Valuing recreation in Italy's protected areas using spatial big data

Michael Sinclair[1,2,3], Andrea Ghermandi[1,2], Giovanni Signorello[4,5*], Laura Giuffrida[4], Maria De Salvo[4,5]

[1] Department of Natural Resources and Environmental Management, University of Haifa, Haifa, Israel.
[2] Natural Resources and Environmental Research Center, University of Haifa, Israel.
[3] Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow, UK
[4] Department of Agriculture Food and Environment, University of Catania, Italy.
[5] Centre for the Conservation and Management of Nature and Agroecosystems (Cutgana), University of Catania, Italy.

*Corresponding author email: giovanni.signorello@unict.it

Abstract

Protected areas offer unique opportunities for recreation, but the non-market nature of these benefits presents a significant challenge when trying to represent value in the decision-making processes. The most common techniques to value recreation are based on resource-intensive primary surveys which are difficult to perform at a large scale or in remote locations. This is true in the case of Italy, where a large and diverse network of protected areas suffers from lack of data. Here, we offer an alternative data source for the valuation of recreation by integrating the metadata of geotagged photographs from social media into single-site, individual travel cost models for 67 Italian protected areas. Count data model results are generally consistent with standard economic and consumer demand theory for ordinary goods, with a zero-truncated Poisson model returning down sloping demand curves for 50 of 67 sites. A significant travel cost coefficient was returned for 33 sites (p-value <0.05) for which consumer surplus estimates were found in the range between €6.33 and €87.16, with a mean value per trip of €32.82. Although not without their own challenges, the results presented highlight the possibilities of new forms of spatial big data as a novel data source for environmental economists.

Keywords: social media; spatial big data; travel cost method; protected areas; Italy; recreation; Flickr; geotagged photographs.

1. Introduction

Protected areas (PA) provide a range of vital ecosystem services (Haines-Young and Potschin-Young, 2018) which generate significant value for society (Balmford *et al*., 2002; Harmon and Putney, 2003; Heagney *et al.*, 2019). These spaces are important for ecology and biodiversity while also providing a range of important cultural ecosystem services (CES) to humans, including offering spaces to engage in recreational activities (de Groot *et al*., 2012; Kettunen and ten Brink, 2013; Balmford *et al*., 2015). Exploring recreation to PA generally requires detailed visitation data (Schägner *et al*., 2017; Spenceley *et al*., 2021). Given the free access nature of the majority of protected areas, the monitoring of access and use is not common and generally only happens as part of a dedicated research remit, often relying on time-consuming and expensive survey-based monitoring which can quickly become outdated. This is particularly true for large, remote, and geographically dispersed protected areas where complexities and costs in data collection are substantial. These issues are true in the Italian context given the number of large sites and diverse geography of its natural areas. Previous research on the recreational use of natural areas and valuation of its benefits relies on the meta-analytical investigation of previous, independently conducted primary valuation efforts (De Salvo and Signorello, 2015) or on a limited number of surveyed sites (Schirpke *et al*., 2018). The currently available attempts for a comprehensive characterization of nature-based recreation at the country level involve probabilistic approaches to the determination of recreation-driven mobility functions, which are primarily informed by indirect, contextual variables such as population density and accessibility (Capriolo *et al.*, 2020), and rely on behavioral data extrapolated from surveys conducted in other contexts (Paracchini *et al.*, 2014). Due to lack of data, economic valuations of country-wide nature-based recreation in Italy were limited to the aggregation of estimated fuel costs for reaching the sites (Capriolo *et al.*, 2020) or based on a simple unit value transfer (or average value transfer) (Balmford *et al.*, 2015; Schägner *et al.*, 2016). In general, there is a lack of systematic monitoring data on recreation and visitation to PA in Italy and the value this creates for society.

To ensure the non-market benefits of CES are adequately represented in the decision-making process, environmental economists implement valuation techniques based on stated or revealed preferences (Champ *et al*., 2013). Of these, the Travel Cost Method (TCM) is the most well established for the valuation of recreational benefits. The TCM extracts welfare estimates based on the assumption that travel behaviour can be used to infer a demand for access to the recreational experience (Parsons, 2013). The TCM comes in a variety of forms, from single-site models focusing only on the intensive margin component of recreation decision, to multi-site models. The latter can be based on random utility theory, which analyses the extensive margin aspect of recreation behavior, or on corner solution frameworks such as Kuhn-Tucker (K-T) models, which exploit both components (intensive and extensive margins) of recreation consumption (Parsons, 2013; Nicita *et al*., 2016). Although the strengths and limitations differ between methods, two consistent challenges lie in the resource-intensive nature of the survey process on which these TCM rely and in relying on cross-section data for analyses over long time periods (Cooper and Loomis, 1990).

The global penetration of smartphones and the integration of GPS technology in mobile phones have generated large volumes of novel behavioural data for use in environmental research (Ilieva and McPhearson, 2018; Ghermandi and Sinclair, 2019; Cui *et al*., 2021). Available data sources range from social media data, such as geotagged photographs (e.g., Flickr, Instagram) or text (e.g., Twitter, Weibo), to GPS data generated when users interact with their mobile devices (e.g., call data records). These new forms of spatial big data enable researchers to expand the understanding of human mobility at wider spatial and temporal scales than traditional forms of survey-based data allow. This is particularly important in the context of human-nature interactions where a lack of mobility data is a challenge especially in large or remote sites. The last decade or so has witnessed substantial growth in the number of studies assessing nature-based recreation using forms of spatial big data (Signorello *et al*., 2018; Ghermandi and Sinclair, 2019; Cui *et al*., 2021), particularly in the context of protected areas (Barros *et al*., 2021; Wilkins *et al*., 2021).

An emerging area of research utilizes spatial big data as a source of revealed preferences for the valuation of nature-based recreation which can overcome some of the limitations of primary data collection techniques (Fisher *et al*., 2018; Legget *et al*., 2017). This opportunity is made possible because GPS data has been found as a suitable proxy for recreational visitation to natural areas (Wood *et al*., 2013) and also as a source of data to estimate the home location of visitors to natural areas (Sinclair *et al*., 2020a). Combining these factors with the TCM technique may allow researchers to generate estimates of recreational value without the need to undertake resource-intensive surveys. Towards this end, spatial big data from mobile phones and social media have recently been utilized to value recreation in the context of national parks (Sinclair *et al*., 2020b), urban greenspace (Cui *et al*., 2021), wetlands (Ghermandi, 2018; Sinclair, Ghermandi and Sheela, 2018), lakes (Keeler *et al*., 2015) and beaches (Jaung and Carrasco, 2020; Kubo *et al*., 2020). Research findings so far are promising, with value estimates comparing well to those generated by representative surveys (Sinclair *et al*., 2020b). However, most research has applied the zonal TCM, i.e., a simple variant of TCM that lacks strong theoretical support (Loomis and Walsh, 1997), and applications have been restricted to a selection of larger more remote sites or a large selection of smaller sites. An opportunity exists to use these new forms of spatial big data to extend the analysis to a large sample of large and dispersed sites while also testing different types of TCM.

In this paper, we explore a novel approach to value recreation at large spatial scales and without the need for expensive field-based surveys by integrating geo-location data from social media photographs into the individual TCM valuation technique, and demonstrate it in application to 67 national parks and protected areas in Italy. Compared to previous research, the novelty of the research is twofold. Firstly, we expand on the literature in terms of the complexity and specifications of TCM tested, demonstrating how such techniques can be applied to the large spatial scale, number and diversity of study sites represented by the entire set of the largest terrestrial protected areas in Italy. Secondly, we go some way to filling an empirical gap in knowledge on recreational value through the case study application of Italy's protected areas and national parks where there is a general lack of systematic data on recreation.

2. Materials and methods

2.1 Extracting Flickr geotagged photographs in Italian protected areas

Boundaries for Italian PA were extracted from The World Database of Protected Areas (WDPA), the most comprehensive global database of marine and terrestrial protected areas[1] (UNEP-WCMC and IUCN, 2020). The spatial boundaries of Italy's 25 national parks as well as terrestrial PA with a designated International Union for Conservation of Nature (IUCN) category and an area over 100 km^2 were extracted from the database for use in the project. We limited the investigation to relatively large PA in the hope that an adequate number of geotagged photographs were available. This limitation can be overcome in future studies, by using data fusion techniques to integrate in the analysis data from multiple social media sources, such as Instagram and Twitter (Ghermandi *et al.*, 2020; Ma, Kirilenko and Stepchenkova, 2020). This resulted in a total of 67 PA, the spatial extension of which are represented in Figure 1 (see Appendix S1 for the full details of the PA). The metadata of public geotagged Flickr photographs (https://www.flickr.com/) taken within these 67 PA was collected using code written in R 3.5.0 (R Core Team, 2020) by calling the "flickr.photos.search" function on the website's application programming interface (API). Data were retrieved for the period between January 1, 2005 and December 31, 2018[2] using a georeferenced bounding envelope for each PA. The data collected includes the GPS coordinates, user ID, photo ID and timestamp. Photo-user-days (PUD), i.e., unique combinations of user ID, date and recreation site, were extracted from the dataset for each PA to correct for the fact that a single user can take multiple photographs during a single visit (Wood *et al.*, 2013).

[1] Protected areas which were entirely marine were not included (coded 2 by the WPDA).
[2] Data was extracted before the change of Flickr ownership, from Yahoo to SmugMug, which occurred between 2018/2019 and led to alterations in the website's free service, ultimately limiting the number of photographs a user could upload to 1000. Data from 2019 and 2020 was not included in the analysis because the changes in data availability would have led to inconsistencies in the database.

To test the suitability of the Flickr data in the study area it is common to compare PUD with visitor numbers from official sources. Doing so allows to test the fit of the data and to facilitate predictions on the number of annual visits for sites where official data is absent. Unfortunately, in the case of Italy, there is a general lack of official visitor data for natural areas. In absence of comprehensive visitor data for the 67 PA, we utilize estimated visitor days to the 25 national parks available for 2016 (latest available data), based on stays at lodging facilities located within the park boundaries[3]. After removing the two marine national parks, justified by the expected reduced use of phones and cameras while undertaking water-based recreation activities (Ghermandi *et al.*, 2020), an ordinary least squares model was administered:

$$logY_i = \beta_0 + \beta_{PUD}\, logPUD_i + \beta_{AREA}\, AREA_i + \alpha + \varepsilon \qquad (1)$$

where Y_i is the natural logarithm of the visitor days for the ith site in 2016; PUD_i is the total PUD count for the ith site (2005-2018[4]); $AREA_i$ is the area in square kilometers for the ith site; β_0, β_{PUD} and β_{AREA} are the intercepts and slope coefficients respectively, and ε is the error term. The assumption of normality for the distribution of residuals was tested with the Shapiro-Wilk diagnostic test and outliers were checked using Cook's d. Potential heteroskedasticity of the residuals was investigated using the Breusch-Pagan test. Testing for outliers returns one park (Parco Nazionale Del Gran Sasso E Monti Della Laga[5]) with a Cook's d of 0.50, which substantially exceeds the conventional threshold of four times the average value of d (mean d = 0.055). This park was removed, and the regression redone with a sample of 22 sites. This final model was used to estimate annual visitation for PA using the total PUD and area for the respective site.

2.2 Home location analysis of social media visitors

For all visitors, a home city was extracted from the social media profile where publicly available by calling the "flickr.people.getInfo" function on the website's API. For users who did not provide a home city, we inferred a home location based on their public geotagged photographs. Research has shown that the public social media data of a user can be used to assign a home region with good

[3] https://annuario.isprambiente.it/ada/downreport/html/7032
[4] We also tested the analysis on Flickr data from 2016 only but found results were weaker than aggregating 2005-2018 data. This could be because aggregating data provides a larger sample for the analysis while also accounting for the changing number of Flickr users when using data from one year only.
[5] This park has the lowest density of accommodation establishments and beds among all national parks (data is for 2016, same source as the visitors data): since the visitors data we have is based on visitor days in accommodation establishments, it stands to reason that the official data might be somewhat off in a site that has a low density of such establishments since it might overlook day-time hikers (according to Iezzi and Zarelli (2015) most visitors in the Gran Sasso are "hikers who know the area", thus presumably including a lot of locals).

accuracy (Ghermandi, 2018; Sinclair *et al.* 2018; Sinclair *et al.*, 2020a). For the subset of users who did not state their home city, we first estimated their home country using the Database of Global Administrative Areas (GADM, v2.8) (https://gadm.org/). The home country was assumed as that where they recorded the most active days, given the location of their entire library of public Flickr photographs. This is the most accurate technique recorded in the literature (Sinclair *et al*., 2020). For the subset of Italian visitors, we subsequently estimated their home region[6] using the same technique that was performed at country level. Finally, for use in the TCM analysis, a unique latitude and longitude home location was estimated for each Italian visitor by mapping the public photographs within their home region and determining the coordinates which minimised the mean distance between the points[7] (Sinclair et al., 2018; Sinclair et al., 2020b).

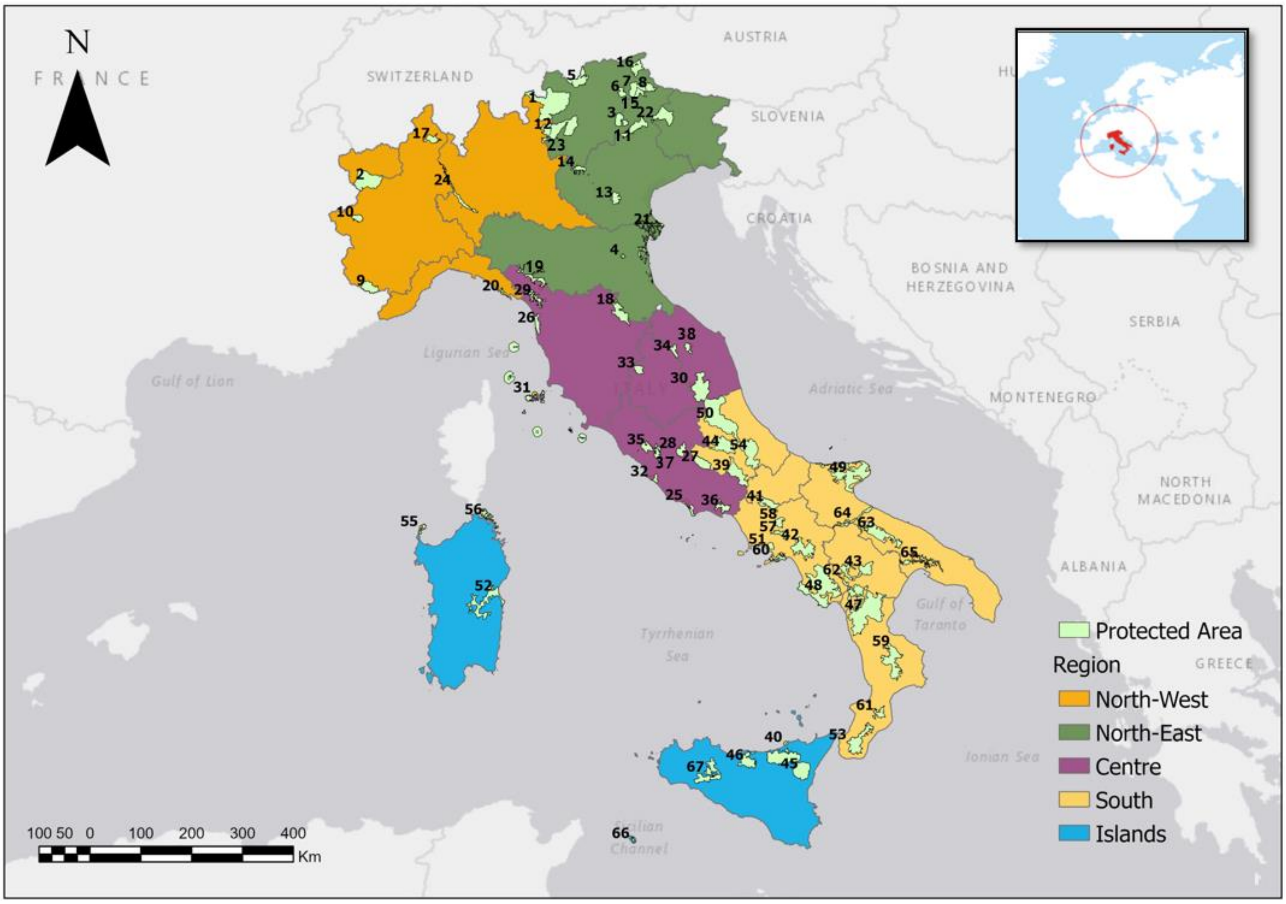

Figure 1: location and ID of protected area study sites in Italy

2.3 Crowdsourced travel cost method

[6] Using the second spatial level of the Database of Global Administrative Areas (GADM, v2.8) (https://gadm.org/) which corresponds in Italy to level two of the EU-Nomenclature of Territorial Units for Statistics (NUTS2).

[7] For users who stated their home city, we took the coordinates from the Google Maps API.

To perform an individual TCM for each site, it was necessary to extract the number of trips an individual made to a site, his or her income (for the estimation of the opportunity cost of time), and the associated travel costs. The number of trips to PA was extracted from the database of PUD for each visitor (see section 2.1). In absence of individual income data for Flickr visitors, we estimated a proxy of individual income using data available at municipality level[8], transformed into an hourly rate (Eurostat, 2021). Estimating a visitor's travel was performed in five stages. First, we extracted the point on the PA boundary which minimized the straight-line distance from a visitor's home location (see section 2.2) using the 'dist2Line' function of the geosphere package (v1.5–10; Hijmans, 2019) in R 3.5.0 (R Core Team, 2020). We considered this as a conservative estimate of the distance given it is the closest point on the PA to a visitor's home. Second, to calculate the round-trip distance and travel time based on the road network between this point and the home location, we called the Google Maps distance matrix API using the R package 'gmapsdistance' (v3.4; Melo et al., 2018). Third, to assign drive cost to a return trip, we used a rate of 0.14 €/km based on an estimated car occupancy rate of two, following analysis done by Capriolo et al (2020). Fourth, income was used to calculate an opportunity cost of travel at a fractional rate of one third of an individual's income which is common in the literature (Parsons, 2003). Finally, the opportunity cost and the return drive cost were combined to complete the travel cost for a trip.

The application of an individual TCM using data from a long period of time, like in the present study, requires careful consideration. Previous research has recommended multiple one-year cross-section models for the analysis of multi-year travel cost data, in light of the potential instability of demand parameters (Hellerstein, 1993; Cooper & Loomis, 1990). In spite of their potential to control for otherwise unobservable individual-specific factors, panel estimators were found to have inherent problems in travel cost analysis, thus limiting their usefulness (Hellerstein, 1993). Both modeling strategies are unfortunately unfeasible in the present study, considering the relatively small annual sample size in several of the investigated sites. Even when tested in the site with the largest number of data points (i.e., Parco Regionale Dei Monti Lattari), the results of multiple one-year models were largely inconclusive, likely due to the small number of data points available for each year (ranging between N=8 and N=76)[9]. Of the 14 yearly models for which we could estimate a truncated Poisson model (each year between 2005 and 2018), only three returned a statistically significant coefficient for the travel cost variable, albeit in all three cases with the expected negative sign. We opted thus to

[8] As a proxy of income, annual gross domestic product per capital at a municipality level (averaged between 2005-2018) was transformed into an hourly rate based on a 38-hour working week (see Appendix S8 and S9).

[9] Although the database covers the entire period 2005-2018, the vast majority of users are only active over a fraction of it. For the Parco Regionale Dei Monti Lattari, for instance, the average time between the first and the last geotagged photo uploaded by individual visitors is 1.4 years, with 93% of the users active for only one year or less.

rely on pooled models as an alternative modeling strategy. The rationale for this is that both aforementioned studies (Hellerstein, 1993; Cooper and Loomis, 1990) observed that, although they were obviously unable to capture the dynamics of temporal variation of demand parameters, pooled models provided good estimates of the average consumer surplus (CS) over the entire considered time period, which is consistent with the objectives of the present study.

Pooled, single-site individual models were thus estimated for each of the 67 PA using various count data regression models (Hilbe, 2011, Hilbe, 2014; Cameron and Trivedi, 2013). Poisson and Negative Binomial (NB1 and NB2) regressions were considered as well as zero-truncated versions of both models, to account for the lack of zero values in the data. The Poisson model is frequently used in the literature for count data (Hellerstein and Mendelsohn, 1993) while the NB1 and NB2 help to account respectively for under and over dispersion, frequently present in travel cost data (Hellerstein and Mendelsohn, 1993; Haab and McConnell, 2002)[10]. No correction for endogenous stratification is applied (Shi and Huang, 2018) since, unlike in onsite surveys, frequent visitors are not more likely to be sampled than occasional visitors when using social media data.

For each PA, we consider visitors who made a return drive of 240 km or less as a conservative cut off for a daytrip (we also test the sensitivity of halving this value to 120 km). The number of PUD observed for each visitor was entered as the dependent variable to each regression model. It was regressed against a visitor's income and their travel cost inclusive of opportunity cost of time[11]. The model for each PA was as follows:

$$logY_i = \beta_0 + \beta_{TC}\, TC_i + \beta_I\, I_i + \alpha + \varepsilon \qquad (2)$$

where Y_i is the natural logarithm of the PUD for the i[th] visitor; β_0 is the intercept of the model; TC_i is the travel cost incurred per trip (in €) for the i[th] visitor; β_{TC} is the coefficient of TC_i; I_i is the individual income (€/hour) for the i[th] visitor; β_I is the regression coefficient of I_i; α is the dispersion parameter which is returned for the NB1 and NB2 regressions, and ε is the error term. The most appropriate model is selected for further analysis based on appropriate statistical criteria - such as the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC) - and Z statistical test on dispersion parameter α (Cameron and Trivedi, 1990; Hilbe, 2011; Hilbe, 2014).

---

[10] Poisson regression model assumes that the variance of the dependent variable is equal to its mean (equidispersion hypothesis). Negative binomial models relax this assumption by including the estimation of a dispersion parameter α, which is set to zero in Poisson models. In the NB1 model the variance is a multiple of the mean. In the NB2 model the variance is quadratic in the mean (Cameron and Trivedi, 2013).

[11] A third variable, return travel cost to the nearest substitute site, was considered but removed because the 67 PA included in the analysis represent a selection of the potential substitute sites for visitors but not all of them. Defining what could be considered a potential substitute was not feasible given the scale of the analysis.

For PA which return a negative and statistically significant TC coefficient (p-value <0.05), we use the selected model to calculate the compensating Marshallian Consumer Surplus[12] per visit (CS_{trip}) following Creel and Loomis (1990):

$$CS_{trip} = -\frac{1}{\beta_{TC}} \qquad (3)$$

where β_{TC} is the regression coefficient of the travel cost variable in the model.

For the selection of PA where a CS value is reported, we also estimate annual CS values using the predicted annual visits returned from the calibration model presented in section 2.1.

3. Results

3.1 Summary of spatial big data in Italian protected area

Extracting geotagged Flickr photographs within the PA boundaries resulted in 231,655 photographs. These photographs represent 45,078 PUD captured by 17,105 unique visitors. Figure 2 shows the number of PUD for each PA and the inferred home locations of Italian visitors. The home city was available for 6240 of the 17,105 visitors, 3184 (51%) of which were Italian. Of the remaining 10,865 visitors, the photograph libraries were available for 10,667. It was possible to estimate a home country for 10,620 of those users based on the metadata of their public Flickr photographic library. After assessing home location of the available sample (16,841), 9862 were Italian (~59%), 4299 were from other European countries (~26%) and 2680 were international (~16%). Figure 3 shows the magnitude of visits to PA, aggregated regionally (see Figure 1 for regional boundaries), based on the home designation of the visitor (see appendix S2 for more details). As expected, Italian visitors make up the majority of visits to PA across all regions, followed by other European and international visitors, respectively. Of the 9862 Italian visitors, a home region could be determined for 9552[13] (97%), the geographic coverage of which is spatially represented in Figure 2.

PUD counts were found to be significantly correlated (Spearman's rho=0.610, p-value<0.01) with visitor days in 2016 based on the calibration model for 22 sites (see section 2.1). This is in the range found by previous studies (see Barros, Gutiérrez and García-Palomares, 2021). The results of the ordinary least squares regression show that PUD counts and area are able to explain most of the variation in the dependent variable ($R^2 = 0.74$; Adj. $R^2 = 0.71$; n = 22). Predicted visits from the model

[12] For PA which return statistically significant coefficient values (p-value <0.05) with the expected sign for both travel cost and income variables (negative and positive respectively), we also calculated compensating variation and equivalent variation as a better approximation of economic benefits (Bockstael et al., 1984; Anderson, 2010). Results showed that income effects were minimal and therefore the Marshallian measure of consumer surplus can serve as a good approximation of the economic value of a recreational trip (Anderson, 2010). To make comparison between values possible in the paper, we only report Marshallian measure of consumer surplus in the results.

[13] Those who could not be determined was because they returned more than one possible home region.

(after retransformation from the log-scale[14]) are significantly correlated with observed visits for the 22 sites (Spearman's rho = 0.802, p-value <0.001). Area and PUD are not highly nor significantly correlated (rho = 0.333, p-value = 0.131) and both variables are significant in the regression (p-value <0.01). Assumptions regarding normality of residuals and homoskedasticity are respected.

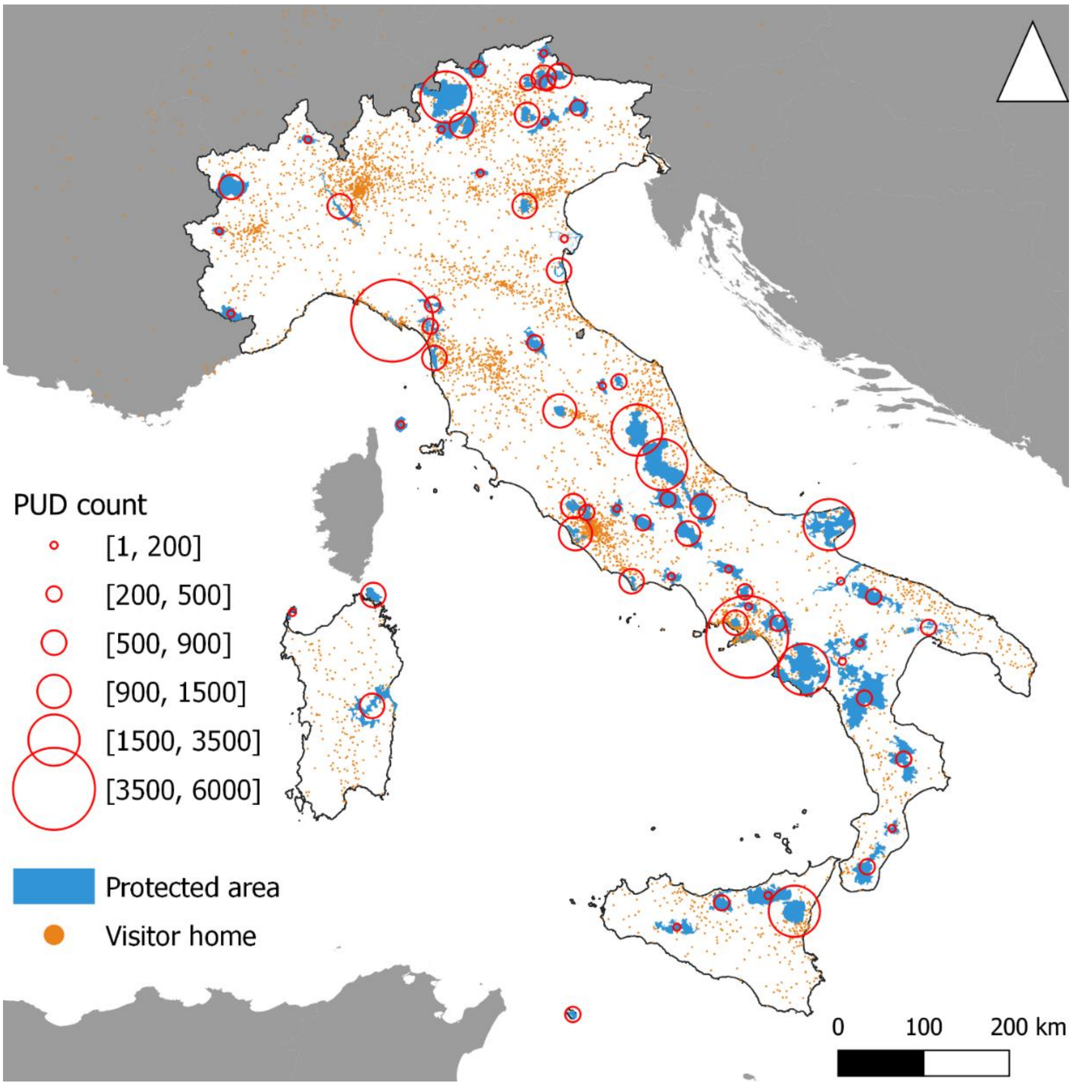

Figure 2: Photo-user-days (PUD) to protected areas (2005-2018) and distribution of estimated home locations for Italian visitors

[14] We also tested Duan's smeaning technique (Sinclair et al., 2019) but this did not improve the fit compared to the naive estimate.

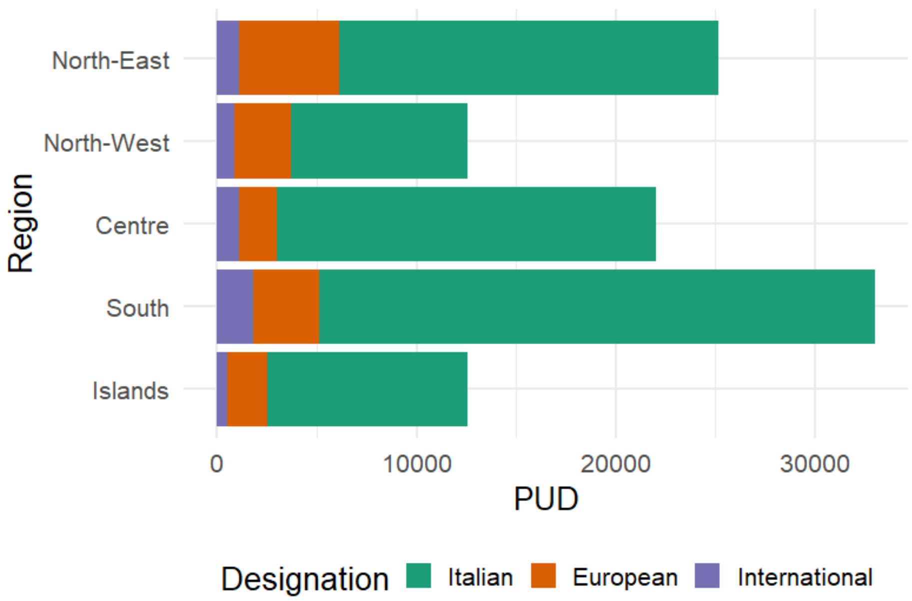

Figure 3: Aggregate photo-user-days to protected areas by region and visitor designation

Note: European visits exclude Italian visits

3.2 Spatial big data and economic estimates

The zero truncated Poisson model was selected as the most appropriate model of the four tested (the results of all models are available in appendix S5). The results of this model are summarized visually in Figure 4 (see Appendix S3 for tabular results). Only PA with a significant travel cost coefficient are presented (p-value <0.05). Of the 67 PA, 55 returned a negative travel cost coefficient as expected from economic theory (representing a down sloping demand curve), of which 33 were significant at the 95% level. The income coefficient was positive and significant for 22 of these PA (as is expected from economic theory) with 13 sites significant at the 95% level. The number of visitors included in the models presented in Figure 4 range between 17 and 488, with the Central regions returning consistently higher samples than the other regions (range: 124 to 318). For 28 of the presented sites (85%) we had a sample size of more than 40 visitors, while 20 sites (61%) had a sample size of over 100 visitors and 8 (24%) had a sample size of over 200 visitors.

Figure 4: results from the travel cost model using spatial big data

Note: results are based on the zero-truncated Poisson model with a 240 km return drive cut off (tabular results are presented in Appendix S3); only PA with a significant travel cost coefficient are included (p-value <0.05); a red bar between coefficients indicates a non-significant income coefficient (p-value >0.05) while a black bar indicated a significant income coefficient (p-value <0.05); the number in brackets is the sample size included in the model.

Figure 5 compares the travel cost model using a sensitivity of drive buffers based on 120 km and 240 km return trips. In comparison to the 240 km cut off, the zero truncated Poisson model based on a 120 km return drive results in less PA with the expected negative and significant travel cost coefficient (28 PA). Combining negative and significant results from both drive cut-off levels would result in 41 PA (~61% of total). Twenty PA are significant at both levels while 21 are significant at one level but not the other (120 km = 8 PA, 240 km = 13 PA). The South region, which has the most PA in the study area, returns the highest number of PA with the expected outputs while also returning the highest number of PA with results consistent across the two drive buffers. In the North-East and North-West, more PA return significant results at the 120 km buffer.

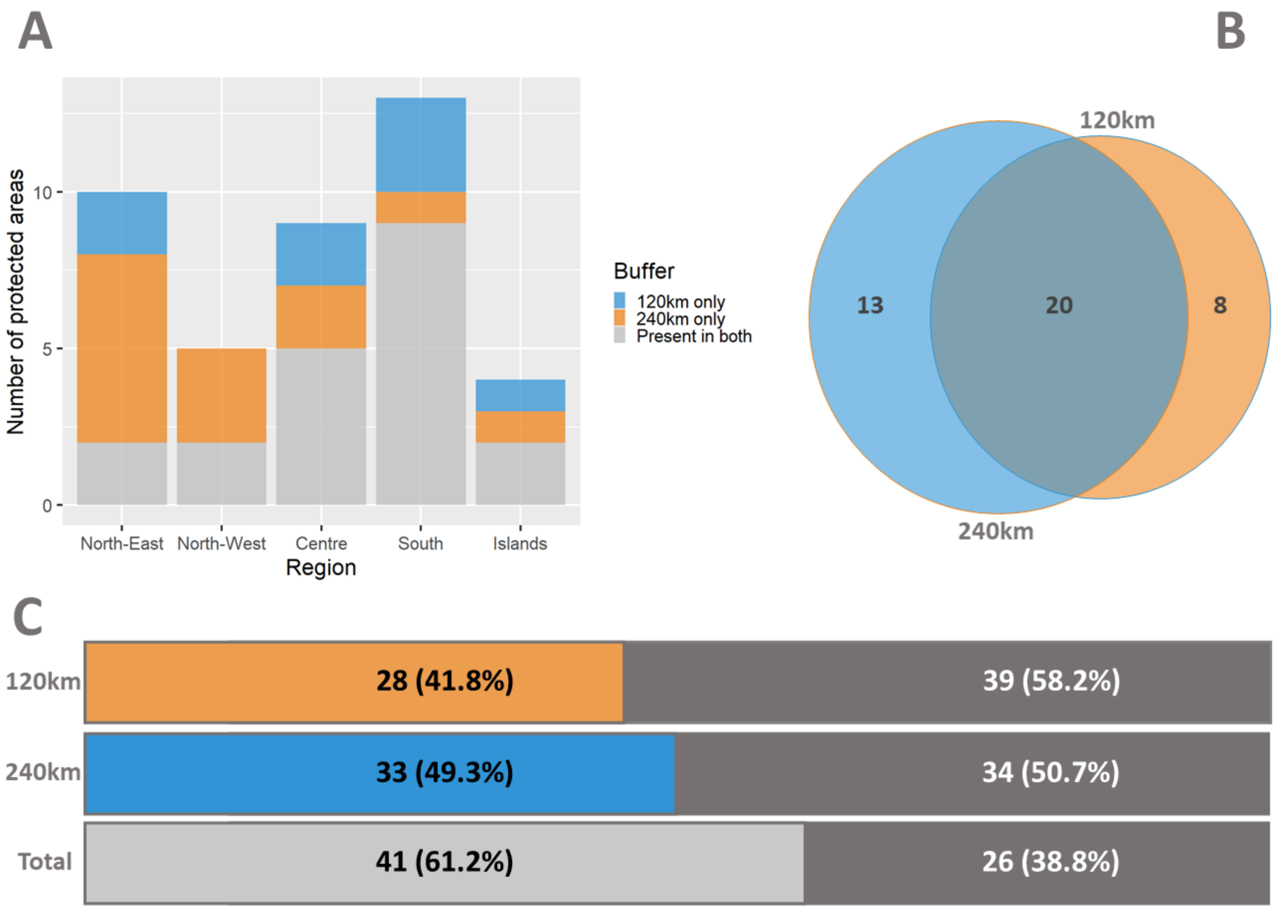

Figure 5: number of protected areas with significant and negative travel cost coefficients using a sensitivity of return drive distances

Note: results are based on the zero truncated Poisson model presented in Figure 4; only parks with return a negative and significant trave cost coefficient are presented (p-value <0.05).

The CS values per trip are presented in Figure 6 for 33 PA based on the results of the zero truncated Poisson model (Figure 4). CS values estimates range between €6.33 and €87.16, with a mean value per trip of €32.82. Parco Naturale Provinciale dell'Adamello Brenta returns the highest value while Parco dei Sicani returns the lowest value. Results should be considered in the context of the sample of visitors included in the model, where Parco dei Sicani has the lowest sample size of the PA (visitors in the model = 17). While there is no clear regional pattern in the results of CS values, the South and North-East regions have 5 out of the top 6 CS values. Using the results of the calibration model presented at the end of section 3.1 we estimate annual visitor days and aggregate consumer surplus values for these 33 PA (Table 1).

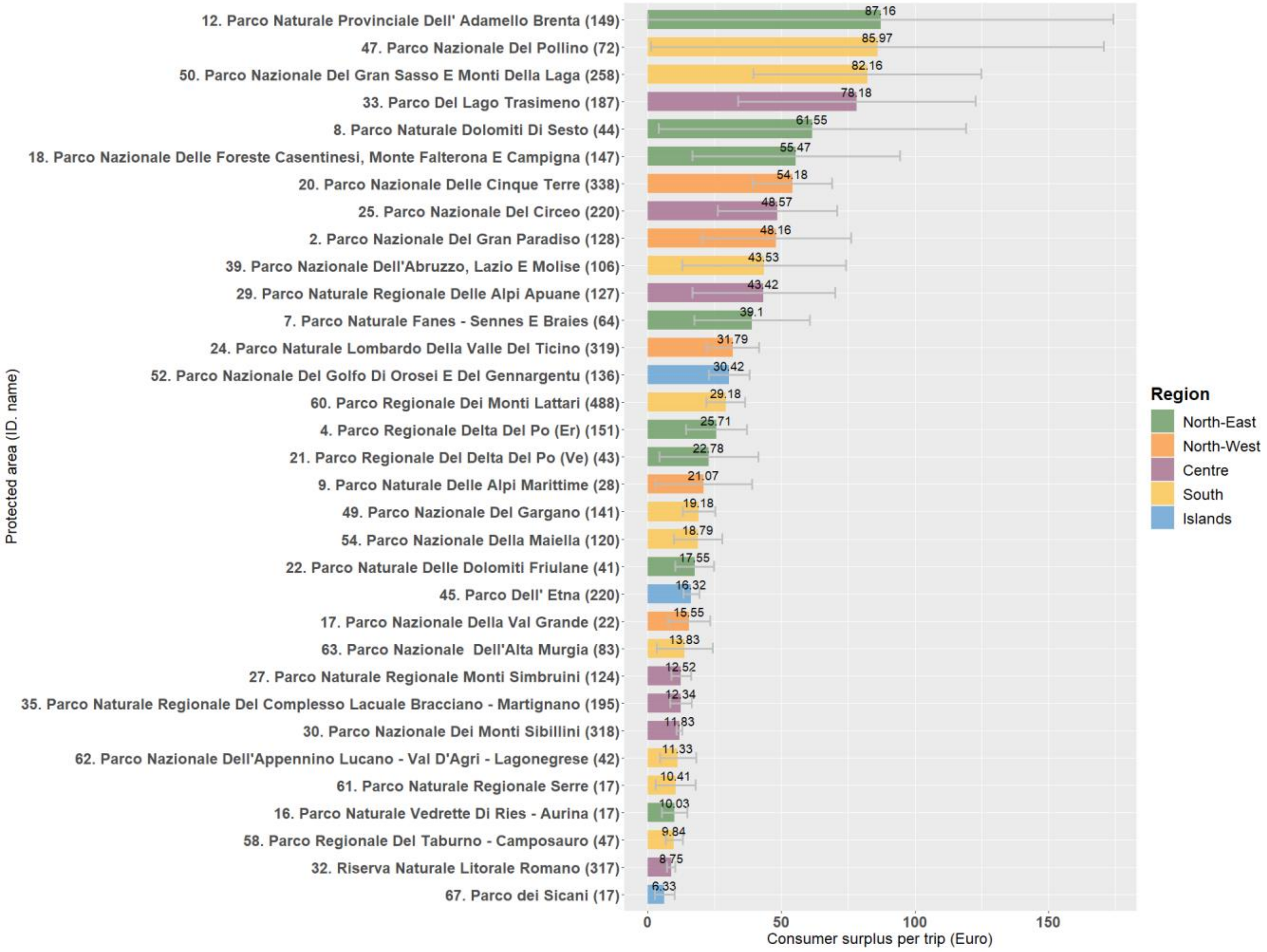

Figure 6: consumer surplus value per trip for Italian protected areas

Note: results are based on the zero truncated Poisson model presented in Figure 3; values are in 2018 €/trip based on a rate of 0.14 €/km.; the numbers in parenthesis reflect the number of visitors in the sample for that PA; grey bars represent the boundaries of the 95% confidence intervals; tabular results are presented in Appendix S6.

Table 1: Estimated annual visitor days and consumer surplus values for select protected areas

ID	Name	Region	Estimated visitor-days	Estimated consumer surplus
2	Parco Nazionale Del Gran Paradiso	North-West	487,095	€ 23,458,485
4	Parco Regionale Delta Del Po (Er)	North-East	191,641	€ 4,927,087
7	Parco Naturale Fanes - Sennes E Braies	North-East	282,479	€ 11,044,933
8	Parco Naturale Dolomiti Di Sesto	North-East	204,734	€ 12,601,365
9	Parco Naturale Delle Alpi Marittime	North-West	90,975	€ 1,916,839
12	Parco Naturale Provinciale Dell' Adamello Brenta	North-East	439,920	€ 38,343,398
16	Parco Naturale Vedrette Di Ries - Aurina	North-East	92,284	€ 925,609
17	Parco Nazionale Della Val Grande	North-West	65,903	€ 1,024,784
18	Parco Nazionale Delle Foreste Casentinesi, Monte Falterona E Campigna	North-East	211,175	€ 11,713,875
20	Parco Nazionale Delle Cinque Terre	North-West	770,095	€ 41,723,725
21	Parco Regionale Del Delta Del Po (Ve)	North-East	60,441	€ 1,376,856
22	Parco Naturale Delle Dolomiti Friulane	North-East	135,759	€ 2,382,570

24	Parco Naturale Lombardo Della Valle Del Ticino	North-West	266,374	€ 8,468,044
25	Parco Nazionale Del Circeo	Centre	227,934	€ 11,070,748
27	Parco Naturale Regionale Monti Simbruini	Centre	160,528	€ 2,009,810
29	Parco Naturale Regionale Delle Alpi Apuane	Centre	164,933	€ 7,161,371
30	Parco Nazionale Dei Monti Sibillini	Centre	952,141	€ 11,263,824
32	Riserva Naturale Litorale Romano	Centre	323,577	€ 2,831,298
33	Parco Del Lago Trasimeno	Centre	291,900	€ 22,820,772
35	Parco Naturale Regionale Del Complesso Lacuale Bracciano - Martignano	Centre	201,999	€ 2,492,670
39	Parco Nazionale Dell'Abruzzo, Lazio E Molise	South	289,274	€ 12,592,092
45	Parco Dell' Etna	Islands	733,807	€ 11,975,722
47	Parco Nazionale Del Pollino	South	1,500,899	€ 129,032,296
49	Parco Nazionale Del Gargano	South	1,642,032	€ 31,494,182
50	Parco Nazionale Del Gran Sasso E Monti Della Laga	South	2,118,573	€ 174,061,994
52	Parco Nazionale Del Golfo Di Orosei E Del Gennargentu	Islands	538,890	€ 16,393,043
54	Parco Nazionale Della Maiella	South	368,523	€ 6,924,547
58	Parco Regionale Del Taburno - Camposauro	South	90,974	€ 895,187
60	Parco Regionale Dei Monti Lattari	South	778,758	€ 22,724,156
61	Parco Naturale Regionale Serre	South	36,202	€ 376,858
62	Parco Nazionale Dell'Appennino Lucano - Val D'Agri - Lagonegrese	South	162,360	€ 1,839,541
63	Parco Nazionale Dell'Alta Murgia	South	271,966	€ 3,761,292
67	Parco dei Sicani	Islands	63,974	€ 404,953

Note: CS values are from figure 6 and only PA with significant TCM results from figure 4 are included. Values are in 2018 € based on a rate of 0.14 €/km in the TCM.

4. Discussion

4.1 Spatial big data for revealed preferences research

This paper presents a novel and low cost approach to estimate the value of nature-based recreation in large and remote natural locations using social media data as an alternative to traditional primary survey-based techniques. Through the case study of 67 Italian PA, we apply the individual TCM technique using the metadata of free and publicly available geotagged photographs as an alternative to visitor surveys and extend the state of the art by testing various model specifications and a sensitivity of drive distances on the TCM results while also extending the size and volume of natural sites included in the analysis for a region where limited data exists.

Previous studies have found spatial big data to be a reliable data source for TCM analysis, returning model results which are consistent with economic theory (Ghermandi, 2018; Sinclair, Ghermandi and Sheela, 2018; Jaung and Carrasco, 2020; Kubo *et al*., 2020) and comparable to results generated based on representative primary surveys (Sinclair et al., 2020b). Here we tested, for the first time at national scale, the potential of using these data in single-site individual TCM, using four different model specifications and across a sample of 67 PA. We find results consistent with economic theory with a down-sloping demand curve for most sites across all models. The zero-truncated Poisson model was selected as the most suitable model from which to perform further analysis, outperforming the other models based on the econometric results (see appendix S5). Using this model with a sample of visitors who made a return drive of 240 km or less, 50 PA (75%) returned a negative travel cost coefficient with 33 PA significant at the 95% level. Testing the sensitivity of different drive distances showed that the larger 240 km buffer returned more sites with the expected findings. This was particularly true for the Northern regions. In the Centre and South, we find that many PA return results which are consistent using both drive buffers with only a few PA in each region returning the expected results at one or the other distance.

Regionally, the North-West which has the smallest sample of PA in the study (8 sites) returned ~63% of sites with a significant TC coefficient result (p-value <0.05). This is somewhat higher than the other regions which have a range between ~38% and 50%. In terms of the sample of visitors included in the models, the Central region has a consistently higher number than other regions which has been shown to generate more reliable results based on previous findings (Sinclair et al., 2020b). A lack of data for some sites meant that we could not generate TCM results. This lack of data could be owing to the fact that social media visitors are less likely to visit certain sites or that they represent less iconic PA and therefore people are less likely to share experiences in these sites, leading to their under-representation in the data. This is something which is generally noted as a limitation in the literature (Ghermandi and Sinclair, 2019).

4.2 Recreation valuation estimates for Italian protected areas

Value estimates based on TCM using geotagged photographs from Flickr have been shown to return value estimates for nature-based recreation which are similar to those obtained using representative surveys (Sinclair *et al*., 2020b). Although we cannot directly compare the value estimates presented here to results based on primary surveys, owing to a lack systematic valuation data for the study sites, we can look to past valuations for some natural sites in Italy as a comparison (Signorello *et al*., 2009; De Salvo and Signorello, 2015). A meta-analysis including 265 estimates of daily CS based on 46 primary studies in Italy reported a mean CS per visit of €9.69 (in 2013 prices), with values in the range of €0.90–63.24, which are generally in the range of our results albeit with an average CS which is around three times lower. We under or overestimate the CS value for some parks, Etna for example, where previous results returned a consumer surplus value of €31.25 per trip which is around twice that found in our analysis (€16.32 per trip) (Signorello *et al*., 2009).

Extending the valuation results to explore aggregate annual values based on the estimated number of trips to sites is important for a wider perspective. These results have the limitation that the calibration of visits is based on all Flickr PUD while the CS values rely only on local visits (within a 240 km return drive). Our results therefore must assume that the value a tourist places on a trip is equivalent to that of a local. The results should therefore be viewed in that context.

4.3 Limitations and future directions

While the research presented here offers a low cost and scalable alternative to traditional primary data collection techniques to perform TCM, some limitations should be noted. A non-trivial complication with the TCM more generally, is how to account for multiday and multipurpose trips. In the application of social media data to the TCM, this complication is amplified. In this study, we have assumed that all trips are daytrips and we have not accounted for potential multipurpose trips, owing to the nature of the data it is difficult to determine these factors accurately at the scale of analysis undertaken here. To compensate, we only include trips which can reasonably be considered as day trips by segmenting trips at a 240 km return drive distance (120 km was also tested). Better understanding the issue of multiday and multipurpose trips when using social media data for TCM metrics investigation but was outside of the scope of this research given the number of large sites included here.

Selection bias is an unavoidable limitation of using social media data in research and the geographic and socio-demographic representation of social media users is an area of ongoing research (Lenormand *et al*., 2018; Ghermandi and Sinclair, 2019). Certain biases are expected to be introduced to the analysis in so far as mobile phone and social media users generally tend to represent a younger and wealthier demographic of visitor (Hausmann *et al*., 2018). Despite this, previous studies into human-nature interactions have shown Flickr data to be geographically comparable to survey data in terms of visitor provenance (Sinclair et al., 2020a) and regional visitation levels (Sinclair *et al*., 2020b). In terms of preferences for nature-based recreation, results revealed by the analysis of Flickr photographic content are not found to be significantly different from preferences stated in primary surveys (Hausmann *et al*., 2018) although it is not clear if this finding is consistent with different natural areas across different countries.

The relatively small annual sample size that was available for the sites under consideration in this study limited the analysis to the investigation of pooled travel cost models, rather than multiple one-year cross-section models. While these can provide an estimated average value of per-trip CS over the considered time period, they are unable to capture the dynamics of temporal variation of demand parameters and as such should be used carefully, if at all, in discussing temporal trends and in forecasting. We suggest that future research will explore whether such limitations can be overcome through the integrated investigation of data from multiple social media sources, which has the potential to greatly enhance the size of the data samples available for analysis. Pooling data across many platforms may also allow for activity-specific travel cost modeling for a range of activities (given the further reduction in sample size at the activity level). Relying on data from multiples sources was also found to increase the correlation between social media data counts and observed visitation in a recently published meta-analysis (Ghermandi, 2022). Currently, however, analytical techniques to control for the different ways in which social media users interact with the individual platforms as well as changes in popularity of specific sites over time are lacking.

One should also point out that potential temporal variation of demand parameters is a frequent concern in TCM applications when the results of the analysis of cross-section data (e.g., from one year or one season of monitoring) need to be extrapolated over long time periods (e.g., for cost-benefit analyses). Cooper and Loomis (1990) have shown, for instance, how reliance on one single year of data can lead to substantial errors in the estimation of future benefit streams. Some authors have argued that because of the costs involved in collecting individual-level data, zonal TCM based on data routinely collected by management agencies may become indispensable for analyses over lengthy timescales (Weber et al., 2012). This paper, albeit with the aforementioned limitations, shows that user-generated content from social media can provide a valuable source of revealed preference data that can be tapped into for individual-level analyses, including at broad temporal and spatial scales. However, despite the great potential of new sources of data, in certain cases the sample size remains a challenge at the site level even for them.

We believe that the results presented here and the expanding literature in this area should garner support for this novel data as a source of revealed preferences for use in environmental economics. Future research can be further extended and analysis developed to include emerging forms of spatial big data generated by the use of mobile phones applications, a data source which is gaining traction for human mobility analysis at high resolutions and wide spatial scales (Khataee *et al*., 2021; Sinclair *et al*., 2021). Combining automated photograph content analysis (Ghermandi *et al*., 2020; Runge *et al*., 2020) with valuation estimates may allow to develop TCM techniques based on specific types of recreation (e.g., cycling or boating) and stratify value estimates accordingly. The nature of these novel data sources also lends well to more complex implementations of TCM including multi-site RUM and K-T models which can help account for some of the previously mentioned limitations of the individual single site TCM.

4.4 Conclusion

Using a large sample of Italy's protected areas and national parks, this research sheds light on the use of spatial big data as a source of revealed preferences which can facilitate non-market valuation at wide spatial scales with limited resources as an alternative to traditional methods of data collection. Testing four TCM models across the wide sample of sites returned consistent results across most PA. The zero-truncated Poisson model was found to perform best with value estimates generated in the range of previous primary valuation studies in Italy, despite some site specific under or ever estimation. Although the limitations inherent with spatial big data merit further investigation, this novel and emerging data source offers a unique potential for environmental economists and other environmental researchers as a low-cost alternative to traditional survey data for revealed preferences analysis. This growing resource of behavioural data extends to wide spatial scales, the potential of which has only begun to be tapped by environmental economists.

Acknowledgements

We would like to thank the three anonymous reviewers and handling editor for the detailed and constructive comments and feedback which helped improve this work into its current form.

References

Anderson, D. M. (2010) 'Estimating the economic value of ice climbing in Hyalite Canyon: An application of travel cost count data models that account for excess zeros', *Journal of environmental management*, 91(4), 1012-1020.

Balmford, A., Bruner, A., Cooper, P., Costanza, R., Farber, S., Green, R.E., Jenkins, M., Jefferiss, P., Jessamy, V., Madden, J., Munro, K.. (2002) 'Economic reasons for conserving wild nature', *Science.* 297(5583):950-3. Doi: 10.1126/science.1073947

Balmford, A. Green, J., Anderson, M., Beresford, J., Huang, C., Naidoo, R., Walpole, M., Manica, A. (2015) 'Walk on the Wild Side: Estimating the Global Magnitude of Visits to Protected Areas', *PLOS Biology*, 13(2), p. e1002074. doi: 10.1371/journal.pbio.1002074.

Barros, C., Gutiérrez, J. and García-Palomares, J. (2021) 'Geotagged data from social media in visitor monitoring of protected areas; a scoping review', *Current Issues in Tourism*, pp. 1–17. doi: 10.1080/13683500.2021.1931053.

Bockstael, N., Hanemann, W.M., Strand, I. (Eds.), (1984) 'Benefit Analysis Using Indirect or Imputed Market Methods', vol. II. Department of Agicultural and Resource Economics, University of Maryland, College Park, MD.

Cameron, A. and Trivedi, P.K. (1990) 'Regression-based tests for overdispersion in the Poisson model', *Journal of Econometrics*, 46, pp. 347-364. doi:10.1016/0304-4076(90)90014-K.

Cameron, A. and Trivedi, P.K. (2013) *'Regression Analysis of Count Data'*, 2rd edition, Cambridge University Press. doi:10.1017/CBO9781139013567

Capriolo, A., Boschetto, R.G., Mascolo, R.A., Balbi, S., Villa, F. (2020) ‘Biophysical and economic assessment of four ecosystem services for natural capital accounting in Italy’, *Ecosystem Services*, 46, p. 101207. doi: 10.1016/j.ecoser.2020.101207.

Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.) (2003) ‘A Primer on Nonmarket Valuation’, Springer. doi: 10.1007/978-94-007-0826-6

Cooper, J., Loomis, J. (1990) ‘Pooled time-series cross-section travel cost models: testing whether recreation behavior is stable over time’. *Leisure Sciences*, 12, pp. 161-171.

Cui, N., Malleson, N., Houlden, V., Comber, A. (2021) ‘Using VGI and Social Media Data to Understand Urban Green Space: A Narrative Literature Review’, *ISPRS International Journal of Geo-Information*, 10(7), p. 425. doi: 10.3390/ijgi10070425.

De Salvo, M. and Signorello, G. (2015) ‘Non-market valuation of recreational services in Italy: A meta-analysis’, *Ecosystem Services*, 16, pp. 47–62. doi: 10.1016/j.ecoser.2015.10.002.

Eurostat (2021). ec.europa.eu/eurostat. Accessed 4 Oct. 2021.

Fezzi, C., Bateman, I. J. and Ferrini, S. (2014) ‘Using revealed preferences to estimate the Value of Travel Time to recreation sites’, *Journal of Environmental Economics and Management*, 67(1), pp. 58–70. doi: 10.1016/j.jeem.2013.10.003.

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*, pp. 465-474, doi 10.1016/j.jenvman.2018.05.045.

Ghermandi, A. (2018) ‘Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands’, *Ecosystem Services*, 31, pp. 351–357. doi: 10.1016/j.ecoser.2017.12.012

Ghermandi, A. (2022) ‘Geolocated social media data counts as a proxy for recreational visits in natural areas: A meta-analysis’, *Journal of Environmental Management*, 317, p. 115325. doi:10.1016/j.jenvman.2022.115325.

Ghermandi, A., Sinclair, M., Fichtman, E., Moshe, G. (2020) ‘Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing’, *Global Environmental Change*, 65, p. 102189. doi: 10.1016/j.gloenvcha.2020.102189.

Ghermandi, A. and Sinclair, M. (2019) ‘Passive crowdsourcing of social media in environmental research: A systematic map’, *Global Environmental Change*, 55, pp. 36–47. doi: 10.1016/j.gloenvcha.2019.02.003.

de Groot, R., Brander, L., van der Ploeg, S., Costanza, R., Bernard, F., Braat, L., Christie, M., Crossman, N., Ghermandi, A., Hein, L., Hussain, S., Kumar, P., McVittie, A., Portela, R., Rodriguez, L.C., ten Brink, P., van Beukering, P. (2012) ‘Global estimates of the value of ecosystems and their services in monetary units’, *Ecosystem Services*, 1(1), pp. 50–61. doi: 10.1016/j.ecoser.2012.07.005.

Haab, T. and McConnell, K. (2002) *Valuing Environmental and Natural Resources*. Edward Elgar Publishing. doi: 10.4337/9781843765431.

Haines-Young, R. and Potschin-Young, M. (2018) 'Revision of the Common International Classification for Ecosystem Services (CICES V5.1): A Policy Brief', *One Ecosystem*, 3, p. e27108. doi: 10.3897/oneeco.3.e27108.

Harmon, D. and Putney, A. (2003) '*The full value of parks : from economics to the intangible'*Rowman & Littlefield Publishers.

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: Social media data in protected areas', *Conservation Letters*, 11(1), p. e12343. doi: 10.1111/conl.12343.

Heagney, E.C., Rose, J.M., Ardeshiri, A., Kovac (2019) 'The economic value of tourism and recreation across a large protected area network', *Land Use Policy*, 88,104084. doi: 10.1016/j.landusepol.2019.104084.

Hellerstein, D. (1993) 'Intertemporal data and travel cost analysis', *Environmental and Resource Economics*, 3(2), 193-207.

Hellerstein, D. and Mendelsohn, R. (1993) 'A Theoretical Foundation for Count Data Models', *American Journal of Agricultural Economics*, 75: 604-611. https://doi.org/10.2307/1243567.

Hijmans, R.J., 2019. geosphere: Spherical Trigonometry. R package version 1.5-10. < https://CRAN.R-project.org/package=geosphere >.

Hilbe, J. (2011) '*Negative Binomial Regression'*, Cambridge University Press. doi:10.1017/CBO9780511973420.

Hilbe, J. (2014) '*Modeling Count Data*', Cambridge University Press. doi:10.1017/CBO9781139236065.

Hindsley, P., Landry, C. E., Schnier, K., Whitehead, J. C., Zarei, M. (2022) Joint estimation of angler revealed preference site selection and stated preference choice experiment recreation data considering attribute non-attendance. *Journal of Environmental Economics and Policy*, 0, 1-19. Doi: 10.1080/21606544.2022.2077840.

Ilieva, R. T. and McPhearson, T. (2018) 'Social-media data for urban sustainability', *Nature Sustainability*, 1(10), pp. 553–565. doi: 10.1038/s41893-018-0153-6.

Iezzi, D. F., & Zarelli, F. (2015) 'What tourists say about the Italian national parks: a web mining analysis', *Rivista Italiana di Economia Demografia e Statistica*, *69*(3), 73-82.

Jaung, W. and Carrasco, L. R. (2020) 'Travel cost analysis of an urban protected area and parks in Singapore: a mobile phone data application', *Journal of Environmental Management*, 261, p. 110238. doi: 10.1016/j.jenvman.2020.110238.

Keeler, B. L., Wood, S.A., Polasky, S., Kling, C.L., Filstrup, C.T. (2015) 'Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes', *Frontiers in Ecology and the Environment*, 13(2), pp. 76–81. doi: 10.1890/140124.

Kettunen, M. and ten Brink (eds) (2013) '*Social and Economic Benefits of Protected Areas An Assessment Guide*', Routledge.

Khataee, H., Sheuring, I., Czirok, A., Neufeld, Z. (2021) 'Effects of social distancing on the spreading of COVID-19 inferred from mobile phone data', *Scientific Reports*, 11(1), p. 1661. doi: 10.1038/s41598-021-81308-2.

Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., Shirayama, Y. (2020) 'Mobile phone network data reveal nationwide economic value of coastal tourism under climate change', *Tourism Management*, 77, p. 104010. doi: 10.1016/j.tourman.2019.104010.

Leggett, C., Horsch, E., Smith, C., Unsworth, R. (2017) '*Estimating recreational visitation to federally-managed lands*'. Cambridge, MA

Lenormand, M., Luque, S., Langemeyer, J., Tenerelli, P., Zulian, G., Aalders, I., Chivulescu, S., Clemente, P., Dick, J., van Dijk, J., van Eupen, M., Giuca, R.C., Kopperoinen, L., Lellei-Kovács, E., Leone, M., Lieskovský, J., Schirpke, U., Smith, A.C., Tappeiner, U., Woods, H. (2018) 'Multiscale socio-ecological networks in the age of information', *PLOS ONE*. Edited by O. Gruebner, 13(11), p. e0206672. doi: 10.1371/journal.pone.0206672.

Loomis, J.B., Walsh, R.G. (1997) Recreation economic decisions: comparing benefits and costs. Venture: State College, PA.

Melo, R.A., Rodriguez, D.T., Zarruk, D., (2018) 'gmapsdistance: Distance and Travel Time Between Two Points from Google Maps'. R package version 3.4. < https://CRAN.Rproject. org/package=gmapsdistance > .

Nicita, L., Signorello, G., De Salvo, M. (2016) 'Applying the Kuhn–Tucker model to estimate the value of recreational ecosystem services in Sicily', *Journal of Environmental Planning and Management*, 59:7, 1225-1237, doi: 10.1080/09640568.2015.1064365

Parsons, G. R. (2013) 'Travel cost methods' In: Shogren, J.F., (ed.) *Encyclopedia of Energy, Natural Resource, and Environmental Economics*, volume 3, pp. 349-358 Amsterdam: Elsevier.

R Core Team (2020) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Runge, C. A., Hausner, V. H., Daigle, R. M., Monz, C. A. (2020) Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis. *Environmental Research Communications*, 2, 075001. Doi:10.1088/2515-7620/ab9c33.

Schägner, J. P., Maes, J., Brander, L., Paracchini, M., Hartje, V., Dubois, G. (2017) 'Monitoring recreation across European nature areas: A geo-database of visitor counts, a review of literature and a call for a visitor counting reporting standard', *Journal of Outdoor Recreation and Tourism*, 18, pp. 44–55. doi: 10.1016/j.jort.2017.02.004.

Schirpke, U., Scolozzi, R., Da Re, R., Masiero, M., Pellegrino, D., Marino, D. (2018) 'Recreational ecosystem services in protected areas: A survey of visitors to Natura 2000 sites in Italy', *Journal of Outdoor Recreation and Tourism*, 21, pp. 39–50. doi: 10.1016/j.jort.2018.01.003.

Shi, W., and Huang, J. C. (2018) ‘Correcting on-site sampling bias: a new method with application to recreation demand analysis’, *Land Economics*, 94(3), 459-474. doi: 10.3368/le.94.3.459

Signorello, G., Englin, J., Longhorn, A., De Salvo, M. (2009) ‘Modeling the Demand for Sicilian Regional Parks: A Compound Poisson Approach’, *Environmental and Resource Economics*, 44(3), pp. 327–335. doi: 10.1007/s10640-009-9288-2.

Signorello, G., Farinella, G.M., Di Silvestro, L., Torrisi, A., Gallo, G. (2018) ‘Exploring geo-tagged photos to assess spatial patterns of visitors in protected areas: the case of park of Etna (Italy)‘, In BURGHARDT, Dirk, Siming CHEN, Gennady ANDRIENKO, Natalia ANDRIENKO, Ross PURVES, Alexandra DIEHL, eds., VGI Geovisual Analytics Workshop, Konstanz, Germany.

Sinclair, M., Ghermandi, A. and Sheela, A. M. (2018) ‘A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India’, *Science of The Total Environment*, 642, pp. 356–365. doi: 10.1016/j.scitotenv.2018.06.056.

Sinclair, M., Ghermandi, A., Sheela, A. M., and Joseph, S. (2019) ‘Recreation and environmental quality of tropical wetlands: a social media based spatial analysis’. *Tourism Management* 71: 179-186.

Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A. (2020a) ‘Using social media to estimate visitor provenance and patterns of recreation in Germany’s national parks’, *Journal of Environmental Management*, 263, p. 110418. doi: 10.1016/j.jenvman.2020.110418.

Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A. (2020b) ‘Valuing nature-based recreation using a crowdsourced travel cost method: A comparison to onsite survey data and value transfer’, *Ecosystem Services*, 45, p. 101165. doi: 10.1016/j.ecoser.2020.101165.

Sinclair, M., Zhao, Q., Bailey, N., Maadi, S., & Hong, J. (2021). ‘Understanding the use of greenspace before and during the COVID-19 pandemic by using mobile phone app data’. UC Santa Barbara: Center for Spatial Studies. http://dx.doi.org/10.25436/E2D59P.

Spenceley, A., Schägner, J. P., Engels, B., Cullinane Thomas, C., Engelbauer, M., Erkkonen, J., Job, H., Kajala, L., Majewski, L., Mayer, M., Metzler, D., Rylance, A., Scheder, N., Smith-Christensen, C., Souza, T. B. and Woltering, M. (2021) ‘*Visitors count! Guidance for protected areas on the economic analysis of visitation*’, United Nations Educational, Scientific and Cultural Organization, Paris, France and German Federal Agency for Nature Conservation, Bonn, Germany.

UNEP-WCMC & IUCN (2020) The World Database on Protected Areas (WDPA). UNEP-WCMC, Cambridge, UK. Available at http://www.protectedplanet.net

Weber, M.A., Mozumder, P., and Berrens, R.P. (2012) ‘Accounting for unobserved time-varying quality in recreation demand: An application to a Sonoran Desert wilderness’. *Water Resources Research* 48, W05515.

Wilkins, E.J., Wood, S.A., and Smith, J.W. (2021) ‘Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review’, *Environmental Management* 67, pp. 120–132 doi 10.1007/s00267-020-01373-7

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(1), p. 2976. doi: 10.1038/srep02976.