Crowson, M., N. J. B. Isaac, A. J. Wade, K. Norris, R. Freeman, and N. Pettorelli. 2023. Using geotagged crowdsourced data to assess the diverse socio-cultural values of conservation areas: England as a case study. Ecology and Society 28(4):28. https://doi.org/10.5751/ES-14330-280428

Research

Using geotagged crowdsourced data to assess the diverse socio-cultural values of conservation areas: England as a case study

Merry Crowson [1,2], *Nick J. B. Isaac* [3], *Andrew J. Wade* [1], *Ken Norris* [4], *Robin Freeman* [2] and *Nathalie Pettorelli* [2]

ABSTRACT. Humanity benefits immensely from nature, including through cultural ecosystem services. Geotagged crowdsourced data provide an opportunity to characterize these services at large scales. Flickr data, for example, have been widely used as an indicator of recreational value, while Wikipedia data are increasingly being used as a measure of public interest, potentially capturing often overlooked and less-tangible aspects of socio-cultural values (such as educational, inspirational, and spiritual values). So far, few studies have explored how various geotagged crowdsourced data complement each other, or how correlated these may be, particularly at national scales. To address this knowledge gap, we compare Flickr and Wikipedia datasets in their ability to help characterize the sociocultural value of designated areas in England and assess how this value relates to species richness. Our results show that there was at least one Flickr photo in 35% of all designated areas in England, and at least one Wikipedia page in 60% of them. The Wikipedia and Flickr data were shown not to be independent of each other and were significantly correlated. Species richness was positively and significantly associated with the presence of at least one geotagged Wikipedia page; more biodiverse designated areas, however, were not any more likely to have at least one Flickr photo within them. Our results highlight the potential for new, emerging datasets to capture and communicate the socio-cultural value of nature, building on the strengths of more established crowdsourced data.

Key Words: *big data; biodiversity; cultural ecosystem services; geotagged data; natural capital; value*

INTRODUCTION

Historically, arguments for conservation have promoted intrinsic values within a “nature for itself” framing (Mace 2014), but contemporary debate emphasizes the specific, quantifiable benefits society receives from nature (Hungate and Cardinale 2017, Pan and Vira 2019). This shift in focus is illustrated by the growth of natural capital accounting both internationally (SEEA 2021) and nationally (Natural Capital Committee 2019, The White House 2022). Natural capital is another term for the stock of renewable and non-renewable natural resources (e.g., plants, animals, air, minerals, freshwaters) that combine to yield a flow of benefits to people (Mace et al. 2015). Natural capital accounting is an umbrella term covering efforts to use an accounting framework to measure and report on natural capital and the flow of benefits we receive in a systematic way (SEEA 2021). As an approach, it emphasizes the process of valuation, namely estimating the relative importance, worth, or usefulness of natural capital to society (Natural Capital Coalition 2016). This usually involves some form of quantification, if not monetization, to be used within decision making and planning. The goal of conducting nature valuations is to determine in which ways nature is valuable and for whom, typically to enable better governance (TEEB 2010, Balvanera et al. 2022). Critics of natural capital accounting highlight that the value of nature can be considered infinite and boiling this down to a series of benefits means essentially “selling out” on nature (McCauley 2006, Schröter et al. 2014). Proponents of the natural capital approach argue that if the benefits provided by nature are not assigned a value they will, by default, be assigned a value of zero, as so often happens within interactions between society and life-supporting ecosystems (Mace 2019).

A full valuation of our natural environment is challenging, however, because it underpins every aspect of human well-being, and different values emerge from different world views (Balvanera et al. 2022). A range of different metrics are needed to reflect the diverse values of nature (Harrison et al. 2017, Balvanera et al. 2022), and many of these metrics need to be developed to operationalize the approach. Cultural ecosystem services, which include non-material benefits such as spiritual enrichment, cognitive development, recreation, and aesthetic experiences (MEA 2005), are usually hard to quantify and are often omitted from the valuation process despite being an important aspect of social-ecological systems. A range of methods have been developed to study cultural ecosystem services and the values associated with them, including contingent valuation (willingness to pay), choice experiments (Cheng et al. 2019), questionnaires (Schirpke et al. 2022), deliberative methods (Allen et al. 2021), interviews, photo elicitation of values (Graves et al. 2017), and participatory mapping (Jaligot et al. 2019, Muñoz et al. 2020). These methods shed important light on people’s relationship with nature, but as “stated preference” methods, they rely on capturing an accurate account of people’s preferences and values, which is not always straightforward (e.g., Nassauer 1983, Häfner et al. 2018). In addition, although these methods have important strengths at a local scale, they are very difficult to apply nationally, because of practical considerations around recruiting enough participants.

User-generated digital data have the potential to characterize diverse socio-cultural values at large scales. Various studies have shown the potential of geotagged, crowdsourced data from social media sites such as Flickr, as an indicator of nature-based recreation at a national and regional scale (Wood et al. 2013, Graham and Eigenbrod 2019, Muñoz et al. 2020). This has led to a body of work on valuing cultural ecosystem services focusing on visitation rates and on the spatial and temporal variation in human engagement with the natural environment (see e.g., van Zanten et al. 2016, Mancini et al. 2018, Calcagni et al. 2019). New opportunities are emerging to identify digital data that have the

[1]Department of Geography and Environmental Science, University of Reading, [2]Institute of Zoology, Zoological Society of London, [3]UK Centre for Ecology & Hydrology, [4]Natural History Museum, London

potential to characterize human perceptions of nature at large scales (Ladle et al. 2016, 2019, Schuetz and Johnston 2021) and capture public interest in species and ecosystems. A range of studies have looked at how people's interest in particular species has varied over time, using data on Wikipedia page views (Millard et al. 2021) or Google Trends (Schuetz and Johnston 2021). Wikipedia, the collaborative encyclopaedia, offers a powerful data source to map public interest at large spatial scales, making use of the activity of a huge community of existing users. Many Wikipedia pages are geotagged and these can be mapped to see what areas or landmarks are of interest to the public. Recent work, for example, modelled public interest in protected areas in Brazil using Wikipedia page views (Guedes-Santos et al. 2021).

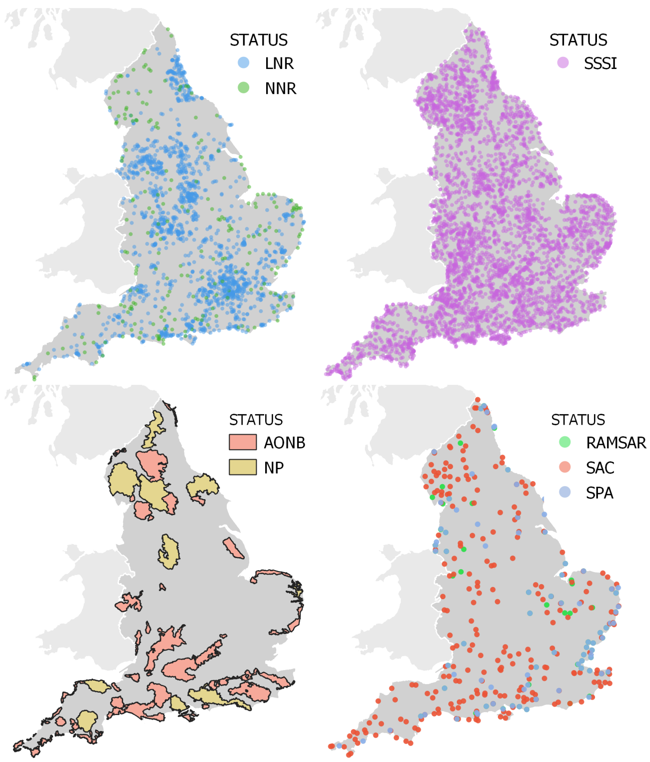
However, it is not currently clear how the public interest in Wikipedia pages relates to different socio-cultural values. There is some evidence that Wikipedia pages relating to attractions and events can predict visitation (Khadivi and Ramakrishnan 2016), but some users may decide to read and write Wikipedia pages because of an interest in a place or topic that is not directly related to visitation. This means that Wikipedia data could potentially capture a range of other non-tangible socio-cultural values, such as educational, inspirational, aesthetic and spiritual value, sense of place, and cultural heritage (Hernández-Morcillo et al. 2013) that are often systematically overlooked due to being intangible or having messy benefits (Milcu et al. 2013, Chan and Satterfield 2020).

There are currently no large-scale studies comparing the informational signature of Wikipedia data with the information contained in other geo-tagged datasets known to directly correlate with visitation rates, such as Flickr. However, doing so would help identify what aspects of the socio-cultural value of nature Wikipedia data are able to capture. To address this gap, we compare Flickr and Wikipedia data in their ability to characterize the socio-cultural value of designated areas in England. We also assess how they each relate to species richness. We chose England because there are good data on the natural environment readily available and widespread use of both Flickr and Wikipedia.

STUDY AREA

The scope of our study is limited to terrestrial ecological systems and includes designated areas on mainland England. The designation types considered are national parks, areas of outstanding natural beauty (AONBs), Ramsar sites, special areas of conservation (SACs), special protection areas (SPAs), local nature reserves (LNRs), national nature reserves (NNRs), and sites of special scientific interest (SSSIs; n = 6349; Fig. 1). These types of designations were chosen as the most relevant for nature conservation in England following Lawton et al. (2010). National parks and AONBs are designated for their cultural, landscape, and (in the case of national parks) recreational value, but also have nature conservation as part of their primary statutory purpose (Lawton et al. 2010). Ramsar sites, SACs, SPAs, LNRs, NNRs, and SSSIs have nature conservation as their primary designation and have a high level of protection. Three of these types of statutory sites are a result of international treaties and obligations (Ramsar sites, SACs, and SPAs). Sites of special scientific interest and NNRs include some of the highest quality wildlife areas, whereas LNRs are designated by local authorities. There are spatial overlaps between some of the designations, for example 24% of the total area of national parks and 12% of the area of AONBs are also designated SSSIs (Lawton et al. 2010). Designated areas with the exact same extent and with more than one designation were only included once. Designated areas with offshore areas were clipped to the coastline (at low tide).

Fig. 1. Map of designated areas considered in this study, broken down by designation type. Designation types considered are local nature reserves (LNR), national nature reserves (NNR), sites of special scientific interest (SSSI), areas of outstanding natural beauty (AONB), national parks (NP), Ramsar sites (RAMSAR), special areas of conservation (SAC), and special protection areas (SPA). Data on designated areas from Natural England (2017, 2019) and the Joint Nature Conservation Committee (2022).



MATERIALS AND METHODS

Data acquisition

To create a metric from the Flickr data, we assessed whether a given designated area had at least one Flickr photo associated with it. To create a metric from the Wikipedia data, we assessed whether a given designated area had at least one geotagged Wikipedia page within it. To assess if each designated area in England had at least one Flickr photo associated with it, the Flickr application programming interface (API) was queried to create a dataset of photos taken within designated areas between 2016 and 2019, which coincided with the period for which reliable Wikipedia data were available. The packages RCurl (Temple Lang 2020a), XML (Temple Lang 2020b), and httr (Wickham 2020) were used in R (R Core Team 2020) to request and download the data. The user and photograph ID, the date when the photo was taken, and the geographic coordinates of where it was taken were downloaded. These data are anonymous, with the user ID not personally identifying the user in any way, in line with data protection regulation and data privacy concerns (Di Minin et al. 2021). Designated areas were categorized into two groups: those with no Flickr photos and those with one or more Flickr photos.

To assess if each designated area in England had at least one geotagged Wikipedia page within it, a dataset was created containing all Wikipedia pages (written in the English language) with geotags within designated areas on the 1st of May 2019. The number of Wikipedia pages has increased from year to year since at least 2004 (Wikimedia Statistics 2021), so the pages that exist on a particular day broadly represent those created in the period leading up to that date. The most recent list of unique pages and their associated geotags is available for the English-language edition of Wikipedia as a dump (https://dumps.wikimedia.org/enwiki/). A spatial filter was applied in QGIS to select the pages within designated areas. As with the Flickr photos, the areas were categorized into two groups, those with no Wikipedia pages and those with at least one Wikipedia page.

For the biodiversity indicator, we used a freely available dataset on the estimated species richness of birds, butterflies, and vascular plants at a 100 km² scale that was compiled for the period 2000 to 2013 (Dyer and Oliver 2016), derived from species occurrence data from the Biological Records Centre. To create a single species richness indicator from the three taxonomic groups considered (birds, butterflies and vascular plants), we scaled each group, selected the maximum value for species richness found within each designated area for each taxonomic group, and added the three values together.

In addition to the variables described, we included variables that have been shown to influence cultural ecosystem service value in previous studies, namely distance to major towns and cities, population density, coastal location, public transport connectivity, elevation, presence of rivers, and presence of waterbodies (Graham and Eigenbrod 2019, Mancini et al. 2019, Muñoz et al. 2020). To estimate the distance from designated areas to urban centers, we obtained the vector boundaries for major towns and cities in England for 2015 (Office for National Statistics 2015) and calculated the minimum distance between each designated area and the closest major town or city. To quantify the connectivity of designated areas to urban centers, we used data from OpenStreetMap (2020) to obtain the number of bus stops and train stations within the "pedestrian shed" of designated areas (a 500 m buffer). A radius of 500 meters is considered in literature as a convenient pedestrian shed to capture proximity dynamics (Carpio-Pinedo 2014). We extracted the maximum population density for each designated area from the SEDAC Gridded Population of the World dataset for 2015, with a resolution of 30 arc-seconds (~1 km; SEDAC 2018).

We used the Office for National Statistics shapefile of the extent of realm (coastline) to identify coastal designated areas (Office for National Statistics 2019). We extracted the maximum height within each designated area from the Shuttle Radar Topography Mission 90 m digital elevation model data (Jarvis et al. 2008). We used data from OpenStreetMap (2020) to create a factor identifying whether each designated area had at least one river within it and at least one waterbody within it; waterbodies considered included lakes, reservoirs, and wetlands.

Analysis

Pearson's chi-squared test was used to test for independence between the Flickr data and the Wikipedia data, and Spearman's Rank correlation was used to assess the direction and strength of correlation between these metrics.

We used binomial generalized linear models with a logit link function to model the probability of obtaining a Flickr picture or a Wikipedia page within a given designated area, implemented using the function "glm" from the core R "stats" package (R Core Team 2020). We chose glm as a method because interpretability of the model is a priority, and the coefficients are a robust way to gain insight into the relationships between the independent variables and the Flickr and Wikipedia data. Fixed covariates considered for both models were the log of the area of the designated area; species richness of birds, butterflies, and vascular plants; the log of the distance to the closest major town or city; the log of the number of public transport links; population density; coastal location (categorical with two levels); maximum height; presence of at least one river (categorical with two levels); and presence of at least one waterbody (categorical with two levels). Akaike's information criterion (AIC) was used as the selection criteria for covariates to be included in our final best models. We used a stepwise approach, starting with a maximal model including all the fixed covariates and conducting backward model selection (Zuur et al. 2009) using the function "step" in the stats package (R Core Team 2020).

All the covariates were standardized so that the coefficients were comparable. Model assumptions were verified by plotting residuals versus fitted values and against each covariate. We assessed the residuals for spatial autocorrelation by calculating Moran's I, creating a map of the residuals for visual inspection and by plotting a distance-based semivariogram. The percentage of variance explained by our best models was calculated using the "rsq" package (Zhang 2020). To evaluate the accuracy of our predictive model, we used the "pROC" package (Robin et al. 2011) to create a receiver-operating characteristic (ROC) curve and calculate the area under the curve (AUC; Zou et al. 2007).

RESULTS

With regard to the distribution of Flickr photos and geotagged Wikipedia pages, in total, 2194 areas had at least 1 Flickr photo (34.6% of all designated areas) and 3829 areas had at least 1 Wikipedia page associated with them (60.3% of all designated areas).

In relation to comparing the Wikipedia and Flickr data to assess what sociocultural values Wikipedia data have the potential to capture, the Pearson's chi-squared test for independence between the Flickr and the Wikipedia data gave a test statistic of χ^2 = 522.64 (df = 1, $p < 0.001$), meaning that the two datasets were not independent of each other. The Spearman correlation coefficient for the relationship between the Flickr and Wikipedia data was $r_s = 0.29$, with the two datasets shown to be significantly and positively correlated ($p < 0.001$). Designated areas with both

Table 1. Contingency table showing the relationship between the binary variables from the Flickr and Wikipedia data for all designated areas (n = 6349). The Flickr data are used to categorize designated areas into those with at least 1 Flickr photo (Y) and those with no photos (N). Likewise, Wikipedia data are used to split designated areas into those with at least 1 geotagged Wikipedia page (Y) and those with none (N).

		Flickr data		
Wikipedia data		Flickr: Y	Flickr: N	
	Wiki: Y	1747	2082	3829
	Wiki: N	447	2073	2520
		2194	4155	6349

a Flickr photo and a Wikipedia page made up 27.5% of the total, whereas 32.6% of all areas had neither a Flickr photo nor a Wikipedia page (Table 1).

In relation to understanding the relationship between the value of designated areas and species richness, our best models for the Flickr and Wikipedia data explained 38% and 27% of the variability in the probability of a given designated area having a Flickr picture or a Wikipedia page associated with it, respectively. The AUC score for the best model of the Flickr data was 0.84 and for the best model of the Wikipedia data it was 0.80 (see Appendix 1 for the corresponding ROC curves). These best models (Table 2 and Table 3) showed that designated areas with high values for species richness were significantly more likely to have at least one geotagged Wikipedia page associated with them (p-value < 0.001), but this was not the case for geotagged Flickr photos (p-value = 0.08; see Appendix 2 for the prediction plots for species richness).

Moran's I analyses suggested that spatial autocorrelation in the residuals of both best models remained significant but was very small in the case of both the Flickr data (observed = 0.03, expected = -0.0002, p-value < 0.001) and the Wikipedia data (observed = 0.04, expected = -0.0002, p-value < 0.001). Analysis of the spatial autocorrelation using semi-variograms and subsampling our data showed that this amount of spatial autocorrelation did not affect our conclusions (see Appendix 3).

DISCUSSION

Although Flickr and Wikipedia data have been used separately to study the relationship between humans and the natural world, we have shown for the first time that these two data sources are not independent of each other. Our results provide evidence that Wikipedia data capture patterns of visitation to designated areas to some extent, which makes sense given that people are likely to be interested in places that they plan to visit or have visited. However, the correlation between the Wikipedia and Flickr datasets was found to be relatively small, so it is likely that some of the signal from the Wikipedia data captures less tangible aspects of the value of nature, such as educational, inspirational, and spiritual values. The results also highlight that the diverse socio-cultural values of nature are closely intertwined and hard to separate into neat categories, with visitation closely linked to public interest as captured by digital, user-generated data. This is relevant to the debate about whether designated areas should be "set aside" for nature or managed as shared spaces between people and biodiversity (Adams et al. 2014) because a lack of access may lead to a fall in public interest in a site. Our work also shows that Wikipedia and Flickr data have different relationships with species richness, providing further support for the idea that the two datasets contain different signals, and highlighting that species richness has a significant positive effect on public interest in designated areas in England.

Table 2. Estimated regression parameters, standard errors, z-values and p-values for the binomial GLM of the Flickr data. Model R2 is 0.38. The independent variables are the log of the geographical extent of the designated area, species richness, coastal location, the presence of a water body, population density in the designated area and the log of the number of public transport links.

	Estimate	Std. error	*z*-value	*p*-value
Intercept	-1.18	0.05	-25.44	< 0.001
LogArea	1.53	0.06	27.48	< 0.001
SpeciesRichness	-0.06	0.03	-1.72	0.08
Coastal : Yes	1.83	0.13	13.70	< 0.001
Waterbody : Yes	0.45	0.07	6.39	< 0.001
PopulationDensity	0.19	0.04	4.13	< 0.001
LogPublicTransportLinks	0.25	0.05	5.34	< 0.001

Table 3. Estimated regression parameters, standard errors, z-values and p-values for the binomial GLM of the Wikipedia data. Model R2 is 0.27. The independent variables are the log of the geographical extent of the designated area, species richness, maximum height of the landscape in the designated area, coastal location, the presence of a river, population density in the designated area and the log of the number of public transport links.

	Estimate	Std. error	*z*-value	*p*-value
Intercept	0.80	0.05	17.17	< 0.001
LogArea	1.56	0.05	30.90	< 0.001
SpeciesRichness	0.38	0.03	11.13	< 0.001
MaxHeight	-0.25	0.04	-6.58	< 0.001
Coastal: Yes	-0.38	0.13	-3.02	0.002
River: Yes	-0.27	0.06	-4.09	< 0.001
PopulationDensity	0.15	0.04	3.51	< 0.001
LogPublicTransportLinks	-0.17	0.04	-3.81	< 0.001

Interestingly, our results suggest that species richness is not generally an important driver of visitor numbers in designated areas in England, as measured by Flickr data. Previous research is mixed in this area, with some studies finding a positive relationship between designated landscapes (protected for their high biodiversity value) and the number of Flickr photos at a regional (Gliozzo et al. 2016) and national (Graham and Eigenbrod 2019) scale, whereas other found no evidence that designation increased the number of Flickr photos (Hornigold et al. 2016, Mancini et al. 2018). In our study, geodiversity such as coastal location and the presence of a water body played an important role (Table 2). Local population density and the number of public transport links both have small positive effects on the probability of finding a Flickr photo in a designated area, whereas distance to the nearest major town or city did not have a significant effect. These findings are surprising because other studies have shown that connectivity and proximity to urban areas have a large positive influence on visitor numbers (e.g., van Zanten et al. 2016 for Europe, Mancini et al. 2019 for Scotland). However, these studies are from countries that, unlike England, are not densely populated, and where most designated areas are relatively inaccessible. Distance to towns was also not important in explaining visitor numbers in Vermont (Sonter et al. 2016), where conserved lands exist throughout the state and the maximum distance between any conserved land and a town is less than 100 km.

Using crowdsourced geotagged data to study the diverse values of high biodiversity areas is in the spirit of other recent work in "culturomics," which uses digital data to capture less tangible aspects of the relationship between humans and nature by, for example, capturing national park visitors' sentiment from social media texts (Hausmann et al. 2020), track species awareness through time using Wikipedia page views (Millard et al. 2021), and using expressions in photographs to reflect an aesthetic judgment of natural areas (Do 2019). The challenge with this approach is that it can be very difficult to validate digital data using independent, non-digital sources (Correia et al. 2021). The evidence that Flickr data are good measures of visitation is strong (Wood et al. 2013, Mancini et al. 2018), but there is less evidence for what precise socio-cultural values Wikipedia data can capture. Our study finds that local population density has a significant effect on the likelihood of finding a geotagged Wikipedia page within a designated area. People are often more interested in local entities, and part of this is likely to be due to the ease with which they can be visited. Indeed, a study of online public interest in birds, measured by Wikipedia pageviews, found that those more commonly encountered in the wild attracted more pageviews (Mittermeier et al. 2021a). Given the lack of information on the motivation of Wikipedia users (Mittermeier et al. 2021b), it is hard to know what proportion of public interest is unrelated to visitation, and indeed it may be unproductive to compartmentalize value into such neat categories.

There are various methodological considerations that need to be considered when interpreting data sourced from Wikipedia that point to other limitations for this study. First, there are various considerations around the geography of Wikipedia data that have implications for the interpretation of our results. Because Wikipedia is organized by language rather than by country, language becomes the best proxy for country (Mittermeier et al. 2021b), which is why we chose to use the Wikipedia pages in English. This leads to two limitations: there are English speakers across the entire world, although those who post georeferenced photographs on Flickr are more likely to have easy access to the area. And by choosing to consider only the English language Wikipedia pages, we overlooked the potential value from people originating from non-English speaking countries (or indeed whose favored language is not English). More generally, there is the known bias in who contributes to Wikipedia and Flickr. In the case of Wikipedia, the demographic of editors is known to be predominantly white and male (Wagner et al. 2015), while high Flickr photograph density has been shown to correlate with high densities of well-educated white people (Li et al. 2013). Indeed, any study using data from internet users excludes those who do not use the internet. Failing to include certain sectors of the population when drawing conclusions about value is a clear limitation of this kind of big data approach because the question "of value to whom?" remains an issue (Milcu et al. 2013, Ghermandi and Sinclair 2019, Wilkins et al. 2021). Although directly addressing these issues is either not technically possible or beyond the scope of this study, taking them into account is important when interpreting the results.

It is possible that the way Wikipedia data are used could affect what is being captured. Wikipedia data have already been used to create metrics and indicators in a range of ways, and our approach of using geotagged pages within designated areas is cutting edge, capturing a broad range of spatial entities within designated areas that contribute to their value, from streams to stone circles. However, future work could develop indicators that make use of additional variables associated with the online encyclopaedia, such as page size (Wong and Rosindell 2021), page name (Chua et al. 2021), page views (Nolan et al. 2022), page edits, and the distribution of languages and users (Mittermeier et al. 2021b).

A further limitation of our study is the species richness data used to model value. The spatial resolution of this data is 100 km² (Dyer and Oliver 2016), which is a coarse resolution given the small size of many designated areas in England (the median size of LNRs in this study is 0.1 km², for example). This means that variation in species richness at a scale relevant to the smaller designated areas may not be captured by the dataset, which can be problematic in cases where these designated areas are small hotspots of biodiversity. In addition, the taxonomic groups included (birds, bees, and vascular plants) exclude other groups likely to be of interest to people (for example mammals). Although the data used are currently the best available at a national scale, there is room for improvement should other data be published in the future. In addition, future work could explore to what extent particular "charismatic species" increase visitor numbers, rather than more general measures of species richness. There is, for example, evidence that people will pay more or stay longer in a protected area if they have the possibility of encountering particular wildlife species (Mustika et al. 2020). It is also possible that the abundance of species influences recreational value because visitors are more likely to be able to see species that are present in large numbers. Exploring further what aspects of biodiversity people care about and are more likely to visit is a promising direction for future work in this area.

Understanding the values of nature is a fundamental step to comprehend and manage the interlinkages between people and other-than-human nature (Díaz et al. 2015). This study, as is the case for all attempts to value nature, emerges from a particular regional context and worldview. The natural capital approach in England pushes for a national scale assessment of the value of stocks and the benefits that flow from them (Natural Capital Committee 2020) and regular reporting of these within national accounts by the government (Office for National Statistics 2022). It is within this type of valuation exercise that it is important to find datasets relevant to national policy and which can give a static snapshot of the state of the environment and the values we derive from nature. However, it is possible that any national assessment of value will remain very limited because people's experience of the natural world is primarily local. There are other ways of setting up valuation exercises that these data could also be used for, at various scales, including valuation exercises that highlight the importance of directing ecosystem service assessment and valuation exercises toward specific trade-offs and decision making (Posner et al. 2016, Chan and Satterfield 2020). In the case of characterizing the diverse values of nature to inform environmental decision making, the culturonomics approach taken within this study is probably best combined with more discursive and deliberative methodologies, both to ensure public participation and reflection (Allen et al. 2021) and to help define the cultural ecosystem services through the lens of the beneficiaries themselves (Katz-Gerro and Orenstein 2015). Finally, the fact that the Flickr and Wikipedia data used in this study have a spatial dimension means they can be used in mapping exercises, and through this, enhance the visibility and communication of cultural ecosystem services, which is another important aspect of valuation exercises (Hernández-Morcillo et al. 2013).

The intrinsic value of high biodiversity areas remains a strong moral argument for their continued conservation, however the focus within policy on including the benefits humans receive from nature into decision making in a formal way requires new approaches to capturing value, including attempts at quantification. This study has shown that there is potential for new, emerging datasets to capture and communicate the socio-cultural value of nature, building on the strengths of more established crowdsourced data. However, in addition to critically assessing new datasets as we have done in this study, it is important to recognize that no single indicator can be expected to represent the value of a landscape and can, in fact, be used against conservation efforts. For example, many of the designated areas did not have any Flickr or Wikipedia data associated with them, but this does not mean that they do not have any value to people. Going forward, it would be interesting to use a series of case studies on different designated areas, at various scales, with different characteristics and designation types, and combine "stated preference" and more deliberative methods with digital data to understand in more detail the processes at work, including people's motivations.

Acknowledgments:

This research was supported by the QMEE CDT, funded by NERC grant number NE/R012229/1. NP is funded by Research England.

Data Availability:

This study was a re-analysis of data that are publicly available. Data derived through the re-analysis undertaken in this study are available from the University of Reading Research Data Archive at https://doi.org/10.17864/1947.000516. The code that supports the findings of this study are openly available at https://github.com/merrycrowson/geotagged_crowdsourced_data.git.

LITERATURE CITED

Adams, W. M., I. D. Hodge, and L. Sandbrook. 2014. New spaces for nature: the re-territorialisation of biodiversity conservation under neoliberalism in the UK. Transactions of the Institute of British Geographers 39:574-588. https://doi.org/10.1111/tran.12050

Allen, K. E., C. Castellano, and S. Pessagno. 2021. Using dialogue to contextualize culture, ecosystem services, and cultural ecosystem services. Ecology and Society 26(2):7. https://doi.org/10.5751/ES-12187-260207

Balvanera, P., U. Pascual, C. Michael, B. Baptiste, and D. González-Jiménez. 2022. Methodological assessment report on the diverse values and valuation of nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES, Bonn, Germany. https://doi.org/10.5281/zenodo.6522522

Calcagni, F., A. Terra Amorim Maia, J. J. T. Connolly, and J. Langemeyer. 2019. Digital co-construction of relational values: understanding the role of social media for sustainability. Sustainability Science 14(5):1309-1321. https://doi.org/10.1007/s11625-019-00672-1

Carpio-Pinedo, J. 2014. Urban bus demand forecast at stop level: space syntax and other built environment factors. Evidence from Madrid. Procedia - Social and Behavioral Sciences 160:205-214. https://doi.org/10.1016/j.sbspro.2014.12.132

Chan, K. M. A., and T. Satterfield. 2020. The maturation of ecosystem services: social and policy research expands, but whither biophysically informed valuation? People and Nature 2 (4):1021-1060. https://doi.org/10.1002/pan3.10137

Cheng, X., S. Van Damme, L. Li, and P. Uyttenhove. 2019. Evaluation of cultural ecosystem services: a review of methods. Ecosystem Services 37(June 2019):100925. https://doi.org/10.1016/j.ecoser.2019.100925

Chua, M. A. H., A. Tan, and L. R. Carrasco. 2021. Species awareness days: do people care or are we preaching to the choir? Biological Conservation 255:109002. https://doi.org/10.1016/j.biocon.2021.109002

Correia, R. A., R. Ladle, I. Jarić, A. C. M. Malhado, J. C. Mittermeier, U. Roll, A. Soriano-Redondo, D. Veríssimo, C. Fink, A. Hausmann, J. Guedes-Santos, R. Vardi, and E. Di Minin. 2021. Digital data sources and methods for conservation culturomics. Conservation Biology 35(2):398-411. https://doi.org/10.1111/cobi.13706

Díaz, S., S. Demissew, J. Carabias, C. Joly, M. Lonsdale, N. Ash, A. Larigauderie, J. R. Adhikari, S. Arico, A. Báldi, A. Bartuska, I. A. Baste, A. Bilgin, E. Brondizio, K. M. A. Chan, V. E. Figueroa, A. Duraiappah, M. Fischer, R. Hill, T. Koetz, P. Leadley, P. Lyver, G. M. Mace, B. Martin-Lopez, M. Okumura, D. Pacheco, U. Pascual, E. Selvin Pérez, B. Reyers, E. Roth, O. Saito, R. J. Scholes, N. Sharma, H. Tallis, R. Thaman, R. Watson, T. Yahara, Z. A. Hamid, C. Akosim, Y. Al-Hafedh, R. Allahverdiyev, E. Amankwah, S. T. Asah, Z. Asfaw, G. Bartus, L. A. Brooks, J. Caillaux, G. Dalle, D. Darnaedi, A. Driver, G. Erpul, P. Escobar-Eyzaguirre, P. Failler, A. M. M. Fouda, B. Fu, H. Gundimeda, S. Hashimoto, F. Homer, S. Lavorel, G. Lichtenstein, W. A. Mala, W. Mandivenyi, P. Matczak, C. Mbizvo, M. Mehrdadi, J. P.

Metzger, J. B. Mikissa, H. Moller, H. A. Mooney, P. Mumby, H. Nagendra, C. Nesshover, A. A. Oteng-Yeboah, G. Pataki, M. Roué, J. Rubis, M. Schultz, P. Smith, R. Sumaila, K. Takeuchi, S. Thomas, M. Verma, Y. Yeo-Chang, and D. Zlatanova 86. 2015. The IPBES conceptual framework: connecting nature and people. Current Opinion in Environmental Sustainability 14:1-16. https://doi.org/10.1016/j.cosust.2014.11.002

Di Minin, E., C. Fink, A. Hausmann, J. Kremer, and R Kulkarni. 2021. How to address data privacy concerns when using social media data in conservation science. Conservation Biology 35 (2):437-446. https://doi.org/10.1111/cobi.13708

Do, Y. 2019. Valuating aesthetic benefits of cultural ecosystem services using conservation culturomics. Ecosystem Services 36 (April):100894. https://doi.org/10.1016/j.ecoser.2019.100894

Dyer, R., and T. Oliver. 2016. UK ecological status map version 2. Environmental Information Data Centre, Lancaster, UK. https://www.data.gov.uk/dataset/2fd5bf6f-1b7b-4f64-8d20-64a5387dcb88/uk-ecological-status-map-version-2

Ghermandi, A., and M. Sinclair. 2019. Passive crowdsourcing of social media in environmental research: a systematic map. Global Environmental Change 55:36-47. https://doi.org/10.1016/j.gloenvcha.2019.02.003

Gliozzo, G., N. Pettorelli, and M. (Muki) Haklay. 2016. Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. Ecology and Society 21(3):6. https://doi.org/10.5751/ES-08436-210306

Graham, L. J., and F. Eigenbrod. 2019. Scale dependency in drivers of outdoor recreation in England. People and Nature 1 (3):406-416. https://doi.org/10.1002/pan3.10042

Graves, R. A., S. M. Pearson, and M. G. Turner. 2017. Species richness alone does not predict cultural ecosystem service value. Proceedings of the National Academy of Sciences 114 (14):3774-3779. https://doi.org/10.1073/pnas.1701370114

Guedes-Santos, J., R. A. Correia, P. Jepson, and R. J. Ladle. 2021. Evaluating public interest in protected areas using Wikipedia page views. Journal for Nature Conservation 63:126040. https://doi.org/10.1016/j.jnc.2021.126040

Häfner, K., I. Zasada, B. T. van Zanten, F. Ungaro, M. Koetse, and A. Piorr. 2018. Assessing landscape preferences: a visual choice experiment in the agricultural region of Märkische Schweiz, Germany. Landscape Research 43(6):846-861. https://doi.org/10.1080/01426397.2017.1386289

Harrison, P., A. Sier, M. Acreman, B. Bealey, M. Fry, L. Jones, L. Maskell, L. May, L. Norton, D. Read, S. Reis, P. Trembath, T. August, N. Bachiller-Jareno, R. Beck, M. Bogdanova, M. Brown, J. Bullock, E. Blyth, E. Carnell, D. Carss, L. Carvalho, C. Carvell, J. Cosby, R. Dunford-Brown, T. Goodall, H. Gweon, J. Hall, C. Harrower, P. Henrys, N. Isaac, K. Kazlauskis, F. Kral, C. Laize, T. Marthews, D. Masante, G. Mills, D. Morton, E. Nemitz, G. Old, A. Oliver, M. O'Hare, J. Redhead, S. Rennie, G. Rhodes, E. Roberts, D. Roy, P. Scholefield, S. Smart, K. Smith, C. Stratford, P. Taylor, M. Vieno, M. Wilson, I. Winfield, H. Woods, D. Wright, and J. Watkins. 2017. Natural capital metrics. Phase 1 final report: central components. CEH Project NEC06063. UK Centre for Ecology and Hydrology, Lancaster, UK. https://www.ceh.ac.uk/sites/default/files/documents/Natural_Capital_Metrics_project_-NEC06063_Final_Report_central_components.pdf" target="_blank

Hausmann, A., T. Toivonen, C. Fink, V. Heikinheimo, R. Kulkarni, H. Tenkanen, and E. Di Minin. 2020. Understanding sentiment of national park visitors from social media data. People and Nature 2(3):750-760. https://doi.org/10.1002/pan3.10130

Hernández-Morcillo, M., T. Plieninger, and C. Bieling. 2013. An empirical review of cultural ecosystem service indicators. Ecological Indicators 29:434-444. https://doi.org/10.1016/j.ecolind.2013.01.013

Hornigold, K., I. Lake, and P. Dolman. 2016. Recreational use of the countryside: no evidence that high nature value enhances a key ecosystem service. PLoS ONE 11(11):0165043. https://doi.org/10.1371/journal.pone.0165043

Hungate, B. A., and B. J. Cardinale. 2017. Biodiversity: what value should we use? Frontiers in Ecology and the Environment 15 (6):283-283. https://doi.org/10.1002/fee.1511

Jaligot, R., S. Hasler, and J. Chenal. 2019. National assessment of cultural ecosystem services: participatory mapping in Switzerland. Ambio 48(10):1219-1233. https://doi.org/10.1007/s13280-018-1138-4

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4. Consortium for Spatial Information (CGIAR.CSI). http://srtm.csi.cgiar.org

Joint Nature Conservation Committee. 2022. Special areas of conservation (SACs), special protection areas (SPAs) and Ramsar sites in the UK. Joint Nature Conservation Committee, Peterborough, UK. https://jncc.gov.uk/our-work/uk-protected-area-datasets-for-download/

Katz-Gerro, T., and D. E. Orenstein. 2015. Environmental tastes, opinions and behaviors: social sciences in the service of cultural ecosystem service assessment. Ecology and Society 20(3):28. https://doi.org/10.5751/ES-07545-200328

Khadivi, P., and N. Ramakrishnan. 2016. Wikipedia in the tourism industry: forecasting demand and modeling usage behavior. Proceedings of the Twenty-eighth Conference on Artificial Intelligence 30(2):4016-4021. https://doi.org/10.1609/aaai.v30i2.19078

Ladle, R. J., R. A. Correia, Y. Do, G.-J. Joo, A. C. M. Malhado, R. Proulx, J.-M. Roberge, and P. Jepson. 2016. Conservation culturomics. Frontiers in Ecology and the Environment 14 (5):269-275. https://doi.org/10.1002/fee.1260

Ladle, R. J., P. Jepson, R. A. Correia, and A. C. M. Malhado. 2019. A culturomics approach to quantifying the salience of species on the global internet. People and Nature 1(4):524-532. https://doi.org/10.1002/pan3.10053

Lawton, J. H., P. N. M. Brotherton, V. K. Brown, C. Elphick, A. H. Fitter, J. Forshaw, R. W. Haddow, S. Hilborne, R. N. Leafe, G. M. Mace, M. P. Southgate, W. J. Sutherland, T. E. Tew, J. Varley, and G. R. Wynne. 2010. Making space for nature: a review of England's wildlife sites and ecological network. Report to Defra.

Defra, London, UK. https://webarchive.nationalarchives.gov.uk/20130402170324/http:/archive.defra.gov.uk/environment/biodiversity/documents/201009space-for-nature.pdf

Li, L., M. F. Goodchild, and B. Xu. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science 40(2):61-77. https://doi.org/10.1080/15230406.2013.777139

Mace, G. M. 2014. Whose conservation? Science 345 (6204):1558-1560. https://doi.org/10.1126/science.1254704

Mace, G. M. 2019. The ecology of natural capital accounting. Oxford Review of Economic Policy 35(1):54-67. https://doi.org/10.1093/oxrep/gry023

Mace, G. M., R. S. Hails, P. Cryle, J. Harlow, and S. J. Clarke. 2015. Towards a risk register for natural capital. Journal of Applied Ecology 52(3):641-653. https://doi.org/10.1111/1365-2664.12431

Mancini, F., G. M. Coghill, and D. Lusseau. 2018. Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. PLoS ONE 13(7):0200565. https://doi.org/10.1371/journal.pone.0200565

Mancini, F., G. M. Coghill, and D. Lusseau. 2019. Quantifying wildlife watchers' preferences to investigate the overlap between recreational and conservation value of natural areas. Journal of Applied Ecology 56(2):387-397. https://doi.org/10.1111/1365-2664.13274

McCauley, D. J. 2006. Selling out on nature. Nature 443 (7107):27-28. https://doi.org/10.1038/443027a

Milcu, A. I., J. Hanspach, D. Abson, and J. Fischer. 2013. Cultural ecosystem services: a literature review and prospects for future research. Ecology and Society 18(3):44. https://doi.org/10.5751/ES-05790-180344

Millard, J. W., R. D. Gregory, K. E. Jones, and R. Freeman. 2021. The species awareness index as a conservation culturomics metric for public biodiversity awareness. Conservation Biology 35 (2):472-482. https://doi.org/10.1111/cobi.13701

Millenium Ecosystem Assessment (MEA). 2005. Living beyond our means: natural assets and human well-being. Statement from the Board. https://www.millenniumassessment.org/documents/document.429.aspx.pdf

Mittermeier, J. C., R. Correia., R. Grenyer, T. Toivonen, and U. Roll. 2021b. Using Wikipedia to measure public interest in biodiversity and conservation. Conservation Biology 35 (2):412-423. https://doi.org/10.1111/cobi.13702

Mittermeier, J. C., U. Roll, T. J. Matthews, R. Correia, and R. Grenyer. 2021a. Birds that are more commonly encountered in the wild attract higher public interest online. Conservation Science and Practice 3(5):e340. https://doi.org/10.1111/csp2.340

Muñoz, L., V. H. Hausner, C. Runge, G. Brown, and R. Daigle. 2020. Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. People and Nature 2(2):437-449. https://doi.org/10.1002/pan3.10083

Mustika, P. L. K., M. Ichsan, and H. Booth. 2020. The economic value of shark and bay tourism in Indonesia and its role in delivering conservation outcomes. Frontiers in Marine Science 7 (April). https://doi.org/10.3389/fmars.2020.00261

Nassauer, J. I. 1983. Framing the landscape in photographic simulation. Journal of Environmental Management 17:1-16.

Natural Capital Coalition. 2016. The path towards the Natural Capital Protocol: a primer for business. Capitals Coalition, The Hague, The Netherlands. https://naturalcapitalcoalition.org/wp-content/uploads/2016/07/NCC_Primer_WEB_2016-07-08.pdf

Natural Capital Committee. 2019. State of Natural Capital annual report 2019: sixth report to the Economic Affairs Committee of the Cabinet. Natural Capital Committee, London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/774218/ncc-annual-report-2019.pdf

Natural Capital Committee. 2020. End of term report: to the Domestic and Economy Implementation Committee of the Cabinet. Natural Capital Committee, London, UK. https://assets.publishing.service.gov.uk/media/5f9c3befe90e070428701525/ncc-end-of-term-report.pdf

Natural England. 2017. SSSI boundaries. Natural England, York, UK. https://naturalengland-defra.opendata.arcgis.com/datasets/Defra::sites-of-special-scientific-interest-england/about

Natural England. 2019. National parks boundaries. Natural England, York, UK. https://www.data.gov.uk/dataset/334e1b27-e193-4ef5-b14e-696b58bb7e95/national-parks-england#licence-info

Nolan, G., A. Kane, and D. Fernández-Bellon. 2022. Natural history films generate more online interest in depicted species than in conservation messages. People and Nature 4(3):816-825. https://doi.org/10.1002/pan3.10319

Office for National Statistics. 2015. Major towns and cities (December 2015) boundaries, V2. Office for National Statistics, Newport, UK. https://geoportal.statistics.gov.uk/datasets/980da620a0264647bd679642f96b42c1/explore

Office for National Statistics. 2019. Countries (December 2019) boundaries UK BFE. Office for National Statistics, Newport, UK. https://geoportal.statistics.gov.uk/datasets/countries-december-2019-boundaries-uk-bfe?geometry=-31.281%2C51.101%2C26.419%2C59.782

Office for National Statistics. 2022. UK natural capital accounts: 2022. Office for National Statistics, Newport, UK. https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/uknaturalcapitalaccounts/2022

OpenStreetMap. 2020. OpenStreetMap data in layered GIS format. Geofabrik GmbH, Karlsruhe, Germany. https://www.geofabrik.de/data/geofabrik-osm-gis-standard-0.7.pdf

Pan, Y., and B. Vira. 2019. Exploring natural capital using bibliometrics and social media data. Ecology and Society 24(4):5. https://doi.org/10.5751/ES-11118-240405

Posner, S. M., E. McKenzie, and T. H. Ricketts. 2016. Policy impacts of ecosystem services knowledge. Proceedings of the National Academy of Sciences 113(7):1760-1765. https://doi.org/10.1073/pnas.1502452113

R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77. https://doi.org/10.1186/1471-2105-12-77

Schirpke, U., R. Scolozzi, and U. Tappeiner. 2022. Not too small to benefit society: insights into perceived cultural ecosystem services of mountain lakes in the European Alps. Ecology and Society 27(1):6. https://doi.org/10.5751/ES-12987-270106

Schröter, M., E. H. van der Zanden, A. P. E. van Oudenhoven, R. P. Remme, H. M. Serna-Chavez, R. S. de Groot, and P. Opdam. 2014. Ecosystem services as a contested concept: a synthesis of critique and counter-arguments. Conservation Letters 7 (6):514-523. https://doi.org/10.1111/conl.12091

Schuetz, J. G., and A. Johnston. 2021. Tracking the cultural niches of North American birds through time. People and Nature 3 (1):251-260. https://doi.org/10.1002/pan3.10173

Socioeconomic Data and Applications Center (SEDAC). 2018. Gridded population of the world (GPW), v4. SEDAC, New York, New York, USA. https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/documentation

Sonter, L. J., K. B. Watson, S. A. Wood, and T. H. Ricketts. 2016. Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. PLoS ONE 11(9):0162372. https://doi.org/10.1371/journal.pone.0162372

System of Environmental Economic Accounting (SEEA). 2021. Natural capital and ecosystem services FAQ. SEEA, Rome, Italy. https://seea.un.org/content/natural-capital-and-ecosystem-services-faq#What%20is%20natural%20capital%20accounting

Temple Lang, D. 2020a. RCurl: general network (HTTP/FTP/...) client interface for R. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/package=RCurl

Temple Lang, D. 2020b. XML: Tools for parsing and generating XML within R and S-Plus. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/package=XML

The Economics of Ecosystems and Biodiversity (TEEB). 2010. Mainstreaming the economics of nature: a synthesis of the approach, conclusions and recommendations of TEEB. https://teebweb.org/publications/teeb-for/synthesis/

The White House. 2022. National strategy to develop statistics for environmental-economics decisions: a U.S. system of natural capital accounting and associated environmental-economic statistics. https://www.whitehouse.gov/wp-content/uploads/2022/08/Natural-Capital-Accounting-Strategy.pdf

van Zanten, B. T., D. B. van Berkel, R. K. Meentemeyer, J. W. Smith, K. F. Tieskens, and P. H. Verburg. 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

Wagner, C., D. Garcia, M. Jadidi, and M. Strohmaier. 2015. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015 454-463. https://doi.org/10.1609/icwsm.v9i1.14628

Wickham, H. 2020. httr: tools for working with URLs and HTTP. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/package=httr

Wikimedia Statistics. 2021. Pages to date. Wikimedia Foundation, San Francisco, California, USA. https://stats.wikimedia.org/#/en.wikipedia.org/content/pages-to-date/normal%7Ctable%7Call%7Cpage_type~content%7Cmonthly

Wilkins, E. J., S. A. Wood, and J. W. Smith. 2021. Uses and limitations of social media to inform visitor use management in parks and protected areas: a systematic review. Environmental Management 67(1):120-132. https://doi.org/10.1007/s00267-020-01373-7

Wong, Y., and J. Rosindell. 2021. Dynamic visualisation of million-tip trees: the OneZoom project. Methods in Ecology and Evolution 13(2):303-313. https://doi.org/10.1111/2041-210X.13766

Wood, S. A., A. D. Guerry, J. M. Silver, and M. Lacayo. 2013. Using social media to quantify nature-based tourism and recreation. Scientific Reports 3:2976. https://doi.org/10.1038/srep02976

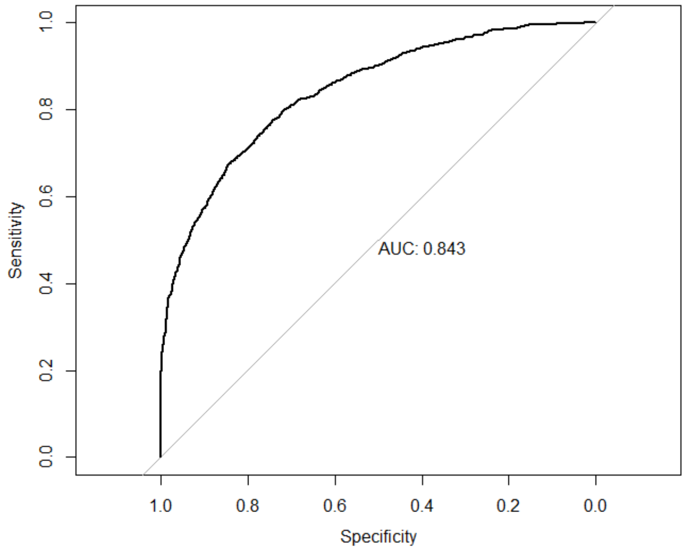
Zhang, D. 2020. rsq: R-squared and related measures. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/package=rsq

Zou, K. H., A. J. O'Malley, and L. Mauri. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation, 115(5):654-657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929

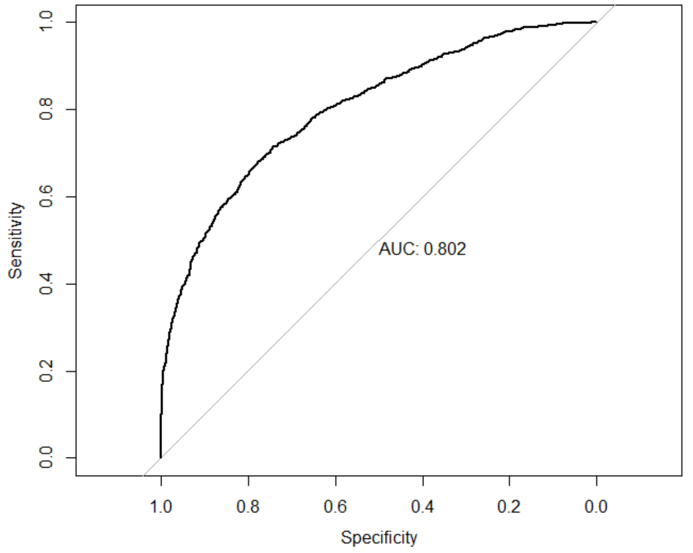
Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, New York, USA. https://doi.org/10.1007/978-0-387-87458-6

Appendix 1

(a)

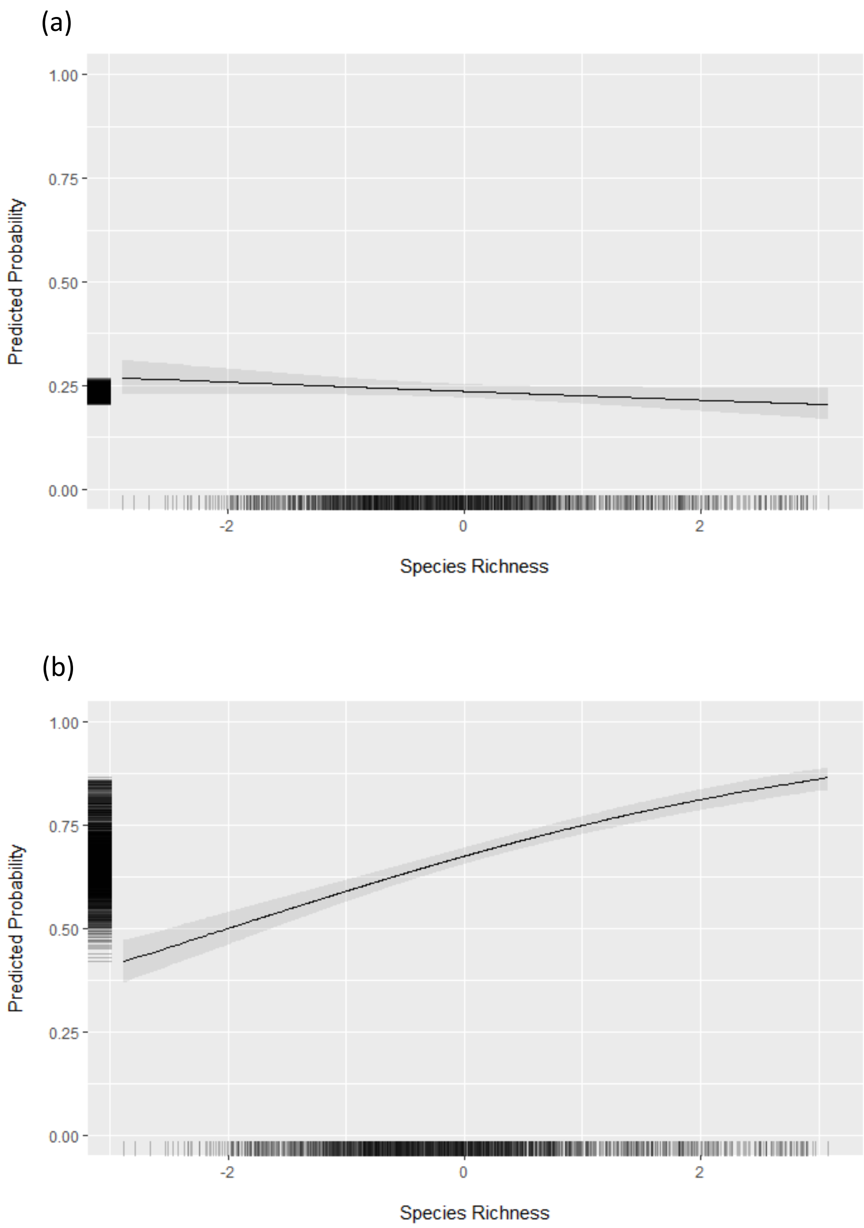


(b)



A1.1: ROC curves for (a) the best model for the Flickr data and (b) the best model for the Wikipedia data. The plots show the sensitivity and specificity of the model at different probability cutoffs. The AUC is an overall summary of diagnostic accuracy and a good model will have a high AUC. AUC equals 0.5 when the ROC curve corresponds to random chance and 1.0 for perfect accuracy (Zou et al. 2007).

Appendix 2



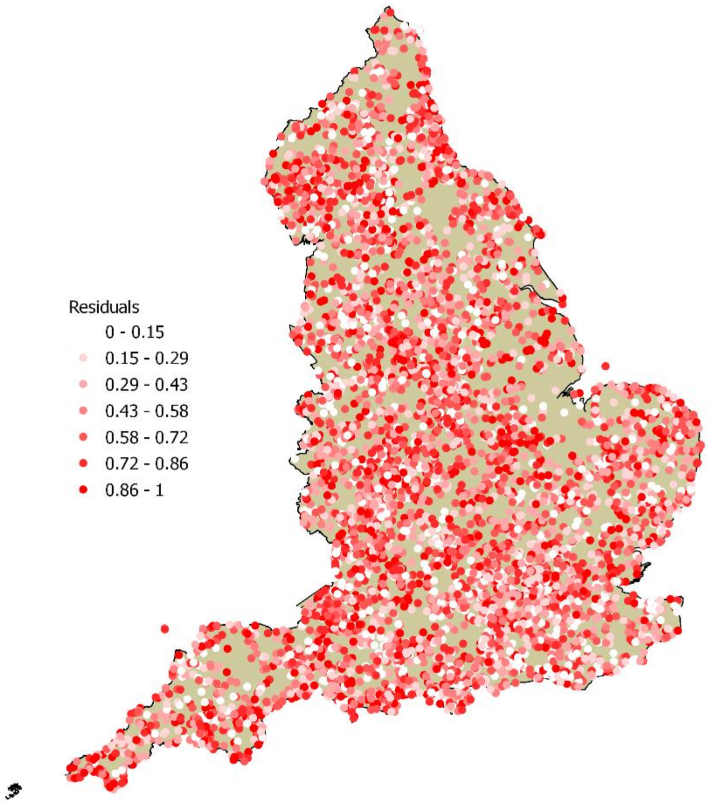
A2.1: Results of the models for (a) the Flickr data and (b) the Wikipedia data. The plots show the predicted probability of finding at least one Flickr photo (a) or Wikipedia page (b) against species richness in designated areas without a coastal location or waterbody, while other variables are kept at their mean. Species richness was centred around the mean and scaled by its SD. Ticks on the plot margins represent the data, the lines represent the predictions from the model and the lighter shaded areas are the 95% confidence intervals.

Appendix 3

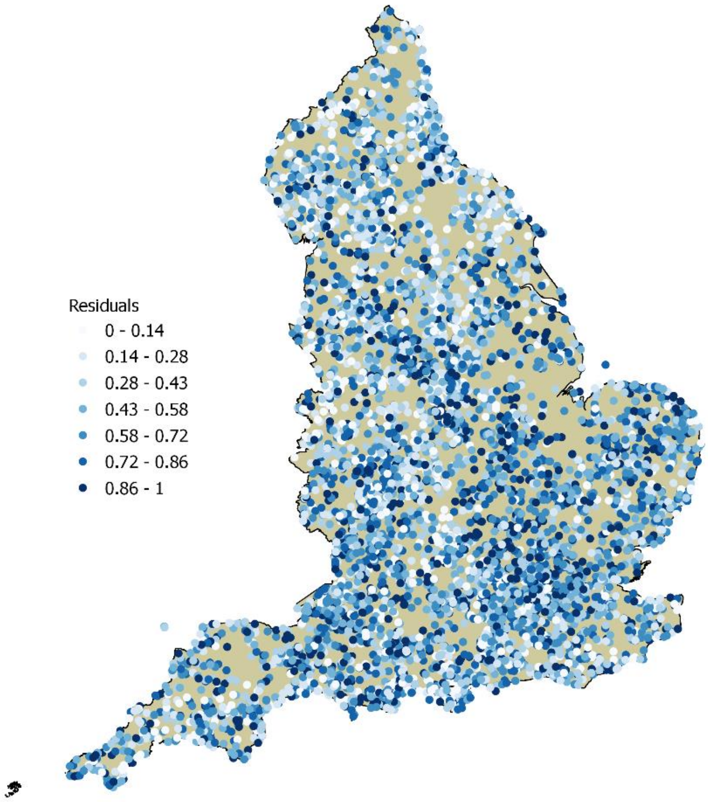
This section briefly explains the additional analysis carried out in order to ensure that the very small amount of spatial autocorrelation found in the residuals of the two binomial glm models is not influencing the conclusions drawn.

Overall, there is no clear spatial autocorrelation visible in the residuals of the two glm models for the Flickr data (A3.1) and the Wikipedia data (A3.2). However, Moran's I for the residuals shows a very small amount of spatial autocorrelation, that is nonetheless significant, for both the model of the Flickr data (observed = 0.03, expected = -0.0002, p-value < 0.001) and the model of the Wikipedia data (observed = 0.04, expected = -0.0002, p-value < 0.001). The semivariograms shows that the spatial autocorrelation is no longer an issue at 4000m, as the semivariograms levels off (A3.3).

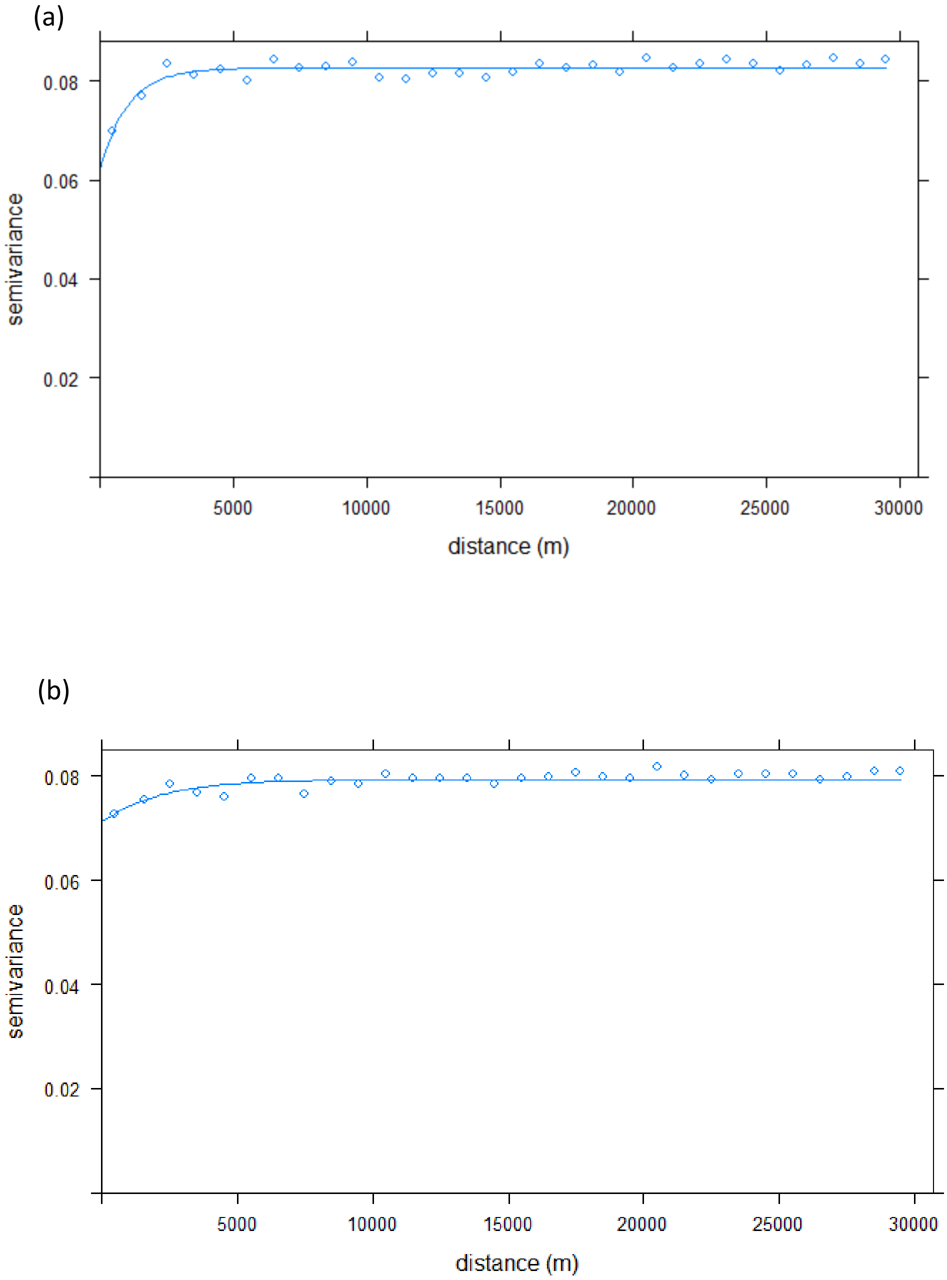
To ensure that our conclusions are not affected by this small amount of spatial autocorrelation we repeatedly took a random sub-sample of the designated areas, ensuring that those included were more than 4000m apart. Each sub-sample included about 2400 designated areas (from the total of 6349), and these were used to run binomial glms for the Flickr and Wikipedia data. The coefficients for the covariates in the models varied by less than 0.1 compared to the glm using all of the data, meaning that our conclusions remain unchanged.



A3.1: Map of the simulated (DHARMa) residuals from the model of the Flickr data. Each point represents the centroid of a designated area. The range of DHARMa residuals is 0 to 1, which makes them easier to visualise and interpret than the “raw” residuals of the model.



A3.2: Map of the simulated (DHARMa) residuals from the best model of the Wikipedia data. Each point represents the centroid of a designated area. The range of DHARMa residuals is 0 to 1, which makes them easier to visualise and interpret than the "raw" residuals of the model.



A3.3: Semivariogram for the simulated (DHARMa) residuals of the best model for the Flickr data (a) and the Wikipedia data (b).