Ecosystem Services xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Ecosystem Services

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

Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands

Andrea Ghermandi [1]

Department of Natural Resources & Environmental Management, Faculty of Management, University of Haifa, 199 Aba Khoushy Ave., Mount Carmel, Haifa 3498838, Israel

ARTICLE INFO

Article history:
Received 22 June 2017
Received in revised form 13 December 2017
Accepted 23 December 2017
Available online xxxx

ABSTRACT

Social media provide a wealth of behavioral data that can be used to investigate the provision of environmental services. In this study, the preferences revealed by photo-sharing social media users are analyzed through travel cost modeling to infer the monetary value of recreation in 115 man-made wetland ecosystems. Photographs' metadata and other publicly available information are used to determine the frequency of recreational trips and the home location of visitors. The mean willingness to pay for access to 115 wetlands is found to range between $5.3 and $374 (2015 international $). The comparison of estimated recreational benefits with the capital, operation and maintenance costs of 74 wetlands reveals that such benefits are considerable and should be taken into account in the design and management of these systems. The approach demonstrates the potential for ecosystem service valuation techniques to incorporate the large amounts of behavioral data available from online resources.

1. Introduction

The rise of social networks has provided environmental researchers with an abundance of volunteered spatial information that was previously unattainable (Goodchild, 2007; Elwood et al., 2012). Online social media networks for photo sharing offer an important platform for crowdsourced geospatial data. This data is made publicly available through the host website by users who thus act as distributed sensors (Singleton, 2010) or "passive mappers" (Heipke, 2010). Such development has been made possible by the concomitant emergence of (1) web 2.0 technologies and (2) progress in portable and low-cost GPS receivers, which are nowadays routinely included in photo cameras, mobile and smart phones (Heipke, 2010). Although users lack a formal training in mapping and follow no quality assurance procedures, the quality of the resulting geospatial data is generally good (Zielstra and Hochmair, 2013). User-generated data uploaded to photo-sharing websites such as Flickr (https://www.flickr.com) has been found to be a useful source of information regarding spatiotemporal patterns of public use and visitation rates in nature-based recreation (Allan et al., 2015; Willemen et al., 2015; Wood et al., 2013; Sonter et al., 2016).

As a crowdsourced alternative or complement to survey-based methodologies (Rieser-Schüssler and Axhausen, 2014; Heikinheimo et al., 2017), this approach offers unique opportunities for integration with revealed preference techniques for the economic valuation of cultural ecosystem services such as travel cost models (TCM). Such integration, however, is still in its infancy (Keeler et al., 2015; Ghermandi, 2016). TCM uses the cost one pays to travel to a site (including the opportunity cost of time) as a surrogate price for the recreational experience. It relies on the insight that visitors respond to cost differences with changes in the quantity of recreation demanded. For single-site individual TCM, the dependent variable – trips made by an individual to the site – is usually estimated with count data models.

Common TCM sampling techniques involve off-site, exogenous sampling of the population (eg phone survey) or on-site sampling of intercepted users (Parsons, 2003). Best practices for the estimation of on-site sampling models involve correcting for demand truncation (only individuals who take trips can be sampled) and endogenous stratification (frequent visitors are more likely to be intercepted) (Englin and Shonkwiler, 1995; Creel and Loomis, 1990). Previous research suggests that social media data are good predictors of home location of visitors to recreational areas (Heikinheimo et al., 2017; Sessions et al., 2016) making the integration with TCM a promising approach.

As a data generating process for TCM, the use of geotagged photos has characteristics that make it unique among sampling procedures. Like on-site sampling, it is a choice-based sampling technique (ie it depends on choices that people have made, in this case to visit a site) and it is subject to truncation. Like off-site sampling, it does not suffer from endogenous stratification of frequent users (ie frequent visitors are not more likely to be social media

[1] Fax: +972 4 8249971.
E-mail address: aghermand@univ.haifa.ac.il

https://doi.org/10.1016/j.ecoser.2017.12.012
2212-0416/

users than occasional visitors). Over both on-site and off-site sampling, crowdsourced data have the advantage of generating large amounts of revealed information without the need to resort to time- and labor-intensive surveys.

This study explores a technique for the monetary valuation of recreation ecosystem services that fully integrates crowd-sourced photographs and TCM. The application focuses on the valuation of the recreation services provided by ecologically engineered wetlands. Such systems, which include constructed wetlands and waste stabilization ponds, are man-made ecosystems that are designed with the primary purposes of (waste) water treatment and water storage and are often indicated with the term "natural treatment systems" (Kadlec and Wallace, 2008). The unique potential of these systems to promote a multi-functional use of the land that includes provision of "designed" ecosystem services, including cultural services, is generally acknowledged (Knight, 1992, 1997; Kadlec and Wallace, 2008; Austin, 2013) but monetary valuations of such benefits are largely missing (Graham and Smith, 2004; Ghermandi and Fichtman, 2015).

This paper first develops a database of geotagged photos and Flickr users associated with 273 engineered wetlands worldwide, including home location for users who self-report this information in their public profiles. Second, five methodologies for inferring home location based on the users' entire set of Flickr photographs are tested and applied to the remaining users. For wetlands with a sufficiently large number of individual users, a Willingness To Pay (WTP) for access is calculated through the estimation of single-site TCMs. For wetlands that are associated with only a small number of users, a WTP for access is inferred through a combination of similarity analysis and benefit function transfer. Finally, the present worth of the capital (CAPEX) and operation and maintenance (OPEX) expenditures for 74 engineered wetlands are compared to the estimated recreational benefits with the goal of promoting a better understanding of the social impacts of these systems.

2. Methods and materials

2.1. Visitation and photo-user-days

The database of geotagged photo metadata in Ghermandi (2016) was extended to include photographs uploaded to Flickr between 23 October 2015 and 2 March 2017. Photographs taken before the construction of the wetlands or prior to 2004 were removed. For each of the 273 wetlands, photo-user-days (PUDs), ie unique combinations of user and recreational site, were calculated after removing multiple photographs taken by one individual user in the same day. A total of 7270 Flickr photographs are available for the analysis, corresponding to 2128 individual users. All personally identifiable user information was removed. The analysis was limited to metadata and did not involve the photographs' content.

2.2. Home location and travel time to wetland

The current location (or hometown if the former is not available) was retrieved from the public profile of 1291 users. Out of these, only 85 provided an approximate location at country, state or regional level (eg South Florida, San Francisco Bay). Five different home detection methods were explored to infer the home location of users who provide no location or an approximate one. They rely on the analysis of the metadata from the entire set of photographs uploaded by individual users, combining the techniques proposed by Bojic et al. (2015) and Li et al. (2013). They involve designating home location as the place where: the user took the largest number of photographs (Method 1); the user took the largest number of photographs and there is at least a 10-day span between the first and last photo (Method 2); the user spent the largest number of PUDs (Method 3); the user spent the largest number of PUDs and there is at least a 10-day span between the first and last photo (Method 4); the timespan between the first and last photo is largest (Method 5). Home location was determined at the level of the lowest administrative sub-division (eg county for the USA), as derived from Hijmans (2012). The analysis was implemented in ArcGIS 10.2.2. The predictive performance of each model is tested on the sample of users who provide only approximate locations. The best-performing method is subsequently implemented to analyze the metadata of more than 900 000 Flickr photographs and infer a home location for 912 Flickr users.

Travel time and distance between the self-reported or imputed home location and the visited wetland polygon's centroid were calculated for 2104 users. The calculation relied on the most efficient routes (ie optimizing travel time) as identified using the web mapping service Google Maps and assuming car as the preferred means of transportation. The nearest points that can be reached by car were taken as the itinerary's starting and end points. The number of unique visitors and PUDs in 115 wetlands were calculated. The remaining 158 wetlands in the database are not associated with any Flickr photograph. Only the 5794 trips taken by the 1347 users with home location within a 150 miles (~241 km) radius from the wetlands were retained for further analysis since they were considered to be representative of single-day trips (Parsons and Kealy, 1992; McConnell and Strand, 1994). Consecutive-day trips from within this radius were treated as separate single-day visits.

2.3. Travel cost modelling and WTP for access

For the 12 wetlands with a sufficiently large number of associated users (ie at least 10 users per variable), a single-site TCM was estimated using a truncated Poisson regression. Truncation in the model accounts for the fact that only strictly positive visitation rates are observed. The endogenous variable in the model is the number of PUDs for each Flickr user in a specific wetland. The model specification includes travel costs–both direct and opportunity costs of travel time–and income. Direct costs associated with a round trip were calculated using the average price of unleaded gasoline in 2015 at country level from Global Petrol Prices (http://www.globalpetrolprices.com/) and a fuel economy of 12 km per liter. Other fixed or operating costs were excluded from the analysis. For the opportunity cost of travel time, the commonly implemented assumption of 1/3 of the wage rate was used. Gross nominal monthly wage rates for the most recent available year were obtained from ILO (2016). They were converted to hourly wage rates using 40 working hours per week as basis for full employment. The sensitivity of the results to using 3/4 of the wage rate (Fezzi et al., 2014) or no opportunity cost of time was evaluated. The willingness to pay (WTP) for access to the wetlands was calculated as the sample mean consumer surplus (Haab and McConnell, 2002).

The median income at the lowest available administrative level was obtained from official statistics bureaus and used as a proxy of the individual visitor's income. Median household income at county and state level for the latest available year was used, respectively, for users in the USA and Malaysia (US Census Bureau, 2016; Malaysia Department of Statistics, 2015). For European countries, the primary income of private households at regional (ie NUTS-2) level in 2013 was obtained from Eurostat (http://ec.europa.eu/eurostat/web/products-datasets/-/tgs00036). All costs and income data were converted to 2015 international $ adjusting for inflation and differences in purchasing power using the appropriate World Bank's World Development Indicators (The World Bank, 2017).

Please cite this article in press as: Ghermandi, A. Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands. Ecosystem Services (2017), https://doi.org/10.1016/j.ecoser.2017.12.012