# 33 Opportunities and limits of social media data for monitoring and valuing visitation in protected areas – evidence from the German national parks

**Marius Mayer[1]**, **Michael Sinclair[2]**, **Manuel Woltering[3]**, **Andrea Ghermandi[2]**, [1]University of Innsbruck, Austria. [2]University of Haifa, Israel. [3]Julius-Maximilians-Universität Würzburg, Germany

As exemplified by the recent review articles by Ghermandi and Sinclair (2019) and Teles de la Mota and Pickering (2020), social media as data source for environmental sciences, nature-based tourism and visitor monitoring is a highly dynamic field. However, what is missing despite many promising results are validations of the results of social media analyses against those obtained with "traditional" onsite visitor monitoring approaches. Therefore, we compare in this contribution social media data (Sinclair et al. 2020a,b) and onsite survey results (e.g. Job et al. 2016) in nearly all German national parks with respect to visitor provenance, visitor type (local, day-tripper, vacationists) and recreational value of national park visits (consumer surplus). Furthermore, we estimate visitation numbers based on social media data for protected areas without systematic visitor counting and provide information about the spatio-temporal visitor behavior in the protected areas.

The methodology is described in detailed in Sinclair et al. 2020a,b. Our approach relies on information from the photo-sharing platform Flickr, from which we downloaded the metadata of all pictures taken between 2005 and 2018 in the GIS-specified spatial envelopes of the 16 German national parks. These pictures were aggregated to photo-user days (PUD), that means we can control for multiple photographs uploaded by individual visitors during a single visit. As only 34% of the identified national park visitors uploading photos on Flickr provided geospatial information on their home city/region in their user profile, we applied several approaches to predict national park visitors' home location (Bojic et al. 2015). This is necessary to group them into visitor types and to estimate the recreational value of national parks based on travel cost models (TCM). Regarding the last-mentioned step, we exactly replicated the TCM approach from Mayer and Woltering (2018) to allow direct comparisons between onsite survey- and social media-based TCM. The visitor structures of the onsite surveys were compared using correlation and regression analyses. All in all, our analyses are based on 71,974 Flickr photos, aggregated to 15,993 PUD.

One important aspect of the research was to compare the representation of different visitor types from social media to that found by onsite surveys which consist of over 24,000 interviews undertaken at the parks. The Pearson correlation values comparing shares of visitor types between social media and onsite survey data are 0.77** for international, 0.83** for local and 0.98** for non-local German visitors (** indicates p-value <0.01). Non-local Germans are underrepresented by 30%, while local and international visitors are both overrepresented by 15%. The overrepresentation of international visitors is higher for more promoted and marketed parks such as Berchtesgaden (see Sinclair et al. 2020a). Regarding the TCM results, most parks show a similar consumer surplus value per trip if we compare survey- and social media-based TCM approaches. The mean absolute error (MAE) per trip is EUR 4.93 and the mean absolute percentage error (MAPE) is 22%. Such results compare favourably with the unit value transfer approach, a common alternative to primary surveys (Sinclair et al. 2020b). Thus, our results successfully validate the TCM approach based on social media data, affirming its potential as an alternative to the valuation of recreation in natural and/or protected areas. This approach can successfully replicate the results of conventional TCM based on representative onsite surveys.

Based on these experiences (and Ghermandi & Sinclair 2019 as well as Teles de la Mota & Pickering 2020), what are the pros and cons of social media as information source about nature-based recreation activities? On the pro-side data can be collected from distance, in a relatively simple way and at low costs; they complement other data sources and can help effectively analyzing the spatial and temporal patterns of visitation; social media data can help evaluate different cultural ecosystem

services, assess visitors' emotions and recreational values; they provide similar results compared to surveys and/or visitor counts; they allow for the timely update of visitation estimates and recreational values. Negative aspects contain varying popularity of social media platforms and limited data access; social media data only represent a relatively small subset of visitors, areas/sites and activities; they often include limited socio-demographic information and require the clarification of ethical issues concerning data security and protection of users' privacy.

To sum up, social media data (like Flickr) can provide important insights for protected area and visitor management/monitoring. Social media can be used as a good proxy to determine visitor preferences for nature-based recreational experiences (Heikinheimo et al. 2017) as these preferences are not significantly different from the general recreational users in natural sites (Hausmann et al. 2017). However, we argue that they are not able to completely replace traditional visitor monitoring approaches for the following reasons: first, in many cases there are not enough social media data to exactly estimate yearly visitation numbers; second, there are (still) few possibilities to derive visitors' socio-demographics, motivations, attitudes and expenditure behavior from social media data. In this way it seems appropriate to suggest using social media approaches where no systematic visitor monitoring would be possible otherwise, be it for financial or organizational reasons (e.g. less strict protected area categories like German nature parks, or unmanaged PAs like nature reserves, Natura 2000-areas). Nevertheless, we recommend including questions about the social media behavior of PA visitors in onsite surveys in order to get representative information about the socio-demographics of social media using PA visitors.

**References**

Bojic I et al. 2015. Choosing the right home location definition method for the given dataset. In Liu TY et al. (Eds.), Social Informatics. Lecture Notes in Computer Science, vol. 9471. Springer, Cham, Switzerland. Ghermandi A & Sinclair M. 2019. **https://doi.org/10.1016/j.gloenvcha.2019.02.003**. Hausmann A et al. 2017. **https://doi.org/10.1111/conl.12343**. Heikinheimo V et al. 2017. **https://doi.org/10.3390/ijgi6030085**. Job et al. 2016. Regionalwirtschaftliche Effekte durch Naturtourismus in deutschen Nationalparken als Beitrag zum Integrativen Monitoring-Programm für Großschutzgebiete. BfN, Bonn-Bad Godesberg. Mayer M & Woltering M. 2018. **https://doi.org/10.1016/j.ecoser.2017.12.009**. Sinclair M et al. 2020a. **https://doi.org/10.1016/j.jenvman.2020.110418**. Sinclair M et al. 2020b. **https://doi.org/10.1016/j.ecoser.2020.101165**. Teles de la Mota V & Pickering C. 2020. **https://doi.org/10.1016/j.jort.2020.100295**.