



Research article

Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs

Carrie Sessions ^{a, b, c, *}, Spencer A. Wood ^{b, c}, Sergey Rabotyagov ^b, David M. Fisher ^{b, c}^a Evans School of Public Policy and Governance, University of Washington, 4100 15th Ave NE, Seattle, WA 98195, USA^b School of Environmental and Forest Sciences, University of Washington, Box 352100, Seattle, WA 98195, USA^c Natural Capital Project, Stanford University, USA

ARTICLE INFO

Article history:

Received 4 May 2016

Received in revised form

31 August 2016

Accepted 3 September 2016

Available online 15 September 2016

Keywords:

Recreation

Visitation rates

Travel cost

Valuation

National parks

Social media

ABSTRACT

Land managers rely on visitation data to inform policy and management decisions. However, visitation data is often costly and burdensome to obtain, and provides a limited depth of information. In this paper, we assess the validity of using crowd-sourced, online photographs to infer information about the habits and preferences of recreational visitors by comparing empirical data from the National Park Service to photograph data from the online platform Flickr for 38 National Parks in the western United States. Using multiple regression analysis, we find that the number of photos posted monthly in a park can reliably indicate the number of visitors to a park in a given month. Through additional statistical testing we also find that the home locations of photo-takers, provided voluntarily on an online profile, accurately show the home origins of park visitors. Together, these findings validate a new method for measuring recreational visitation, opening an opportunity for land managers worldwide to track and understand visitation by augmenting current data collection methods with crowd-sourced, online data that is easy and inexpensive to obtain. In addition, it enables future research on how visitation rates change with changes in access, management or infrastructure, weather events, or ecosystem health, and facilitates valuation research, such as travel cost studies.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recreation is a major use of public lands, yet managers lack detailed, up-to-date, and accessible data regarding visitor use (Loomis, 2000). Visitation data is often limited in spatial and temporal coverage by the expense and time required to survey visitors (Heberling and Templeton, 2009). However, in this age of big data, there may be opportunities for researchers to leverage online information to better understand patterns of recreational use (Carter et al., 2015; Wilde and Pope 2013). One possibility is that photographs posted by recreationalists on social media websites such as Flickr might convey valuable information on recreational park use and the on demographics of visitors (Wood et al., 2013). While this idea is gaining popularity (Arkema et al., 2015; Griffin et al., 2015;

Sonter et al. in press), little work has been done exploring the statistical relationships between photos and traditional visitor use data (Wood et al., 2013; Keeler et al., 2015).

Land managers rely upon current visitor use data to evaluate the economic costs and benefits of proposed policies and management alternatives (Hotelling, 1947; Loomis, 2000). However, the demand for this data far outstrips the supply—few land agencies track visitor use comprehensively and consistently. The National Park Service (NPS) has a relatively extensive visitor monitoring program and publishes monthly counts of visitors, in addition to conducting detailed visitor-use surveys every few years. While this system of measuring visitor use provides valuable information, there is still need for the NPS to have more comprehensive and representative data (Benson et al., 2013; Heberling and Templeton, 2009; Poudyal et al., 2013). The NPS panel data are commonly limited by a lack of data on demographics (Poudyal et al., 2013), a surveyed population that may not be representative of the whole visitor population (Benson et al., 2013; Neher et al., 2013), and aggregated monthly counts of visitors to parks that preclude analyses of visitation

* Corresponding author.

E-mail addresses: carrie.sessions@gmail.com (C. Sessions), woodsp@stanford.edu (S.A. Wood), rabotyag@uw.edu (S. Rabotyagov), davefisher@stanford.edu (D.M. Fisher).

changes on finer spatial and temporal scales (Freeman et al., 2014).

Using social media as a source for visitor use information opens the opportunity to study visitation at sites and times that currently lack comprehensive visitor use data and to provide supplementary information that addresses the limitations of the data currently available. Social media such as Flickr photographs are inexpensive to process, making it easier and less costly to study visitation at remote sites. Many photographs are linked with an online user profile, allowing researchers to further investigate where visitors live. In addition, images are taken and posted throughout the year, making them potentially a source of data with richer temporal coverage and better resolution than traditional survey methods (Freeman et al., 2014). Wood et al. (2013) studied the relationship between Flickr photographs at over 800 recreation sites around the world and found there was a correlation between visitation rates measured by the photos and by standard reporting methods. Building off of that work, Keeler et al. (2015) used photos to model visitation to Minnesota and Iowa lakes and applied the estimated visitation rates in a travel cost study examining recreational demand for clean water.

In this paper, we explore the statistical relationship between the density of Flickr photographs and visitor use data and discuss the accuracy and potential utility of these new methods. Specifically, we assess the validity of using metadata from photos posted on Flickr to infer information about visitors to National Parks in the western United States. By comparing the photo data to statistics published by the NPS, we examine the statistical relationship between the two datasets and explore whether the two approaches yield similar information. First, we evaluate whether the number of images posted online can accurately indicate the recorded number of visitors to a site. Second, we assess whether the hometowns of photo-takers indicated by Flickr metadata portrays an accurate representation of the hometowns of park visitors. Our goal is to show how these new data and approaches can be used to accurately infer visitor use information and to open new opportunities for research on the recreational value of lands.

2. Methods

We study visitation at 38 National Parks in the western United States (Table A1), drawing upon data from two sources: the NPS and Flickr. First we compare monthly counts of recreational visitors to each park published by the NPS with estimates derived from photos posted on the website Flickr – an online platform that allows users to publicly share their photographs (www.flickr.com). As of 2013, there were over six billion photographs posted on Flickr, uploaded by over 71 million users (Wood et al., 2013). Approximately 197 million of these photographs are geo-tagged, meaning that they include the geographical coordinates of where the image was taken; 40% of the geo-tagged photographs are taken in the U.S. (Wood et al., 2013). Uploaded photographs also include the date that the image was taken. All metadata for public photos is accessible via the Flickr Application Programming Interface (<https://www.flickr.com/services/api/>). Second, we assess how home locations of park visitors published by the NPS compare with the origins of park visitors who post photos on Flickr. The NPS periodically conducts Visitor Use Surveys at parks, a survey that asks visitors their home zip code, among other questions. On Flickr, a portion of users volunteer their home location as part of their public profile, which is accessible via the profile's URL. In our study area, 30% of visitors who posted a photo provided their home location.

2.1. Modeling visitor counts

We analyze monthly visitation counts from 2007 to 2012 for all

38 of the western U.S. National Parks excluding Hawaii (Table A1 and Fig. A1). From the NPS, we obtain publicly available Visitor Use Statistics, a monthly count of the number of recreational visitors to each park (NPS, 2015). From Flickr, we query the global set of geo-tagged photos using GIS to retrieve all photographs taken within the boundaries of each park. Then, from the set of images within a park, we count unique combinations of users and dates per month and call the sum monthly “photo-user-days” (PUD sensu Wood et al., 2013). To clarify, if five people visit Yellowstone National Park on a day and each upload one photo, the resulting PUD equals five, whereas if one person uploads five photos on a single day, the PUD value equals one. Monthly PUD data were computed with the InVEST software package (Sharp et al., 2016). We analyze the monthly PUD in each park from January 2007 to December 2012.

Visual inspection of the data shows that there is strong correlation between average monthly Flickr PUD and the NPS' visitor counts (Fig. 1). To quantify this relationship, we use multiple regression analysis to construct a predictive model that uses PUD and other control variables to estimate the number of visits observed by the NPS to a park. We model the number of observed visits to park i in month j using a negative binomial regression,

$$NPS_{ij} \sim \text{Negative Binomial}(\exp(\mathbf{x}_{ij} \beta), \theta) \quad (1)$$

where \mathbf{X} is a vector of covariates: the natural log of PUD, year, the month of visit, an interaction term between PUD and June/July/August, and a fixed effect for the park; beta is the regression coefficient and theta represents the error term that is a random variable drawn from the gamma distribution. We use April as the reference case for month and Arches National Park as the reference case for park. The model analyzes monthly data between 2007 and 2012 for National Parks in the western contiguous United States. The use of the negative binomial model is consistent with statistical practice to estimate count data (Cameron and Trivedi, 1998), characterized by observations that have a lower bound of zero and can only take integer values. The model is similar to the Poisson regression but is more general in allowing overdispersion, and is often used in recreation demand models (Haab and McConnell, 2002). The interaction term between PUD and June/July/August controls for the observation that in some of the parks, PUD tracks with visitation though most of the year but underestimates it during summer months (Fig. 1). A fixed effect for park allows for a hierarchical model in which observations within each park are correlated and non-randomly related. Including the fixed effect gives each park a unique intercept, thereby controlling for differences among parks that may alter its estimated visitation. Year accounts for the variability in the number of photos posted annually on Flickr (Fig. A2).

Using the same model specification as above, we also build models specific to each park, allowing each park to have a unique relationship with PUD. While the objective of the pooled model is to assess the overall ability of PUD to infer visitor counts, the objective of the park-specific models is to increase the accuracy of predicted counts. Each of the park-specific regression models estimates the number of NPS-measured visits to a park in month j using a negative binomial regression,

$$NPS_j \sim \text{Negative Binomial}(\exp(\mathbf{x}_j \beta), \theta) \quad (2)$$

where \mathbf{X} is a vector of covariates: the natural log of PUD, year, the month of visit, and an interaction term between PUD and June/July/August. Beta is the regression coefficient and theta represents the error term that is a random variable drawn from the gamma distribution.

Download English Version:

<https://daneshyari.com/en/article/5117206>

Download Persian Version:

<https://daneshyari.com/article/5117206>

[Daneshyari.com](https://daneshyari.com)