Applied Geography 79 (2017) 115–126

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas

Noam Levin [a, c, *], Alex Mark Lechner [b, d], Greg Brown [c, e]

[a] Department of Geography, The Hebrew University of Jerusalem, Mt Scopus, Jerusalem, 91905, Israel
[b] Regional Water and Land Resources, Sustainable Minerals Institute, The University of Queensland, St Lucia, Brisbane, Queensland, 4072, Australia
[c] School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, Queensland, 4072, Australia
[d] School of Environmental and Geographical Sciences, University of Nottingham Malaysia Campus, Semenyih, Selangor, 43500, Malaysia
[e] Natural Resources Management & Environmental Sciences Department, California Polytechnic State University, San Luis Obispo, California, 93407, USA

A R T I C L E   I N F O

Article history:
Received 10 March 2016
Received in revised form
11 October 2016
Accepted 22 December 2016

Keywords:
Flickr
OpenStreetMap
Place values
Public participation GIS (PPGIS)
Volunteered geographic information (VGI)
Wikipedia

A B S T R A C T

Parks and protected areas provide a wide range of benefits, but methods to evaluate their importance to society are often ad hoc and limited. In this study, the quality of crowdsourced information from Public Participation GIS (PPGIS) and Volunteered Geographic Information (VGI) sources (Flickr, OpenStreetMap (OSM), and Wikipedia) was compared with visitor counts that are presumed to reflect social importance. Using the state of Victoria, Australia as a case study, secondary crowdsourced VGI data, primary crowdsourced (PPGIS data) and visitor statistics were examined for their correspondence and differences, and to identify spatial patterns in park popularity. Data completeness—the percent of protected areas with data—varied between sources, being highest for OSM (90%), followed by Flickr (41%), PPGIS (24%), visitation counts (5%), and Wikipedia articles (4%). Statistically significant correlations were found between all five measures of popularity for protected areas. Using stepwise multiple linear regression, the explained variability in visitor numbers was greater than 70%, with PPGIS, Flickr and OSM having the largest standardized coefficients. The social importance of protected areas varied as a function of accessibility and the types of values (direct or indirect use) expressed for the areas. Crowdsourced data may provide an alternative to visitor counts for assessing protected area social importance and spatial variability of visitation. However, crowdsourced data appears to be an unreliable proxy for the full range of values and importance of protected areas, especially for non-use values such as biological conservation.

1. Introduction

The advent and combination of several technologies, namely the global positioning system (GPS), the internet, Google Maps, smartphones and social media, have transformed the ways we generate, consume and interact with geographic information (Goodchild, 2007; Haklay, Singleton, & Parker, 2008). In the world of Web 2.0, where citizens can be seen as sensors (as termed by Goodchild, 2007), and where people voluntarily contribute geographic data, either knowingly or unknowingly, new avenues open for spatial research. Whereas in the past, most geographic information and maps were generated by governments, non-governmental organizations, and major companies, in the recent decade the sharing of volunteered geographic information (VGI; Fast & Rinner, 2014) is rising, and concerns are being raised about its credibility (e.g., Flanagin & Metzger, 2008; Spielman, 2014). The quality of user-generated or crowdsourced data is an obvious challenge when using big (geo) data, which is characterized by volume, velocity, and variety (Goodchild, 2013).

Volunteered geographic information (Goodchild, 2007) is a form of *crowdsourcing* (Howe, 2006) where ideas or content are solicited for a certain project, from a large group of people, especially from an online community, with enabling technical infrastructure including hardware, software, and the Geoweb – online location-based services (Fast & Rinner, 2014). *Public participation GIS* (PPGIS) is a term that originated in 1996 to describe how GIS technology could support public participation processes (NCGIA 1996a, 1996b). Although the term *crowdsourcing* did not originate

* Corresponding author. Department of Geography, The Hebrew University of Jerusalem, Mt Scopus, Jerusalem, 91905, Israel.
E-mail address: noamlevin@mscc.huji.ac.il (N. Levin).

http://dx.doi.org/10.1016/j.apgeog.2016.12.009
0143-6228/

Table 1
Sources used for examining the popularity and values associated with protected areas.

Source	Description
Visitation	Official visitation statistics, available for 342 parks from Parks Victoria (2011).
PPGIS	Study participants (n = 1905) were sampled, recruited, and requested to identify place values using markers on a Google Map of Victoria, Australia. The place values were a typology of values (biological/conservation, heritage/cultural, intrinsic/existence, learning/education/research, recreation, scenic/aesthetic, therapeutic/health, wilderness/pristine (see Brown, Weber and De Bie, 2014 for the value definitions). Participants could place as many of the different types of markers as needed to identify places values. Place values (also called *landscape values*) are a type of "relationship" or "transactive" value that bridges both "held" values (what is personally important) and "assigned" values (the relative importance of external objects) (Brown & Weber, 2012).The number and type of place values were counted within each protected area.
Flickr	The total number of georeferenced Flickr photos (between 2004 and 2012) within each protected area (further details on this dataset are provided in Levin et al., 2015). Flickr is a photo-sharing website, and Flickr photos have been shown to be well-correlated with official visitation statistics (Levin et al., 2015; Wood et al., 2013).
Wikipedia	The total number of Wikipedia edits which were found for Wikipedia articles on protected areas. The numbers of Wikipedia edits were counted in October 2015. Most protected areas did not have a Wikipedia article.
OpenStreetMap	The total number of vertices within each protected areas as found in the OpenStreetMap roads, tracks, and trails layer provided by Geofabrik (http://www.geofabrik.de/data/download.html). OpenStreetMap started in 2004 and provides a free digital map of the world created collaboratively by internet users (Haklay, 2010). The density of vertices within protected areas is expected to reflect the physical infrastructure, as well as the interest and familiarity in a protected area by the public. Parks with more facilities for visitors (e.g., trails, tracks, roads) and which are planned to attract more visitors, will likely have more features which can be mapped using OSM. Thus, parks with more OSM vertices are an indication of more mapped features, but also of greater familiarity and interest by the public in mapping it. In addition, the density of vertices can indicate the accuracy (i.e. scale) in which the mapping was done. The number of vertices used to digitize a certain feature depends both on the geometric complexity of that feature, and also on the levels of generalization and simplification used by the person who digitized that feature. Thus, more vertices can also indicate more care and details in the mapping process.

until a decade later, PPGIS methods that collect spatial data from a large group of people for public participation are considered a type of crowdsourcing (Brown, 2015). The related fields of VGI and PPGIS, as forms of crowdsourcing for geographic knowledge production, usually differ in their sampling and mapping methods, and both have experienced significant growth as evidenced in the number of real-world applications and academic publications (Brown & Kyttä, 2014; Mukherjee, 2015; Sui, Elwood, & Goodchild, 2012).

While the majority of the examples of VGI mapping have been for land cover mapping (e.g., roads, land use), crowdsourced data can also offer original insights as preferences for geographic locations (Brown, 2015; Lechner et al., 2014). Locational information can be manually entered (Gao, Li, Li, Janowicz, & Zhang, 2014) or extracted from devices such as smartphones (Birenboim & Shoval, 2016; Shoval, 2007). Associations with places can be identified through geotagged photos (e.g., Li, Crandall, & Huttenlocher, 2009), along with methods that analyze and classify the images (e.g., Yanai, Yaegashi, & Qiu, 2009). Wikipedia articles offer a wide array of options for investigating cultural values associated with sites using quantitative text analysis (as done by Michel et al., 2011 for millions of scanned books, or by Yasseri, Spoerri, Graham, & Kertész, 2014 for millions of Wikipedia articles), as well as by examining the edits to Wikipedia articles (Graham, Hogan, Straumann, & Medhat, 2014; Hardy, Frew, & Goodchild, 2012). The density of OpenStreetMap (OSM) edits has been shown to be highly correlated with socioeconomic status of inhabitants (Haklay, 2010).

Addressing uncertainty associated with crowdsourced data quality such as completeness and accuracy is critical to ensure the reliability of geographic analyses (Comber et al., 2013; Devillers & Jeansoulin, 2010; Shi, 2010). For VGI data, mutually reinforcing observations, independently derived from different user-generated sources, can serve as a method of data triangulation (Denzin, 2006) to test the correspondence between and completeness of crowdsourcing information. Indeed, integrating multi-sourced volunteered geographic data has been suggested as an approach for enriching the information available and overcoming some of the limitations associated with a single source of user generated data (Sester, Arsanjani, Klammer, Burghardt, & Haunert, 2014).

In this paper, our aim is to compare user-generated information from several sources with official observations to determine the degree to which different Web 2.0 sources correspond and complement each other. The specific theme pursued here is the importance of protected areas as perceived by the public. A protected area is a clearly defined geographical space, recognised, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values (Dudley, 2008). While protected areas may be highly important for their ecological or biological features, of which the public may not be fully aware, our aim is to examine a range of social values associated with protected areas. Because park visitation is a volitional human behaviour, there is clearly personal and social importance (from benefits) associated with visitation activity or people would not visit protected areas. Further, protected area funding (an indirect measure of importance) is often linked to visitation. Thus, the social value of protected areas can be assessed in multiple ways ranging from visitation, a type of popularity measure, to the assessment of specific place values (e.g., scenery, recreation) that people associate with protected areas from crowdsourced methods. In this study we have assessed social value using PPGIS data which accounts for both direct and indirect, and use and non-use values.

Official statistics concerning visitors' numbers can represent a real world reference for the popularity of protected areas; however, visitation statistics are missing for most protected areas globally (Eagles et al., 2002), and even when visitation statistics are available, additional information on what people actually do during their visit is often lacking (Buckley, Robinson, Carmody, & King, 2008). Recently there have been studies whose aim was to predict visitation rates using statistical modeling to estimate the economic value generated by protected areas (Balmford et al., 2015), to quantify human presence in protected areas using crowdsourced and remote-sensing data (Levin, Kark, & Crandall, 2015), and to assess the potential attractiveness of protected areas based on the species which are present in them using the Flickr photo-sharing website (Willemen, Cottam, Drakou, & Burgess, 2015). PPGIS and VGI methods have been used to identify both use and non-use values in protected areas in multiple global settings including the U.S. (Brown & Alessa, 2005; Brown, Kelly, & Whitall, 2014; Van Riper and Kyle, 2014), Canada (Beverly, Uto, Wilkes, & Bothwell, 2008), and Australia (Brown, Weber and De Bie, 2014; Van Riper

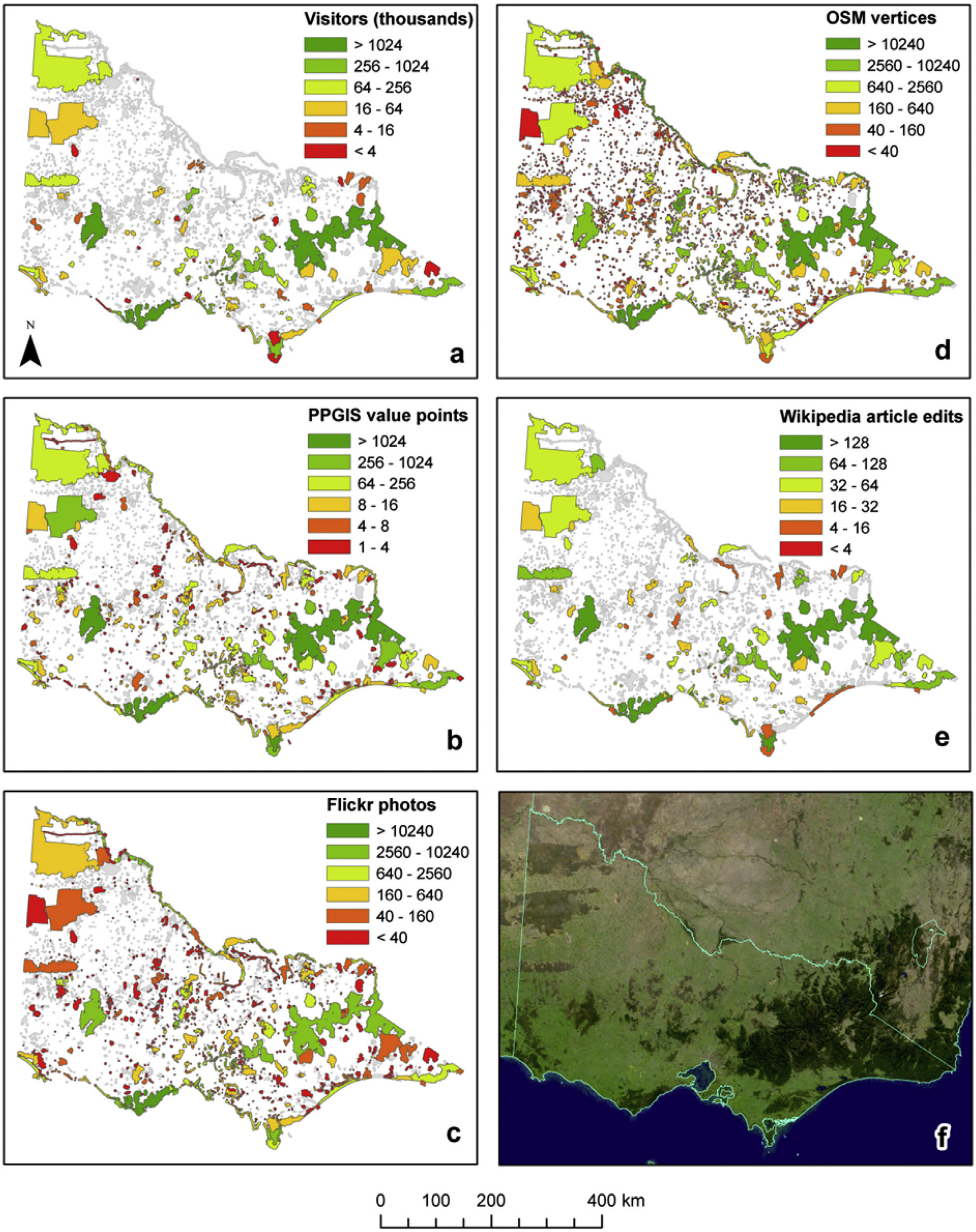

Fig. 1. Spatial patterns of Victoria's protected areas popularity as measured using: a) official visitation statistics; b) PPGIS value points; c) Flickr photos; d) OpenStreetMap vertices; and e) Wikipedia article edits. f) True-color satellite image of Victoria.

et al., 2012).

In this study, place values identified with a range of primary crowdsourced data (PPGIS) are compared to secondary crowdsourced VGI data (Flickr counts, Wikipedia edits, OSM edits), and to visitation statistics. Using the state of Victoria, Australia as a case study the following research questions were asked: (1) how strongly correlated are the various sources of crowdsourced data and visitation statistics and do they provide similar or different information, (2) does the type of protected area (e.g., national versus metropolitan park) influence the utility of crowdsourced

Table 2
Spearman's rank correlation coefficients between total number of protected areas (N), total area and counts of Visitors (official number of visitors), PPGIS (number of PPGIS value points), Flickr (number of Flickr photos), OSM (number of OSM vertices) and Wikipedia (number of edits to Wikipedia articles), where the observations units are protected area types (n = 36; Table S1). In the correlations shown above the diagonal, observations for which no data was available were given the value of 0.

	N	Total area	Visitors	PPGIS	Flickr	OSM	Wikipedia
N		**0.554****	0.304	**0.899*****	**0.940*****	**0.994*****	0.114
Total area	**0.554****		**0.484****	**0.653*****	**0.586*****	**0.543****	**0.447****
Visitors	0.076	0.351		**0.556****	**0.494****	0.309	**0.685*****
PPGIS	**0.894*****	**0.625*****	0.318		**0.953*****	**0.900*****	0.281
Flickr	**0.936*****	**0.552****	0.187	**0.949*****		**0.937*****	0.241
OSM	**0.994*****	**0.543****	0.091	**0.897*****	**0.936*****		0.102
Wikipedia	−0.252	0.181	**0.773*****	0.017	−0.095	−0.263	

*p < 0.05, **p < 0.01, ***p < 0.001.
The bold values represent statistically significant correlations.

Table 3
Spearman's rank correlation coefficients between percentage of number of protected areas (N), area and total Visitors (official number of visitors), PPGIS (number of PPGIS value points), Flickr (number of Flickr photos), OSM (number of OSM vertices) and Wikipedia (number of edits to Wikipedia articles), where the observations units are protected area types (n = 36; Table S2). In the correlations shown above the diagonal, observations for which no data was available were given the value of 0.

	N	Total area	Visitors	PPGIS	Flickr	OSM	Wikipedia
N		**0.554****	0.087	**0.353***	**0.429***	**0.663**	0.080
Total area	**0.554****		0.298	**0.621**	**0.532***	**0.703**	**0.345**
Visitors	0.116	0.325		**0.774**	**0.661**	**0.470****	**0.647**
PPGIS	**0.370***	**0.631**	**0.732**		**0.880**	**0.740**	**0.525****
Flickr	**0.388***	**0.610**	**0.698**	**0.912**		**0.837**	0.326
OSM	**0.663**	**0.703**	**0.412***	**0.771**	**0.827**		0.317
Wikipedia	0.059	0.245	**0.565***	**0.503***	0.344	0.226	

*p < 0.05, **p < 0.01, ***p < 0.001.
The bold values represent statistically significant correlations.

information, (3) which categories of perceived protected area values (e.g., recreation, scenic, biological) are best represented in crowdsourced data, and (4) can crowded-sourced data be used to build a robust, predictive model of protected area visitation? In addition, the influence of parks' accessibility from populated areas on their crowdsourced information was assessed using artificial night-light data. We conclude the article by discussing the potential of crowdsourced data for assessing protected area importance from the perspective of spatial data quality.

2. Methods

2.1. Study area

This study covered the state of Victoria, the sixth largest state or territory in Australia with an area of 237,629 km^2 and an estimated population of 5,768,600 (Australian Bureau of Statistics, 2013). Most of Victoria's population is concentrated near the capital city of Melbourne, Australia's second-largest city. About 35% of Victoria is public land estate ("Crown" lands), including state managed national parks, state forests, federally managed commonwealth lands, metropolitan and regional parks, and specialized reserves for the protection of historic and cultural resources. Parks and conservation reserves make up about 50% of all Crown land while state forests comprise about 40% (DEPI, 2013).

Table 4
Spearman's rank correlation coefficients between the variables Visitors (official number of visitors), PPGIS (number of PPGIS value points), Flickr (number of Flickr photos), OSM (number of OSM vertices) and Wikipedia (number of edits to Wikipedia articles), calculated for each of the protected areas (n = 3072). In the correlations shown above the diagonal, observations for which no data was available were given the value of 0.

	Visitors	PPGIS	Flickr	OSM	Wikipedia
Visitors		**0.446*****	**0.364*****	**0.296*****	**0.737*****
PPGIS	**0.623*****		**0.584*****	**0.507*****	**0.430*****
Flickr	**0.465*****	**0.590*****		**0.673*****	**0.317*****
OSM	**0.569*****	**0.559*****	**0.679*****		**0.264*****
Wikipedia	**0.523*****	**0.343*****	**0.342*****	**0.295****	

*p < 0.05, **p < 0.01, ***p < 0.001.
The bold values represent statistically significant correlations.

2.2. Data sets

We used a polygon layer of protected areas in Victoria from the public land classification geodatabase (PLMGEN) maintained by the Victorian Department of Environment and Primary Industries (DEPI). The protected area dataset included 36 protected area types (Table S1). A spatial buffer of 1 km was calculated for each of the protected areas to account for edge effects and to capture nearby areas when intersecting this layer with crowdsourcing layers with lower spatial precision. The popularity and perceived importance of these protected areas was examined using the following sources (Table 1): Visitation, PPGIS, Flickr, Wikipedia and OpenStreetMap.

The mean and maximum values of artificial night-lights as of May 2014 were mapped within each protected area using the Suomi National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) (available from http://ngdc.noaa.gov/eog/viirs/download_monthly.html, accessed January 26th, 2016) as an indicator of human activity such as population centers and industrial activity (Román & Stokes, 2015; Zhang, Levin, Chalkias, & Letu, 2015). VIIRS Day/Night Band acquires visible light at night-time hours at a spatial resolution of 0.75 km (Miller et al., 2013) with radiance values provided in nanoWatts/cm^2/sr.

2.3. Analyses

For each of the protected areas, we identified the number of observations where crowdsourced data were available as a measure of data completeness. Data completeness refers to the presence/

Table 5
Spearman's rank correlation coefficients between the variables of Visitors (official number of visitors), PPGIS (number of PPGIS value points), Flickr (number of Flickr photos), OSM (number of OSM vertices) and Wikipedia (number of edits to Wikipedia articles) calculated for each of the protected areas (n = 3072), normalized by area (dividing the value of each variable by the size of the protected area). In the correlations shown above the diagonal, observations for which no data was available were given the value of 0.

	Visitors/area	PPGIS/area	Flickr/area	OSM/area	Wikipedia/area
Visitors/area		**0.410*****	**0.302*****	**0.150*****	**0.728*****
PPGIS/area	**0.744*****		**0.543*****	**0.356*****	**0.379*****
Flickr/area	**0.705*****	**0.688*****		**0.607*****	**0.241*****
OSM/area	**0.708*****	**0.605*****	**0.658*****		**0.077*****
Wikipedia/area	**0.728*****	**0.649*****	**0.647*****	**0.722*****	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
The bold values represent statistically significant correlations.

Table 6
Spearman's rank correlation coefficients between the variable of Visitors/area (official number of visitors) and the variables of: PPGIS (number of PPGIS value points), Flickr (number of Flickr photos), OSM (number of OSM vertices) and Wikipedia (number of edits to Wikipedia articles) calculated for each of the protected areas (n = 3072), normalized by area (dividing the value of each variable by the size of the protected area). Correlations were calculated only for those categories of protected areas where there was data in each of the VGI variables for at least seven parks (Table S1).

	National park - schedule 2, National parks act (n = 39)	State park - schedule 2B, National parks act (n = 22)	Other park - schedule 3, National parks act (n = 6)	Marine national park - schedule 7, National parks act (n = 5)	Metropolitan park (n = 9)	Reservoir park (n = 6)
PPGIS/area	**0.870*****	**0.763*****	**0.636***	0.029	0.228	0.429
Flickr/area	**0.882*****	**0.639****	0.600	0.571	−0.007	0.071
OSM/area	**0.874*****	0.406	**0.791****	0.657	0.179	0.393
Wikipedia/area	**0.668*****	**0.732*****	0.029	0.429	0.167	0.143

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
The bold values represent statistically significant correlations.

absence of feature entities as well as to the attributes associated with the feature entities (Veregin, 1999). Completeness can thus be determined by comparison of actual entities/attributes to the abstract or theoretical set of all feature objects or attributes of interest. Data was aggregated and analyzed by protected area type, in addition to the analysis at the level of individual protected areas.

As a surrogate for the accessibility of protected areas from Victoria's largest city, the mean distance of all grid cells within each of the protected areas was calculated from the brightest grid cell in Melbourne, Victoria. The urban system of Australia is very centralized, with just a small number (five) of cities which are very large (Pumain & Moriconi-Ebrard, 1997) including the capital city of Melbourne that contains 76% of the state's population (Australian Bureau of Statistics, 2016). In addition, Australian cities are spread out with most suburbs comprising low-density development (Black, 1996). Therefore, the CBD of Melbourne, whose location was determined here based on its night-time brightness, can serve as a good measure for identifying the population center of the State of Victoria.

Spearman rank correlations (R_S) were calculated in XLSTAT (2015, Addinsoft®) to quantify and test the significance of the correspondence between the variables of visitation, PPGIS, Flickr, Wikipedia, OSM, VIIRS night lights, and distance from the brightest grid cell in Melbourne. Correlations were examined for both raw values (e.g., number of Flickr photos) and standardized values as determined by dividing the raw values by protected area size (e.g., number of Flickr photos/km^2). Finally, stepwise multiple regression models were run using XLSTAT (2015, Addinsoft®) to assess the potential for VGI variables to be used as proxies for estimating visitor counts to protected areas. The probability for entry into the stepwise regression model was set as 0.05, and the probability for removal from the stepwise regression model was set as 0.1 (these are commonly used default values; Colman & Pulford, 2011). Models were run with and without log transforming the variables, and using both raw and standardized counts by protected area size.

3. Results

3.1. Crowdsourcing information by type of protected area (n = 36)

Overall in our dataset, there were 3072 protected areas with a total area of 87,733 km^2 and a median area of 6.6 km^2, consisting of 36 different types of protected areas. National parks comprised 45.2% of the total area (but just 1.5% of the total number of protected areas), and bushland reserves comprised 11.4% of the total area (but 47.3% of the total number of protected areas) (Table S1). Out of the 3072 protected areas, VGI data was available at varying degrees of completeness for analysis: OSM (n = 2774 protected areas), Flickr photos (n = 1264 protected areas), PPGIS (n = 749 protected areas), formal visitation statistics (n = 163 protected areas for which there was a match between the protected areas GIS dataset and the names of protected areas given by Parks Victoria, 2011) and Wikipedia articles (n = 127 protected areas) (Fig. 1; Table S1). The most abundant crowdsourced information by protected area type was for National Parks and Metropolitan Parks (Table S2). Most of the visitors (32.4 million according to available official visitation data) were in National Parks (46.9%) and Metropolitan Parks (31.3%). Most of the PPGIS place value points (26,100) were mapped in National Parks (50.4%) and Metropolitan Parks (12.2%). Most of the Flickr photos (589,944) were identified in Port and Coastal Facilities (35%), Metropolitan Parks (25.2%), and National Parks (13.1%). Most edits of Wikipedia articles (6,966) were related to National Parks (47.7%) and State Parks (13.7%). Most vertices within OSM data (742,968) were found in Metropolitan Parks (17.9%), National Parks (15.9%), and Bushland Reserves (15.8%) (Table S2).

An examination of the correspondence between crowdsourcing data by protected area type (n = 36) showed that the variables of PPGIS, Flickr and OSM were highly correlated and significant ($R_s \geq 0.9$; Table 2). Official visitation statistics were significantly correlated with all crowdsourcing variables; however, Wikipedia was not significantly correlated with the other three crowdsourcing data types (Table 2). Similar patterns emerged when examining the

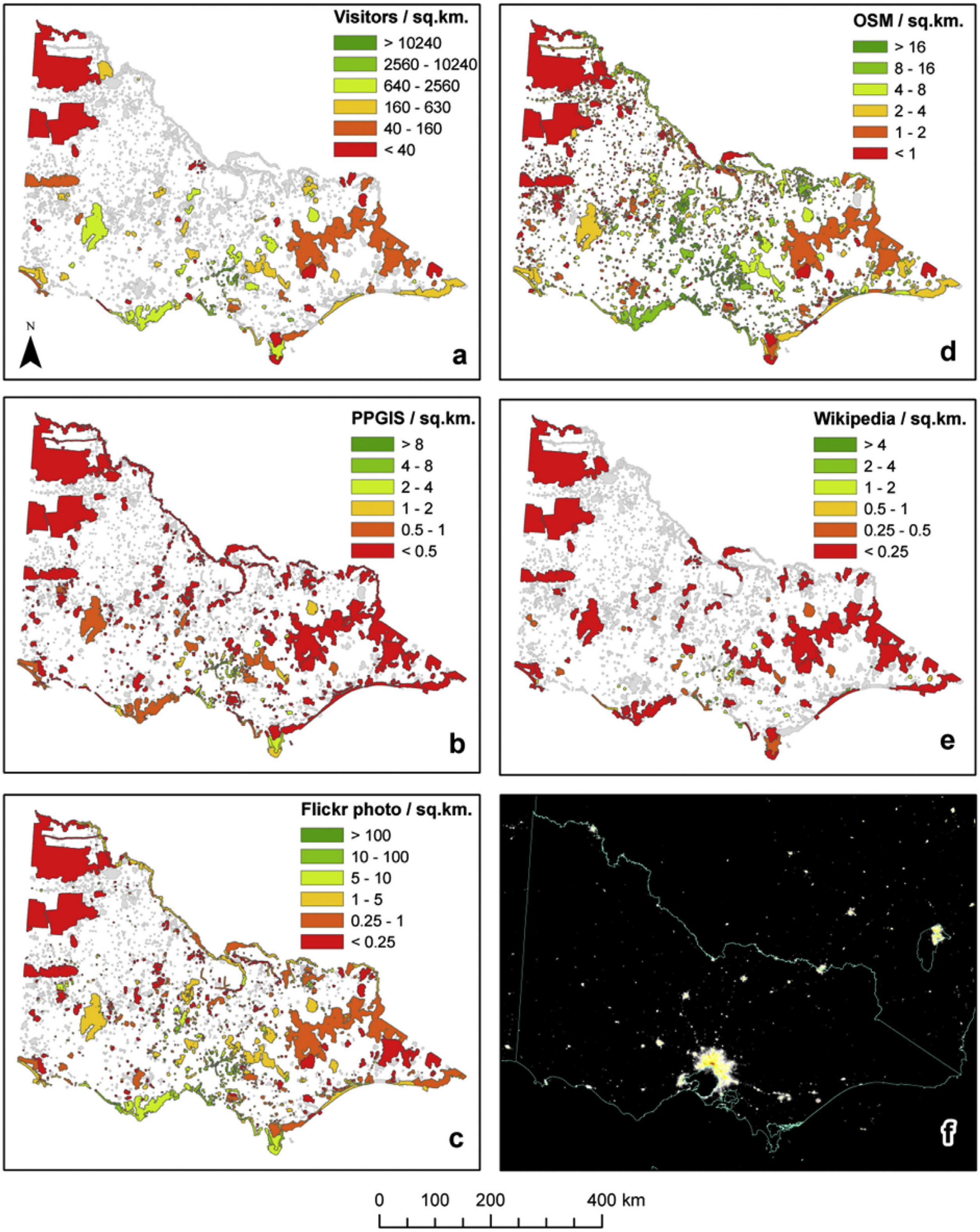

Fig. 2. Spatial patterns of Victoria's protected areas popularity as measured using: a) official visitation statistics per area; b) PPGIS value points per area; c) Flickr photos per area; d) OpenStreetMap vertices per area; and e) Wikipedia article edits per area. f) VIIRS May 2014 night-time satellite image of Victoria.

total numbers of crowdsourced data (e.g., the total number of Flickr photos within a certain type of protected areas, not the number of protected areas of a certain type with Flickr photos) with Wikipedia being the crowdsourcing data type least correlated with other crowdsourcing information types (Table 3).

Table 7
Spearman's rank correlation coefficients between the variables Visitors (official number of visitors), PPGIS (number of PPGIS value points), Flickr (number of Flickr photos), OSM (number of OSM vertices) and Wikipedia (number of edits to Wikipedia articles) calculated for each of the protected areas (n = 3072), and the mean brightness (from the VIIRS composite image of May 2014) and the distance from the brightest grid cell in the May 2014 image.

	Mean brightness, VIIRS May 2014	Distance from brightest grid cell
Visitors	**0.179***	**−0.212****
PPGIS	**0.313*****	**−0.247*****
Flickr	**0.598*****	**−0.351*****
OSM	**0.865*****	**−0.395*****
Wikipedia	**0.227***	−0.133
Visitors/area	**0.606*****	**−0.687*****
PPGIS/area	**0.604*****	**−0.564*****
Flickr/area	**0.676*****	**−0.466*****
OSM/area	**0.885*****	**−0.472*****
Wikipedia/area	**0.581*****	**−0.625*****

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
The bold values represent statistically significant correlations.

3.2. Crowdsourcing information analyzed for individual protected areas (n = 3072)

Statistically significant correlations were found between all data sources for protected areas, either when examined with raw numbers (Table 4) or when normalized by area (Table 5). The statistical correlations between official visitation statistics and the four crowdsourcing variables were higher after normalizing by area and excluding protected areas without data (Table 5). Statistical correlations between visitor counts and the four VGI variables varied as a function of the type of protected area (Table 6); correlations were larger and statistically significant between all variables for National Parks and statistically significant between three variables for State Parks (Table 6). The correlations with metropolitan and reservoir parks were smaller (less than 0.5) and non-significant (Table 6).

Using raw counts to map the popularity of Victoria's protected areas, clear spatial patterns were not obvious (Fig. 1). However, after normalizing the five VGI measures by protected area size, a clear pattern emerged, with protected areas that were closer to Melbourne (the capital of Victoria, and home to 74% of Victoria's population) being more popular (Fig. 2; Melbourne is the large bright area in the bottom-center of Fig. 2f). Indeed, statistically significant correlations were found showing that protected areas further away from Melbourne were less popular (Table 7), and protected areas which were brighter (i.e., near artificial night-light sources) were more popular (Table 7). When analyzing the percent of PPGIS value points within protected areas by the types of values mapped, protected areas were more popular (i.e., more visited, having more Flickr photos, more OSM vertices and more Wikipedia article edits) if they were perceived as providing direct use values as identified in PPGIS (therapeutic/health, scenic/aesthetic, recreation, learning/education/research, heritage/culture), whereas protected areas which were perceived as providing indirect use values (wilderness/pristine, intrinsic/existence, biological/conservation) were less popular (Fig. 3).

In the stepwise multiple regression models, the explanatory power of the crowdsourced variables for predicting visitor counts was high, ranging between 0.48 and 0.79 (Table 8; Fig. 4). Out of the 10 models tested, the crowdsourced variable most frequently

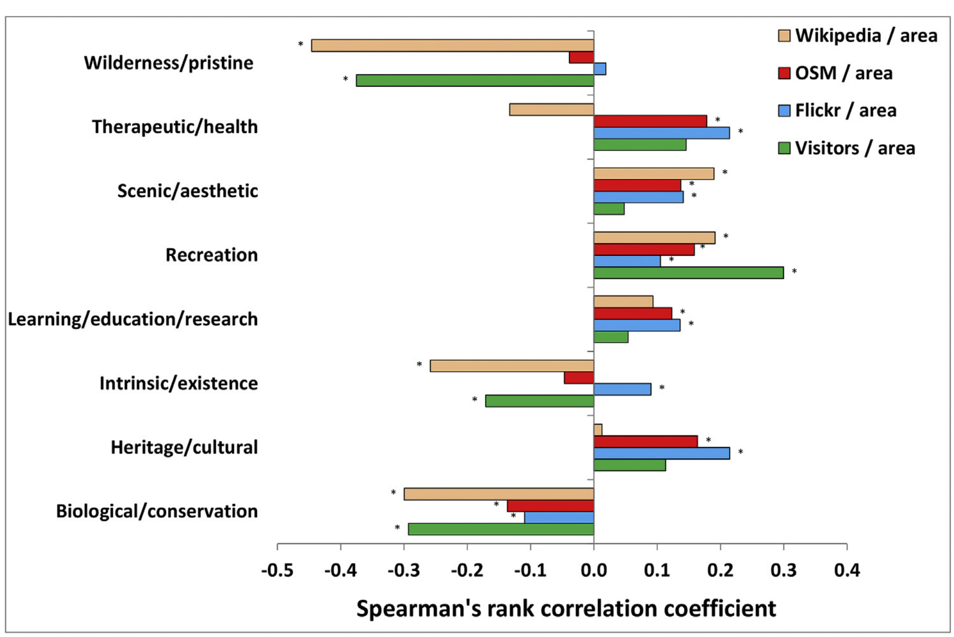

Fig. 3. Spearman's rank correlation coefficients for PPGIS place values and counts of Wikipedia (number of edits to Wikipedia articles), OSM (number of OSM vertices), Flickr (number of Flickr photos) and Visitors (official number of visitors), calculated for each of the protected areas, normalized by area (count divided by size of protected area). Statistically significant correlations are shown by an asterisk.

Table 8
Multiple stepwise linear regression models, using VGI variables to predict visitor numbers in protected areas. The number of observations in all models was n = 102. Models vary by the number of variable used, whether they were log transformed or not, and whether they were normalized by area size of the protected areas. The values for each of the independent variables in the table show their standardized coefficients in the model (if they entered it). Blank cells represent variables which were not included within a multiple regression analysis. Cells with values of 0.000 represent variables which were included in the statistical analysis but did not enter the stepwise regression model.

	Dependent variable	Visitors	Log visitors	Visitors / area	Log visitors / area	Visitors	Log visitors	Visitors / area	Log visitors / area	Visitors	Visitors	Visitors	Visitors
	Independent variables	First 4 variables	First 4 variables log transformed	First 4 variables / area	First 4 variables / area log transformed	All 12 variables	All 12 variables, first 4 variables log transformed	All 12 variables, first 4 variables / area	All 12 variables, first 4 variables / area log transformed	11 variables, PPGIS values as percentages	11 variables, PPGIS values as counts	8 variables, PPGIS values as percentages	8 variables, PPGIS values as counts
	Adjusted R^2	**0.780**	**0.482**	**0.788**	**0.686**	**0.780**	**0.571**	**0.788**	**0.718**	**0.780**	**0.780**	**0.013**	**0.637**
	Total PPGIS	0.000	**0.379**	0.499	**0.548**	0.000	**0.563**	0.499	**0.538**				
	Flickr	0.416	0.000	**0.620**	0.000	0.416	0.000	**0.620**	0.000	0.416	0.416		
	OSM	**0.485**	0.213	-0.282	0.000	**0.485**	0.000	-0.282	0.000	**0.485**	**0.485**		
	Wikipedia	0.172	0.306	0.160	0.356	0.172	0.379	0.160	0.319	0.172	0.172		
PPGIS counts	% Biological					0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	% Heritage					0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	% Intrinsic					0.000	0.000	0.000	0.000	0.000	0.000	0.000	**-1.243**
	% Learning					0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.498
	% Recreation					0.000	0.226	0.000	0.189	0.000	0.000	0.000	0.697
	% Scenic					0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.102
	% Therapeutic					0.000	0.000	0.000	0.000	0.000	0.000	**0.141**	0.739
	% Wilderness					0.000	-0.213	0.000	0.000	0.000	0.000	0.000	**-1.242**

The bold values show the predictor variable with the highest coefficient in each column. Grey cells with values of 0.000 represent variables which were included in the statistical analysis but did not enter the stepwise regression model.

included in the model was Wikipedia (10), followed by PPGIS (7), OSM (7), and Flickr (6) (Table 8). The crowdsourced variable which consistently had the largest standardized coefficient was PPGIS (4; between 0.38 and 0.56), followed by OSM (4; 0.48), and Flickr (2; 0.62) (Table 8). When including the percent of PPGIS value points within protected areas by the type of values mapped, only two PPGIS values were significant: recreation values showing a positive relationship to visitation, and wilderness values showing a negative relationship to visitation (Table 8). When constructing a regression model with PPGIS values alone, six out of eight PPGIS values entered the model, with intrinsic and wilderness values showing a negative relationship to visitation (Table 8). However, in the model with the largest adjusted R^2 (0.79), none of the PPGIS values entered the stepwise regression model (Table 8).

4. Discussion

Our study found strong and significant correlations between all crowdsourced data and visitation statistics, demonstrating the potential to use crowdsourced data to characterize the social and perceived importance of protected areas and as a proxy for visitation statistics. However, there were differences in the strength of correlations, data quality, and information characterized by the different types of crowdsourced data. These differences can be attributed to the range of methods for deriving the data, from social survey methods using PPGIS to geotagged Flickr “big data” as proxies of public preferences. These differences and the advantages and disadvantages of the crowdsourced data tested in this study are summarized in Table 9.

4.1. Strength of correlation between crowdsourced data and visitation

Our results suggest that a wide range of crowdsourced data can potentially be used as a surrogate for visitation statistics. While previous studies have shown that Flickr photo density are correlated with visitation statistics (e.g., Levin et al., 2015; Wood, Guerry, Silver, & Lacayo, 2013), this is the first study to demonstrate the potential utility of OSM and Wikipedia crowdsourced data for this task. However, Wikipedia generally had lower correlations with other crowdsourced data and visitation statistics suggesting lower suitability for this purpose. The patterns of spatial distribution of visitation and crowdsourced data were also similar with protected areas closer to Melbourne and populated areas showing larger counts.

The relatively high correlations and reasonably strong, predictive models from crowdsourced data suggest their potential utility for assessing protected area visitation. However, these findings must be tempered by data limitations, the most important being the completeness of crowdsourced data (Balmford et al., 2015). The coverage of crowdsourced data across Victoria was spotty, with coverage diminishing as a function of distance from Melbourne, the primary population center. For example, Flickr data was available for less than half of the protected areas in the study area and for most protected areas, no PPGIS place values were mapped.

In this study, the visitation statistics were assumed to be accurate and reliable, but obtaining and accurately quantifying visitation is more challenging than it may appear. As noted by Buckley (2009), few countries have accurate visitor numbers, reliable counts require continuously staffed access roads with no other entry, automated counters are expensive and inaccurate, and visitation numbers can vary daily by orders of magnitude making the comparison of single-day counts unreliable.

4.2. Utility of crowdsourced information by protected area type (e.g., national versus metropolitan parks)

Crowdsourced data was most complete for national parks that attract greater crowdsourcing effort by virtue of their size, status, and familiarity, and more limited for state, and especially metropolitan parks. Indeed, using multiple evaluation criteria, national parks in Victoria were rated higher than state parks (Deng, King, &

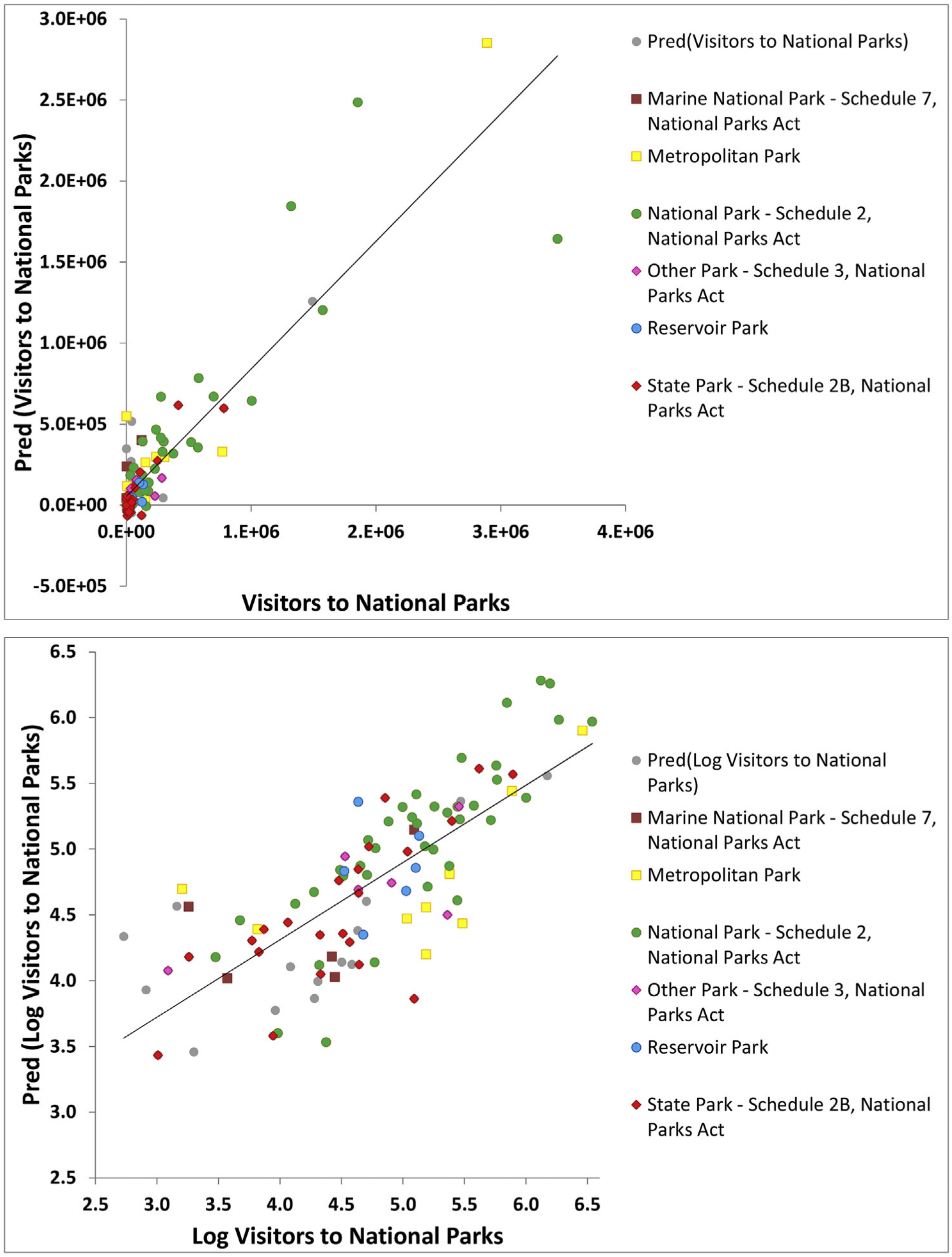

Fig. 4. Scatter plots showing observed vs. predicted numbers of visitors (before and after log transformation), based on 12 variables in stepwise multiple linear regression.

Bauer, 2002). Thus, the potential utility of crowdsourced data for identifying protected area importance appears highest for popular national parks, and lower for state and metropolitan parks. Visitation to national and state parks was significantly correlated with crowdsourced variables, but metropolitan parks were not significantly correlated despite some metropolitan parks being among the most visited. While crowdsourced data is often evaluated for its potential use in data poor regions, in the case of protected areas, crowdsourced data would appear to be more useful for protected areas that are likely to be more data rich, in this case, national parks.

4.3. *Types of protected area values (e.g., recreation, scenic, biological) best represented in crowdsourced data*

The strong correlations between crowdsourced data and visitation statistics suggest crowdsourced data may provide proxies for assessing direct use values associated with protected areas such as the enjoyment of scenery, recreation, learning opportunities, appreciation of history and culture, and therapeutic benefits. These findings are similar to other studies showing that perceived public land values associated with direct uses (e.g., recreation) tend to be located closer to populated areas than non-use values (Brown, Reed, & Harris, 2002; Brown, Weber, et al., 2014). Indirect-use values (biological conservation and wilderness) and non-use values (intrinsic/existence) were found to be negatively correlated with crowdsourced data and thus non-PPGIS crowdsourced data appear to be an unreliable proxy for the full range of values and importance associated with protected areas. Given that many protected areas such as national parks have a dual and potentially conflicting statutory mandate to provide for both visitor enjoyment (direct use values) and natural and cultural landscape protection (Brown & Weber, 2011), a distinctive advantage of PPGIS data collection over other crowdsourcing types is the capacity to assess

Table 9
Summary of data quality and advantages and disadvantages of VGI, PPGIS and visitation variables for characterizing social values for protected areas. Each of the variables is denoted whether it can be characterized as "Big Data" (BD), active or passive collection method (A, P), structured or unstructured data types (S, US) – see Discussion.

Variable	Target	Original spatial unit	Spatial resolution/ precision	Thematic resolution/fidelity	Completeness (% of protected areas with data)	Advantages	Disadvantages
OSM (BD, P, S)	Proxy for protected area importance representing accessibility	Vector polylines	GPS <10 m or map scale	On-screen digitization using satellite images	90.3%	• Free • Spatially explicit • High completeness	• Unclear meaning
Flickr (BD, P, S)	Proxy for protected area importance represented by actual visits and photos uploaded	Point data	GPS <10 m	Observation	41.1%	• Free • Spatially and temporally explicit • Visitation hotspots • Actual visitation • Allows for image analysis and content analysis of tags • User characteristic analysis (i.e. visitor origins) • High completeness	• Unclear meaning • Confounding factors –Accessibility
PPGIS (A, S)	Place values held by sampled individuals and groups	Point data	Mapping scale dependent	Multiple values, special places, development preferences, activities, or any spatial attribute of research interest	24.4%	• Spatially explicit • Value hotspots • Robust scientific methods e.g. sampling design • Multiple values (use and non-use values) • Clear meaning	• Cost
Visitation count (A, S)	Proxy for protected area importance represented by registered visitation	Protected area boundary	Protected area boundary	Official records	5.3%	• Actual visitation • Clear meaning	• Cost • Confounding factors – visitor numbers affected by park capacity • Accessibility
Wikipedia(BD, P, US)	Proxy for protected area importance	Protected area boundary	Protected area boundary	Text	4.1%	• Free • Allows for content analysis	• Unclear meaning • Low completeness

non-use values of protected areas, which in some cases, may exceed the value of direct-use values. Our analysis also showed that the type of protected area, which partly reflect these dual use values (e.g., national versus metropolitan parks), affect the strength of correlations. However, this analysis is confounded by the low sample size associated with some categories of protected areas (e.g., metropolitan parks). Whereas previous PPGIS research assessed landscape values to identify potential biophysical correlates with place values, this research examined whether behavioural variables associated with place (i.e., park visitation, VGI contributions about place) were also related to mapped landscape values in PPGIS.

4.4. Crowdsourced data as a predictor of protected area visitation

While the predictive model of visitation statistics had large R^2 values, the explanatory power of crowdsourced variables identified by the model varied when using either raw or normalized data (log and area). The differences in the models are related to the distribution of data values which included a small number of very popular parks (mainly national parks). Log transforming or normalizing the data by area reduces the influence of these parks with larger visitor counts on correlations.

4.5. Crowdsourcing data quality considerations

Most of the research on the quality/credibility of VGI have focused on either accuracy in the traditional GIS land cover mapping data quality sense (Foody et al., 2013; Haklay, 2010) or credibility-as-perception, where VGI data is evaluated in terms of the expertise and perspectives of the volunteers (Flanagin & Metzger, 2008; Spielman, 2014). The use of VGI data in this study represents a departure from the original intention of crowdsourcing for accessing "collective intelligence" to one of identifying "collective values" or the perceived importance of protected areas. As noted by Brown, Weber, and de Bie (2015), the credibility of crowdsourced data cannot be fully addressed without reference to the intended purpose and potential use of the spatial data. When assessing the quality of crowdsourced data for the purpose of measuring the importance of protected areas, the concepts of sample representativeness and sampling error become critically important, along with data completeness and spatial accuracy (Devillers & Jeansoulin, 2010; Lechner et al., 2014).

Data from VGI are less structured in comparison to data gathered using PPGIS methods that often use systematic population sampling. Given the non-systematic, passive sampling that characterizes VGI systems, it is not possible to know the representativeness of the data collected. In contrast, PPGIS data collection is typically an active and structured data collection process that seeks to identify place values using pre-defined survey questions for a given study population. An important advantage of PPGIS is its foundation in social survey practice where measurement validity, response bias, and representativeness of data can be addressed (Brown, 2004). PPGIS often involves active data collection in specific projects and thus is more expensive to implement (Table 9). Indeed, identifying less expensive correlates of PPGIS data from VGI was a key motivation for undertaking the analyses in this paper.

The advantage of passive data collection approaches, utilizing big data (OSM and Flickr), is the ability to collect greater quantities of data, resulting in greater coverage and thus completeness. Data gathered through VGI reflect a broad range of place-based features and visitor experiences. The importance of specific place values (e.g., scenery) could potentially be inferred from VGI data such as Flickr (e.g. Li et al., 2009; Yanai et al., 2009) and there is also the potential to identify the characteristics of contributors through conducting content analysis of user contributions.

Spatial precision/resolution was much higher for OSM, Flickr and PPGIS data than for visitation statistics and Wikipedia. Higher spatial resolution data allows for the identification of hot-spots of interest within protected areas, including the locations where people actually visit. In contrast, visitation statistics and Wikipedia provide information at a coarse spatial scale for the entire protected area. However, using gazetteers (Gao et al., 2014), the spatial information available within Wikipedia articles can be enhanced to identify specific locations within protected areas.

A key source of uncertainty common to all crowdsourced methods is collection bias, where accessibility, proximity, and familiarity strongly influence what people contribute to crowdsourced databases. The natural tendency for individuals to engage in geographic or spatial discounting (mapping attributes closer to home), in combination with the uneven spatial distribution of human settlements, will necessarily result in clustered and non-uniform crowdsourced data vis-à-vis protected areas. Although web interfaces allow users to enter observations anywhere in the world, the amount of user-contributed data in all cases (PPGIS, OSM and Wikipedia) was biased toward populated areas within the state of Victoria. Indeed, accessibility is known to be a key factor driving visitation rates (Balmford et al., 2015; Levin et al., 2015) while familiarity and proximity are key variables that determine what people map in PPGIS systems (Brown & Kyttä, 2014), as well as the contributions to Wikipedia articles (Hardy et al., 2012). Thus, crowdsourcing data for protected areas will be biased toward direct use values in locations that are proximate, accessible, and familiar.

5. Conclusion

Crowdsourced data through VGI has the potential to provide valuable information on place values and visitation for conservation planners and protected area managers. OSM and Flickr, in particular, provide a free, up-to-date, and high spatial and temporal resolution information source. However, as our analyses revealed, each crowdsourced database has limitations in terms of spatial data quality and sampling bias. While PPGIS and reported visitation data provide less ambiguous measures for assessing the importance of protected areas, VGI data have greater coverage (in particular Flickr and OSM). Validating these data sources and addressing uncertainty in VGI data represents an important area of future research that will be necessary before crowdsourced data achieves acceptance for use in protected area planning and management, and for quantifying and qualifying the characteristics and values of protected areas.

Acknowledgments

N.L. was partly supported by ARC CEED.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeog.2016.12.009.

References

Australian Bureau of Statistics. (2013). *QuickStats*. Available from: http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/2?opendocument&navpos=220 (Accessed 27 April 2014).
Australian Bureau of Statistics. (2016). *3218.0-Regional population growth, Australia, 2014-15*. Available from: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/3218.0Main%20Features252014-15?opendocument&tabname=Summary&prodno=3218.0&issue=2014-15&num=&view= (Accessed 09 October 2016).
Balmford, A., Green, J. M., Anderson, M., Beresford, J., Huang, C., Naidoo, R., et al.

(2015). Walk on the wild side: Estimating the global magnitude of visits to protected areas. *PLoS Biology, 13*(2), e1002074.

Beverly, J. L., Uto, K., Wilkes, J., & Bothwell, P. (2008). Assessing spatial attributes of forest landscape values: An internet-based participatory mapping approach. *Canadian Journal of Forest Research, 38*(2), 289–303.

Birenboim, A., & Shoval, N. (2016). Mobility research in the age of the smartphone. *Annals of the American Association of Geographers, 106*(2), 283–291. http://dx.doi.org/10.1080/00045608.2015.1100058.

Black, J. (1996). Higher density housing and transport in Australian cities. In Y. Hayashi, & J. Roy (Eds.), *Transport, land-use and the environment* (pp. 55–80). U.S.: Springer.

Brown, G. (2004). Mapping spatial attributes in survey research for natural resource management: Methods and applications. *Society & Natural Resources, 18*(1), 1–23.

Brown, G. (2015). Engaging the wisdom of crowds and public judgement for land use planning using public participation geographic information systems. *Australian Planner, 52*(3), 199–209.

Brown, G., & Alessa, L. (2005). A GIS-based inductive study of wilderness values. *International Journal of Wilderness, 11*(1), 14–18.

Brown, G., Kelly, M., & Whitall, D. (2014). Which 'public'? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. *Journal of Environmental Planning and Management, 57*(2), 190–214.

Brown, G., & Kyttä, M. (2014). Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. *Applied Geography, 46*, 122–136.

Brown, G., Reed, P., & Harris, C. C. (2002). Testing a place-based theory for environmental evaluation: An Alaska case study. *Applied Geography, 22*(1), 49–77.

Brown, G., & Weber, D. (2011). Public participation GIS: A new method for use in national park planning. *Landscape and Urban Planning, 102*(1), 1–15.

Brown, G., & Weber, D. (2012). Measuring change in place values using public participation GIS (PPGIS). *Applied Geography, 34*, 316–324.

Brown, G., Weber, D., & De Bie, K. (2014). Assessing the value of public lands using public participation GIS (PPGIS) and social landscape metrics. *Applied Geography, 53*, 77–89.

Brown, G., Weber, D., & de Bie, K. (2015). Is PPGIS data good enough? An empirical evaluation of the quality of PPGIS crowd-sourced spatial data for conservation planning. *Land Use Policy, 43*, 228–238.

Buckley, R. (2009). Parks and tourism. *PLoS Biology, 7*, 6–7. http://dx.doi.org/10.1371/journal.pbio.1000143.

Buckley, R., Robinson, J., Carmody, J., & King, N. (2008). Monitoring for management of conservation and recreation in Australian protected areas. *Biodiversity and Conservation, 17*(14), 3589–3606.

Colman, A. M., & Pulford, B. D. (2011). *A crash course in SPSS for Windows: Updated for versions 14, 15, and 16*. John Wiley & Sons.

Comber, A., See, L., Fritz, S., Van der Velde, M., Perger, C., & Foody, G. (2013). Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation, 23*, 37–48.

Deng, J., King, B., & Bauer, T. (2002). Evaluating natural attractions for tourism. *Annals of Tourism Research, 29*(2), 422–438.

Denzin, N. (2006). *Sociological methods: A sourcebook* (5th ed.). Chicago, IL: Aldine.

DEPI—Department of Environmental and Primary Industries. (2013). *Victorian Crown land area statement*. Available from: http://www.depi.vic.gov.au/__data/assets/pdf_file/0006/199068/FactSheet_CrownLandVictoria_20130821FINAL.pdf (Accessed 22 April 2014).

Devillers, R., & Jeansoulin, R. (2010). *Fundamentals of spatial data quality*. Hoboken: Wiley.

Dudley, N. (Ed.). (2008). *Guidelines for applying protected area management categories*. Gland, Switzerland: IUCN, 86pp.

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

Fast, V., & Rinner, C. (2014). A systems perspective on volunteered geographic information. *ISPRS International Journal of Geo-Information, 3*, 1278–1292. http://dx.doi.org/10.3390/ijgi3041278.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal, 72*(3–4), 137–148.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., et al. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS, 17*, 847–860. http://dx.doi.org/10.1111/tgis.12033.

Gao, S., Li, L., Li, W., Janowicz, K., & Zhang, Y. (2014). Constructing gazetteers from volunteered big geo-data based on Hadoop. *Computers, Environment and Urban Systems*. http://dx.doi.org/10.1016/j.compenvurbsys.2014.02.004.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

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

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers, 104*(4), 746–764.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordinance Survey datasets. *Environment and Planning B Planning & Design, 37*(4), 682.

Haklay, M., Singleton, A., & Parker, C. (2008). Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass, 2*(6), 2011–2039.

Hardy, D., Frew, J., & Goodchild, M. F. (2012). Volunteered geographic information production as a spatial process. *International Journal of Geographical Information Science, 26*(7), 1191–1212.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine, 14*(6), 1–4.

Lechner, A. M., Raymond, C. M., Adams, V. M., Polyakov, M., Gordon, A., Rhodes, J. R., et al. (2014). Characterizing spatial uncertainty when integrating social data in conservation planning. *Conservation Biology, 28*, 1497–1511. http://dx.doi.org/10.1111/cobi.12409.

Levin, N., Kark, S., & Crandall, D. J. (2015). Where have all the people gone? Enhancing global conservation using night lights and social media. *Ecological Applications, 25*(8), 2153–2167.

Li, Y., Crandall, D. J., & Huttenlocher, D. P. (2009, September). Landmark classification in large-scale image collections. In *ICCV* (pp. 1957–1964).

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182.

Miller, S. D., Straka, W., Mills, S. P., Elvidge, C. D., Lee, T. F., Solbrig, J., et al. (2013). Illuminating the capabilities of the suomi national polar-orbiting partnership (NPP) visible infrared imaging radiometer suite (VIIRS) day/night band. *Remote Sensing, 5*(12), 6717–6766.

Mukherjee, F. (2015). Public participatory GIS. *Geography Compass, 9*(7), 384–394.

National Center for Geographic Information and Analysis (NCGIA). (1996a). *Summary report: GIS and Society Workshop*. SouthHaven, MN, 2–5 March, 1996.

National Center for Geographic Information and Analysis (NCGIA). (1996b). *Summary report: Public participation GIS Workshop*. Orono, ME, 10–13 July, 1996.

Parks Victoria. (2011). *Parks Victoria visitation data*. Parks-Victoria-Visitation-Data.xlsx Excel file Available from: http://parkweb.vic.gov.au (Accessed 16 January 2016).

Pumain, D., & Moriconi-Ebrard, F. (1997). City size distributions and metropolisation. *Geojournal, 43*(4), 307–314.

Van Riper, C. J., & Kyle, G. T. (2014). Capturing multiple values of ecosystem services shaped by environmental worldviews: A spatial analysis. *Journal of Environmental Management, 145*, 374–384.

Van Riper, C. J., Kyle, G. T., Sutton, S. G., Barnes, M., & Sherrouse, B. C. (2012). Mapping outdoor recreationists' perceived social values for ecosystem services at Hinchinbrook Island National Park, Australia. *Applied Geography, 35*(1), 164–173.

Román, M. O., & Stokes, E. C. (2015). Holidays in lights: Tracking cultural patterns in demand for energy services. *Earth's Future, 3*(6), 182–205.

Sester, M., Arsanjani, J. J., Klammer, R., Burghardt, D., & Haunert, J. H. (2014). Integrating and generalising volunteered geographic information. In *Abstracting geographic information in a data rich world* (pp. 119–155). Springer International Publishing.

Shi, W. (2010). *Principles of modeling uncertainties in spatial data and spatial analyses*. Boca Raton, FL: CRC Press.

Shoval, N. (2007). Sensing human society. *Environment and Planning B: Planning and Design, 34*(2), 191–195.

Spielman, S. E. (2014). Spatial collective intelligence? Credibility, accuracy, and volunteered geographic information. *Cartography and Geographic Information Science, 41*(2), 115–124.

Sui, D., Elwood, S., & Goodchild, M. (2012). *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice*. Springer Science & Business Media.

Veregin, H. (1999). Data quality parameters. Geographical information systems. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Principles and technical issues* (2nd ed., Vol. 1, pp. 177–189). New York: John Wiley & Sons.

Willemen, L., Cottam, A. J., Drakou, E. G., & Burgess, N. D. (2015). Using social media to measure the contribution of Red List species to the nature-based tourism potential of African protected areas. *PloS one, 10*(6), e0129785.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports, 3*, 2976. http://dx.doi.org/10.1038/srep02976.

Yanai, K., Yaegashi, K., & Qiu, B. (2009). Detecting cultural differences using consumer-generated geotagged photos. In *Proceedings of the 2nd international workshop on location and the web*. LOCWEB '09, pages 12:1–12:4, New York, NY, USA. ACM.

Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). The most controversial topics in Wikipedia: A multilingual and geographical analysis. In P. Fichman, & N. Hara (Eds.), *Global Wikipedia: International and cross-cultural issues in online collaboration*. Scarecrow Press.

Zhang, Q., Levin, N., Chalkias, C., & Letu, H. (2015). Nighttime light remote sensing – monitoring human societies from outer space. In P. S. Thenkabail (Ed.), *Chapter 11 in remote sensing handbook* (Vol. 3, pp. 289–310).