetheless, 23.7% of all photos taken in protected areas within the contiguous United States were taken outside of the network of roads, tracks, and paths (as mapped by OpenStreetMap). In Yellowstone National Park, 31.8% of all photos were taken from primary roads, 24.5% of the photos were taken from walking paths, 14.5% of the photos were taken outside of the network of roads, tracks, and paths, and 10.8% of the photos were taken from unclassified roads (Fig. 7). In Grand Canyon National Park, 43.2% of all photos were taken from walking paths, 13.8% of the photos were taken from tracks, 12.7% of the photos were taken from tertiary roads, and 11.9% of the photos were taken outside of the network of roads, tracks, and paths (Fig. 7). Examining all protected areas, as well as only those protected areas located within the tropical and subtropical moist broadleaf forest biome, we found no correlation between deforestation rates within protected areas and their visitation statistics.

Worldwide, only seven protected areas (over 25 km^2 in size) scored the maximum possible value of 100% human impact (entire area both lit and photographed). These included Ma On Shan (Hong Kong), Central Catchment (Singapore), Valle del Lombro (Italy), the Sonian Forest (Belgium), and three parks in the Netherlands (Nh-stichting Gooisch, Zuid-Kennemerland, and Meijendel en Berkheide; detailed results in Appendix: Table A7). We found that the smaller the protected area, the more of its area tends to be photographed ($R^2 = 0.34$, $P < 0.0001$), and it is likely under greater human pressures overall.

Biodiversity hotspots, Last of the Wild, and unprotected areas

While Earth's 35 biodiversity hotspots (Mittermeier et al. 2005) cover less than 16% of the Earth's land surface, nearly a quarter of all Flickr photos were captured within these biodiversity-rich and threatened hotspots. We defined areas with high human presence as those having over 10% of their total area brightly lit ($L > 15$) and over 10% of their area containing Flickr photos. The three biodiversity hotspots with the highest human presence were the California Floristic Province (with 23.6% of the hotspot area covered by Flickr photos and 13.5% lit), Japan (15.7% photographed, 26.2% lit), and

TABLE 1. Spearman's rank correlation coefficients matrix at the country level ($n = 193$) between the variables tested for explaining human activity as quantified by night-light brightness and Flickr photos. Table continues on next page.

Variable	Lit area ($L > 0$)		Lit area ($L > 15$)		Pixels with Flickr pgs.	Area with Flickr pgs. (%)
	Size	Lit (%)	Size	Lit (%)		
Area	0.720**	−0.565**	0.584**	−0.416**	0.528**	−0.713**
Population	0.793**	−0.210**	0.714**	−0.091	0.650**	−0.402**
Population density	−0.214**	0.709**	−0.065	0.612**	−0.073	0.690**
GDP	0.874**	0.275**	0.942**	0.435**	0.864**	0.057
GDP per capita	0.174*	0.680**	0.365**	0.745**	0.357**	0.662**
Area with Flickr pgs. (%)	−0.226**	0.846**	−0.041	0.766**	0.106	
$L > 0$ lit area size		0.043	0.943**	0.171*	0.845**	−0.226**
$L > 0$ lit area (%)	0.043		0.226**	0.925**	0.175*	0.846**
$L > 15$ lit area size	0.943**	0.226**		0.405**	0.862**	−0.041
$L > 15$ lit area (%)	0.171*	0.925**	0.405**		0.304**	0.766**
IP (%)	0.269**	0.673**	0.448**	0.751**	0.468**	0.646**

Notes: Two-tailed statistical significance of correlation coefficients is shown as * $P < 0.05$; ** $P < 0.01$. Night-light brightness values (L) were quantized between 0 and 63. IP refers to internet penetration and pgs refers to photographers.

the Mediterranean Basin (11.4% photographed, 12.5% lit; Fig. 8).

Last of the Wild regions (as defined by Sanderson et al. 2002) cover 38% of the Earth's land surface, and indeed only 0.7% (1.3 million) of all Flickr photos originate from these areas. However, we found that 40 of the 661 Last of the Wild areas had over 10% of their area photographed by multiple Flickr photographers (over 30 000 photographers; Fig. 8).

We detected 425 non-lit locations outside currently protected areas with 25 or more Flickr photographers, representing high visitation in unprotected areas (Fig. 9). These areas, such as Salar de Uyuni (Bolivia; 185 photographers), the Pantanal (Brazil; 206), and fresh-

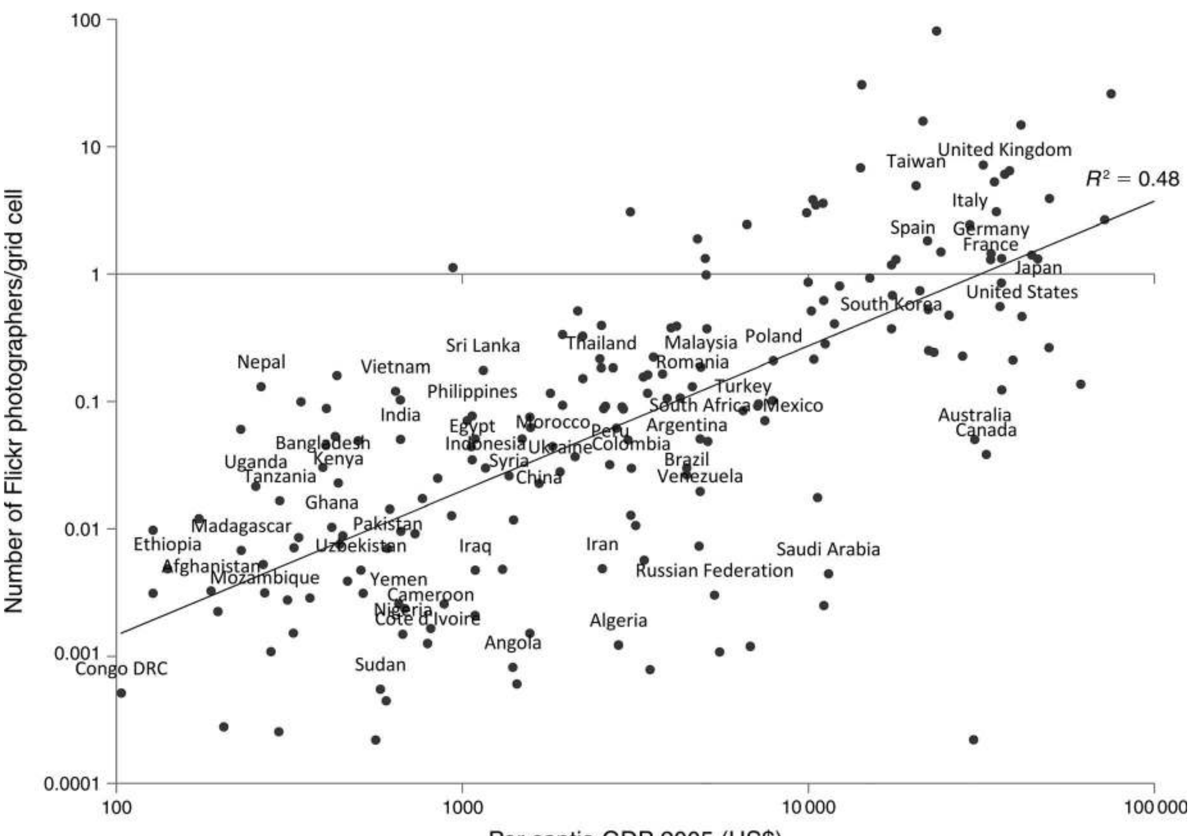

FIG. 2. The relationship between 2005 national per capita GDP and mean number of Flickr photographers (per 0.01° grid cell) at the country level. Note the log–log axis scales.

TABLE 1. Extended.

Mean no. of Flickr pgs.	Flickr pgs. in non-lit areas (%)		IP (%)
	$L = 0$	$L \leq 15$	
−0.655**	0.371**	−0.014	−0.201**
−0.341**	0.057	−0.248**	−0.106
0.667**	−0.604**	−0.335**	0.251**
0.140	−0.403**	−0.634**	0.488**
0.703**	−0.670**	−0.553**	0.859**
0.986**	−0.658**	−0.305**	0.646**
−0.151*	−0.157*	−0.435**	0.269**
0.854**	−0.879**	−0.587**	0.673**
0.040	−0.363**	−0.625**	0.448**
0.802**	−0.908**	−0.743**	0.751**
0.690**	−0.664**	−0.596**	

water Lake Manassarovar (Tibet, China; 207) were located in nonurban remote areas and clearly attract many visitors.

DISCUSSION

Monitoring human impact and visitation rates to protected areas, and their shifts over space and time, is important for conservation planning and management (Watson et al. 2014; Balmford et al. 2015). But with over 100 000 protected areas worldwide covering more than 12% of the Earth's land surface, it is very difficult (practically impossible) to monitor these globally across all protected areas, particularly for remote areas. Furthermore, information on areas that are not yet protected but attract many visitors is also important for management and future prioritization of conservation efforts. While most photos are taken by people outside protected areas, the millions of Flickr photos uploaded to the internet combined with night-light imagery allows us to map and quantify, for the first time, worldwide visitation of both protected and unprotected areas, which is an important type of human activity outside urban (and highly lit) areas. This enables the identification of visitation hotspots (and coldspots) for multiple countries and ecoregions across the world (Fig. 4). Lit sections within protected areas indicate infrastructure which facilitates visitation, whereas skyglow of light into protected areas indicates proximity to population centers (Gaston et al. 2014, 2015) and thus a large source of potential visitors. The photography-based visitation metrics developed and quantified here can be useful toward assessing future protected area gaps, strategies, and effectiveness of protected area management in relation to pressures created by visitors (Chape et al. 2005). When a protected area is declared internationally (such as a UNESCO World Heritage

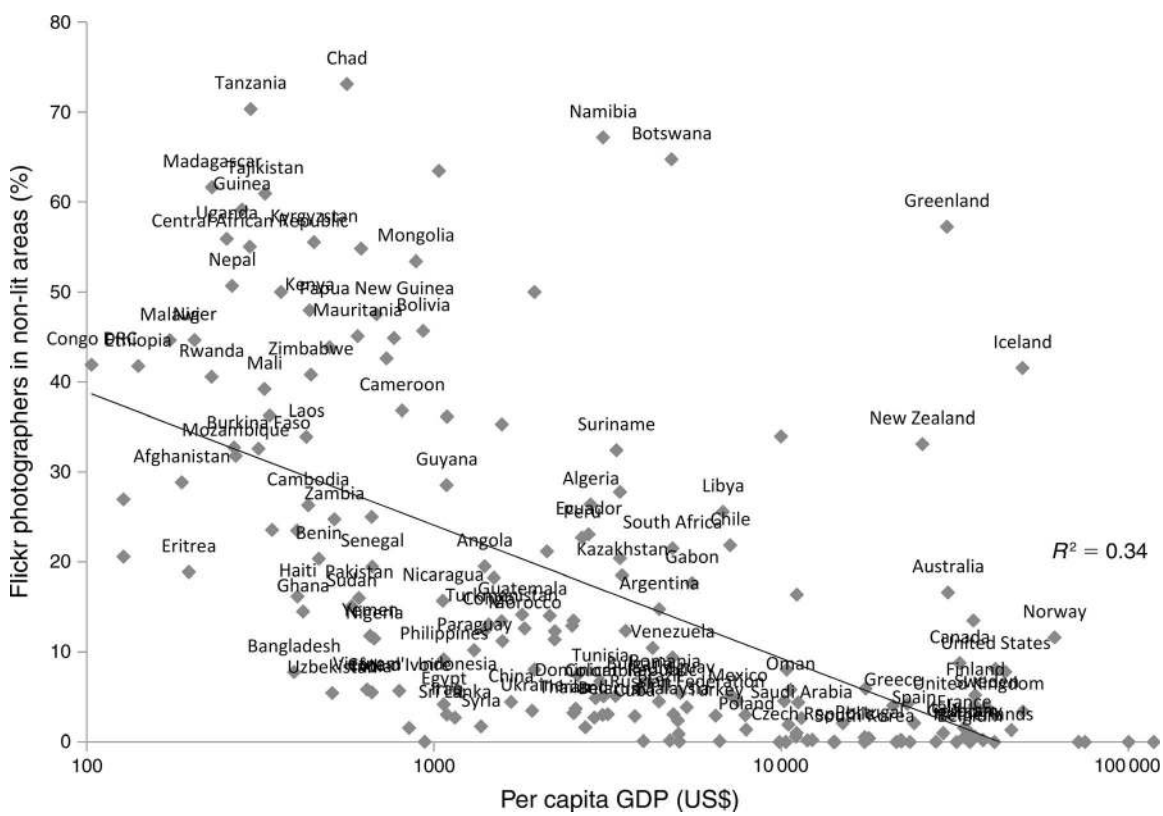

FIG. 3. The relationship between national per capita GDP and percentage of Flickr photographers in non-lit areas (defense meteorological satellite program values = 0). Countries above the regression line are likely to be countries in which tourists and visitors are attracted to non-lit locations, which are often protected areas. Note the *x*-axis log scale.

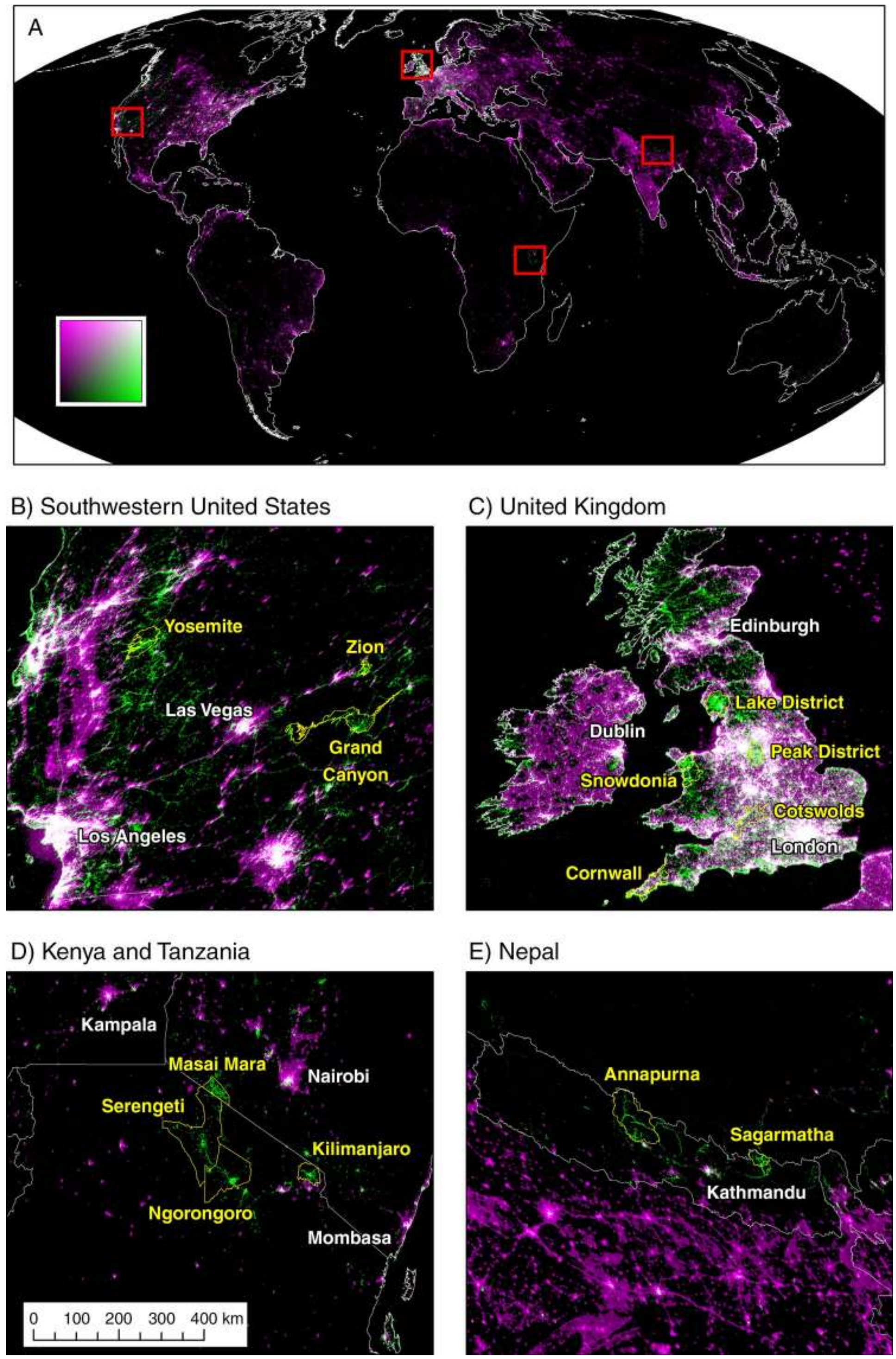

FIG. 4. (A) Global distribution of satellite-derived night lights averaged between 2004 and 2012 (in purple) and Flickr photographers (in green). Areas with both Flickr photos and night lights are shown in white, unlit areas with Flickr photos are shown in green, lit areas with few or no Flickr photos are shown in magenta, and unlit areas with no Flickr photos are shown in black. Four regions are shown in more detail at the same spatial scale, including, from left to right: (B) the southwestern United States, (C) United Kingdom and Ireland, (D) Kenya and Tanzania, and (E) Nepal. The most-photographed protected areas within each of these four areas based on Flickr (years) are shown by yellow outline. All maps are aligned north–south.

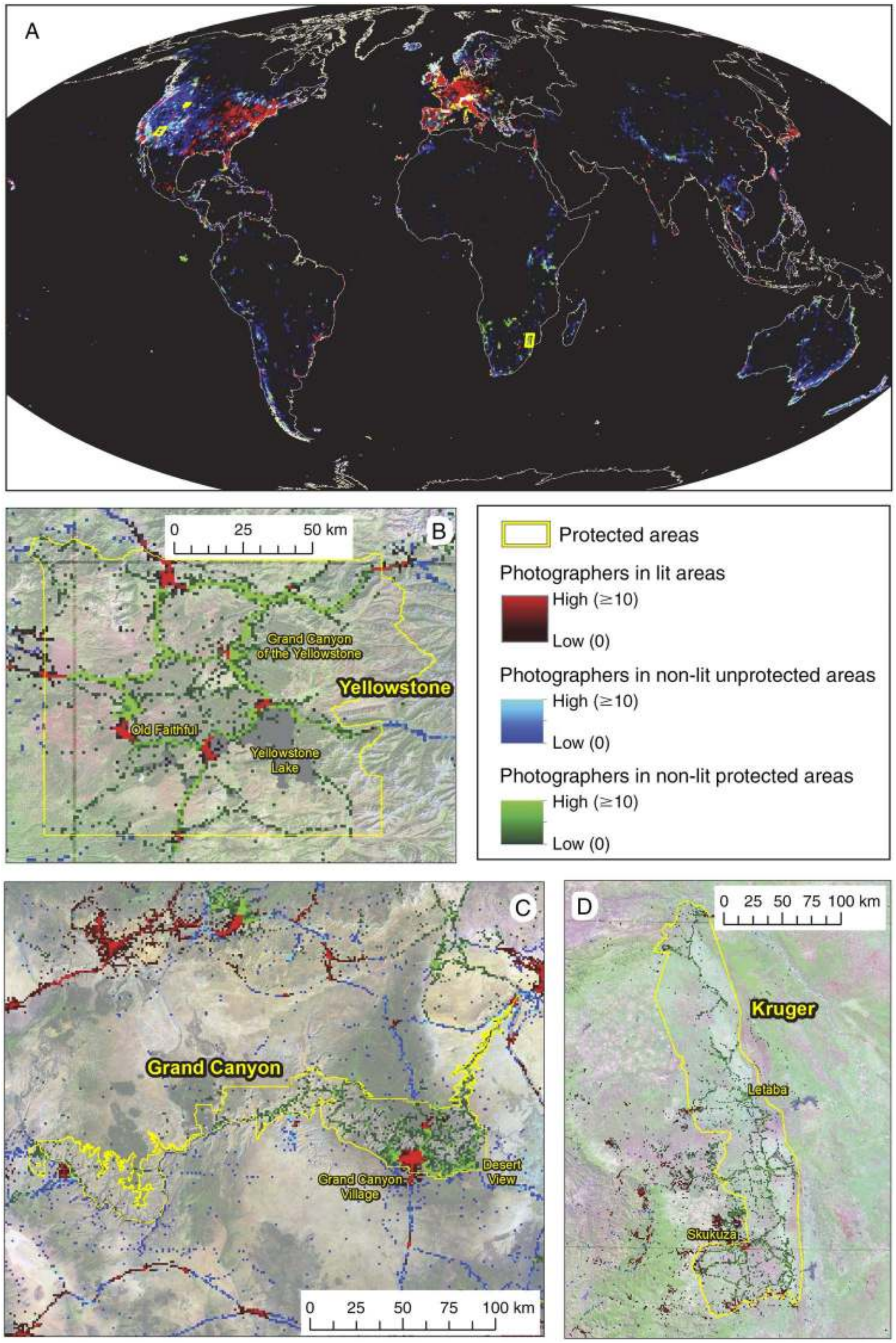

Fig. 5. (A) Hotspots of Flickr photographers in lit areas (in red), non-lit protected areas (in green), and non-lit unprotected areas (in blue). Zooming in on three famous national parks, notice the unequal distribution of visitors to different parts of (B) Yellowstone National Park, Wyoming, Idaho, Montana, USA (123 430 photos, 16.9% of the area photographed), (C) Grand Canyon National Park, Arizona, USA (140 690 photos, 29.1% of the area photographed), and (D) Kruger National Park, South Africa (21 851 photos, 12.5% of the area photographed). Flickr data are shown on top of false-color Landsat images. All maps are aligned north–south. Number of photographers is shown per grid cell.

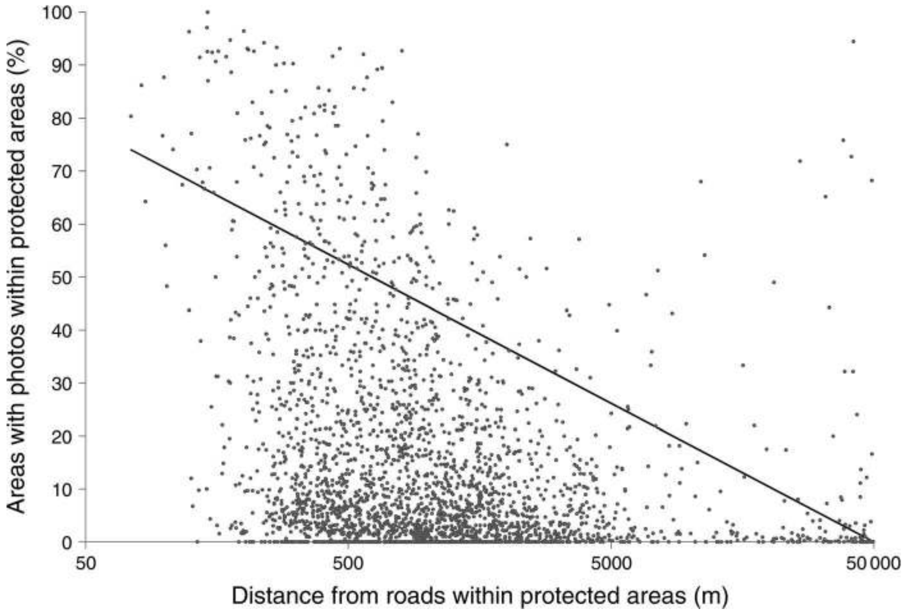

FIG. 6. Percentage of area with photos within protected areas, as a function of mean distance from roads and tracks within protected areas in the contiguous United States. The line is the 0.9 quantile regression line (pseudo $R^2 = 0.59$). Note the *x*-axis log scale.

Site or a Ramsar Site), one of the aims is to highlight its universal importance for both the country's residents and visitors (Buckley 2004); the analysis here confirms that indeed more people are drawn to internationally recognized protected areas, and that protection is generating tourism activity. Remote wilderness areas, also defined by Sanderson et al. (2002) as the Last of the Wild, are highlighted in some conservation efforts (Brooks et al. 2006, Craigie et al. 2014). Using Flickr-based visitation statistics, the definition of wilderness areas could be refined and redefined dynamically over time with user-based information that allows identification of regions that have few visitors roaming them and less light pollution resulting from urban and industrial activities.

With some national parks and coastal areas being overcrowded, conservation efforts aimed at protecting their biodiversity and natural ecosystems may be jeopardized (Eagles et al. 2002, Christ et al. 2003). The approach and methods proposed here allow managers to identify overcrowded areas as well as areas that visitors do not reach. Using the approach presented here to map where people are active outside urban and populated areas, we were also able to discover highly visited, yet unprotected, natural areas. These sites should be considered as potential sites for prioritizing future global conservation efforts and resource allocation, so that visitors' activity in these areas can be managed. However, these areas often have competing interests that might stand in the way of declaring a new protected area. For instance, in the Salar de Uyuni (Bolivia), there are conflicts between agriculture, tourism, and lithium mining (Aguilar-Fernandez 2009), whereas in the Pantanal (Brazil) there are conflicts between deforestation for cattle, river damming, population growth, and tourism (Lourival et al. 2011). However, many of these unprotected visitation hotspots may well require visitor-related management and conservation-related planning. Governmental and international conservation organizations may use the approach presented here to help identify potential sites for investing further conservation resources.

New data policies of space agencies promote free and open access to data from governmental satellites (e.g., Landsat, MODIS, Sentinel, and others; Kark et al. 2008, Malenovský et al. 2012, Wulder et al. 2012, Turner et al. 2015), and there is an increase in the availability of freely distributed global products derived from satellite images (e.g., SRTM and ASTER-derived digital elevation models, or the Landsat-derived data set of deforestation; Hansen et al. 2013, Rexer and Hirt 2014). The integrated approach presented in our study, combining satellite imagery with large social media databases, could help revolutionize how we dynamically prioritize conservation planning and management efforts of protected and unprotected areas at multiple scales. The approach represents a novel shift from generating statistics using top-down approaches (e.g., population censuses done by governments), toward bottom-up approaches, in which everyday users reporting their activities using social media are used as sensors (Goodchild 2007, Giles 2012) to generate and collect long-term aggregated data that cannot be collected by traditional means (Giles 2005). The relationship found here between the number of visitors and Flickr photos corresponds with recent work that examined selected recreational sites (e.g., Disney-

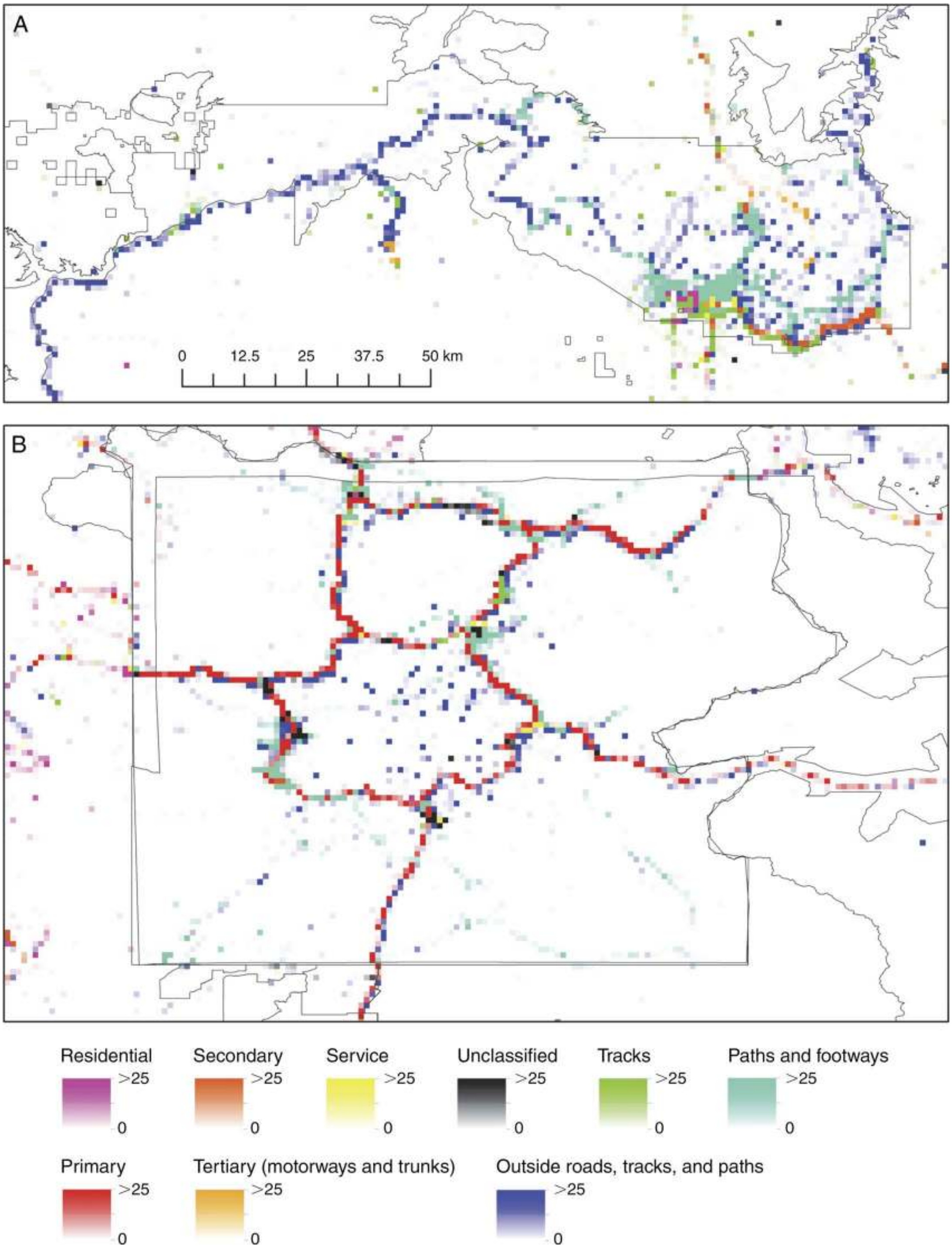

FIG. 7. Distribution of Flickr photos taken in (A) Grand Canyon National Park and (B) Yellowstone National Park, classified as number of photos per road type as defined within OpenStreetMap; see *Methods: Spatial analysis* for further details on OpenStreetMap. Both panels are aligned north–south.

land, California, USA; Wood et al. 2013). In spite of the strong correspondence found between visitation statistics and Flickr data, we acknowledge that this early use of online social data still includes a range of possible biases that need to be overcome with time. Internet user-generated data is characterized by uneven and clustered geographies (e.g., for Wikipedia; see Graham et al. 2014). For instance, Internet penetration rates vary between countries, with some governments practicing different levels of web censorship (Warf 2011). Flickr

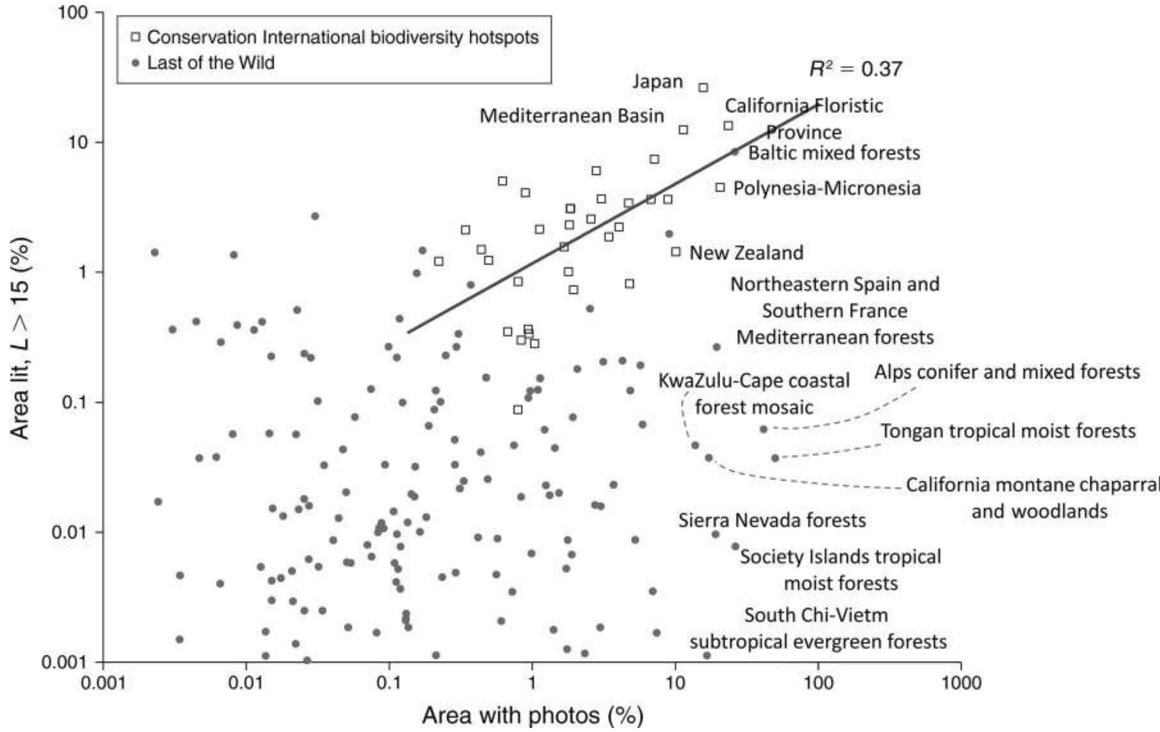

FIG. 8. Percentage of lit area ($L > 15$) and percentage of area with photos as calculated for the Earth's 35 biodiversity hotspots based on Conservation International and the world's Last of the Wild areas. Names of biodiversity hotspots and of Last of the Wild areas are shown for regions in which more than 10% of the area had photos (at a spatial resolution of 0.01°). The regression line and coefficient in the figure refer to the 35 biodiversity hotspots. No significant correlation was found for the Last of the Wild regions. Note the log–log axis scales.

users are not representative of the overall human population or of all visitors to an area; social media users are typically better educated, younger, and wealthier than average (Li et al. 2013), and Flickr in particular is still most popular in the United Stated and Western Europe. Errors may occur in photo GPS tags and timestamps, as well as when scenery photos of one protected area are taken from an observation point located in an adjacent protected area. Nevertheless, this tool provides a new approach to quantify actual visitation rates across all protected areas globally in a way impossible in the past, prior to the emergence of social media data. Additional analyses can be performed using data tagged to the photo, such as analyzing the spatial patterns of foreign vs. domestic visitors (Straumann et al. 2014), or even using the visual content of the photo itself (Zhang et al. 2012).

Citizen science offers a powerful tool for generating and analyzing data for ecology and biodiversity research, and social media Application Programming Interfaces (APIs) can be used to perform data mining and analyze various research questions using open source tools, which also allow high reproducibility (Catlin-Groves 2012). We expect that with time, as mobile sensors improve and associated technology continues to permeate society, estimates of human activity from social media data will be refined and improved. Statistical methods have been developed and applied by the ecological scientific community to correct for sampling effects when assessing species richness (e.g., using species accumulation curves; Colwell et al. 2004). In a similar fashion, the scientific community is gradually gaining more confidence in using user-generated data, and developing new analytical techniques for quality assurance, geospatial statistics, and for quantifying possible biases of information that is derived by data mining from the internet (Dickinson et al. 2010). The increasing availability of free ecological data and open source statistical and geospatial algorithms and software go hand in hand with Web 2.0 user-generated data, enhancing the ability to conduct global-scale research and to replicate peer-reviewed functions (Rocchini and Neteler 2012).

Anthropogenic threats to protected areas include many other factors possibly more important than human visitors, such as encroachments by urban and agricultural land uses (Hamilton et al. 2013), deforestation (Hansen et al. 2013), poaching, resource extraction (Laurance et al. 2012), introduction of invasive species (Gibbons et al. 2000), pollution, and more. Although global data sets of population and land use are available, they are mostly available on yearly intervals, whereas social media data are generated continuously and can be analyzed at temporal intervals of minutes, hours, and

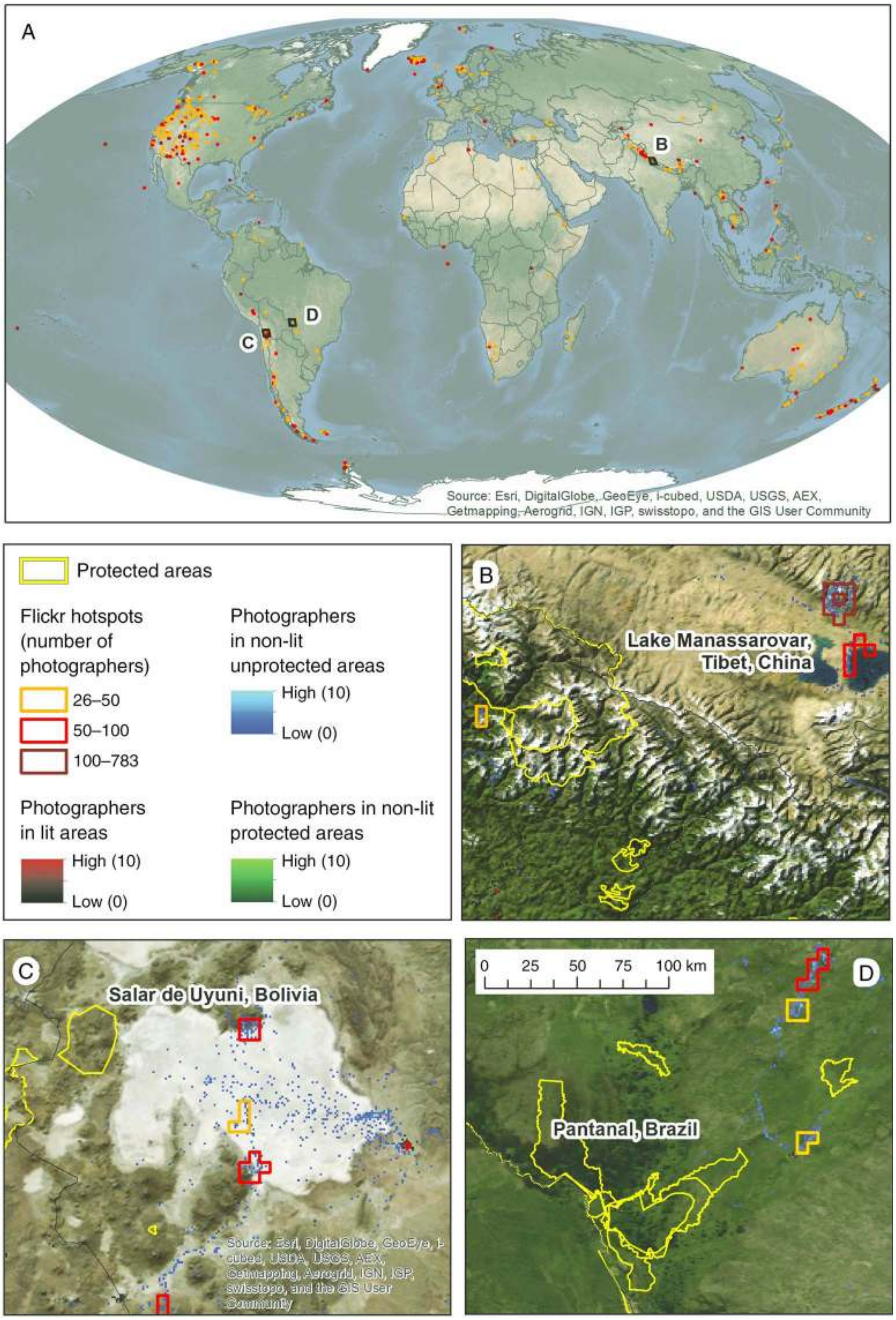

FIG. 9. (A) Hotspots of Flickr photographers in non-lit areas that are located outside currently declared protected areas. These sites draw many visitors and provide potential locations for declaring future protected areas. Three hotspots areas are shown in greater detail: (B) Lake Manassarovar, Tibet, China, (C) Salar de Uyuni, Bolivia, and (D) the Pantanal, Brazil. Number of photographers is shown per grid cell.

days (Sakaki et al. 2010). We offer our approach as a complementary method to traditional estimates for assessing human impact (as in the human footprint of Sanderson et al. [2002]). While we did not find any correspondence between deforestation and visitation within protected areas (although one might speculate that the presence of many visitors, and infrastructure for visitors, is a deterrent to unauthorized deforestation), we did find that accessibility is one of the key factors explaining highly visited (photographed) protected areas. Over-visitation and related disturbances should be taken into account when aiming to achieve conservation goals (Laurance 2013) and are often crucial for good management planning. The combined use of satellite imagery with social media data provides an invaluable tool for future conservation planning and prioritization that should be further explored.

Acknowledgments

The authors are thankful to Jeremy Kark and Hugh Possingham for helpful comments on an earlier version of the manuscript. We also thank two anonymous reviewers, whose suggestions helped improve the clarity of the manuscript. S. Kark is supported by the Australian Research Council. D. Crandall was supported in part by the National Science Foundation through a CAREER award (IIS-1253549).

Literature Cited

Aguilar-Fernandez, R. 2009. Estimating the opportunity cost of lithium extraction in the Salar de Uyuni. Thesis. Duke University, Durham, North Carolina, USA.

Asner, G. P., and R. E. Martin. 2008. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. Frontiers in Ecology and the Environment 7:269–276.

Balk, D. L., U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson. 2006. Determining global population distribution: methods, applications and data. Advances in Parasitology 62:119–156.

Balmford, A., J. M. H. Green, M. Anderson, J. Beresford, C. Huang, R. Naidoo, M. Walpole, and A. Manica. 2015. Walk on the wild side: estimating the global magnitude of visits to protected areas. PLoS Biology 13(2):e1002074.

Balmford, A., J. L. Moore, T. Brooks, N. Burgess, L. A. Hansen, P. Williams, and C. Rahbek. 2001. Conservation conflicts across Africa. Science 291:2616–2619.

Bastiaanssen, W. G., D. J. Molden, and I. W. Makin. 2000. Remote sensing for irrigated agriculture: examples from research and possible applications. Agricultural Water Management 46:137–155.

Bennie, J., T. W. Davies, J. P. Duffy, R. Inger, and K. J. Gaston. 2014. Contrasting trends in light pollution across Europe based on satellite observed night time lights. Scientific Reports 4:3789.

Brockmann, D., L. Hufnagel, and T. Geisel. 2006. The scaling laws of human travel. Nature 439:462–465.

Brooks, T. M., R. A. Mittermeier, G. A. da Fonseca, J. Gerlach, M. Hoffmann, J. F. Lamoreux, C. G. Mittermeier, J. D. Pilgrim, and A. S. Rodrigues. 2006. Global biodiversity conservation priorities. Science 313:58–61.

Buckley, R. 2004. The effects of World Heritage listing on tourism to Australian National Parks. Journal of Sustainable Tourism 12:70–84.

Cantú-Salazar, L., C. D. L. Orme, P. C. Rasmussen, T. M. Blackburn, and K. J. Gaston. 2013. The performance of the global protected area system in capturing vertebrate geographic ranges. Biodiversity and Conservation 22:1033–1047.

Catlin-Groves, C. L. 2012. The citizen science landscape: from volunteers to citizen sensors and beyond. International Journal of Zoology 2012:349630.

Chape, S., J. Harrison, M. Spalding, and I. Lysenko. 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. Philosophical Transactions of the Royal Society B 360:443–455.

Chinn, M. D., and R. W. Fairlie. 2007. The determinants of the global digital divide: a cross-country analysis of computer and internet penetration. Oxford Economic Papers 59(1):16–44.

Christ, C., O. Hillel, S. Matus, and J. Sweeting. 2003. Tourism and biodiversity: mapping tourism's global footprint. Conservation International, Washington, D.C., USA.

Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. Ecology 85:2717–2727.

Craigie, I. D., R. L. Pressey, and M. Barnes. 2014. Remote regions—the last places where conservation efforts should be intensified. Biological Conservation 172:221–222.

Crandall, D., L. Backstrom, J. Kleinberg, and D. Huttenlocher. 2009. Mapping the world's photos. Pages 761–770 *in* J. Quemada, G. León, Y. Maarek, and W. Nejdl, editors. Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, April 20–24, 2009. Association for Computing Machinery, New York, New York, USA.

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. Citizen science as an ecological research tool: challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41:149–172.

Eagles, P. F. J., S. F. McCool, C. D. Haynes, A. Phillips, and United Nations Environment Programme. 2002. Sustainable tourism in protected areas: guidelines for planning and management. International Union for Conservation of Nature and Natural Resources, Gland, Switzerland.

Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis. 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing 18:1373–1379.

Elvidge, C. D., D. Ziskin, K. E. Baugh, B. T. Tuttle, T. Ghosh, D. W. Pack, E. H. Erwin, and M. Zhizhin. 2009. A fifteen year record of global natural gas flaring derived from satellite data. Energies 2:595–622.

Gaston, K. J., J. P. Duffy, and J. Bennie. 2015. Quantifying the erosion of natural darkness in the global protected area system. Conservation Biology 29(4):1132–1141.

Gaston, K. J., J. P. Duffy, S. Gaston, J. Bennie, and T. W. Davies. 2014. Human alteration of natural light cycles: causes and ecological consequences. Oecologia 176:917–931.

Gibbons, J. W., et al. 2000. The global decline of reptiles, déjà vu amphibians. BioScience 50:653–666.

Giles, J. 2005. Internet encyclopaedias go head to head. Nature 438:900–901.

Giles, J. 2012. Making the links. Nature 488:448–450.

Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69:211–221.

Goodchild, M. F. 2013. The quality of big (geo) data. Dialogues in Human Geography 3:280–284.

Graham, M., B. Hogan, R. K. Straumann, and A. Medhat. 2014. Uneven geographies of user-generated information: patterns of increasing informational poverty. Annals of the Association of American Geographers 104:746–764.

Gröchenig, S., R. Brunauer, and K. Rehrl. 2014. Digging into the history of VGI data-sets: results from a worldwide study on OpenStreetMap mapping activity. Journal of Location Based Services 8(3):198–210.

Haklay, M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design 37:682–703.

Hamilton, C. M., S. Martinuzzi, A. J. Plantinga, V. C. Radeloff, D. J. Lewis, W. E. Thogmartin, P. J. Heglund, and A. M. Pidgeon. 2013. Current and future land use around a nationwide protected area network. PLoS ONE 8:e55737.

Hansen, M. C., et al. 2013. High-resolution global maps of 21st-century forest cover change. Science 342:850–853.

Hölker, F., C. Wolter, E. K. Perkin, and K. Tockner. 2010. Light pollution as a biodiversity threat. Trends in Ecology & Evolution 25:681–682.

Jeffries, A. 2013. The man behind Flickr on making the service 'awesome again'. The Verge, March 20, 2013. http://www.theverge.com/2013/3/20/4121574/flickr-chief-markus-spiering-talks-photos-and-marissa-mayer

Kark, S., N. Levin, and S. Phinn. 2008. Global environmental priorities: making sense of remote sensing. Trends in Ecology & Evolution 23:181–182.

Koenker, R., and G. Bassett, Jr. 1978. Regression quantiles. Econometrica 46:33–50.

Koenker, R., and J. A. Machado. 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94:1296–1310.

Laurance, W. F. 2013. Does research help to safeguard protected areas? Trends in Ecology & Evolution 28:261–266.

Laurance, W. F., et al. 2012. Averting biodiversity collapse in tropical forest protected areas. Nature 489:290–294.

Levin, N., and Y. Duke. 2012. High spatial resolution night-time light images for demographic and socio-economic studies. Remote Sensing of Environment 119:1–10.

Levin, N., A. Shmida, O. Levanoni, H. Tamari, and S. Kark. 2007. Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. Diversity and Distributions 13:692–703.

Li, L. N., M. F. Goodchild, and B. Xu. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science 40:61–77.

Lourival, R., M. Drechsler, M. E. Watts, E. T. Game, and H. P. Possingham. 2011. Planning for reserve adequacy in dynamic landscapes; maximizing future representation of vegetation communities under flood disturbance in the Pantanal wetland. Diversity and Distributions 17:297–310.

Malenovský, Z., H. Rott, J. Cihlar, M. E. Schaepman, G. García-Santos, R. Fernandes, and M. Berger. 2012. Sentinels for science: potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. Remote Sensing of Environment 120:91–101.

Margules, C. R., and R. L. Pressey. 2000. Systematic conservation planning. Nature 405:243–253.

Miller, S. D., W. Straka, S. P. Mills, C. D. Elvidge, T. F. Lee, J. Solbrig, A. Waither, A. K. Heidinger, and S. C. Weiss. 2013. Illuminating the capabilities of the Suomi National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) day/night band. Remote Sensing 5:6717–6766.

Mittermeier, R. A., P. R. Gil, M. Hoffman, J. Pilgrim, T. Brooks, C. G. Mittermeier, J. Lamoreux, and G. A. B. da Fonseca. 2005. Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions. Cemex, Mexico City, Mexico.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858.

Rexer, M., and C. Hirt. 2014. Comparison of free high resolution digital elevation data sets (ASTER GDEM2, SRTM v2.1/v4.1) and validation against accurate heights from the Australian National Gravity Database. Australian Journal of Earth Sciences 61:213–226.

Rocchini, D., and M. Neteler. 2012. Let the four freedoms paradigm apply to ecology. Trends in Ecology & Evolution 27:310–311.

Sakaki, T., M. Okazaki, and Y. Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. Pages 851–860 *in* M. Rappa, P. Jones, J. Freire, and S. Chakrabarti, editors. Proceedings of the 19th International Conference on World Wide Web, Raleigh, North Carolina, USA, April 26–30, 2010. Association for Computing Machinery, New York, New York, USA.

Sanderson, E. W., M. Jaiteh, M. A. Levy, K. H. Redford, A. V. Wannebo, and G. Woolmer. 2002. The human footprint and the last of the wild. Bioscience 52:891–904.

Seiferling, I. S., R. Proulx, P. R. Peres-Neto, L. Fahrig, and C. Messier. 2012. Measuring protected-area isolation and correlations of isolation with land-use intensity and protection status. Conservation Biology 26:610–618.

Seto, K. C., B. Guneralp, and L. R. Hutyra. 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of Sciences USA 109:16083–16088.

Song, C., T. Koren, P. Wang, and A. L. Barabási. 2010. Modelling the scaling properties of human mobility. Nature Physics 6:818–823.

Straumann, R. K., A. Çöltekin, and G. Andrienko. 2014. Towards (re)constructing narratives from georeferenced photographs through visual analytics. Cartographic Journal 51:152–165.

Turner, W., C. Rondinini, N. Pettorelli, B. Mora, A. K. Leidner, Z. Szantoi, and C. Woodcock. 2015. Free and open-access satellite data are key to biodiversity conservation. Biological Conservation 182:173–176.

Warf, B. 2011. Geographies of global Internet censorship. GeoJournal 76:1–23.

Watson, J. E., N. Dudley, D. B. Segan, and M. Hockings. 2014. The performance and potential of protected areas. Nature 515:67–73.

Wood, S. A., A. D. Guerry, J. M. Silver, and M. Lacayo. 2013. Using social media to quantify nature-based tourism and recreation. Scientific Reports 3:2976.

Wulder, M. A., J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock. 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment 122:2–10.

Zhang, H., M. Korayem, D. Crandall, and G. LeBuhn. 2012. Mining photo-sharing websites to study ecological phenomena. Pages 749–758 *in* A. Mille, F. Gandon, J. Misselis, M. Rabinovich, and S. Staab. Proceedings of the 21st International Conference on World Wide Web, Lyon, France, April 16–20, 2012. Association for Computing Machinery, New York, New York, USA.

Zielstra, D., H. H. Hochmair, and P. Neis. 2013. Assessing the effect of data imports on the completeness of OpenStreetMap—a United States case study. Transactions in GIS 17:315–334.

Supplemental Material

Ecological Archives

Appendix A is available online: http://dx.doi.org/10.1890/15-0113.1.sm