Social media as a means of understanding park visitors: An evaluation of user-generated content related to two low use, remote national parks

by

Danielle T. Dagan

B.A., University of California, Santa Cruz, 2014

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Horticulture and Natural Resources
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Ryan L. Sharp, Ph.D.

Copyright

Abstract

Using two remote and sparsely visited backcountry and wilderness areas in Alaska as a case study, this manuscript evaluates data from one season of online user-generated content. Social media posts from Twitter and Flickr were collected, coded for content, and mapped where geographic data was available. The resulting dataset was assessed for its quality, and results were compared with visitor use information reported by commercial use operators. Overall, out of 4,008 users in the dataset, 223 were on-site visitors to one or both parks, and 144 visited the backcountry. Social media content overrepresented wildlife viewing behaviors, and very little spatial data emerged. User-generated content in the study of very remote wilderness may have some application for managers, but the entire volume of related user-generate content may not represent on-site visitors, especially on Twitter. More research is needed to understand limitations in application of publicly available social media data.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to thank my advisor, Dr. Ryan Sharp, not just for shepherding me through this process despite an uncertain start, but also for introducing me to an entire research area and profession. The opportunity to embed myself in the world of visitor use management, made possible by your knowledge and network, has been hugely impactful on the development of this thesis and the direction of my research interests overall. Thank you also to the remainder of my committee for your guidance, patience, and willingness to come on board late in the game. I would not have been able to stay on track if you had not jumped in without hesitation.

I would also like to acknowledge Emily Wilkins and Asher Vollmer, without whom I would still be struggling to string together the code used to collect this data.

Of course, I would not be anywhere if it were not for my family, who support me unfailingly. You always said I could and should be whoever I want, and the influence of that message cannot be overstated. There is freedom in the assurance that someone always has my back, and confidence in knowing that my judgement is trusted. Thank you.

Finally I must acknowledge my pup, who cannot read, for ensuring I go on at least one walk every day, play outside when it snows, and come home at a reasonable hour. She's the best.

Chapter 1: Introduction

Overview

Park managers use information about visitors when managing resources, visitor experiences, and other dimensions related to public land administration and agency mission. Useful visitor information for managers includes data about visitor counts, type of visitor use (e.g. primary activity), timing of use (e.g. season), indicators and thresholds of quality, rule compliance, among others (Manning, 2010; Marion, 2016; Cahill et al., 2018). Commonly used methods for collecting this data include infrared sensors and cameras, on-site observations, permit systems, and questionnaires (Cessford & Muhar, 2003; Manning, 2010). Each of these methods can tax both monetary and human resources, especially considering that government agencies, which serve as caretakers of backcountry and wilderness resources in the United States, often experience budgetary or personnel constraints (Fay, 2010; Weimer and Vining, 2017).

Thus, user-generated content, in the form of social media, is being explored as an alternative, perhaps more cost-effective data source for understanding visitors to parks and protected areas. Since Wood (2013) determined that posts from the photo-sharing website Flickr could approximate visitation rates at parks and protected areas, researchers have been using social media data to evaluate visitation at various scales, preferences for biodiversity, visitor distributions, and socioeconomic concerns (e.g. Fisher, 2018; Hausmann et al., 2018; Walden-Schreiner et al., 2018a; Hamstaed et al., 2018). To the author's knowledge, there is no published research that closely evaluates the content of posts at an extremely low-visitation, remote site, nor published research that evaluates content of posts to identify in-person visitors.

This thesis serves to investigate these new methods, specifically their application in remote and low-use areas. It includes a methods paper that evaluates areas of concern with the application of social media data, and broadly contributes to the park management literature by responding to common issues regarding the quality of social media data in remote areas, and by introducing case-specific conclusions that can become generalizable with additional research. While the methods do not lend themselves to broadly applicable results, there is value in deeply investigating one case-study to uncover potential problems and nuances in accessing and interpreting results.

Objectives and Hypotheses

Objectives

1. To understand to what degree social media users who share content related to the parks have recently visited the parks
2. To examine whether social media data can be used to understand the spatial patterns of on-site visitors on a site- and season- specific scale
3. To examine the limitations of using social media data to understand visitors, specifically whether it overrepresents certain populations of users based on destination or activity type

Hypotheses

1. On Flickr, bear viewing will be overrepresented relative to other data sources
2. The distribution of activities represented on Twitter will not be significantly different from other data sources

Chapter 2: Manuscript

Social media data in remote and low-use wilderness areas: Applications and limitations
To be submitted to the International Journal of Wilderness

Abstract

This paper evaluates data from user-generated content related to two Alaska National Parks, which are mostly comprised of remote and sparsely visited backcountry and wilderness. Social media posts were collected, coded for content, and mapped if geographic data was available. The dataset was assessed for quality and compared with information reported by commercial operators. Of 4,008 users, 223 were on-site visitors and 144 visited the backcountry. Social media overrepresented wildlife viewing behaviors, and spatial data was limited. User-generated content in the study of low-use, remote backcountry remains useful, but more research is needed to understand limitations and practitioner applications at a site-specific scale.

Introduction

Visitor use management and monitoring

Information about visitor use in parks and protected areas is an essential component of effective management (Cessford and Muhar, 2003). Data about recreation users has long been recognized as important for several reasons, including the ability to manage budgets, personnel, protect resources, and to provide visitors with enjoyable experiences (Manning, 2010). Useful visitor information for managers includes data on visitor counts, type of visitor use (e.g. primary activity), timing of use (e.g. season), indicators and thresholds of quality, rule compliance, among others (Manning, 2010; Marion, 2016; Cahill et al., 2018).

Previous studies indicate that it is beneficial for managers to understand characteristics of the users, including activity type (Manning, 2010). Activity type can influence visitor temporal and spatial distribution, impacts on natural resources, or user conflicts (Jacob and Schreyer, 1980; Kyle et al, 2003). While activity may not necessarily influence users' normative thresholds for park or protected area conditions (Donnelly et al., 2000), settings related to activity types do

relate to normative standards as well as norm strength (Kuentzel et al., 2007). Additionally, the amount of engagement in various outdoor recreation activities may influence management decisions, especially because planners often address individual activities types through different management actions (for an example see Van Huizen et al., 2013).

Commonly used methods for collecting this data include infrared sensors and cameras, on-site observations, permit systems, and questionnaires (Cessford & Muhar, 2003; Manning, 2010). These methods can be costly, in both monetary and human capital resources, especially considering that government agencies, which serve as caretakers of backcountry and wilderness resources in the United States, often experience budgetary or personnel constraints (Fay, 2010; Weimer and Vining, 2017). Questionnaires are commonly used to understand park visitors, including their demographics, behaviors, normative standards for natural or social conditions, values, and satisfaction (Vaske, 2008).

Data collected in visitor questionnaires is often used to assess both current and desired conditions by evaluating demographic information, reported visitor behaviors, desired and acceptable conditions for crowding and other indicators of quality (Manning, 2010). However, while commonly used, questionnaires may not be the ideal tool for every context. For one, survey response rates may be declining broadly (Vaske, 2008), and some have argued that the number of surveys being distributed across disciplines has been influencing people's willingness to participate overall (e.g. Olson, 2014). While there are methods for improving response rates, such as sending reminder emails for online surveys (Van Mol, 2017), the overall downward trend of response rates signals a need for alternative or supplemental methods. This is especially true for researchers and managers with limited resources operating in environments where collecting questionnaires on site is prohibitively challenging. Distributing questionnaires is often expensive

and time-consuming, especially in protected areas without controlled access points or in remote locations (Fay, 2010).

In the United States, where this study is located, land designated as wilderness under the Wilderness Act of 1964 receives the highest level of protection. While wilderness can be understood as a social concept that includes several social, political, and physical considerations, "federally designated wilderness" is legally required to include specific characteristics, such as opportunities for solitude, primitive or unconfined recreation, and limited motorized use (Dawson and Hendee, 2009). The term "backcountry" may refer to similar social and physical characteristics, and may include areas that are both designated wilderness and not designated wilderness. The U.S. National Park Service (NPS), which administers the parks discussed in this study, uses the term backcountry to refer generically to "primitive" or "undeveloped" areas (NPS, 1991). The NPS generally limits backcountry infrastructure to trails, unpaved roads, and facilities associated with dispersed use, and administers backcountry such that human recreational use is most often dispersed (NPS, 1991). While all designated wilderness is considered backcountry, not all backcountry is designated wilderness.

Visitor use monitoring in backcountry and wilderness is uniquely challenging, especially in terms of cost (Hollenhorst, 1992). A report from the U.S. Forest Service regarding a National Forest in Alaska describes a number of challenges directly related to the dispersed nature of use in backcountry areas, as identified by land managers (Fay et al., 2010). For example, without a concentrated area of visitor use, a location through which all or most visitors pass, or a permit system, survey distribution becomes extremely expensive. Fay et al. (2010) share that managers describe survey distribution in dispersed backcountry settings that receive little use as "cost-prohibitive," stating that combining multiple data sources can significantly improve the sample.

Additionally, when these challenges result in smaller sample sizes, bias in introduced to the study (Hollenhorst, 1992).

Often managers lack the required data about visitors to make informed decisions in a useful geographic scale; and while expediency with which data can be collected is an important factor for managers, traditional methods can be slow, relative to the need for managers to act or plan (Fay et al. 2010; Cerveny et al., 2011; Cahill, 2018; Mancini et al., 2018). In light of these challenges, researchers have begun investigating methods for utilizing user-generated online content, such as social media posts, to assess visitor use in parks and protected areas (e.g. Sessions et al., 2016; Fisher et al., 2017; Levin et al., 2017; Mancini et al., 2018; Hausman et al., 2018; Hamstead et al., 2018; Tenkanen et al., 2017).

Social media studies

Social scientists have been exploring online data sources since the early days of the internet. As early as 1997, with the emergence of web blogs, those who study marketing and related disciplines began exploring how this type of data could be used to understand consumers (Khang et al. 2012). By 2007, geographers were utilizing web data to understand populations on a landscape scale (Goodchild 2007). This was termed *volunteered geographic information* (VGI) by Goodchild (2007), and that term continues to be used in the scientific literature. Studies utilizing VGI have used a range of internet sources (e.g. Twitter, OpenStreetMap, Strava fitness mobile app) to address a variety of social and natural research topics, fitness activities, urban planning, and natural disaster relief (Arapostathis, et al., 2018, Haworth and Bruce 2015, Griffin and Jiao 2015). However, only recently have researchers begun applying these concepts to visitor use in parks and protected areas, and even less commonly in backcountry or wilderness.

Recently, researchers have begun working to understand how user-generated content can reveal visitor use and preferences in parks and protected areas, which has the potential to improve planning and management of these areas. Indeed, planners have considered volunteered geographic information, such as mobile fitness tracker data, when identifying visitor capacity (NPS, 2018). An early application of social media data for visitor use research by Wood (2013) evaluated Flickr data from parks and protected areas around the world and determined that it could serve as a proxy for visitation rates. Several studies have used social media data to identify geographic hotspots for cultural ecosystem services (e.g. Albert at al. 2014, Pastur et al. 2016, Richards and Friess 2015, Tenerelli et al. 2016, van Zanten et al. 2016). Additionally, social media has been used to understand how visitors move through parks by utilizing geotagged photos (e.g. Chua 2016, Heikinheimo et al. 2017). However, few studies have utilized social media data to analyze recreation use by activity type, or evaluated the content of posts to understand visitor behaviors, as opposed to their preferences or locations. Few studies have used these methods in low-use or remote outdoor settings. Heikinhemo at al. (2017) demonstrated that social media content can serve as a proxy for on-site questionnaire responses regarding activity type, although they conducted the study in an urban-proximate park. Tenkanen et al. (2017) assessed the usability of Instagram, Flickr, and Twitter for monitoring visitors in both heavily used and scarcely used parks. While they did identify a correlation between the volume of social media user days and park popularity, they found that this correlation underestimates popularity of less visited parks.

Wood (2013) describes many possible limitations to using social media data in the context of parks and protected areas. He questions whether there are biases because certain activities are more photogenic or easy to photograph, because perceived value affects posting

behavior, or because local visitors may be less inclined to share their experience online. Users may be tailoring shared content on social media to elicit a positive response from social media followers, which may influence the data (Boley et al. 2018). Biases and limitations to user-generated data have indeed been demonstrated. Tweets and other VGI data sources have been shown to have an urban bias (Hecht and Stephens, 2014). Additionally, social media is not used equally by all demographics. In Kruger National Park, Hausmann et al (2018) asked visitors about their social media use. They found that older and higher income tourists were less likely to use social media. Moreover, they found that different social media channels elicit different types of post content, specifically Flickr photos showed more biodiversity while Instagram showed more photos of people.

Previous studies were primarily conducted in urban-proximate areas, relatively densely used areas, or conservation areas with cellular signal (e.g. Hamstead et al, 2018; Fisher et al., 2018; Hausmann et al., 2018), although Mancini et al. (2018) analyzed spatial data from six years of Flickr posts in the relatively low-visitation Cairngorms National Park in Scotland. Tenkanen et al. (2017) compared social media data and official park visitation statistics in national parks in Finland and South Africa, and showed that while social media was a good proxy for visitation in heavily populated parks, it may underestimate popularity of less-visited parks. Improved understanding of social media data in remote parks and protected areas can clarify the applicability of user-generated content for managers and researchers. Walden-Schreiner et al. (2018b) evaluated geotagged Flickr photos on a site-specific scale. They used seven years of data from two low-use protected areas, in Argentina and Australia, to assess seasonal patterns. They used a visual inspection of post content to determine whether posts

contained appropriate geographic information; however they did not use the content to assess visitor behavior.

Objectives

Overall, the objective of this study is to evaluate the utility of using user-generated content to understand visitors on a site-specific scale in remote and low-use backcountry and wilderness areas. Within these spatial contexts, this study aims to:

1. understand to what degree social media users who share content related to the parks are in-person visitors
2. examine whether social media data can be used to understand the behavior of on-site visitors on a site- and season- specific scale
3. examine the limitations of using social media data to understand visitors, specifically whether it overrepresents certain populations of users

Methods

Study area

Katmai and Lake Clark National Parks and Preserves (KATM and LACL), located in southwest Alaska, are two of the least visited National Parks. The two parks have a combined eight million acres, nearly six million of which is federally designated wilderness (NPS, 2016). The parks are almost entirely considered backcountry according to the NPS guidelines described above. All of LACL can be considered backcountry, while the only area which is not considered backcountry is Brooks Camp in KATM, which is a relatively small area renowned for its brown bear viewing opportunities, and contains a visitor center, picnic area, a lodge and restaurant,

seasonal ranger housing, trails, bear viewing platforms, and other infrastructure. Brooks Camp is also well-known for its "bear cams," which livestream brown bears throughout their active season (see https://explore.org/livecams/brown-bears/brown-bear-salmon-cam-brooks-falls). Neither park is accessible by road, and are typically accessed by small fixed-wing aircrafts or boats. Once in the parks, many visitors are accompanied by commercial guides. There is little to no cellular connectivity in the park, although some visitors who stay in backcountry lodges have access to WiFi. While the commercial use authorized (CUA) pilots and guides report data to the parks, there is currently no reservation system that the parks utilize to monitor backcountry use.

The National Park Service Southwest Alaska Inventory and Monitoring Network (SWAN) compiles data reported by CUAs for each park (SWAN, 2018a; SWAN, 2018b). CUA data can be useful for counting visitors, and have been used in similar remote parks, such as Kenai Fjords National Park and Gates of the Arctic National Park and Preserve (Fay et al., 2010). According to these reports, visitor use days reported by CUAs to KATM has fluctuated between 25,000 and 30,000 between 2008 and 2017, while the number of visitor use days reported to LACL has grown from 4,000 days to over 15,300 days over the same time period. This is slightly under the visitation reported by NPS (which reports number of visitors, in contrast to visitor use days). The NPS reports 37,818 KATM recreation visitors and 22,755 LACL recreation visitors in 2017, the most recent year of CUA data summarized by SWAN. In 2018, KATM reported the same number of recreation visitors, while LACL reported a decrease, with 14,479 recreation visitors. KATM uses 20 different user counts to calculate recreational visitation, including visitors staying at lodges, commercial use visitors, and visitor center visitor counts. LACL uses the number of observed visitors entering the park to determine recreation visits (NPS visitor statistics available at https://irma.nps.gov/Stats/).

While raw data is not available, it appears from figures in the KATM report that the area with highest number of visitor use days is Brooks Camp, the only front country area in both parks. It also appears that the sum of visitor use days in the remaining areas in KATM approximates the number at Brooks Camp alone. Furthermore, because all of LACL is considered backcountry, when these parks are combined there are more backcountry visitor use days than front country days (Figure 1, Figure 2).

Visitor activities are also reported by SWAN (2018a, 2018b). These documents do not report values, either in percentages or frequencies, except in figures (Figures 1-4). For the purposes of this study comparisons will be made to estimated visitor use days based on visual inspection of these figures. For both parks, bear viewing and recreational angling were the most frequently reported activities by a significant margin. In Lake Clark, photography was also a very highly reported activity, although not as commonly reported as bear viewing or fishing. Activities are not mutually exclusive. At both parks, bear viewing use has been increasing in recent years relative to sport fishing. In 2016, the most recent year of data reported, there were approximately 42% more bear viewing days than sport fishing days in LACL (~3,600 angling; 5,100 bear viewing), while there were similar numbers of each reported at KATM.

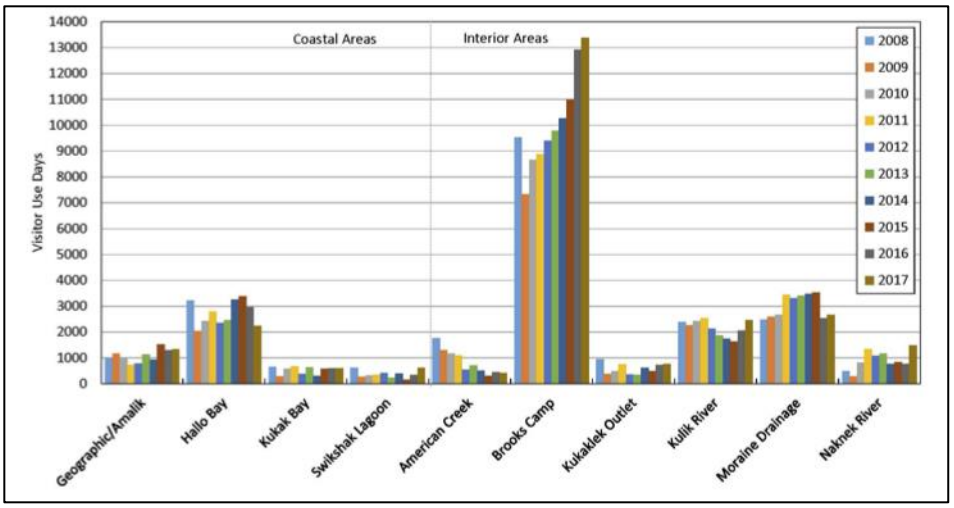

Figure 1. 2008-2017 visitor use days in Katmai National Park and Preserve for select locations, as reported to the park by commercial use operators. Source: Southwest Alaska Inventory and Monitoring Network (2018a).

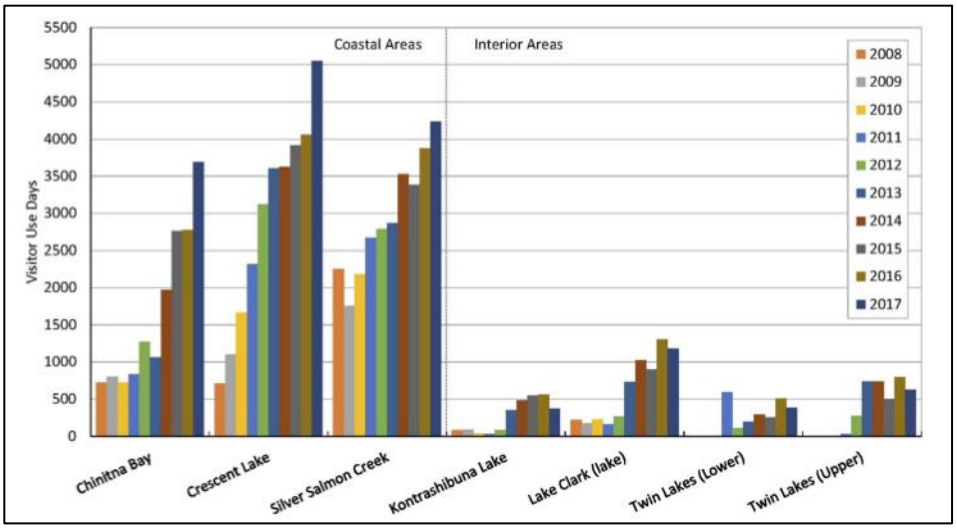

Figure 2. 2008-2017 visitor use days in Lake Clark National Park and Preserve for the most visited coastal and interior areas, as reported to the park by commercial use operators. Source: Southwest Alaska Inventory and Monitoring Network (2018b).

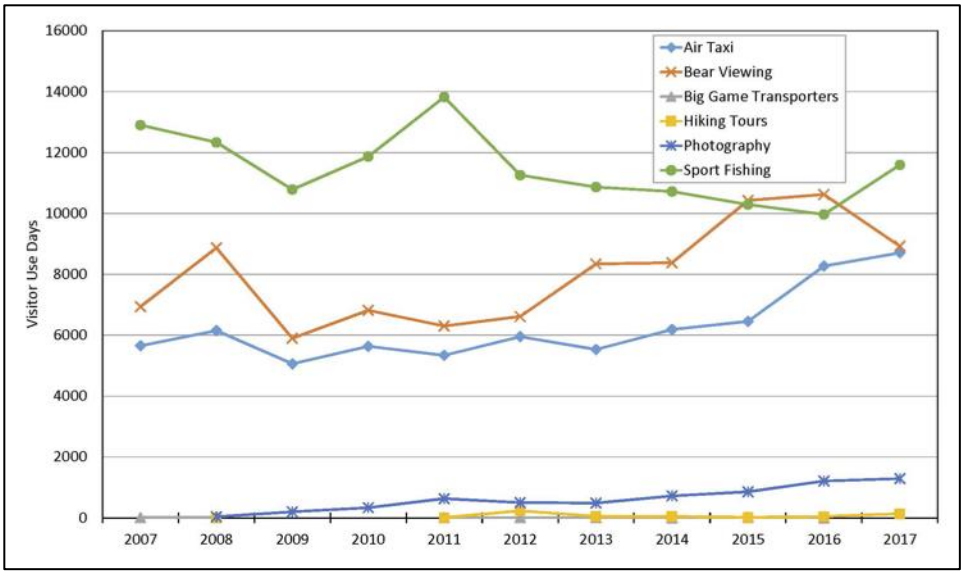

Figure 3. 2008-2017 visitor use days for each main activity in Katmai National Park and Preserve, as reported to the park by commercial use. Source: Southwest Alaska Inventory and Monitoring Network (2018a).

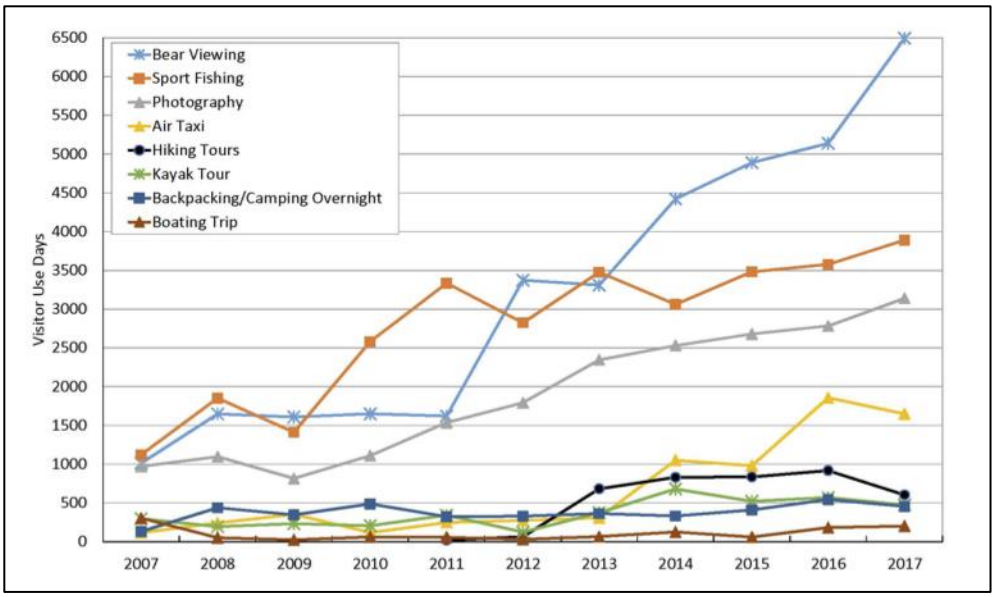

Figure 4. 2008-2017 visitor use days for each main activity in Lake Clark National Park and Preserve, as reported to the park by commercial use operators. Source: Southwest Alaska Inventory and Monitoring Network (2018b).

Data collection

Researchers commonly use Twitter, Flickr, and Instagram as social media data sources for a variety of reasons, including their popularity and ease of data access (Tenkanen et al., 2017). However, this evaluates only Twitter and Flickr. Ideally, Instagram would be used as it has been shown to be most accurate (Tenkanen et al., 2017; Heikinheimo, 2018), however Instagram has begun implementing substantial limitations to their application programming interface (API), which is what allows external queries to access the platform. Many API features were limited as of the end of 2018, with further depreciation planned for 2020, making further evaluation of Instagram as a data sources less applicable for management and future research (see https://www.instagram.com/developer/).

Twitter data was collected through the Twitter API using a custom code (Hausmann et al. 2018) which was written in Python programming language. Twitter's API only allows users to query back in time seven days, so the code was set up to download Tweets in real time between June 15, 2018 and October 24, 2018, which approximates the peak visitation season at both parks (SWAN, 2018a; SWAN, 2018b). Additionally, tweets continued to download for two months after peak season, until October 24, 2018, to account for a lag time between visiting and posting. All tweets were saved into a database in real time. The database includes tweet text, photo URLs, profile information, and geographic data in a CSV file.

Flickr data posted during the same time frame (June 15, 2018 through October 24, 2018) was collected using a custom Python code, which downloaded photos with unique identifier file names, and created a CSV file including the unique identifier and any caption text, tags, date and time the photo was taken, date and time it was posted, and geographic information if available (e.g. Heikinheimo, 2018). The Flickr API allows developers to collect photos and corresponding

metadata by specifying a time frame and bounding box, which is a geographic area defined by two latitudes and two longitudes. (Hausmann et al, 2018). Additionally, photos were queried by manually inputted tags (e.g. Mancini, 2018).

Due to concerns that low overall visitation would result in too few posts for meaningful analysis, the initial data collection was executed with very liberal search queries, in order to capture as many posts as possible. In order to evaluate data collection methods in remote areas with little-to-no cellular connectivity, this wide query "net" was combined with detailed cleaning and coding methods.

Both Twitter and Flickr posts were queried using key words and a geographic bounding box. If a post contained any individual keyword, or was geotagged within the bounding box, it was downloaded into a raw dataset. The bounding box includes a range of latitudes and longitudes that comprise an imaginary rectangle on the map (e.g. Hausmann et al., 2018), and the one used in this study encompassed both KATM and LACL. Typically, a bounding box is used in the initial query for scraping posts, and then area boundaries are further defined in ArcGIS, and posts within the bounding box but outside the study area are discarded (e.g. Heikenheimo et al., 2018). However, because an objective of this research is to investigate the efficacy of using geotags to collect user-generated content in areas without cellular connectivity, and because a visual inspection was used to determine if a user visited the park, this method was not appropriate. For both Flickr and Twitter, the bounding box for geotags was set at [-156.713358, 57.681894, -153.00000, 61.586310]. Twitter keywords queried included park names as both regular post text and hashtags, account handles, and names of two key locations as both tweet text and hashtags ("katmai", "#katmai", "lake clark", "#lakeclark", "hallo bay", "#hallobay", "proenneke's cabin", "#proennekescabin", "@katmainps", "@lakeclarknps"). However, queries

related to these locations, Hallo Bay and Proenneke's Cabin, were later discarded because of low return upon initial examination of results, leaving queries for the park names as tweet text and hashtags, as well as the park account handles. Because users on Flickr cannot tag other accounts, and do not typically use hashtags, Flickr was queried for posts including tags "katmai" or "lake clark."

Data cleaning

Data was cleaned for duplicates, and each post was manually inspected to code for content (Richards and Friess, 2015). Retweets and replies were removed from the dataset, and posts in a language that the reviewer did not understand were assessed using Google Translate and by inspecting media attachments. Each post was coded according to content using a visual inspection of both textual content and attached media, such as photos, videos, or links to external websites, and done by a person familiar with the area. Posts were coded as unrelated to the parks, non-visit (posts about the parks but not indicating a visit, e.g. news stories), by location if the post was in Brooks Camp or in an unclear location. Posts that were in backcountry and wilderness were coded by activity (Figure 5).

While it is possible that, given this codebook, a user coded as non-visit was a visitor who did not include necessary visit information (i.e. a type II error), it was very clear whether an individual *post* was discussion about the park, rather than sharing a park visit. The vast majority of non-visit posts were clearly discussion about the bear livestream at Brooks Camp, and nearly all the rest were about park-driven media campaigns and a small number of viral news stories. The only content that could have been confused between a visit and a non-visit park-related post were instances where images that required an in-person visit were shared without any description

of events, and the user account was not explicit about sharing non-original content. This was uncommon, and when it occurred the reviewer conducted a reverse image search in those instances to ascertain whether the photo was original content. However, it is possible that an in-person visitor only shared content about the park, e.g. shared a news story, but did not share any original content or post about visiting. This would result in a visitor being coded as a non-visitor. The possibility of this occurrence is a limitation of this method and a minor threat to internal validity.

In order to limit bias resulting from a small number of active users, only one post per backcountry visitor was coded, however all of the user's posts were considered. If they appeared to engage in both fishing and bear viewing, they were coded as engaging in both. If none of the user's posts indicated a bear viewing or fishing activity, it was coded as other (Table 1). In cases where the user shared content that fell into multiple categories, they were coded according to the activity closest to a backcountry visit (e.g. if posts were both unrelated and non-visit, the user was coded as non-visit; if they shared non-visit and Brooks Camp content, they were coded as Brooks Camp; if posts shared both Brooks Camp and backcountry activities, they were coded by their backcountry activity).

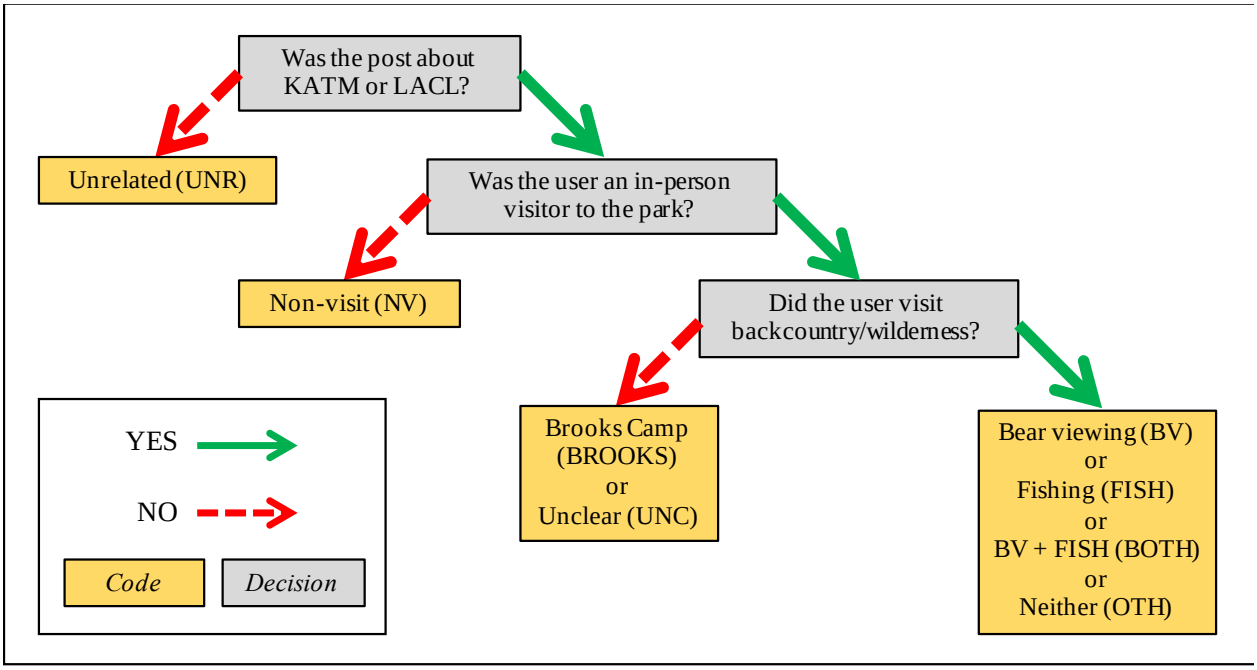

Figure 5. Decision tree used to determine code for social media users through a visual inspection. Users' posts were evaluated collectively, and the user was coded according to the category furthest to the right in this table (for example, if a user shared non-visit content and backcountry bear viewing, they were coded as bear viewing). Table 1 includes detailed description of each decision point.

Table 1. Codebook, which was used in combination with the decision tree in Figure 5 to evaluate user activities and locations.

Was the post about KATM or LACL?	**Yes if at least one of the user's posts:** • refers to the at least one of the parks by name • refers to at least one location in the parks • refers to a commercial entity that operates exclusively in at the parks **No if none of the above**
Was the user and in-person park visitor?	**Yes if at least one of the user's posts:** • Specifies that the user was in a park or a specific park location **No if all of the user's posts fall into these categories, AND no posts qualify as "yes" using the criteria above:** • indicate that the user was watching KATM's bear cameras, include content watermarked with the explore.org name and logo (i.e. screenshots of KATM's bear cameras) • share a news story • discuss KATM's fat bear week contest, make reference to contentious current events (e.g. Pebble Mine controversy) • text suggests that post content belongs to another user (e.g. sharing a famous photographer's photo) • or if user account belongs to a news outlet, guide company, a park, a non-profit, or similar organization

Did the user visit backcountry or wilderness?	**Yes if at least one of the user's posts:** • specifies that the user was at LACL • specify that the use was in a specific backcountry location (including the coast, broadly) • includes both indication that they were in one or both parks AND content of a feature that does not exist at Brooks Camp (e.g. bears eating clams) in the same post **No if all of the user's posts include images or describe features that only exist at Brooks Camp (see below)**
Did the user visit Brooks Camp, or was their location in the park unclear?	**Brooks Camp if posts:** • include platforms, Brooks Falls, Brooks Lodge • include photo angles that can only be taken from Brooks platforms (i.e. photos taking from above a bear) **Unclear if posts:** • include content that indicate the user was in LACL and/or KATM, but information is insufficient to confirm a backcountry/wilderness or Brooks Camp location
Was the user bear viewing, fishing, both, or neither?	**Bear viewing if no posts that indicated sport fishing activity, AND at least one of the user's posts include:** • photos of bears • photos of people photographing bears • description of bear viewing • indication of a guide company that only provides bear viewing tours **Fishing if there were no posts that indicated bear viewing activity, AND at least one of the user's posts include:** • photos of casting fishing line or holding a fish • description of fishing activities **Both if:** • the user shared at least one post that indicates each fishing and bear viewing **Neither if:** • none of the user's posts indicated either bear viewing or fishing

Analysis

Descriptive statistics were used to evaluate the volume of social media users that were in-person visitors, as well as the frequencies of activity types. In addition to a Person's Chi-Squared test to determine if there was a difference between the distributions of activities shared on each platform, a Cramer's V test was conducted to evaluate . Geographic information was mapped and categorized based on location and location type.

Results

Twitter and Flickr content

In all, 46,768 Tweets and 758 Flickr individual posts were collected. After discarding retweets and replies, 11,795 Tweets remained in the dataset. In this dataset, there were 6,087 Twitter users, most of whom were not visitors. A total of 2,113 Twitter users (34.7%) shared content unrelated to the parks, and 3,784 (62.2%) shared content about one or both parks, but not related to an in-person visit. Finally, 190 Twitter users visited the park. Unless otherwise noted, non-visitors are excluded from all further analysis, and analysis will only include the 190 users coded as in-person visitors. The majority of Twitter users shared content from LACL or KATM backcountry visits (122 users), and of those users, 88 shared content related to bear viewing (Table 3, Table 4).

Across 789 Flickr posts, there were 35 unique users. Only one post was not an in-person visit; 34 Flickr users were park visitors. Unless otherwise noted, non-visitors are excluded from all further analysis, and analysis will only include the 34 users coded as in-person visitors. The majority of Flickr users shared content from LACL or KATM backcountry visits (22 users), and of those users, 18 shared content related to bear viewing (Table 3, Table 4).

Ultimately, 144 backcountry visitors were captured between Twitter and Flickr, along with 54 Brooks Camp visitors and 25 visitors in an unidentifiable park location. The majority of backcountry users (73.6%) shared content related to bear viewing.

Platform comparison

A Pearson's Chi-Squared test was conducted to determine whether significant differences exist between Twitter and Flickr users when considering both user locations (Table 3) and

backcountry activities (Table 4). When evaluating locations, users who did not visit were excluded. Significant differences do not exist in how activities (bear viewing, fishing, both, or neither) were distributed (Table 2). The difference between locations approaches significance at 0.052. The major difference between the distributions is that nine Twitter users only fished, while no Flickr users only fished (one engaged in both bear viewing and fishing). When the Chi-Square analysis is run grouping fishing and those who engaged in both bear viewing and fishing, the p-value rises to 0.644 (Chi-square = 0.882). Because of the small number of observations, the Chi Squared value may not be meaningful on its own, however it was used as a rudimentary test of whether datasets could be combined. As a result, datasets were combined and social media users are evaluated as one group for the remainder of this paper, unless otherwise noted.

Table 2. Pearson's Chi-Squared test of the relationship between distribution of visitor activities and location (backcountry, front country, unclear) and backcountry activity (bear viewing, fishing, both, neither) and social media platform.

	Chi-Square value	Asymptotic Significance	Cramer's V
Location	0.548	0.760	0.050
Backcountry activity*	7.716	0.052	0.234
Collapsed backcountry activity**	0.882	0.644	0.079

**Two empty cells, see Table 4*
***Combined fishing-only users with those who shared fishing and bear viewing content*

Table 3. Frequencies of unique user locations of social media users who shared content related to Katmai and Lake Clark National Parks and Preserve June-October 2018.

	Twitter	Flickr	Both
Backcountry visit	122	22	144
Brooks Camp	47	7	54
Location within park unclear	20	5	25
Non-visit park content	3,784	1	3,785
Total	*3,973*	*35*	*4,008*

Table 4. Frequencies (percentages) of unique user backcountry activities of social media users who shared content related to Katmai and Lake Clark National Parks and Preserve June-October 2018.

	Twitter	Flickr	Both
Bear viewing	88 (61.11%)	18 (12.5%)	106 (73.61%)
Neither	25 (17.36%)	3 (2.08%)	27 (18.75%)
Fishing	9 (6.25%)	0 (0%)	9 (6.25%)
Bear viewing and fishing	0 (0%)	1 (0.69%)	1 (0.69%)
Total	*122 (84.72%)*	*22 (15.28%)*	***144 (100%)***

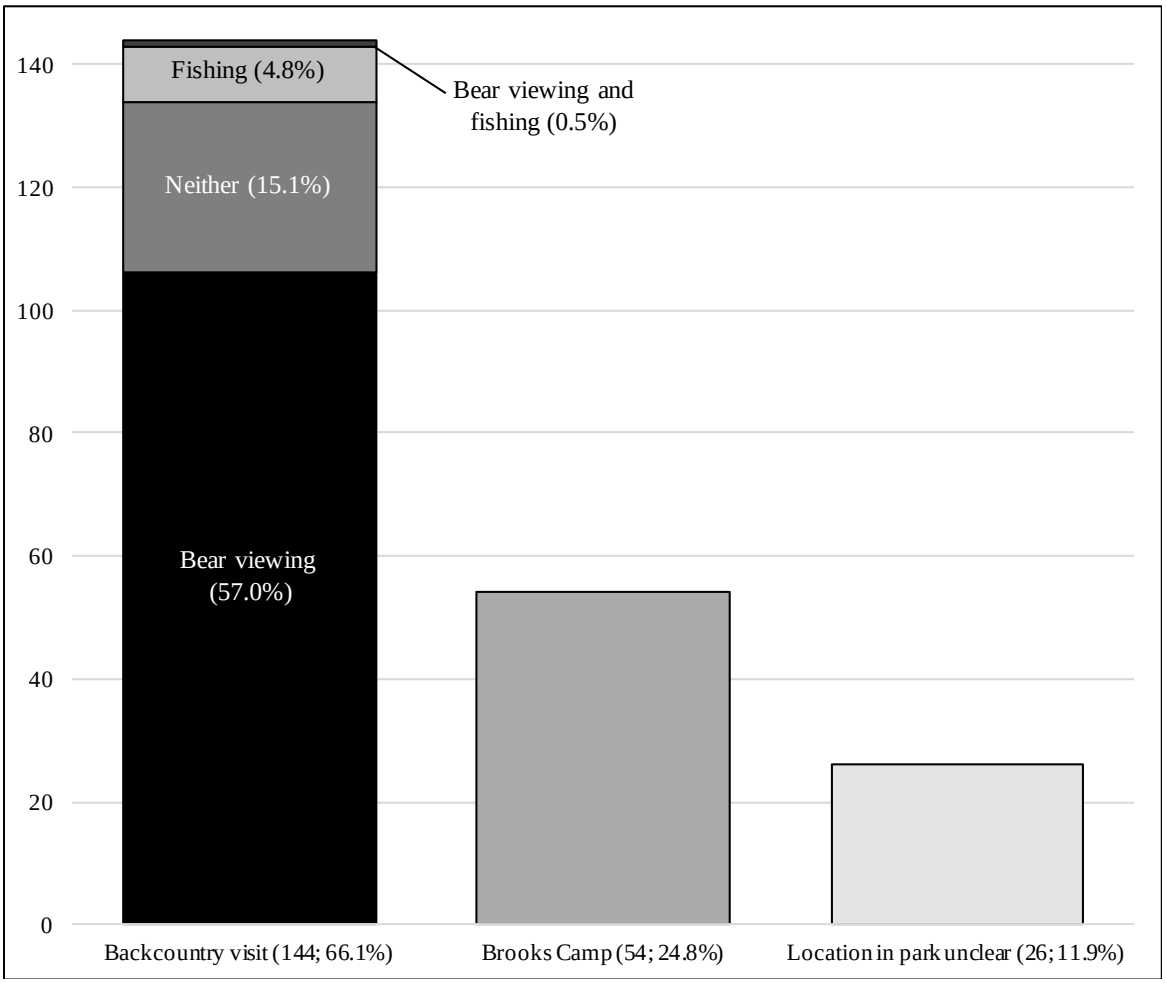

Figure 6. Locations of unique Twitter and Flickr users who visited Katmai and Lake Clark National Park and Preserve, and activities of backcountry visitors.

Spatial results

The majority of users with associated geographic information did not share content related to an in-person visit. While there were 1,834 social media posts with associated geographic information, only 158 unique users comprise that number. There were 35 users coded as sharing non-visit park-related content, and 92 shared content unrelated to the parks.

Of the users who visited the parks in person, only 31 had available geographic information associated with their posts. Of those 31, four were Flickr users and the remainder were Twitter users. This ratio closely reflects the ratio of total Twitter and Flickr visitors (185 and 34, respectively). A total of 31 observations is not sufficient for statistical analysis to have adequate power, but observational analysis can suggest potential characteristics of spatial data in this remote, low-visitation park.

Of the 31 total users with associated geographic data, 16 points were outside the parks (Table 5). Of those, 14 were associated with towns outside of the parks, one in Alaska wilderness over 200 miles northeast of LACL, and one was in Alagnak Wild River (an NPS unit that geographically connects to KATM). Of the 15 points that lay in the parks, three were in Brooks Camp, three were in the frequently visited backcountry areas of Chinitna Bay and Hallo Bay (SWAN, 2017a; SWAN 2017b), one was along a river near the KATM coast, and eight were clustered in one point in KATM. These eight points were all derived from Twitter cross-posts of Instagram content that was manually tagged with Katmai National Park as the location. Because posts with a manual tag on Instagram select the same random point in the area to associate with, it is highly unlikely that the users actually visited this location (Figure 6).

Table 5. Visitor locations based on volunteered geographic information of users coded as in-person visitors to Katmai and Lake Clark National Parks and Preserves

	Number of users
City or town	**14**
Port Alsworth	*11*
Seward	*2*
Anchorage airport	*1*
LACL or KATM	**15**
Instagram point (KATM)	*8*
Chinitna Bay (LACL)	*1*
Hallo Bay (KATM)	*2*
Coastal river (KATM)	*1*
Brooks Camp (KATM)	*3*
Other undeveloped area	**2**
Alagnak Wild River	*1*
Interior Alaska, on edge of Denali NP	*1*
Total	**31**

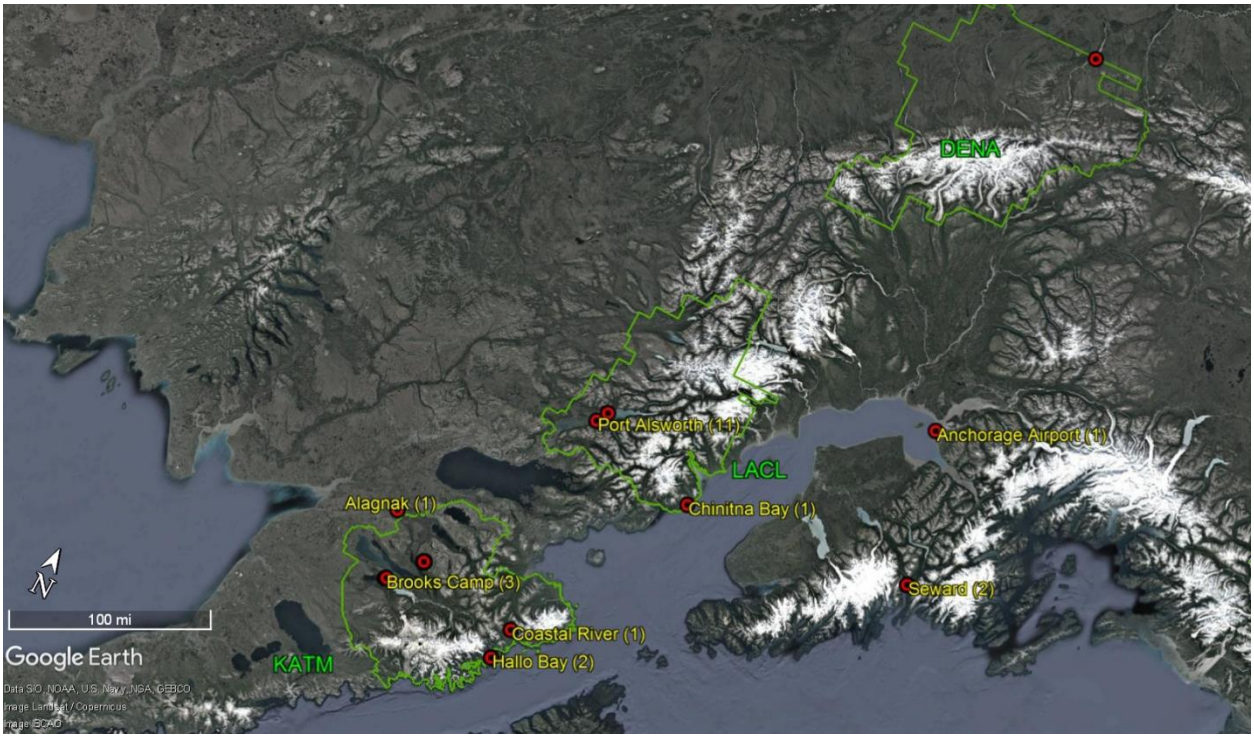

Figure 7. Map of geographic information associated with in-person visitors. Labels and frequencies correspond with Table 5.

Discussion

While past studies have shown correlations between visitation, visitor activities, and spatial density in parks and protected areas, many studies note that these correlations are weaker in lower use, more remote areas (e.g Heikenheimo et al., 2017; Tenkanen et al., 2017; Sessions et al., 2016; Fisher et al., 2018; Walden-Schreiner et al., 2018; Levin et al., 2017). By evaluating the content of posts related to two remote, low-visitation parks, this study investigates how well social media content in these areas represents in-person visitors. Capturing data using a wide net and manually identifying Tweets from actual visitors revealed that only 189 Twitter users who post about the parks are actual in-person visitors, compared with 3,784 users who shared content related to the parks but not related to an in-person visit. This suggests that while Twitter can be used to understand *public* perceptions of the park, it may not be a useful tool for understanding *visitors*. The Flickr data revealed that while there were over 700 posts from the parks, there were only 35 users sharing photos.

This past summer (June-Sept 2018) there were 37,018 visitors to KATM and 12,952 visitors to LACL according to park statistics. When combining both parks, and with a confidence interval of 95%, a sample size of 381 would be required to produce results in the 5% confidence interval. Therefore, a sample of 35 is not adequate to understand visitors' spatial density in the parks. The total volume of users is also not an adequate sample. When the two social media platforms are combined, the result is a sample of 223 visitors total, and 144 confirmed backcountry visitors, which can be used to make very limited conclusions about visitor use. While the small number of observations reduces the power of some quantitative analysis, and makes others impossible, it indicates that social media has very limited application for understanding visitor use in this setting and spatiotemporal scale. While nearly all Flickr users

were in-person visitors, Flickr yielded a very small sample. Twitter's sample was significantly larger, but the amount of time required to identify users who were in-person visitors is likely too great to be of any managerial value.

Overall, there were enough social media users who were in-person visitors to make limited conclusions about social media use by visitors to KATM and LACL. However, 95% of Twitter users who posted about the parks were not visitors. While this does not suggest that user-generated content cannot be used to estimate visitor volume, it does highlight the need to apply quantitative conversion factors if used at a fine scale; additionally, this conversion may vary depending on park site. For example, because KATM has a very popular web livestream and associated "fat bear week" social media campaign, there is potentially more online content shared about the park by non-visitors than places with less active online presences. There were 1,735 users who shared content related to the livestream, and 896 who shared content related to fat bear week (there may be overlap). Social media content from parks with low visitation may be comprised of a higher proportion of in-person visitors at parks with correspondingly low media attention; however, KATM and their livestream has a very active social media presence, and a New York Times article about LACL was published during the study period (see https://www.nytimes.com/2018/07/16/travel/lake-clark-national-park-alaska.html). If researchers or managers choose to evaluate visitors based on social media content where a user manually tags or references the park (in contrast to geotags), the amount of external media or park-driven social media presence should be considered in interpretation of the results. Additionally, these results suggest that filtering user-generated content using this, or another, method is required in order to engage in further analysis of social media user characteristics, such as demographic information, to understand in-person visitors.

The distribution of activity types shared by social media users, compared with data reported by commercially authorized users (CUAs), relatively overestimates bear viewers relative to anglers (SWAN 2018a, SWAN 2018b). There are several possible reasons for this. For one, photography is more often combined with bear viewing than fishing (Sharp et al., 2019). This may also be because bear viewing is a more charismatic activity, i.e. users may perceive that it will return more positive feedback on social media. It is certainly possible that this is due to a limitation in the method; people who view bears and fish only share bear content for any of the reason previously discussed. Future research may benefit from using the primary CUA data to make statistical comparisons between frequencies of CUA-reported activities and activities shared on social media.

While the distribution of activity types reported by CUAs and shared on social media may not align, the ratio of Brooks Camp versus backcountry visitor seems to be align between CUA reported data and social media content. It is not possible to quantify the similarity using just the reports generated by SWAN (2018a, 2018b), however it appears that in 2017 there were approximately 13,500 user days at Brooks Camp and 25,000 user days at the remaining areas of KATM and all of LACL (numbers approximated from graphs, see Figures 1 and 2); in other words, ~35% of CUA reported visitor use days were in Brooks Camp. Social media analysis revealed 54 Brooks Camp visitors and 144 backcountry visitors, or 27.2% Brooks Camp. Future research may also use primary CUA data to evaluate whether significant similarities or differences in locations exist.

Limitations

Several limitations to this study exist. For one, as this is an in-depth study of just one season at just two parks, it is certainly not generalizable to other wilderness areas. Rather, it

serves to draw attention to possible applications and limitations related to using social media data to understand visitors in very remote contexts and at a fine spatial scale.

Additionally, technological barriers to social media data collection continue to limit researchers' and managers' ability to use it. While the data in this study was collected prior to March 12, 2019, Flickr will be severely restricting the number of photos they will host for free users on that date; users who do not pay for their accounts will only be allowed to host 1,000 photos, and photos above that limit will be deleted from their site (Austin, 2019). While only 3% of free users have shared more than 1,000 photos on Flickr, it is important to recognize that social media companies are private entities that can, and do, change what data is available at any time. This has already impacted researcher's ability to use Instagram, a platform that arguably works best for understanding park visitors (Tenkanen et al., 2017; Heikenheimo at al., 2017). Instagram has, in the past, suddenly shut down access to its Application Programming Interface (API), which is the interface that allows external parties to use programming codes to extract data (Volpicelli, 2018). This caused third party apps to malfunction, and they quickly reenacted the API; however, they are continuously making changes to what data is available (see: https://www.instagram.com/developer/).

Implications

Despite the limitations in using social media to understand in-person visitors, high social media attention from non-visitors may be useful for the park in terms of extending the reach of their education or conservation messages (Skibins and Sharp, 2018). In KATM, it is especially relevant to consider engagement with non-visiting online users, as the park's bear-viewing livestream reaches more than 10 million people. Given the difficulty in reaching this remote site,

when evaluating livestream viewers, these online-only "visitors" are economically valued at more than twice that of in-person visitors (Loomis, et al., 2018).

Overall, the results of this study suggest that using social media data at a fine scale in a remote, low-visitation area has some limited use. When casting a wide query net, it is possible to collect sufficient observations to evaluate some visitor characteristics. However, the vast majority of users are not in-person park visitors. Therefore, caution should be used when interpreting user-generated content, since it may not represent in-person visits. This study does not conflict with the work that other researchers have done, evaluating whether there is a correlation between reported visitation and social media volume when comparing many sites or over many years.

Rather, it suggests that consideration need to be paid to scale and interpretation of results. While volume of posts may correlate with visitation volume, managers should use caution if evaluating characteristics of social media users, because users are not necessarily visitors. Qualitatively understanding the content of user-generated content can enhance other metrics, like was done in this study, and can contribute to a well-rounded understanding of what visitors do in parks and protected areas. However, there may be a bias toward certain activities, perhaps those that more naturally include photography or are more photogenic.

Future research may further investigate how visitors' perceptions of social media likability influences what they choose to share. Additionally, future research may adapt the visual inspection methods used in this analysis to investigate whether automated content analysis or machine learning can identify posts shared by in-person visitors. Finally, future research may explore how proximity to cellular signal or internet influences the volume of user-generated content created and shared by in-person visitors.

References

Albert, C., Hauck, J., Buhr, N., & von Haaren, C. (2014). What ecosystem services information do users want? Investigating interests and requirements among landscape and regional planners in Germany. Landscape Ecology, 29(8), 1301-1313.

Arapostathis, S. G., Spyrou, N., Drakatos, G., Kalabokidis, K., Lekkas, E., & Xanthopoulos, G. (2018). Mapping information related to floods, extracted from VGI sources, for effective disaster management within the Greek territory; the floods of West Attica (November 2017 Greece) case study. In 11th International Conference of the Hellenic Geographical Society.

Austin, P.L. (2019, February 7). Flickr is about to delete tons of photos. Here's how to save yours before they're gone. *TIME*. Retrieved from http://time.com/

Boley, B. B., Jordan, E. J., Kline, C., & Knollenberg, W. (2018). Social return and intent to travel. Tourism Management, 64, 119-128.

Cahill, K., Collins, R., McPartland, S., Pitt, A., & Verbos, R. (2018). Overview of the Interagency Visitor Use Management Framework and the Uses of Social Science in its Implementation in the National Park Service. In The George Wright Forum (Vol. 35, No. 1, p. 32). George Wright Society.

Cerveny, L. K., Blahna, D. J., Stern, M. J., Mortimer, M. J., Predmore, S. A., & Freeman, J. (2011). The use of recreation planning tools in US Forest Service NEPA assessments. Environmental management, 48(3), 644-657.

Cessford, G., & Muhar, A. (2003). Monitoring options for visitor numbers in national parks and natural areas. Journal for nature conservation, 11(4), 240-250.

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. Tourism Management, 57, 295-310.

Dawson, C. P., & Hendee, J. C. (2009). Wilderness management: Stewardship and protection of resources and values. Fulcrum Pub.

Donnelly, M. P., Vaske, J. J., Whittaker, D., & Shelby, B. (2000). Toward an understanding of norm prevalence: A comparative analysis of 20 years of research. *Environmental Management*, *25*(4), 403-414.

Haworth, B., & Bruce, E. (2015). A review of volunteered geographic information for disaster management. Geography Compass, 9(5), 237-250.

Hecht, B. J., & Stephens, M. (2014). A Tale of Cities: Urban Biases in Volunteered Geographic Information. ICWSM, 14, 197-205.

Fay, G., Colt, S., & White, E. M. (2010). Data survey and sampling procedures to quantify recreation use of national forests in Alaska. Gen. Tech. Rep. PNW-GTR-808. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 59 p., 808.

Fisher, D. M., Wood, S. A., White, E. M., Blahna, D. J., Lange, S., Weinberg, A., Tomco, M., & Lia, E. (2018). Recreational use in dispersed public lands measured using social media data and on-site counts. Journal of environmental management, 222, 465-474.

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

Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. Journal of Transport & Health, 2(2), 238-247.

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. Computers, Environment and Urban Systems.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. Conservation Letters, 11(1), e12343.

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. ISPRS International Journal of Geo-Information, 6(3), 85.

Heikinheimo, V., Tenkanen, H., Hiippala, T., & Toivonen, T. (2018). Digital Imaginations of National Parks in Different Social Media. On the Way to Platial Analysis: Can Geosocial Media Provide the Necessary Impetus?: Proceedings of the First Workshop on Platial Analysis . Heidelberg university, Heidelberg

Hollenhorst, S. J., Whisman, S. A., & Ewert, A. W. (1992). Monitoring visitor use in backcountry and wilderness: A review of methods. Gen. Tech. Rep. PSW-134. Berkeley, Calif.: Pacific Southwest Research Station, Forest Service, US Department of Agriculture. 10 p, 134.

Jacob, G. R., & Schreyer, R. (1980). Conflict in outdoor recreation: A theoretical perspective. Journal of leisure research, 12(4), 368-380.

Khang, H., Ki, E. J., & Ye, L. (2012). Social media research in advertising, communication, marketing, and public relations, 1997–2010. Journalism & Mass Communication Quarterly, 89(2), 279-298.

Kuentzel, W. F., Laven, D., Manning, R. E., & Valliere, W. A. (2008). When do normative standards matter most? Understanding the role of norm strength at multiple national park settings. Leisure Sciences, 30(2), 127-142.

Kyle, G., Graefe, A., Manning, R., & Bacon, J. (2003). An examination of the relationship between leisure activity involvement and place attachment among hikers along the Appalachian Trail. *Journal of leisure research*, *35*(3), 249-273.

Levin, N., Lechner, A. M., & Brown, G. (2017). An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. Applied geography, 79, 115-126.

Loomis, J., Richardson, L., Huber, C., Skibins, J., & Sharp, R. (2018). A method to value nature-related webcam viewing: The value of virtual use with application to brown bear webcam viewing. Journal of Environmental Economics and Policy, 7(4), 452-462.

Mancini, F., Coghill, G. M., & Lusseau, D. (2018). Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. PloS one, 13(7), e0200565.

Manning, R. E. (2011). Studies in outdoor recreation: Search and research for satisfaction. Corvallis, OR: Oregon State University Press.

Marion, J. L. (2016). A review and synthesis of recreation ecology research supporting carrying capacity and visitor use management decision-making. Journal of Forestry, 114(3), 339-351.

National Park Service. (2018). Petroglyphs National Monument Visitor Use Management Plan / Environmental Assessment. U.S. Department of the Interior. Retrieved from https://parkplanning.nps.gov/document.cfm?parkID=89&projectID=66887&documentID=91665

National Park Service. (2016). Parks With Wilderness. Retrieved from https://www.nps.gov/subjects/wilderness/wilderness-parks.htm

National Park Service. (1991). Natural Resource Management Reference Manual #77. NPS-77. National Park Service. Washington, DC.

Obed Wild and Scenic River. (2002). Obed Wild and Scenic River Final Climbing Management Plan. *U.S. Department of the Interior*.

Olson, C. A. (2014). Survey burden, response rates, and the tragedy of the commons. Journal of Continuing Education in the Health Professions, 34(2), 93-95.

Pastur, G. M., Peri, P. L., Lencinas, M. V., García-Llorente, M., & Martín-López, B. (2016). Spatial patterns of cultural ecosystem services provision in Southern Patagonia. Landscape ecology, 31(2), 383-399.

Richards, Daniel R., and Daniel A. Friess. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. Ecological Indicators 53: 187-195.

Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. M. (2016). Measuring recreational visitation at US National Parks with crowd-sourced photographs. Journal of environmental management, 183, 703-711.

Sharp, R.L., Brownlee, M.T.J., Dagan, D.T. (2019). Visitor Use and Associated Indicators at Katmai National Park and Preserve. Katmai National Park and Preserve, National Park Service, U.S. Department of the Interior.

Skibins, J. C., & Sharp, R. L. (2018). Binge watching bears: Efficacy of real vs. virtual flagship exposure. Journal of Ecotourism, 1-13.

Southwest Alaska Inventory and Monitoring Network. (2017a). Katmai National Park and Preserve Visitor Use. National Park Service.

Southwest Alaska Inventory and Monitoring Network. (2017b). Lake Clark National Park and Preserve Visitor Use. National Park Service.

Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: a geographically weighted approach for mountain landscapes. Ecological Indicators, 64, 237-248.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific reports, 7(1), 17615.

Van Huizen, A., Chruch, C., Thuerk, E., Malin, M., Wojcik, D., Whissen, S., Sharp, R.L., Henderson, J., Blodgett, N.J., Stein, J., Colvin, C., Notzon, C., Turina, F. & McVeigh, M. (2013). Ozark National Scenic Riverways General Management plan/Wilderness Study/Environmental Impact Statement. National Park Service, Van Buren, MO.

Van Mol, C. (2017). Improving web survey efficiency: the impact of an extra reminder and reminder content on web survey response. International Journal of Social Research Methodology, 20(4), 317-327.

van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data. Proceedings of the National Academy of Sciences, 113(46), 12974-12979.

Vaske, J. J. (2008). Survey research and analysis: Applications in parks, recreation and human dimensions. Venture Publ.

Volpicelli, G. (2018, April 28). Can Instagram keep its nose clean? *The Guardian*. Retrieved from http://www.theguardian.com/

Walden-Schreiner, C., Leung, Y. F., & Tateosian, L. (2018a). Digital footprints: Incorporating crowdsourced geographic information for protected area management. Applied Geography, 90, 44-54.

Walden-Schreiner, C., Rossi, S. D., Barros, A., Pickering, C., & Leung, Y. F. (2018b). Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. Ambio, 47(7), 781-793.

Weimer, D. L., & Vining, A. R. (2017). Policy analysis: Concepts and practice. Routledge.

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.

Chapter 3: Reflection

Research Objectives

In writing this thesis, I set out to achieve three general objectives. I believe I began to answer both questions to a certain degree, although the results of this study are not generalizable to other areas or park contexts.

Objective 1: To understand to what degree social media users who share content related to the parks are in-person visitors

Capturing data using a wide net and manually identifying Tweets from actual visitors revealed that only 189 Twitter users who post about the parks are actual in-person visitors, compared with 3,784 users who shared content related to the parks but not related to an in-person visit. This suggests that while Twitter can be used to understand *public* perceptions of the park, it may not be a useful tool for understanding *visitors*. The Flickr data revealed that while there were over 700 posts from the parks, there were only 35 users sharing photos. This past summer (June-Sept 2018) there were 37,018 visitors to KATM and 12,952 visitors to LACL according to park statistics. When combining both parks, and with a confidence interval of 95%, a sample size of 381 would be required to produce results in the 5% confidence interval. Therefore, a sample of 35 is not adequate to understand visitors' spatial density in the parks. The total volume of users is also not an adequate sample. When the two social media platforms are combined, the result is a sample of 223 visitors total, and 144 confirmed backcountry visitors, which can be used to make very limited conclusions about visitor use. While the small number of observations reduces the power of some quantitative analysis, and makes others impossible, it indicates that social media has very limited application for understanding visitor use in this setting and

spatiotemporal scale. While nearly all Flickr users were in-person visitors, Flickr yielded a very small sample. Twitter's sample was significantly larger, but the amount of time required to identify users who were in-person visitors is likely too great to be of any managerial value.

Objective 2: To examine whether social media data can be used to understand the recreation activities of on-site visitors on a site- and season- specific scale

My conclusion is that the answer to this question depends on context. While previous research demonstrates that visitation estimates derived from user-generated content correlate with more traditional visitation estimate techniques, this has been most effective using larger temporal or geographic scales, or areas that are more densely visited. However, in this case, using social media to evaluate park-specific user behaviors may not be feasible due to a combination of low use and fineness of scale.

Objective 3: To examine the limitations of using social media data to understand visitors, specifically whether it overrepresents certain populations of users

In addition to the limitations having to do with the small sample and cell size, described above, this thesis also supports the idea that the type of shared content on social media is not representative of actual behaviors. These results indicate a bias toward posting photos of wildlife, leading to an overrepresentation of bear viewing behaviors relative to fishing, the other primary activity type in these parks. Specifically, social media content was comprised of a higher proportion of bear viewers than National Park Service reports. There are several possible reasons for this. For one, photography is more often combined with bear viewing than fishing (Sharp et al., 2019). This may also be because bear viewing is a more charismatic activity, i.e. users may

perceive that it will return more positive feedback on social media. It is certainly possible that this is due to a limitation in the method; people who view bears and fish only share bear content for any of the reason previously discussed.