Applied Geography 90 (2018) 44–54

Contents lists available at ScienceDirect

Applied Geography

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

Digital footprints: Incorporating crowdsourced geographic information for protected area management

Chelsey Walden-Schreiner[a,b,*], Yu-Fai Leung[a,b,c], Laura Tateosian[b,c]

[a] North Carolina State University, Dept. of Forestry & Environmental Resources, Campus Box 8008, Raleigh, NC, 27695, USA
[b] North Carolina State University, Dept. of Parks, Recreation, & Tourism Management, Campus Box 8004, Raleigh, NC, 27695, USA
[c] North Carolina State University, Center for Geospatial Analytics, Campus Box 7106, Raleigh, NC, 27695, USA

1. Introduction

Environmental changes and biodiversity loss attributed to human activities increased more rapidly in the last century than during any other time in recorded history (Millenium Ecosystem Assessment, 2005; Pimm et al., 2014). This unprecedented rate and scale of human impact and ecological footprint ushered proposals to name a new geological epoch, the Anthropocene (Malhi, 2017; Steffen et al., 2011). The extent and complexity of human-environment interactions requires innovative approaches to address multidimensional human-environment interactions, fostering a paradigm shift in how such interactions are measured, monitored, and addressed (Harden et al., 2014). Dynamic feedback and interacting effects leading to increased complexity have resulted in studies analyzing such interactions as coupled systems (An & López-Carr, 2012).

Geography and Geographic Information Science (GIScience) are central to exploring, understanding, and modelling the human footprint (Sanderson et al., 2002) and human-environment interactions (Zvoleff & An, 2014), coupling social and ecological datasets through geographic location for integrated analysis. Increasing amounts, access, and accuracy of spatial data available for integrated analysis has fostered a paradigm shift in GIScience (Elwood, Goodchild, & Sui, 2012). Specifically, how spatial data are produced and consumed have rapidly expanded to new users and platforms (Goodchild, Aubrecht, & Bhaduri, 2017; Sui, Elwood, & Goodchild, 2013).

The ubiquitous and accessible nature of big (geo)data facilitated a proliferation of non-traditional spatial data products and innovative data collection methods in research exploring human-environment interactions (Longley & Adnan, 2016; See et al., 2016; Senaratne, Mobasheri, Ali, Capineri, & Haklay, 2017). Mobile phones and other portable electronic devices provide individuals with sensors of their environments capable of generating and sharing large amounts of geospatial data. Digital footprints, or the "digital traces each of us leaves behind as we conduct our lives," (Weaver & Gahegan, 2007, p. 324), can be considered as a form of volunteered geographic information (VGI). Digital footprints captured by mobile devices and shared through social media platforms, like geotagged photos, offer an opportunity to study human-environment interactions central to geography at multiple spatial and temporal resolutions (Wang, Guo, Fu, & Li, 2014). This research explores VGI originating outside of formal research endeavors to quantify human presence and model infrastructure and environment interactions in areas designated for biodiversity conservation.

Protected areas (PAs) are designated terrestrial and marine landscapes formally managed to conserve nature, ecosystem services, and cultural values. PAs are core components in the strategy to minimize biodiversity loss globally, and human (i.e., visitor) data are often limited due to myriad competing research needs and challenges associated with visitor data collection. Sampling difficulties and time intensive methods in natural areas can serve as barriers to detailed visitor data (Cessford & Muhar, 2003), especially in large or topographically complex PAs. Additionally, staffing or financial constraints may necessitate prioritizing other objectives or concentrating visitor monitoring to select locations.

The dearth of visitor data in PAs is credited as a main limitation in implementing proactive management strategies to minimize visitor impact on resources (Hadwen, Hill, & Pickering, 2008) and to measure ecosystem services (Schägner, Brander, Maes, Paracchini, & Hartje, 2016). Encroaching development or resource extraction near PA boundaries, climate change, and limited budgets can challenge efforts to balance use with biodiversity. To understand the spatial implications of environmental changes on visitor use patterns and support sustainable and effective management of PAs, managers require timely data on resource conditions and visitor use at site specific and landscape scales, as well as analytical tools to conduct integrated analyses to understand human-environment (i.e., visitor-environment) interactions.

This study contributes to the spatial-temporal understanding of human visitation and distribution patterns in PAs by leveraging digital footprints, manifested as publicly shared geotagged photos, as an efficient spatial data source to (1) model spatial distributions of visitor use to identify patterns of visitation at multiple temporal scales and (2) assess relationships between infrastructure, environmental factors, and visitor distribution patterns. The novel combination of geotagged photos and machine learning can identify annual and seasonal

* Corresponding author. North Carolina State University, Dept. of Parks, Recreation, & Tourism Management, Campus Box 7106, Raleigh, NC 27695, USA.
E-mail address: cawalden@ncsu.edu (C. Walden-Schreiner).

https://doi.org/10.1016/j.apgeog.2017.11.004
Received 18 May 2017; Received in revised form 17 November 2017; Accepted 19 November 2017
0143-6228/

importance of infrastructure and environmental factors and inform management strategies for conserving natural resources, while providing opportunities for tourism and recreation.

2. Literature review

2.1. Spatial visitor use monitoring

Visitor use and impact monitoring within PAs focuses on use levels, locations, and associated environmental and social impacts to identify trends and alert management to departures from desired conditions (Cessford & Burns, 2008; Leung & Monz, 2006; Tomczyk & Ewertowski, 2011). The number of visitors to an area represents the most basic yet vital information for PAs, important for planning facility design and capacity, aligning with PA missions, and identifying potential for negative impacts (Cessford & Muhar, 2003). A variety of counting methods exist, ranging from observation studies, registries, permits, and on-site surveys to the use of electronic traffic counters, sensors, and photographic or videographic methods (Arnberger, Haider, & Brandenburg, 2005; Marvin et al., 2016; O'Connor, Zerger, & Itami, 2005; Xia & Arrowsmith, 2008). Global positioning system (GPS) tracking and computer modelling represent increasingly leveraged methods for understanding spatial use patterns and provide high-resolution data to quantify use intensities, explore changes in spatial-temporal patterns, and understand resource impacts (Beeco & Brown, 2013; Hallo et al., 2012; Meijles, de Bakker, Groote, & Barske, 2014; Van Kirk, Martin, Ross, & Douglas, 2014). Collectively, traditional and emerging approaches provide necessary and important information for PA conservation. Cost, efficiency, or the environment can still hinder wider or repeat implementation (Eagles, 2014). Additionally, data collection is often the responsibility of the research partner or managing organization and subject to staffing and budget constraints.

Sampling strategies prioritizing high use areas or peak visitation periods are vital in protecting ecological and site integrity at well-known destinations. However, this approach may miss emerging trends or seasonal variations. Emerging trends can lead to unintended negative impacts. Recreation ecology literature has shown that impacts can happen quickly, with small amounts of use causing disproportionate amounts of impact (Hammitt, Cole, & Monz, 2015; Monz, Cole, Leung, & Marion, 2010). Furthermore, although changing climate patterns vary geographically in intensity and manifestations, changes in the location or quality of resources may affect utilization and visitor-environment interaction patterns (Jones & Scott, 2006). Recent studies demonstrate that visitors are already changing their decisions about the timing, frequency, or duration of PA visits (Buckley & Foushee, 2012; Loomis & Richardson, 2006; McEvoy, Cavan, Handley, McMorrow, & Lindley, 2008; Richardson & Loomis, 2005; Scott, Jones, & Konopek, 2007), further complicating the sampling schemes of visitor data collection.

Temporal changes in visitation patterns, coupled with changing environmental conditions, also suggests the possibility of spatial shifts in visitor-environment interactions. Changes in average precipitation and temperature have altered some species distributions (Chen, Hill, Ohlemüller, Roy, & Thomas, 2011) and phenology (Lesica & Kittelson, 2010; Menzel et al., 2006), together with changing seasonal access in PAs, affords opportunities for temporal and spatial changes in use patterns. These changing patterns may not be congruent with existing infrastructure (i.e., managed trails) or management strategies, resulting in new or increasing negative impacts. Quantifying and visualizing visitor distributions is especially relevant when the landscape itself is also dynamic or unique, as patterns may change quickly or unexpectedly. Therefore, methods are also needed to collect visitor use data at a landscape scale without unduly burdening data collection and analysis efforts.

2.2. Research applications of geotagged photographs

VGI has helped address data collection constraints, with increasing work on accuracy, reliability, and credibility (Connors, Lei, & Kelly, 2012; Flanagin & Metzger, 2008; Levin, Lechner, & Brown, 2017; Senaratne et al., 2017). Including the public in scientific data collection (citizen science) continues as an active area of research in geography and GIScience, though less attention has focused on crowdsourced data arising outside of a formal research collaboration or citizen science effort (Connors et al., 2012). These platforms offer a potential complement to informing monitoring efforts of emerging trends or time-sensitive spatial data needs beyond the capacity of traditional monitoring.

Photo sharing sites like Flickr, Panoramio, Instagram, and others allow for cloud storage of user photos and map-based visualization of geotagged photo locations (van Zanten et al., 2016). In accordance with a site's privacy and use policy, researchers can query photo metadata using the site's application programming interface (API). APIs allow requests to be made of a server, app, or software, which then responds with data. The use of APIs to extract human movement patterns from geotagged media has progressively been documented in tourism and recreation literature (see Zheng, Zha, & Chua, 2010), with Flickr offering a popular platform due to the number of photos and accessible API (Levin et al., 2017). By 2007, one year after the inclusion of geotagging functionality, Flickr hosted over 20 million geotagged photos (Zheng et al., 2010). That number does not take into account the large volume of photos without geotags, for which researchers are also developing processes to assign spatial attributes (Kalogerakis, Vesselova, Hays, Efros, & Hertzmann, 2009; Liu, Yuan, Cong, & Xu, 2014).

Research also suggests that Flickr data can be spatially accurate (Zandbergen & Barbeau, 2011; Zielstra & Hochmair, 2013) and timely (Antoniou, Morley, & Haklay, 2010). Flickr users have the option of manually geotagging a photo by placing it on a map if the camera or device does not have internal GPS capabilities, or leveraging an external GPS device to record the geographic location (Senaratne et al., 2017). Of the photos uploaded to Flickr in 2014, the most common devices included popular smart phone brands (Dove, 2015). Mobile phones with GPS are capable of achieving horizontal accuracy similar to recreational grade GPS units, often within 10 m of true position (Zandbergen & Barbeau, 2011).

The volume of geotagged photos suggests the impact of incorrectly tagged photos can be minimal, and crowdsourced correction capacity exists on sharing sites. Zielstra and Hochmair (2013) examined the spatial accuracy of geotagged photos through a random selection of geotagged photos from Flickr and Panoramio. Their results indicated median positional errors of 15 m and 46 m in North America for Panoramio and Flickr images, respectively, when compared to manually corrected position based on image content. Antoniou et al. (2010) found geotagged photos in the UK were usually also uploaded within a few weeks following capture. In their sample, only 8.4% of Flickr photos were uploaded more than a year later, suggesting that data from Flickr can be both accurate and timely enough for some applications.

Geotagged photo data related to visitors in PAs have found the number of photos uploaded were positively correlated with other visitor monitoring methods (Sonter et al., 2016; Wood, Guerry, Silver, & Lacayo, 2013). Wood et al. (2013) used data from Flickr to estimate visitation rates at 836 natural and cultural recreation sites in 31 countries and found the number of uploaded photos was positively correlated with empirical visitation counts. Sonter et al. (2016) also found correlations between photo numbers and survey visits to PAs in Vermont, USA, and estimated economic contribution of nature-based tourism and landscape characteristics influencing the spatial-temporal patterns observed using Flickr photos.

Beyond counts, Flickr photos also have the potential to document visitor movements and distribution within a PA. Addressing concerns about the relatively low density of geotagged photos in natural areas,