

## Validating the Use of Social Media Data to Measure Visitation to Public Lands in Utah

Hongchao Zhang  
*Utah State University*

Jordan W. Smith Ph.D.  
*Utah State University*

Follow this and additional works at: <https://scholarworks.umass.edu/ttra>

---

Zhang, Hongchao and Smith, Jordan W. Ph.D., "Validating the Use of Social Media Data to Measure Visitation to Public Lands in Utah" (2020). *Travel and Tourism Research Association: Advancing Tourism Research Globally*. 39.

[https://scholarworks.umass.edu/ttra/2020/research\\_papers/39](https://scholarworks.umass.edu/ttra/2020/research_papers/39)

This Event is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Travel and Tourism Research Association: Advancing Tourism Research Globally by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

# Validating the Use of Social Media Data to Measure Visitation to Public Lands in Utah

## Introduction

The rise of new technologies such as social media, smartphones, and GPS, provide publicly-available user-generated data to researchers, scientists, practitioners, NGOs, and governments. Social media in particular, provides publicly-available data with geographic information that covers large spatial and temporal scales (Wood et al., 2013; Rashidi et al., 2016; Stock, 2018). Recently, researchers have investigated the use of social media data to develop a better understanding of how many tourists visit specific destinations (Wood et al., 2013; Toivonen et al., 2019). Recent work has also begun to use social media data as a way to understand tourists landscape preferences and on-site experiences (van Zanten et al., 2016; van Berkel et al., 2018). This work suggests social media data can be used to replace, or at least compliment, data collected through traditional methods such as on-site surveys and visitor interviews.

The advantages of using social media data in tourism research have been identified in previous research (Wood et al., 2013; van Zanten et al., 2016; Yoshimura and Hiura, 2017; Walden-Shreiner et al., 2018; Barros, Moya-Gómez and Gutiérrez, 2019). Using social media data for tourism research can reduce the financial costs of, and time required for, data collection (Wood et al., 2013). More importantly, social media data cover larger spatial and temporal scales that allow researchers to answer questions which cross-sectional designs do not (Wood et al., 2013; van Zanten et al., 2016). These benefits can be used to inform tourism planning, destination management, and on-site operational needs.

## Literature Review

Within the context of nature-based tourism in the U.S., visitation data are often collected and reported by government agencies such as the USDA Forest Service, and the National Park Service. This type of data is captured through visitor registrations, trail and vehicle counters, or on-site interviews of systematically sampled destinations. These data are often limited in their temporal and spatial coverage, and are relatively expensive to collect (Fisher et al., 2018). Social media data, by comparison, are available for large spatial and temporal extents and are freely available. While these benefits are appealing, the ability of social media data to accurately measure visitation is questionable. Several studies have compared visitation data collected through traditional means and social media data (Wood et al., 2013; van Zanten et al., 2016; Fisher et al., 2018; Toivonen et al., 2019). This work suggests social media data can serve as a valid proxy to traditional visitation data.

For tourism planners, destination managers, and on-site operators, the acquisition and use of social media data to estimate visitation can be challenging. Individuals first need to confirm the social media data have geographic metadata that can be used to show exactly where an individual was when they took a photo or sent a tweet (Fisher et al., 2018). The majority of social media data do not have geotags, making them useless in efforts to estimate visitation. Social media data also need to be cleaned, so they represent a single visitor day. Wood and his colleagues (2013) proposed the measure of “photo-users day” to address this issue. This measure only captures one photo per user per day at a certain location. Finally, geographically-specific and cleaned social media data should be compared with visitation data collected through other

means (if only for a portion of the area or time under investigation) to ensure they are valid. The purpose of this study is to determine the ability of social media data to proxy visitation data collected through traditional means on public land within Utah, USA.

## **Methodology**

### *Study Area*

Our study region consists of all public land within Utah. We define public land as areas managed by federal agencies such as the National Park Service and the Forest Service, as well as the state of Utah's primary park management agency, the Utah Division of Parks and Recreation. Utah contains five national parks (Arches, Bryce Canyon, Capitol Reef, Canyonlands, and Zion) as well as seven other national park units. The state also contains five National Forests (Ashley, Dixie, Fishlake, Manti-La Sal, and Uinta-Wasatch-Cache) and 45 state parks. Collectively, these areas support a vast array of outdoor recreation opportunities ranging from sailing to alpine skiing. The governing bodies for all of the federally- and state-managed outdoor recreation destinations noted above report the number of visitors using their sites each year. For the National Park Service, the agency's Social Science Program is responsible for establishing counting protocols for how each park unit records a recreation visit (Ziesler, 2019). These protocols vary widely by park unit. Park visitation is aggregated to a monthly time scale. The Forest Service estimates visitation for each national forest at 5-year intervals using data collected through on-site interviews conducted at systematically sampled recreation settings within each forest. The program is referred to as the National Visitor Use Monitoring Program. Finally, the Utah Division of Parks and Recreation reports annual visitation to each state park unit based upon internal protocols established by each park unit. In this study, we use annual unit-specific visitation estimates reported by the National Park Service, the Forest Service, and Utah State Parks to develop a validation model which assesses the validity of using social media data as a proxy to estimate reported visitation to public land within Utah.

### *Data Collection*

Annual visitation data were collected from the National Park Service, the Forest Service, and the Utah Division of Parks and Recreation for the period of time between 2006 and 2014. Annual visitation data of National Park Service units were collected from the agency's Integrated Resource Management Applications (IRMA) portal. Data on annual recreation visits to national forests within Utah were collected from the Forest Service's Natural Resource Manager web portal (<https://www.fs.fed.us/recreation/programs/nvum/>). Visitation data for all Utah State Parks were collected from the Division of Parks and Recreation's website (<https://stateparks.utah.gov/resources/park-visitation-data/>).

We compiled two social media datasets, one containing all posts uploaded to the *Panoramio* platform and the other containing posts to the *Flickr* platform. *Panoramio* was a social media platform, active between 2005 and 2016, which allowed its users to upload geotagged photos and archive posts to a central database. At the time the Panoramio platform was discontinued in 2016, the database consisted of 120 million photos (Toivonen et al., 2019). *Flickr* is a photo-sharing platform that has been in continuous operation since 2004. By the end of 2017, the platform had received 6.5 billion uploads from users.

Posts made to both the *Panoramio* and *Flickr* platforms are accessible through each platform's Application Programming Interfaces (APIs), which allow anyone to download both post content (i.e., the image uploaded, comments, etc.) and metadata (e.g., post coordinates, user identification, upload date, etc.) (Di Minin, Tenkanen, and Toivoen, 2015). We collected data from both platforms through their respective APIs, filtering data by the geographic boundaries of national parks, national forests, and Utah State Parks. The *Panoramio* and *Flickr* data covers the same period of time (2006 to 2014) as the observed visitation reported by each land management agency. We limited time period to the years between 2006 to 2014 because the *Panoramio* API only provides data between 2006 to 2014. In order to contrast the validity of the two platforms, *Flickr* data were also only collected between 2006 and 2014.

### *Data Analysis*

***Photo-user days.*** All social media data were processed and filtered in *R* following similar data cleaning procedures used by Wood et al. (2013) and van Zanten et al. (2016). Specifically, we filtered posts by randomly selecting one post from each user for each day. This results in a dataset comprised of photo-user days, as opposed to all uploaded posts. The filtering process is necessary because multiple uploads per day should be attributed to the same recreation visit, as opposed to multiple visits (Wood et al., 2013; van Zanten, et al., 2016).

***The Validation Model.*** We used OLS regression models to determine if social media data can be used as a proxy for traditional visitation data. The models were developed to examine the statistical relationship between the total annual visits to an area in a specific year ( $y_{it}$ ) with total number of photo-user days within that same area for the same year ( $x_{it}$ ). The model can be specified as:

$$y_{it} = x_{it} + \epsilon_{it}$$

where the subscripts  $i$  and  $t$  refer to each study areas and each year respectively. The standard residual error is denoted as  $\epsilon_{it}$ .

## **Results**

### *Reported Visitation and Photo-user Days*

Summary statistics for both reported visitation and photo-user days are presented in Table 1. For national parks and state parks, which report visitation for each park unit annually, visitation gradually increased between 2006 and 2014. The same trend is not observable with the Forest Service data given they only report data for each forest, which vary considerably in their use levels, every five years. None of the national forests in Utah were surveyed in 2010. For the National Park Service, photo-user days also increased each year between 2006 and 2014. Over this time, photo-user days ranged from 4 for Rainbow Bridge National Monument and Timpanogos Cave National Monument in 2006 to 5,515 for Arches National Park in 2013. There are no obvious trends in the amount of photo-user days for either national forests or Utah State Parks units. Variation in photo-user days for national forests ranged from 32 for the Fishlake National Forest in 2006 to 7,442 for Uinta-Wasatch-Cache National Forest in 2011. The range of photo-user days for Utah State Parks spanned 0 for Anasazi State Park in 2006 to 766 for Great Salt Lake State Park in 2013.

As shown in Table 2, the average ratio of annual photo-user days to reported visitation gradually increased between 2006 and 2014, with a range from 0.0002 in 2006 to 0.0022 in 2013. For each type of public lands, there is an observable increasing trend from 2006 to 2014. For national parks, the variation in ratio covered a range from 0.0002 in 2006 to 0.0022 in 2013. This ratio ranged from 0 in 2010 to 0.0252 in 2011 for national forests (more variation exists for national forests because visitation data are only collected every five years). The Utah State Parks units received a lowest ratio of annual photo-user days to reported visitation, with a range spanning 0.0001 in 2006 to 0.0009 in 2014.

### *Model Validity*

Goodness of fit ( $R^2$ ) statistics generated through OLS regression models using data from each social media platform independently as well as data from both platforms combined, are provided in Table 3. In general, the models explained about two-thirds of the variance in visitation to national parks, national forests, and state parks in Utah ( $R^2=0.68$ ). Both the *Panoramio* and *Flickr* data explained comparable proportions of the variance in reported visitation. Given this, and previous research documenting relatively little differences in the spatial variation in posts across platforms (van Zanten et al., 2016), subsequent analysis utilizes the combined social media datasets.

The OLS models revealed substantial differences in the ability of photo-user days to proxy reported visitation across the three types of public lands. Specifically, photo-user days were a substantially better predictor of visitation to national forests ( $R^2=0.79$ ) and national parks ( $R^2=0.73$ ) relative to Utah State Parks units ( $R^2=0.29$ ).

Results from our validation model suggest a strong relationship between photo-user days and reported visitation (Table 4). Combined photo-user days were significantly and positively related to reported visitation to public lands managed by the National Park Service, the Forest Service, and state parks in Utah (Coef.  $\leq 0.56$ ;  $p \leq 0.01$ ). Collectively, results from our validation model present differences across three types of public lands (Table 4). Similar to the  $R^2$  statistics, our OLS validation model shows that photo-user days were a significantly better predictor of reported visitation to national forests (Coef.  $\leq 0.78$ ;  $p \leq 0.01$ ) and national parks (Coef.  $\leq 0.73$ ;  $p \leq 0.01$ ) than Utah State Parks units (Coef.  $\leq 0.24$ ;  $p \leq 0.01$ ). The significant differences exist among national parks, national forests, and state parks can be explained by the ratio of photo-user days to reported visitation (Table 2). The ratio of photo-user days to reported visitation at state parks was about half the average ratio for all three types of public lands.

### **Conclusion and Discussion**

Our study suggests social media posts in national parks, national forests, and state parks are significantly related to onsite observed visitation data. Although social media posts and onsite observed visitation to state parks were also significantly related to reported visitation, the relationship was weaker relative to national parks and national forests. We suggest that social media posts can be used by tourism planners, destination managers, and on-site operators to understand tourism demand. Within Utah, we have demonstrated one of the ways this can be done by identifying visitation 'hot-spots' (concentrations of geotagged social media posts) and presenting them to tourism planners, destination managers, and on-site operators in a series of

workshops and a public website ([www.utahsmostvisited.com](http://www.utahsmostvisited.com)). Destination planning and management needs to be informed by a scientifically-grounded estimate of tourism demand; this work demonstrates how social media data can be used towards those ends. The research also highlights the important “next steps” that researchers can take to ensure tourism planners, destination managers, and on-site operators can make more-informed management decisions in the future.

**Table 1. Annual observed visitation and social media post**

Unit Types	Annual Visitation								
	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b><i>Reported Visitation</i></b>									
National Park Units	8,182,501	8,472,139	8,742,098	9,047,488	9,276,527	9,606,786	9,877,368	9,329,851	10,910,966
National Forests	672,000	7,330,000	531,000	561,000	0*	352,000	7,924,000	337,000	787,000
State Parks Units	4,465,294	4,751,582	4,540,957	4,822,777	4,842,918	4,803,770	5,081,558	4,044,215	3,720,873
<b><i>Photo-user Days</i></b>									
National Park Units	1,307	4,193	7,646	9,395	10,332	12,646	15,360	18,685	14,494
National Forests	1,297	3,800	6,361	5,102	4,698	8,870	8,993	7,660	5,967
State Parks Units	413	914	1,923	1,528	1,419	2,327	3,545	4,368	3,412

Note: \* None of the national forests in Utah were surveyed in 2010.

**Table 2. Results of the ratio of annual photo-users-day to reported visitation data for National Park Service units, national forests, and state parks in Utah between 2006-2014.**

Unit Types	Ratio of Annual Photo-users-day to Reported Visitation								
	2006	2007	2008	2009	2010	2011	2012	2013	2014
Average Ratio	0.0002	0.0004	0.0012	0.0011	0.0012	0.0016	0.0012	0.0022	0.0015
National Park Units	0.0002	0.0005	0.0009	0.0010	0.0011	0.0013	0.0016	0.0020	0.0013
National Forest Units	0.0019	0.0005	0.0120	0.0091	0.0000*	0.0252	0.0011	0.0227	0.0076
Utah State Parks Units	0.0001	0.0002	0.0004	0.0003	0.0003	0.0005	0.0007	0.0011	0.0009

Note: \* None of the national forests in Utah were surveyed in 2010.

**Table 3. R-square Table**

Dataset	Site Types			
	NPS	USFS	Utah State Parks	Overall
<b>Panoramio</b>	0.74	0.76	0.32	0.73
<b>Flickr</b>	0.72	0.79	0.38	0.68
<b>Combined</b>	0.73	0.79	0.29	0.68

Note: NPS = National Park Service; USFS = U.S. Forest Service

**Table 4. Results of the validation model examining the relationship between annual photo-user days and reported visitation for National Park Service units, national forests, and state parks in Utah between 2006-2014.**

	Coef.	SE	T-value	Sig.
<b>Overall Photo-user Days</b>				
Intercept	265.92	48.60	5.47	0.00***
Year	-0.13	0.02	-5.27	<0.00***
Post	0.56	0.04	16.101	<0.00***
<b>NPS Photo-user Days</b>				
Intercept	302.59	73.69	4.11	0.00***
Year	-0.15	0.04	11.83	<0.00***
Post	0.73	0.06	11.83	<0.00***
<b>USFS Photo-user Days</b>				
Intercept	342.01	155.00	2.21	0.06*
Year	-0.17	0.08	-2.15	0.07*
Post	0.78	0.16	4.88	0.00***
<b>Utah State Parks Photo-user Days</b>				
Intercept	178.83	63.53	2.82	0.01**
Year	-0.08	0.03	-2.64	0.01**
Post	0.24	0.06	4.28	0.00***

Note: NPS = National Park Service; USFS = U.S. Forest Service; \*  $p$ -value = 0.1; \*\*  $p$ -value = 0.05; \*\*\*  $p$ -value = 0.01.

## References

- Barros, C., Moya-Gómez, B., & Gutiérrez, J. (2019). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. *Current Issues in Tourism*, 0(0), 1–20. <https://doi.org/10.1080/13683500.2019.1619674>
- Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, 3, 63.
- Fisher, D. M., Wood, S. A., White, E. M., Blahna, D. J., Lange, S., Weinberg, A., ... 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. <https://doi.org/10.1016/j.jenvman.2018.05.045>
- Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75, 197–211. <https://doi.org/10.1016/j.trc.2016.12.008>
- Stock, K. (2018). Mining location from social media: A systematic review. *Computers, Environment and Urban Systems*, 71, 209–240. <https://doi.org/10.1016/j.compenvurbsys.2018.05.007>
- Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., ... Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, 233, 298–315. <https://doi.org/10.1016/j.biocon.2019.01.023>
- Van Berkel, D. B., Tabrizian, P., Dorning, M. A., Smart, L., Newcomb, D., Mehaffey, M., ... Meentemeyer, R. K. (2018). Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. *Ecosystem Services*, 31, 326–335. <https://doi.org/10.1016/j.ecoser.2018.03.022>
- Van Zanten, B. T. van, Berkel, D. B. V., 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. <https://doi.org/10.1073/pnas.1614158113>
- Walden-Schreiner, C., Rossi, S. D., Barros, A., Pickering, C., & Leung, Y.-F. (2018). Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. *Ambio*, 47(7), 781–793. <https://doi.org/10.1007/s13280-018-1020-4>
- 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.
- Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services*, 24, 68–78. <https://doi.org/10.1016/j.ecoser.2017.02.009>
- Ziesler PS. 2019. [Statistical abstract: 2018](#). Natural Resource Data Series. NPS/NRSS/EQD/NRDS—2019/1219. National Park Service. Fort Collins, Colorado