ORIGINAL ARTICLE

Geodiversity Supports Cultural Ecosystem Services: an Assessment Using Social Media

Nathan Fox[1,2] · Laura J. Graham[3,4] · Felix Eigenbrod[1] · James M. Bullock[1,2] · Katherine E. Parks[1]

Abstract

Geodiversity is under threat from both anthropogenic activities and environmental change which therefore requires active management in the form of geoconservation to minimise future damage. As research on the role of geodiversity on ecosystem service (ES) provision has been limited, there is a need to improve our understanding of which aspects are most important to providing ES to better inform approaches to its conservation. Here, we focus on the cultural ES of hiking in Wales, UK. Harnessing big data from the social media website Flickr, we used the locations of geotagged images of hiking and a range of spatial layers representing geodiversity, biodiversity and anthropogenic predictor variables in habitat suitability models. To gain a deeper understanding of the role of geodiversity in driving the distribution of this cultural service, we estimated the strength and nature of the relationship of each geodiversity, biodiversity and anthropogenic indicator with hiking. Our models show that three geodiversity (distance from coast, range in slope and range in elevation) and two anthropogenic (distance from greenspace access point and distance from road) variables were the most important drivers of hiking. Furthermore, we assessed the content of the images to understand which features of geodiversity people interact with while hiking. We found that people generally take images of geomorphological and hydrological features, such as mountains and lakes. Through understanding the geodiversity, biodiversity and anthropogenic drivers of hiking in Wales, as well as identifying the geodiversity features people interact with while hiking, this analysis can help to inform future geoconservation methods by focusing efforts on these important features.

Keywords Gedoviersity · Cultural ecosystem services · Social media · Geoconservation

Introduction

Globally, geodiversity—the diversity of geological structures and processes, including rocks and minerals; geomorphology, including landforms and topography; sediments and soils, including formation processes; and hydrology, including marine, surface and subsurface waters—is under increasing pressures from anthropogenic activities and environmental change (Gray 2008a, 2013; Hjort et al. 2015; Fox et al. 2020a). Threats such as urbanisation, mining and land-use change can damage geodiversity both in situ, such as damage to landforms, and across wider spatial and temporal scales, such as the contamination of hydrological systems and changes to soil processes (Hjort et al. 2015). Activities such as tourism and recreation can cause damage to geodiversity features, for example, through the erosion and removal of material (Gray 2008a). Furthermore, environmental change such as anthropogenic induced sea-level rise, landslides and changes to weather patterns can impact geodiversity features and processes (Prosser et al. 2010; Brazier et al. 2012). These wide-reaching anthropogenic and environmental impacts emphasise the need for the wider adoption of geoconservation.

First introduced by Sharples (1993), the concept of geoconservation is any action intended to conserve geodiversity features, processes, sites and specimens (Gray 2018).

✉ Nathan Fox
 nf2g13@soton.ac.uk

1 School of Geography and the Environment, University of Southampton, University Road, Southampton SO17 1BJ, UK

2 Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB, UK

3 School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

4 Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

Geoconservation includes the following: the creation of protected areas such as the UNESCO Global Geoparks Networks; the promotion of education on geodiversity and its conservation; and the in situ management of sites such as the construction of physical barriers to restrict public access (Gray 2008a; Henriques et al. 2011). Though the concept of geoconservation is starting to be adopted globally, a deeper understanding of the value of the contribution of geodiversity to society would promote greater uptake of geoconservation practices (Gordon et al. 2018).

Geodiversity has value also because it plays an integral role in the delivery and maintenance of ecosystem services (ES)— the benefits we receive from nature (Gray 2011; Fox et al. 2020b). First, geodiversity underpins ES, providing the foundations for all other services to occur (Parks and Mulligan 2010). Second, geodiversity actively contributes, through interactions with biodiversity, people and society, to provisioning services (e.g. food, fibre and fuels), regulating services (e.g. dispersal and dilution of pollutants) and cultural ecosystem services (CES, e.g. aesthetic views) (Fox et al. 2020b). Third, geodiversity can provide services in the absence of any interactions with biodiversity. These geosystem services provide a range of goods and benefits for society, including provisioning services (e.g. construction materials and rare-earth metals), regulating services (e.g. the regulation of thermal flows) and cultural services (e.g. religious sites and recreational activities) (Gray 2011; Van Ree and van Beukering 2016; van Ree et al. 2017; Fox et al. 2020b).

Here, we will focus on the relationship between geodiversity and the CES subcategory of recreational activities (Millennium Ecosystem Assessment 2005; Milcu et al. 2013), which provide physical health and psychological well-being benefits obtained through interactions with the natural environment (Hermes et al. 2018). In particular, we focus on hiking, a recreational activity that is generally described as the act of walking for an extended amount of time through natural or rural areas (Mitten et al. 2018). King et al. (2017) identified six pathways from which CES benefits can arise: cognitive, creative, intuitive, retrospective, regenerative and communicative. Here, we classify hiking under the regenerative pathway, which includes opportunities for recreation, leisure and tourism that provide direct restorative benefits like reducing emotional stress (King et al. 2017). We also acknowledge that some of the benefits of recreational activities can also be linked to other CES through other pathways, including accessing aesthetic views, spiritual and religious motivations, or a sense of place (Collins-Kreiner and Kliot 2017; Wilcer et al. 2019).

Some CES that geodiversity provides, such as recreational activities, may, in turn, exacerbate anthropogenic threats to geodiversity (Figueroa-Alfaro and Tang 2017). For example, iconic and geodiverse landscapes, such as the Grand Canyon which attracts thousands of visitors annually, provide a range of CES including as a popular destination for hiking. Through the provision of these services, the landscapes are at higher risk of damage from overuse and exploitation (Gray 2008a). However, quantifying the scale of the threat of human activity on geodiversity and the services it provides is difficult. First, the relationship between geodiversity and ES, and also the service being assessed, varies over the spatial scale assessed as well as across different locations (Alahuhta et al. 2018). Second, indirect damage to geodiversity, such as impacts on downstream hydrological and geomorphological processes, may not be easily assessed (Hjort et al. 2015; Fox et al. 2020b). Therefore, we need to develop a deeper understanding of the complex relationship between geodiversity, ecosystem services and anthropogenic threats.

To develop our understanding of the relationship between geodiversity, ecosystem services and anthropogenic threats over large spatial and temporal scales, there is a need for suitable datasets. However, current data collection methods for ES assessments, including geodiversity, biodiversity and social-demographic datasets, have several limitations not only at larger spatial and temporal scales but also across smaller scales (Hjort et al. 2012). Many traditional CES assessments, such as monetary assessments or social surveys, are expensive and time-intensive to implement (Tenerelli et al. 2016; Figueroa-Alfaro and Tang 2017). Furthermore, as CES vary based on the individual experiencing them, quantifying their perceived benefits and values is more difficult than for provisioning and regulating services (Daniel et al. 2012; Lee et al. 2019; Havinga et al. 2020). Here, data from social media websites provide advantages over traditional ecological assessments and social surveys, providing large spatial and temporal datasets relatively quickly and at minimal financial cost (Barve 2014; Fox et al. 2020a). These approaches can provide a more objective approach to assessing CES compared to survey data, as they show revealed preferences as opposed to stated preferences. Social media websites have therefore started to become established as a reliable source of data for a vast array of CES studies (Ghermandi and Sinclair 2019).

Data from social media sites can be used in a range of methods. For example, using geolocated posts to assess the spatial variation in CES (Tieskens et al. 2017); using content analysis to assess human-nature interactions depicted in photographs (Richards and Tunçer 2018); and using textual analysis to better understand opinions on CES (Becken et al. 2017; Wilson et al. 2019). Previous work in the field has examined the effectiveness of a variety of different social media platforms. Flickr, an image and video hosting website, has been used to assess aesthetic ecosystem services (Van Zanten et al. 2016; Van Berkel et al. 2018) and recreational ecosystem services (Graham and Eigenbrod 2019; Mancini

et al. 2019). Twitter, a microblogging site, has been used to assess urban greenspace services (Roberts 2017; Johnson et al. 2019). Reddit, a discussion-based forum, has been used to look at recreational, aesthetic and spiritual CES (Fox et al. 2021a). The different platforms all have different strengths in assessing CES and these strengths and weaknesses should be acknowledged and accounted for in analyses. Fox et al. (2021a) demonstrated that Flickr data is more suited to assessing spatial variation and image content analysis, while Reddit is more suited to assessing textual metadata.

Geolocated posts from social media can be analysed using species distribution modelling methods to gain an understanding of the distributions and drivers of CES. Richards and Tunçer (2018) used maximum entropy modelling (MaxEnt) to plot the potential distributions of CES from Flickr photographs. Their study assessed four criteria: distance from attractions, presence of parks, forest cover and managed vegetation cover. They found that distance from attraction had the largest contribution to the distribution of photographs of plants and animals. Walden-Schreiner et al. (2018) used MaxEnt to assess visitor distributions to national parks, including both infrastructure and environmental features as factors, highlighting that infrastructure variables such as visitor centres are more important than environmental variables. Arslan and Örücü (2020) also used MaxEnt, finding that CES distribution based on geotagged Flickr images is most influenced by roads, religious places and historical and cultural areas. As there is a range of different distribution models available, all with different predictive performance, choosing one individual model may not provide robust results, and therefore, many studies have now taken an ensemble approach by combing the outputs of multiple model algorithms (Hao et al. 2019). However, the authors are not aware of any studies that use multiple modelling algorithms to assess CES from social media data.

Here, we focus on Wales, UK, which is known for its high levels of geodiversity. This geodiversity is protected through various geoconservation instruments, for example, through statutorily regulated areas such as Geological Sites of Special Scientific Interest as well as European Geoparks, such as GeoMôn on the island of Anglesey and Fforest Fawr in the Brecon Beacons (Prosser et al. 2010). The geodiversity of Wales, in particular the high geological and landscape diversity, makes it a popular tourist destination (Burek 2012). Through funding and community engagement, geotourism and education are promoted in the iconic landscapes and areas of cultural significance in Wales (Evans et al. 2018). Previous work using social media data to assess CES in Wales (Gliozzo et al. 2016) found that views of geodiversity features, such as peaks and beaches, the presence of historic human structures and formal biodiversity protection areas, are important drivers of CES in non-urban areas. Furthermore, studies assessing social media posts have found

relationships between recreational activities and geodiversity features, in particular, geomorphological and hydrological features (Van Zanten et al. 2016; Oteros-Rozas et al. 2018; Van Berkel et al. 2018; Muñoz et al. 2020). However, due to the multifaceted nature of CES, these analyses, while useful, do not enable us to identify how the use of the landscapes for CES could affect the maintenance of geodiversity.

In this paper, we aim to be the first study to apply an ensemble species distribution modelling approach to understanding what natural and human features drive the distribution of CES derived from geolocated social media data. We aim to quantify which aspects of geodiversity are most important in contributing to the distribution of hiking in Wales, UK. Furthermore, through analysis of the contents of images of hiking, we aim to understand which geodiversity features people interact with to inform focused geoconservation efforts on these potentially at-risk features.

Methods

Social Media Data

Launched in 2004, Flickr is a photograph and video hosting website, with a large and diverse user base that contributes over 25 million new uploads a day (Oteros-Rozas et al. 2018; Ding and Fan 2019). Photographs uploaded to Flickr have a range of available metadata including temporal information, in the form of the time and date the image was taken, as well as spatial information in the form of the latitude and longitude at which the image was taken (Fox et al. 2020a). The high availability of metadata from Flickr means that it has become the most widely used social media site for assessing CES (Langemeyer et al. 2018; Ghermandi and Sinclair 2019). Data from Flickr can be accessed through its Application Planning Interface (API), a computing interface that allows researchers to access the application. Here, data from Flickr was obtained using the "photosearcher" package (Fox et al. 2020a) within the R environment (R Core Team 2020). Our image search was limited to any georeferenced images falling within Wales, delimited by a shapefile (Ordnance Survey 2020), taken between the 1st of January 2010 and the 1st of January 2021, uploaded before the 1st of February 2021 and containing a given keyword in the images title, tags or description. To ensure that all images of hiking were captured our keyword search included synonyms of hiking: "hike", "hiking", "walk", "walking", "trek", "trekking", "ramble", "rambling".

Not all posts returned from social media sites are useful for CES assessments. Many posts to social media sites contain images taken inside buildings or images of people, while other posts may relate to a negative experience (Fox et al. 2021b). To extract only images relating to the natural

environment, the contents of images were automatically tagged using the Google Cloud Vision API (Google Cloud Vision API 2020) within the R environment using the imgrec package (Schwemmer 2019). The Google Cloud Vision API is a machine learning model which can label the contents of images with over 1 million different tags based on a pre-trained dataset. Here, we returned labels for the first 10 features detected in each image. Each label is returned with a given confidence score, scaled between 0 and 1. To select accurate labels without manual validation, we only retained labels with a confidence score of greater than 0.6 (Gosal et al. 2019). We classified the automatically generated labels into biophysical aspects of nature, such as features of biodiversity (e.g. tree or bird) or features of geodiversity (e.g. mountain or lake) and non-biophysical nature labels (e.g. building, car, person, sky) (for a full list of classification see Supporting Information 2). To remove images that are not predominantly of human-nature interactions (e.g. photographs focused on buildings or people), we calculated the ratio of biophysical nature labels to non-biophysical natural labels. Images containing more nature labels than non-nature labels were deemed to be an image focused on nature and were retained for further analysis. Choosing a threshold of at least 50% biophysical nature labels provides a suitable dataset representing human-nature relationships (Fox et al. 2021b). Furthermore, to ensure that users were experiencing a benefit from the hiking experience, we carried out a textual sentiment analysis on the title, description and tags of each image. Here, we used the AFINN dictionary (Nielsen 2011), a collection of words ranked from $+5$ (positive words) to -5 (negative words), to calculate the sum sentiment expressed in the textual metadata of each image. Images in which the user expresses an overall positive sentiment towards the activity were retained for further analysis.

Predictor Variables

To assess the contribution of geodiversity to recreational activities, we chose spatial layers to represent geodiversity features where the relationship with CES has previously been highlighted (Van Zanten et al. 2016; Oteros-Rozas et al. 2018; Van Berkel et al. 2018), as well as features that have not been previously assessed. These are as follows: count of lakes; count of rivers; Euclidian distance to coast; range in elevation; range in slope; count of landscape types; count of bedrock types; count of geosites; and count of soil types (Table 1). Geosites were defined as any Regionally Important Geological and Geomorphological Sites (RIGS) (NRW 2021). Furthermore, as previous studies have found that infrastructure and accessibility are associated with CES distribution (Richards and Tunçer 2018; Muñoz et al. 2020), we also included variables representing these Euclidian distance from roads and Euclidian distance from greenspaces access points. Here, greenspace access points are the entrance to any allotments or community growing spaces, bowling green, cemetery, religious grounds, golf course, other sports facility, play space, playing field, public park or garden and tennis court (Ordnance Survey 2020). Other studies have found that historic sites can also influence the CES distribution obtained from Flickr images (Gliozzo et al. 2016; Van Berkel et al. 2018), so we included the count of scheduled monuments, which are sites of archaeological importance such as burial mounds, castles and churches (NRW 2021). Furthermore, as recreation and CES can be influenced by

Table 1 Predictor variables chosen for species distribution modelling

Grouping	Category	Variable	Description
Geodiversity variables	Geomorphology	elevRange	Range in elevation derived from a DEM (Copernicus 2021)
		slopeRange	Range in slope derived from a DEM (Copernicus 2021)
		landCount	Count of different landscape types (NRW 2021)
	Geology	rockCount	Count of different bedrock types (Ordnance Survey 2020)
		geoDist	Count of differnt RIGs (NRW 2021)
	Hydrology	lakeCount	Count of differnt lakes (NRW 2021)
		riverCount	Count of differnt rivers (NRW 2021)
		coastDist	Euclidean distance from the coast (Ordnance Survey 2020)
	Soil	soilCount	Count of different soil types (Cranfield University 2021)
Non-geodiversity variables	Accessibility and infrastructure	roadDist	Euclidean distance from a road (Ordnance Survey 2020)
		greenDist	Euclidean distance from a greenspace access point (Ordnance Survey 2020)
	Heritage	monuDist	Count of different scheduled monuments (NRW 2021)
	Biodiversity	natvegArea	Area of natural or semi-natural vegetation derived from a landcover map (Copernicus 2021)
		paDist	Euclidean distance from a protected area (UNEP-WCMC and IUCN 2021)

biodiversity and areas designated for its protection, we also included the area of natural vegetation (Copernicus 2021) and distance to a protected area (Protected Planet 2021). Natural and semi-natural vegetation were any areas of broad-leaved forest, coniferous forest, mixed forest, natural grassland, moors and heathland, transitional woodland-shrub, sparsely vegetated areas, inland marshes, peat bogs and salt marshes (Copernicus 2021). As recreational activities can be impacted by landscape diversity at a spatial scale of around 10 km (Graham and Eigenbrod 2019), here, predictor variables were summarised into 10 km^2 grid cells, e.g. elevation range within the 10 km^2 grid cell. To transform the individual variables raster to a comparable scale, each of the raster maps was normalized to a 0–1 scale (Supporting information 3 and 4).

Distribution Modelling

Within the R environment, the BIOMOD2 package (Thuiller et al. 2009, 2012) can be used to perform ensemble distribution modelling using 10 models: generalized linear model (GLM), generalized additive model (GAM), generalized boosting model (GBM), classification tree analysis (CTA), artificial neural network (ANN), surface range envelop (SRE), flexible discriminant analysis (FDA), multiple adaptive regression splines (MARS), random forest (RF), and MaxEnt. The settings for each model were set to the BIOMOD2 default (Hodd et al. 2014). The different models stem from different mathematical backgrounds, for example, CTA is a classification-based model, GLM a regression-based model and MaxEnt is a machine learning-based model (Guisan et al. 2002; Thuiller et al. 2003; Phillips et al. 2006).

Many of the models require both presence and absence data to model the distribution. However, as geotagged photographs from Flickr are a presence-only dataset, there is a need to generate pseudo-absences. The accuracy of each type of model can be influenced by the number of pseudo-absences used, with each model type having a potential optimal number. For example, Barbet-Massin (et al. 2012) suggest using a large number of pseudo-absences for regression models and an equal number of pseudo-absences to presences for classification techniques. However, using the individual optimal number of pseudo-absences for each model type cannot be applied to the ensemble as the models cannot be compared in an unbiased manner unless the same data is used on all the models (Barbet-Massin et al. 2012). Čengić et al. (2020), found that the number of pseudo-absences does not have a strong effect on model performance across multiple model types and recommend choosing a fixed value. Therefore, we generated an equal number of pseudo-absences to the number of presence points. Furthermore, model performance can be influenced by the prevalence (weighting of presences/

pseudo-absences). Here, the modelled prevalence was set at 0.50, which gives the presence and pseudo-absence points equal weighting in the models, as this is recommended for most model types (Barbet-Massin et al. 2012).

To assess each model's performance the data was split into two groups, 80% for training and 20% for testing (Hodd et al. 2014). We used three common metrics to evaluate model performance, the kappa statistic, the true skill score (TSS) and the area under the receiver operating characteristic curve (AUC). The kappa statistic measures the difference in the observed agreement from the model and the expected agreement on a standardized scale of -1 to + 1 scale, with 1 representing perfect agreement, 0 agreement expected by chance, and negative values indicate agreement less than chance (Viera and Garrett 2005). The TSS assesses model performance through the model's sensitivity (probability the model correctly classifies a presence) and specificity (probability the model correctly classifies an absence) (Allouche et al. 2006). The TSS is also on a standardized scale between -1 and + 1, with higher scores indicating better model performance and scores close to or less than zero indicating that the model is no better than random (Allouche et al. 2006; Kaky et al. 2020). The AUC curve is the plot of the sensitivity against (1-specificity) across a series of cut-off points, and the AUC calculates a single number across all thresholds with 0 representing a model where the prediction is 100% incorrect and a value of 1 a model where the predictions are 100% correct (Lobo et al. 2008). To minimise uncertainties arising from subsampling, we carried out 10 replications of each algorithm for cross-validation, with the mean kappa statistic, TSS score and AUC for the testing data for each algorithm used to assess model performance (Kaky et al. 2020). The ensemble models were then built using any model runs with kappa and TSS scores of > 0.6 and an AUC of > 0.8, the minimum acceptable standard of accuracy for the metrics (Hodd et al. 2014).

Here, we wish to assess the importance of the predictor variables for the distribution of recreational activity. However, methods for assessing variable contribution are model-specific and therefore limits the comparison between models. Instead, BIOMOD2 calculates the variable contribution to the model independent of the model algorithm (Thuiller et al. 2009). Variable importance is calculated by Pearson's correlation between the standard prediction (fitted values) and prediction where the variable being assessed is randomly generated. Predictor variables are considered not important when there are higher correlation values—indicating that there is little variation in the standard predictions and the randomly permutated predictions (Thuiller et al. 2009). Here, we assessed the variable importance based on the model algorithms with a mean kappa and TSS score of > 0.6 and an AUC of > 0.8. Furthermore, for these model

algorithms, we generated response curves for the predictor variables.

Mapping Geodiversity Indices

A spatial index for geodiversity can be calculated by adding up normalised partial geodiversity indices e.g. normalised elevation range and normalised bedrock type count (Melelli et al. 2017). These geodiversity index maps can be useful indicators for informing geoconservation efforts (Melelli et al. 2017). However, mapping all geodiversity features with equal importance as each other may introduce biases and not highlight areas with high conservation value (e.g. areas of geodiversity with high CES value). Therefore, a common approach to mapping geodiversity is to weight indices based on the conservation goal (Jankowski et al. 2020). Often geodiversity partial indices are combined into themes; e.g., the count of different lakes and rivers are combined into a hydrology partial index (dos Santos et al. 2020). However, this may decrease the relative importance of each feature type to the final geodiversity index. Therefore, we create normalised partial geodiversity indices for each of the 9 geodiversity variables: count of lakes; count of rivers; Euclidian distance to coast; range in elevation; range in slope; count of landscape types; count of bedrock types; count of geosites; and count of soil types (Fig. 1).

We made two overall geodiversity indices, one where the partial indices had equal weighting and another where we weighted the partial indices based on their mean variable contribution to the distribution model algorithms with
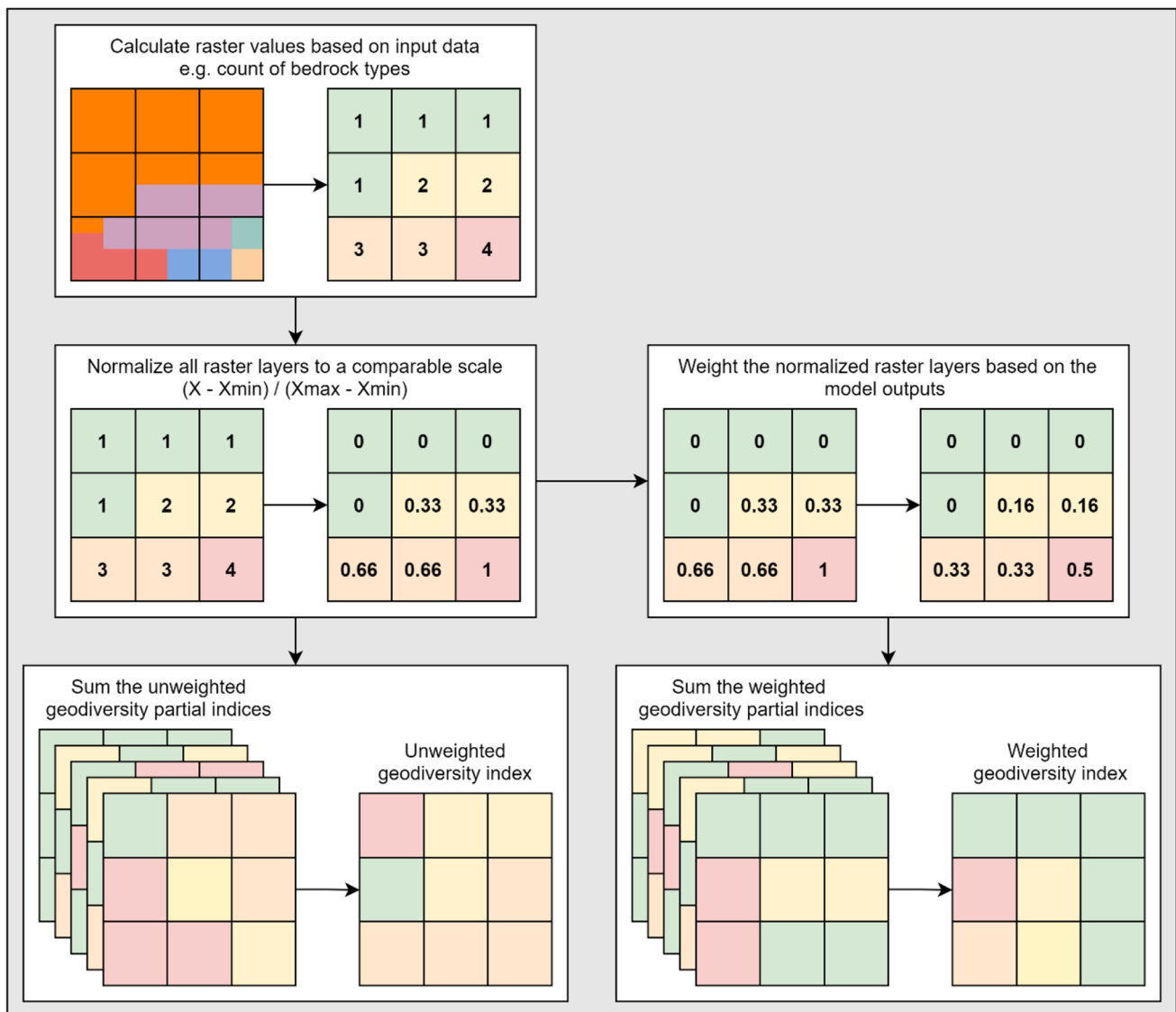

Fig. 1 Methods for creating unweighted and weighted geodiversity indices

a mean kappa and TSS score of > 0.6 and AUC > 0.8. To compare between the equally weighted geodiversity index and the weighted geodiversity index, the two geodiversity indices were reclassified into five classes defined by equal intervals: very low, low, medium, high and very high geodiversity (Pereira et al. 2013; dos Santos et al. 2020; Jankowski et al. 2020).

Understanding Human-Nature Interactions

As well as understanding where people hike, it is also important to understand which aspects of the natural and human-made environment people interact with while hiking in these locations. Purely assessing photograph distribution does account for what the subject of the image is (Yan et al. 2019). For example, people hiking in areas of high elevation may not be interested in topography, but instead they are in that location to take photographs of large-scale vegetation such as a forest (Aiba et al. 2019). To reduce biases introduced by overactive users, following a method similar to the photo user days metric (Wood et al. 2013), we grouped all the images a single user took on a single day as one. For each unique grouping of images by a single user on a single day, we summarised the unique labels returned by the Google Vision Cloud API (e.g. if a single user took 60 images labelled "mountain" a single day, this would be reduced to one "mountain" label). We calculated the frequency of the labels across all users and days and ranked them based on the number of images. Furthermore, to understand which

features provide a more positive hiking experience, for each label we calculated the mean textual sentiment value of the images containing that feature.

Results

Social Media Data

There were 20,910 images taken in Wales between the 1st of January 2010 and the 1st of January 2021 which had the chosen hiking synonyms found within the images title, tags or description. Of these, 16,591 (79.34%) were images we deemed to have a focus on natural or semi-natural areas with some biophysical nature features. Of these, 4,919 (23.52% of the full hiking dataset) images had an overall positive sentiment expressed in the image's textual metadata (Fig. 2).

Distribution Models

Model performance varied between model algorithms, with RF performing the best based on the kappa statistic, TSS and AUC, while the SRE algorithm performed the worst. Out of all the model's RF, CTA, ANN and GMB had a mean kappa statistic and TSS > 0.6 and AUC > 0.8. If using the AUC metric alone, FDA, GAM, MARS and MAXENT would have also been selected (Supporting Information 3).

The order of the most important predictor variables varied with the model algorithm used, though overall the
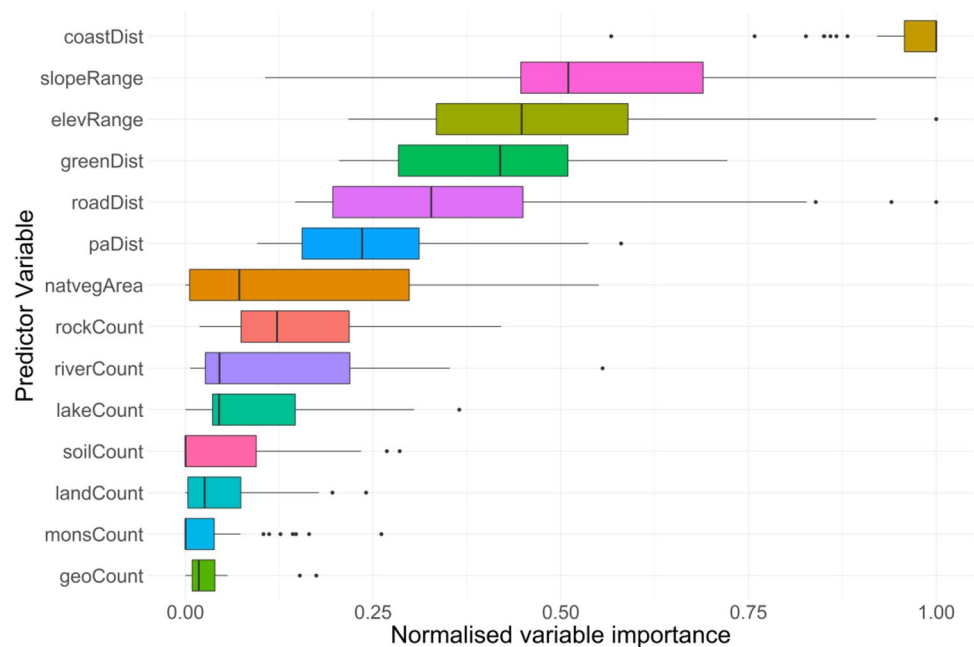

Fig. 2 Distribution of hiking images in Wales; these images are classed as a CES as they have both a focus on biophysical features of nature and contain a positive textual sentiment expressed in the metadata. The inset map shows the location of Wales in comparison to the rest of the UK
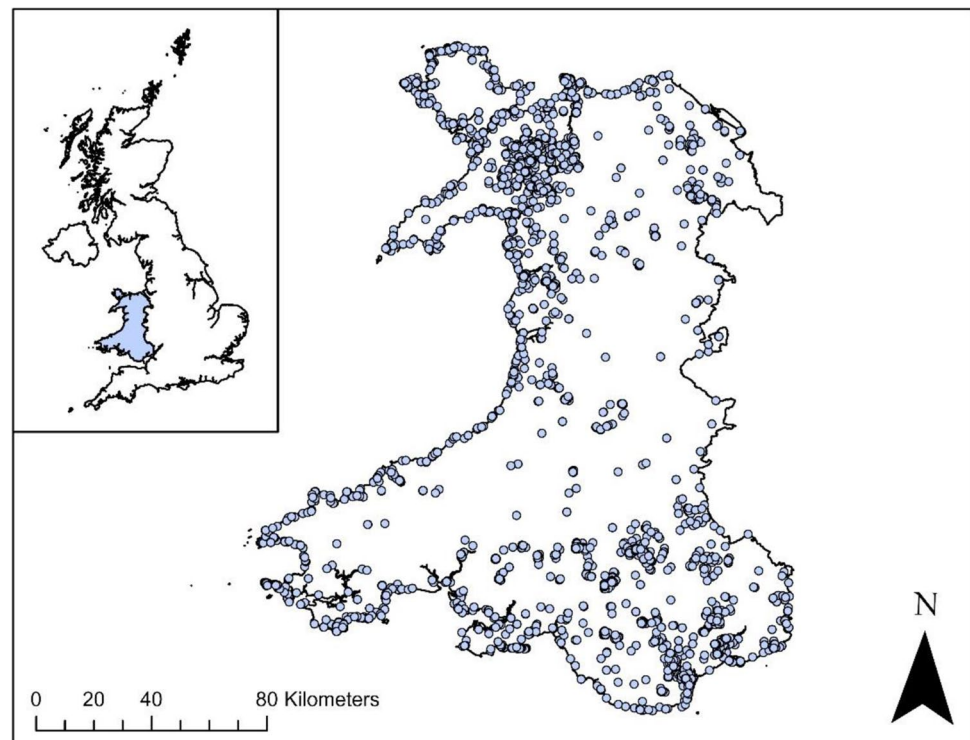
distance to the coast was consistently the most important predictor (Fig. 3; for individual models see Supporting Information 5). The other most important variables were range in elevation, range in slope, distance from a greenspace entrance and distance from a road. Count of different landscape types, bedrock types, soil types, rivers, lakes and monuments all had relatively low importance across the models. Though the biodiversity measures, area of vegetation and distance to protected area, had a relatively higher mean importance than some of the geodiversity measures, they also had a relatively low overall importance.

Inspection of the response curves demonstrates that the probability of images of hiking decreases with distances further away from the coast (Fig. 4, for individual model responses, see Supporting Information 6). Furthermore, all model algorithms show an increase in the probability of hiking images with larger ranges in slope and elevation. For the accessibility predictors, the probability of hiking images increases when closer to a greenspace entrance, however, the probability of images of hiking increases further away from roads. The models' response curves also show that being closer to a greenspace access point also increases the probability of hiking images. There is a small increase in the probability of hiking images in areas with more lakes and geosites. Furthermore, for biodiversity, there is some increase in probability when closer to protected areas, but there is no real change depending on how much natural or semi-natural vegetation is present. There seems to be no change in response when the varying, count of rivers, count of bedrock types, count of landscape types, count of soil types and count of scheduled monuments.
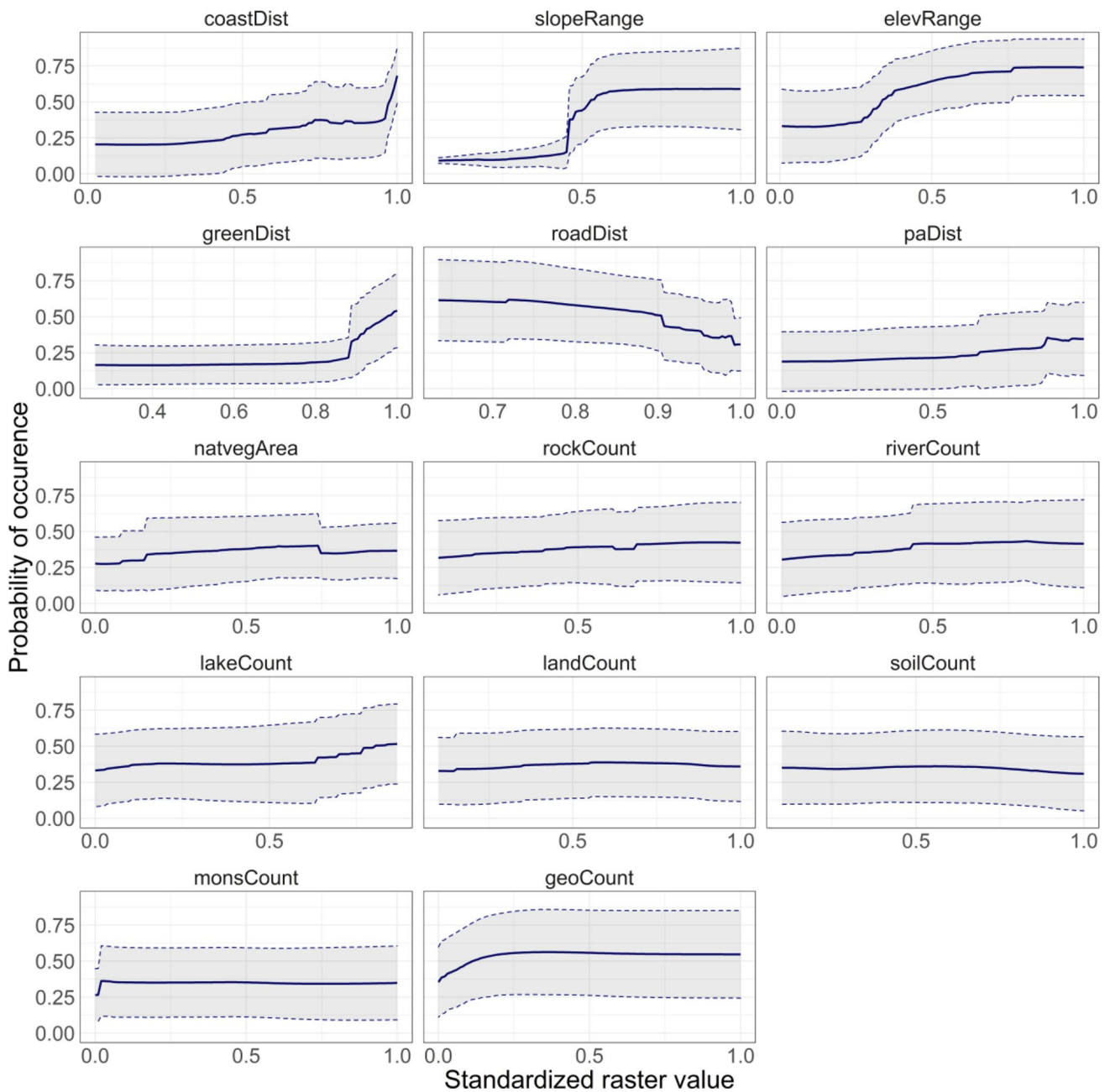
Geodiversity Map

The final geodiversity weightings vary based on the method chosen (Table 2). When all the geodiversity partial indices (count of lakes; count of rivers; Euclidian distance to coast; range in elevation; range in slope; count of landscape types; count of bedrock types; count of geosites; and count of soil types) are given equal weighting, there are small pockets of high geodiversity scattered across Wales (Fig. 5). In general, the two indices agree (45.41% of pixels remained unchanged), with both indices showing very high geodiversity in the Snowdonia National Parks. However, when weighted to the mean variable importance for the distribution models, there is also a relatively large number of areas showing higher geodiversity, with 48.19% of pixels showing an increase. When the geodiversity index is weighted, the higher geodiversity values tend to be within the GeoMôn Geopark or the Pembrokeshire Coast National Park. However, some of the higher geodiversity areas along coastal areas fall outside of any of the larger protected area boundaries. Some areas experienced a decrease in the geodiversity index, with 6.40% pixels having a decline in geodiversity when using the model weightings.

Human-Nature Interactions

In total, there were 1105 different image content labels returned by the Google Vision Cloud API. While hiking, people take photographs of both geodiversity and biodiversity features (Fig. 6). For geodiversity, people more frequently photograph elements of topography and water bodies. Furthermore, images containing river landforms

Fig. 3 Normalised variable importance for ANN, CTA, GBM and RF (model algorithms with a mean kappa statistic and TSS value > 0.6 and AUC value > 0.8)

Fig. 4 Mean, ± 1 SD response curves for all variables from all runs of the ANN, CTA, GBM and RF models. The normalized raster values were created for comparable scale for each variable summarised within the 10km² square: for distance variables, the larger normalized values are, the closer to the feature in question; for the count variables the larger the value the more different types of the feature; for range variable the larger the value the greater the range of that feature and for the area variable the larger the value the greater the area of coverage by that feature

(e.g. "fluvial landforms of streams") and lakes (e.g. "lake") appear relatively frequently (in the most frequent 30 labels out of a possible 1,105). The features of biodiversity photographed most by hikers were primarily of flora, with none of the most frequently photographed features relating to animals. Furthermore, there were some images containing human features, with the most photographed features being "building" (234 images) and "people in nature" (253 images).

There was little variation in the mean sentiment score between images relating to the 30 most frequently photographed features (Fig. 6). Images associated with "building", "nautre" and "coastal and oceanic landforms" have the large associated mean sentiment. Images of topographic features

**Table 2** Weightings used on the geodiversity indices

Variable	Equally weighted map	Weighted based on distribution models
elevRange	1.00	0.50
slopeRange	1.00	0.57
landCount	1.00	0.05
rockCount	1.00	0.15
geoCount	1.00	0.03
lakeCount	1.00	0.09
riverCount	1.00	0.13
coastDist	1.00	0.96
soilCount	1.00	0.05

such as "highland" and "mountain", and those relating to rivers, such as "fluvial landforms of streams", also have relatively high associated sentiment scores. Images containing features labelled as "grassland", "terrestrial plant" and "grass" have the lowest mean sentiment values. We note that the mean sentiment score of most labels has a relatively large associated standard deviation.

Discussion

By using big data from social media sites, we have begun to untangle the complex relationship between geodiversity and CES. First, we have highlighted which natural and human-made features are important in driving the distribution of the locations where people choose to go hiking. Second, we have demonstrated what aspects of the natural and built environment are interacted with while hiking. These are key findings because they can provide information on where geo-conservation efforts should be focused and which features of geodiversity should be prioritised. By highlighting areas of high geodiversity in our geodiversity index maps, policy and decision-makers can combine these outputs with other sources of data (e.g. visitation rates) in order to highlight areas that are potentially at risk of loss to the integrity of geodiversity (Bétard and Peulvast 2019). These findings could then guide sustainable decision making to ensure that geodiversity and the ES that it provides can be consumed sustainably, such as the goals of the UNESCO Global Geopark Program (Prosser 2013).

We found that distance from the coast was the biggest driver of the distribution of hiking in Wales. The importance of the coast on the distribution of CES is similar to other studies using social media data to assess CES. Van Berkel et al. (2018) found that beaches were the most visited area in a coastal-inland gradient, while Ghermandi et al. (2020b) found a higher positive sentiment expressed in coastal images when compared to other landscape types.
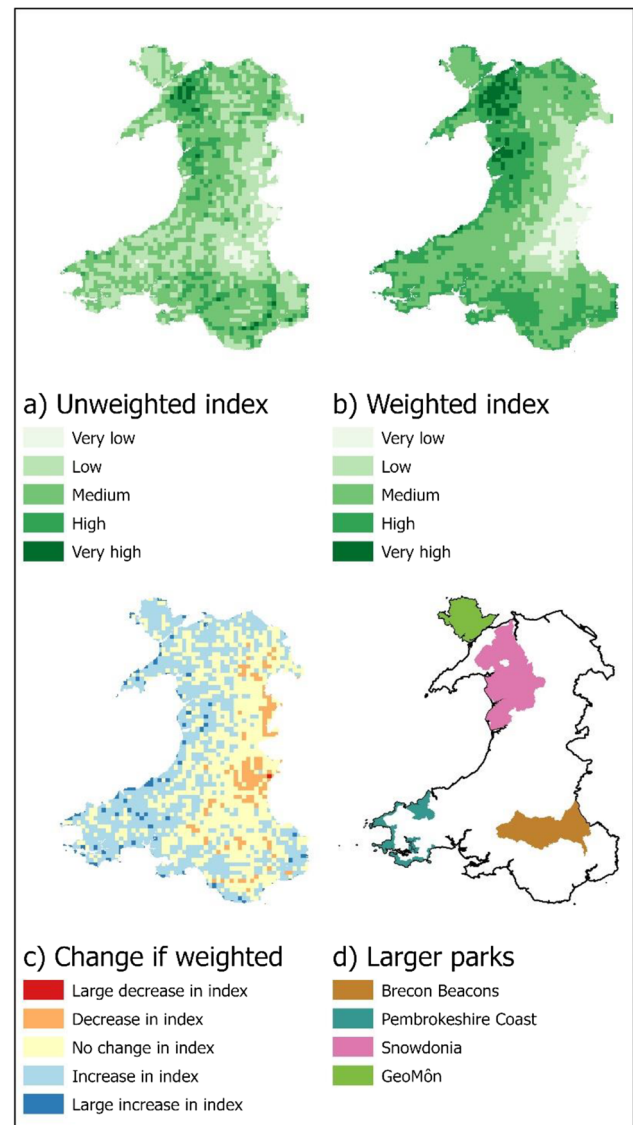
Fig. 5 Normalised geodiversity index—partial indices: count of lakes; count of rivers; Euclidian distance to coast; range in elevation; range in slope; count of landscape types; count of bedrock types; count of geosites; and count of soil types. **a** Partial indices unweighted; **b** partial indices weighted by their mean variable contribution to the ANN, CTA, GBM and RF models; **c** difference in weightings (weighted – unweighted); **d** location of national parks, and GeoMon Geopark (the Fforest Fawr Geopark is located within the boundaries of the Brecon Beacons National Park)

Furthermore, survey data has demonstrated that people in the UK are happiest in natural coastal environments compared to other natural and non-natural areas (MacKerron and Mourato 2013) and that coastal parks in South Africa offer more opportunities for recreational activities such as hiking (Roux et al. 2020). Our results highlight the importance of coastal regions in driving CES, while the image content and sentiment analysis revealed that many people take photographs of coastal and oceanic landforms and that

Fig. 6 The 30 most frequently photographed features based on the Google Vision Cloud API labels, and the mean textual sentiment of all images containing those features
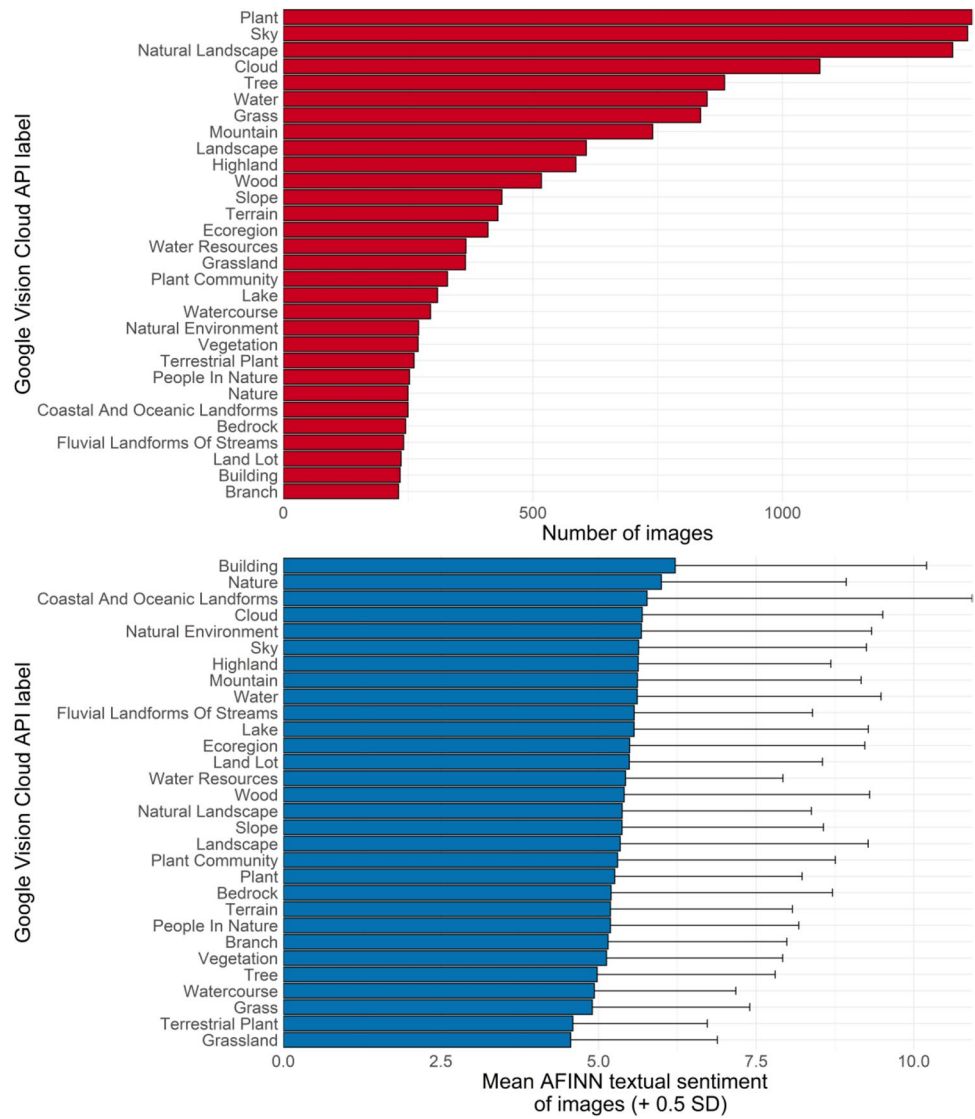
these photographs have a relatively high associated sentiment value. This aligns with Ghermandi et al. (2020a), who found that coastal landforms such as lagoons play an important role in CES and tourism. Future studies should build upon this and assess coastal sites at a finer resolution to further untangle smaller scale relationships between geodiversity and CES.

Our findings also highlight that geomorphological features such as elevation and slope are important determinants in the distribution of recreation. In our study, highly varied slope and elevation increased the probability of hiking images. Other studies have found similar relationships, such as (Van Zanten et al. 2016) who found that features such as hills and mountains were the best predictors of recreational value and (Aiba et al. 2019) who found that the height of mountains plays an important role in the hiking experience. However, these relationships may vary between individuals; for example, some people may find low geomorphological

variation boring for hiking, while less experienced hikers may prefer flatter terrain that does not present too great a challenge (Chhetri 2015). As not all demographics are captured by social media data, with those of lower socioeconomic status often underrepresented (Oteros-Rozas et al. 2018; Hargittai 2020), these results may not be reflective of the entire population and management decisions need to ensure that the opinions of theses demographics are also represented (Graham and Eigenbrod 2019).

Previous studies have also found that water bodies such as lakes and rivers can be an important driver of recreational activities such as hiking (Oteros-Rozas et al. 2018; Schirpke et al. 2018). Nevertheless here, we found that the number of lakes and river was less important in driving distribution compared to other variables. The difference in hydrological features driving recreational activities may be due to the location of the study, demographics of users, or the scale at which we assessed the interactions. For

example, Ghermandi et al. (2020a) found that locals were more likely to take images of coastal habitats than international tourists, which may account for the importance of the coast when compared to freshwater features. Furthermore, though many of the hiking images were taken close to rivers or lakes, here we assessed the relationship between these as geodiversity and ES at a landscape scale (10km^2). As the relationship between geodiversity and ES varies over the spatial scale assessed, the relationship between rivers and hiking may be better explained at a smaller spatial scale (Alahuhta et al. 2018).

We also found that two accessibility variables were important in determining the distribution of hiking in Wales. First, our models suggested that people are more likely to hike closer to the entrance to a greenspace. This further highlights that recreational opportunities and CES are facilitated with access to nature (Roux et al. 2020). Second, here there was a greater increase in the probability of images taken further away from roads. This has also been found by other studies which find high levels of CES in areas that are not accessible by, or close, to roads (Muñoz et al. 2020). However, other social media and CES studies, or those using citizen science data (e.g. nature observation applications such as eBird and iNaturalist), where observations are found closer to infrastructure such as roads (Jacobs and Zipf 2017; Havinga et al. 2020; Muñoz et al. 2020). This complex relationship may be due to the differences in motivations for hiking (Wilcer et al. 2019); for example, novice hikers may not travel far from where they parked or people seeking a tranquil experience may travel much further. These relationships with human infrastructure further demonstrate that CES are co-produced through complex human-nature interactions (Fischer and Eastwood 2016).

Other studies have often found that heritage sites are related to social media posts. For example, Gliozzo et al. (2016) found that viewpoints of historic human structures can influence the distribution of social media uploads in Wales and Van Berkel et al. (2018) found that within coastal environments cultural heritage features were important drivers of CES. However, here, heritage sites were less important at driving CES than natural features such as geomorphology and hydrology, which is consistent with other studies (Kim et al. 2019). Our findings agree with Gliozzo et al. (2016), who suggested that distance to protected areas is relatively important in driving the distribution of CES in Wales. Furthermore, as with other studies, we also found that vegetation cover is generally not the most important driving of hiking (Aiba et al. 2019). Overall, many factors are contributing to the distribution of hiking in Wales, demonstrating that CES are not just influenced by the interactions of geodiversity and biodiversity, but co-produced through complex interactions with society, built infrastructure and heritage

(Fischer and Eastwood 2016; Haines-Young and Potschin 2018; Fox et al. 2020b).

As expected, there were differences between the drivers of where people go hiking (based on the location of the images) and what they photograph when there (Yan et al. 2019). Many of the images do contain features found to be important in driving CES distribution such as geomorphological features. However, some of the drivers of distribution were less represented in the image content. For example, though still being one of the most frequently photographed features (in the top 30), there were relatively few images of coastal landforms compared to other features such as trees and mountains. This may indicate that though people generally go to coastal areas, there may be other smaller-scale natural or human features that contribute to this pattern. Conversely, though scheduled monuments were not an important driver of distribution, we found that images of buildings had the highest relative sentiment score. As we filtered images to represent natural or semi-natural areas, this may suggest that the overall number of human-made features may not be important in driving the hiking experience, but rather individual historic or iconic buildings can contribute to sentimental value and provide extremely positive experiences.

We found that the most photographed feature of geodiversity was water, with many images being of rivers and lakes. There were also a relatively high number of images concerning geological and hydrological landforms, including bedrock and fluvial landforms. As well as being highly photographed images of coastal and oceanic landforms and fluvial landforms of streams had relatively high associated sentiments. This suggests that though variety in geodiversity features, e.g. many different lakes or geological landforms, may not drive the general distribution of hiking in Wales when people are on their hike they appreciate and interact with individual features, e.g. a single lake or geological landform. The larger-scale distribution of hiking may therefore be driven by people choosing to hike in areas of high geomorphological variation to receive restorative benefits from physical exercise, while at the local scale, people receive additional creative benefits from hiking in areas with aesthetic views of hydrological and geological features (King et al. 2017; Oteros-Rozas et al. 2018; Schirpke et al. 2018; Wilcer et al. 2019).

As recreational activities such as hiking can cause damage to landforms and hydrological systems (Gray 2008b; Hjort et al. 2015; Wu et al. 2021), these findings have important geoconservation implications. Suitable geoconservation strategies in Wales should continue to promote hiking in a way that maximises the CES benefits, but in a sustainable way that minimise damage caused to the geodiversity features people most interact with. For example, hiking trails could benefit from the construction of educational signage

that encourages hikers not to touch at risk landforms and refrain from littering to prevent the contamination of hydrological systems, or from the creation of lookout points that physically restrict hikers from interacting with at-risk features while still providing views that contribute to the CES benefits (Gray 2008a).

A range of different spatially explicit methods of calculating geodiversity indices exist across the literature, from counts of different elements to specific diversity indices (Zwoliński et al. 2018). For example, some studies calculated partial indices by counting up the variety of a feature (e.g. geological units and soil orders) within a grid (e.g. Hjort et al. 2012; Pereira et al. 2013), some calculate geodiversity using standard diversity measures, such as Simpson's Diversity Index and Simpson's Evenness Index (e.g. Benito-Calvo et al. 2009), while other studies apply variations of the geodiversity index introduced by Serrano and Ruiz-Flaño (2007). Since our objective was to understand what drives the distribution of where people go hiking, some of our geodiversity partial indices were calculated using count or distance and not a specific diversity measure. For example, by using distance as a partial index measure, we were able to untangle the finer importance of being close to the coastline for recreational hiking in Wales that would not have been possible if we calculated diversity across all hydrological features. Further work should therefore quantify how the differences in geodiversity measures impact their usefulness to differing study contexts (Zwoliński et al. 2018).

Though here we have demonstrated the main geodiversity drivers of hiking (distance to coast and topography) are unmanageable, these results can still be useful for guiding future geoconservation strategies (Fox et al. 2020b). By combining information on where people go hiking, with information on what they interact with when hiking we can better inform geoconservation methods. Here, our results indicate that geoconservation efforts to mitigate against any damage from hiking in Wales may be best focused in coastal and mountainous areas, with targeted management strategies for protecting geomorphological, geological and hydrological features and landforms. Furthermore, as many hikers took photographs of flora, any geoconservation strategies should be undertaken holistically to ensure the future protection of biodiversity as well (Anderson et al. 2015; Lawler et al. 2015). Many areas of high geodiversity already fall within protected areas, primarily in the larger sites such as the Snowdonia National Park, the Pembrokeshire Coast National Park and the GeoMôn Geopark (particularly when weighted based on the distribution models), suggesting that Wales is undertaking good steps in protecting geodiversity from damage caused by recreational activities (Prosser et al. 2010; Evans et al. 2018). However, there are several areas with a high geodiversity, where potential damage is not yet mitigated. Here, these areas may benefit from the creation of geoconservation in the form of a protected area

that promotes the sustainable use of geodiversity for tourism and recreation activities, such as the goals of some UNESCO Global Geoparks (Henriques and Brilha 2017; Gordon 2018; Gray 2019). Though here we have assessed geoconservation in Wales with a focus on promoting sustainable recreation, these methods are transferable to studies with different regions and different management goals. For example, one could use geodiversity variables to predict species distribution (Bailey et al. 2017) and use the results of the analysis to inform geoconservation management strategies that also benefit biodiversity conservation (Anderson et al. 2015; Lawler et al. 2015).

Conclusion

Geodiversity plays an integral role in the delivery of hiking as a CES, both driving the general trends in its distribution as well as the features of nature that people interact with while hiking. Here, we have shown that geomorphological features including the range in slope and the range in elevation and hydrological features such as distance to the coast can play an important role in determining the distribution of hiking. We also note the importance of the co-production of CES through human-nature interactions, with access to nature being key to driving recreational activity distribution. Here, both distance to roads and distance to greenspace access points contributed highly to the distribution of hiking in Wales. While hiking, people tend to interact with geomorphological, geological and hydrological landforms. Geoconservation management strategies should therefore focus on promoting hiking in a suitable manner that maximises the CES benefits received while ensuring the future of the geodiversity features that contribute to these. Future work should apply these methods to different activities, conservation goals and study sites to help inform more tailored geoconservation management plans.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Aiba M, Shibata R, Oguro M, Nakashizuka T (2019) The seasonal and scale-dependent associations between vegetation quality and hiking activities as a recreation service. Sustain Sci 14:119–129. https://doi.org/10.1007/s11625-018-0609-7

Alahuhta J, Ala-Hulkko T, Tukiainen H (2018) The role of geodiversity in providing ecosystem services at broad scales. Ecol Indic 91:47–56. https://doi.org/10.1016/j.ecolind.2018.03.068

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43:1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Anderson MG, Comer PJ, Beier P (2015) Case studies of conservation plans that incorporate geodiversity. Conserv Biol 29:680–691. https://doi.org/10.1111/cobi.12503

Arslan ES, Örücü ÖK (2020) MaxEnt modelling of the potential distribution areas of cultural ecosystem services using social media data and GIS. Environ Dev Sustain 23:2655–2667. https://doi.org/10.1007/s10668-020-00692-3

Bailey JJ, Boyd DS, Hjort J (2017) Modelling native and alien vascular plant species richness: at which scales is geodiversity most relevant? Glob Ecol Biogeogr 26:763–776. https://doi.org/10.1111/geb.12574

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3:327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barve V (2014) Discovering and developing primary biodiversity data from social networking sites: a novel approach. Ecol Inform 24:194–199. https://doi.org/10.1016/j.ecoinf.2014.08.008

Becken S, Stantic B, Chen J (2017) Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. J Environ Manage 203:87–97. https://doi.org/10.1016/j.jenvman.2017.07.007

Benito-Calvo A, Pérez-González A, Magri O, Meza P (2009) Assessing regional geodiversity: the Iberian Peninsula. Earth Surf Proc Land 34(10):1433–1445

Betard F, Peulvast JP (2019) Geodiversity hotspots: concept, method and cartographic application for geoconservation purposes at a regional scale. Environ Manage 63(6):822–834

Brazier V, Bruneau PMC, Gordon JE, Rennie AF (2012) Making space for nature in a changing climate: the role of geodiversity in biodiversity conservation. Scottish Geogr J 128:211–233. https://doi.org/10.1080/14702541.2012.737015

Burek C (2012) The Role of LGAPs (Local Geodiversity Action Plans) and Welsh RIGS as local drivers for geoconservation within Geotourism in Wales. Geoheritage 4:45–63. https://doi.org/10.1007/s12371-012-0054-4

Čengić M, Rost J, Remenska D (2020) On the importance of predictor choice, modelling technique, and number of pseudo-absences for bioclimatic envelope model performance. Ecol Evol 10:12307–12317. https://doi.org/10.1002/ece3.6859

Chhetri P (2015) A GIS methodology for modelling hiking experiences in the Grampians National Park, Australia. Tour Geogr 17:795–814. https://doi.org/10.1080/14616688.2015.1083609

Collins-Kreiner N, Kliot N (2017) Why do people hike? Hiking the Israel National Trail. Tijdschr Voor Econ En Soc Geogr 108:669–687. https://doi.org/10.1111/tesg.12245

Copernicus (2021) Copernicus land portal

Cranfield University (2021) The Soils Guide

Daniel TC, Muhar A, Arnberger A (2012) Contributions of cultural services to the ecosystem services agenda. Proc Natl Acad Sci U S A 109:8812–8819. https://doi.org/10.1073/pnas.1114773109

Ding X, Fan H (2019) Exploring the distribution patterns of flickr photos ISPRS. Int J Geo-Information 8. https://doi.org/10.3390/ijgi8090418

dos Santos FM, de La Corte BD, Saad AR, da Silva Ferreira AT (2020) Geodiversity index weighted by multivariate statistical analysis. Appl Geomatics 12:361–370. https://doi.org/10.1007/s12518-020-00303-w

Evans BG, Cleal CJ, Thomas BA (2018) Geotourism in an Industrial Setting: the South Wales Coalfield Geoheritage Network. Geoheritage 10:93–107. https://doi.org/10.1007/s12371-017-0226-3

Figueroa-Alfaro RW, Tang Z (2017) Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. J Environ Plan Manag 60:266–281. https://doi.org/10.1080/09640568.2016.1151772

Fischer A, Eastwood A (2016) Coproduction of ecosystem services as human-nature interactions-an analytical framework. Land Use Policy 52:41–50. https://doi.org/10.1016/j.landusepol.2015.12.004

Fox N, August T, Mancini F (2020a) "photosearcher" package in R: an accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX 12

Fox N, Graham LJ, Eigenbrod F (2020b) Incorporating geodiversity in ecosystem service decisions. Ecosyst People 16:151–159. https://doi.org/10.1080/26395916.2020.1758214

Fox N, Graham LJ, Eigenbrod F, Bullock JM, Parks KE (2021a) Reddit: a novel data source for cultural ecosystem service studies. Ecosystem Services 50:101331

Fox N, Graham LJ, Eigenbrod F, Bullock JM, Parks KE (2021b) Enriching social media data allows a more robust representation of cultural ecosystem services. Ecosystem Services 50:101328

Ghermandi A, Sinclair M (2019) Passive crowdsourcing of social media in environmental research: a systematic map. Glob Environ Chang 55:36–47. https://doi.org/10.1016/j.gloenvcha.2019.02.003

Ghermandi A, Camacho-Valdez V, Trejo-Espinosa H (2020) Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tour Manag 77

Ghermandi A, Sinclair M, Fichtman E, Gish M (2020) Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing. Glob Environ Chang 65

Gliozzo G, Pettorelli N, MukiHaklay M (2016) Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales UK Ecol Soc 21. https://doi.org/10.5751/ES-08436-210306

Google Cloud Vision API (2020) Documentation for the Google Cloud Vision API

Gordon JE, Crofts R, Díaz-Martínez E, Woo KS (2018) Enhancing the role of geoconservation in protected area management and nature conservation. Geoheritage 10:191–203. https://doi.org/10.1007/s12371-017-0240-5

Gordon JE (2018) Geoheritage, geotourism and the cultural landscape: enhancing the visitor experience and promoting geoconservation. Geosci 8. https://doi.org/10.3390/geosciences8040136

Gosal AS, Geijzendorffer IR, Václavík T (2019) Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosyst Serv 38

Graham LJ, Eigenbrod F (2019) Scale dependency in drivers of outdoor recreation in England. People Nat 1:406–416. https://doi.org/10.1002/pan3.10042

Gray M (2008) Geodiversity: developing the paradigm. Proc Geol Assoc 119:287–298. https://doi.org/10.1016/S0016-7878(08)80307-0

Gray M (2008) Geodiversity: the origin and evolution of a paradigm. Geol Soc Spec Publ 300:31–36. https://doi.org/10.1144/SP300.4

Gray M (2011) Other nature: geodiversity and geosystem services. Environ Conserv 38:271–274. https://doi.org/10.1017/S0376892911000117

Gray M (2019) Geodiversity, geoheritage and geoconservation for society. Int J Geoheritage Park 7:226–236. https://doi.org/10.1016/j.ijgeop.2019.11.001

Gray M (2013) Geodiversity: valuing and conserving abiotic nature, 2nd Edition. John Wiley & Sons

Gray M (2018) Geodiversity: the backbone of geoheritage and geoconservation. Elsevier Inc

Guisan A, Edward TCJ, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Modell 157:89–100

Haines-Young R, Potschin MB (2018) Common international classification of ecosystem services (CICES) V5.1 and guidance on the application of the revised structure

Hao T, Elith J, Guillera-Arroita G, Lahoz-Monfort JJ (2019) A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Divers Distrib 25:839–852. https://doi.org/10.1111/ddi.12892

Hargittai E (2020) Potential Biases in Big Data: Omitted Voices on Social Media. Soc Sci Comput Rev 38:10–24. https://doi.org/10.1177/0894439318788322

Havinga I, Bogaart PW, Hein L, Tuia D (2020) Defining and spatially modelling cultural ecosystem services using crowdsourced data. Ecosyst Serv 43

Henriques MH, dos Reis RP, Brilha J, Mota T (2011) Geoconservation as an emerging geoscience. Geoheritage 3:117–128. https://doi.org/10.1007/s12371-011-0039-8

Henriques MH, Brilha J (2017) UNESCO Global Geoparks: a strategy towards global understanding and sustainability. Episodes 40:349–355. https://doi.org/10.18814/epiiugs/2017/v40i4/017036

Hermes J, Van Berkel D, Burkhard B (2018) Assessment and valuation of recreational ecosystem services of landscapes. Ecosyst Serv 31:289–295. https://doi.org/10.1016/j.ecoser.2018.04.011

Hjort J, Heikkinen RK, Luoto M (2012) Inclusion of explicit measures of geodiversity improve biodiversity models in a boreal landscape. Biodivers Conserv 21:3487–3506. https://doi.org/10.1007/s10531-012-0376-1

Hjort J, Gordon JE, Gray M, Hunter ML (2015) Why geodiversity matters in valuing nature's stage. Conserv Biol 29:630–639. https://doi.org/10.1111/cobi.12510

Hodd RL, Bourke D, Skeffington MS (2014) Projected range contractions of European protected oceanic montane plant communities: focus on climate change impacts is essential for their future conservation. PLoS ONE 9:1–14. https://doi.org/10.1371/journal.pone.0095147

Jacobs C, Zipf A (2017) Completeness of citizen science biodiversity data from a volunteered geographic information perspective. Geo-Spatial Inf Sci 20:3–13. https://doi.org/10.1080/10095020.2017.1288424

Jankowski P, Najwer A, Zwoliński Z, Niesterowicz J (2020) Geodiversity assessment with crowdsourced data and spatial multicriteria analysis. ISPRS Int J Geo-Information 9:716. https://doi.org/10.3390/ijgi9120716

Johnson ML, Campbell LK, Svendsen ES, McMillen HL (2019) Mapping urban park cultural ecosystem services: a comparison

of twitter and semi-structured interview methods. Sustain 11. https://doi.org/10.3390/su11216137

Kaky E, Nolan V, Alatawi A, Gilbert F (2020) A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: a case study with Egyptian medicinal plants. Ecol Inform 60:101150. https://doi.org/10.1016/j.ecoinf.2020.101150

Kim Y, Kim C, ki, Lee DK (2019) Quantifying nature-based tourism in protected areas in developing countries by using social big data. Tour Manag 72:249–256. https://doi.org/10.1016/j.tourman.2018.12.005

King HP, Morris J, Graves A (2017) Biodiversity and cultural ecosystem benefits in lowland landscapes in southern England. J Environ Psychol 53:185–197. https://doi.org/10.1016/j.jenvp.2017.08.002

Lee H, Seo B, Koellner T, Lautenbach S (2019) Mapping cultural ecosystem services 2.0 – potential and shortcomings from unlabeled crowd sourced images. Ecol Indic 96:505–515. https://doi.org/10.1016/j.ecolind.2018.08.035

Lobo JM, Jiménez-valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17:145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Lawler JJ, Ackerly DD, Albano CM (2015) The theory behind, and the challenges of, conserving nature's stage in a time of rapid change. Conserv Biol 29:618–629. https://doi.org/10.1111/cobi.12505

Langemeyer J, Calcagni F, Baró F (2018) Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. Land Use Policy 77:542–552. https://doi.org/10.1016/j.landusepol.2018.05.049

MacKerron G, Mourato S (2013) Happiness is greater in natural environments. Glob Environ Chang 23:992–1000. https://doi.org/10.1016/j.gloenvcha.2013.03.010

Mancini F, Coghill GM, Lusseau D (2019) Quantifying wildlife watchers' preferences to investigate the overlap between recreational and conservation value of natural areas. J Appl Ecol 56:387–397. https://doi.org/10.1111/1365-2664.13274

Melelli L, Vergari F, Liucci L, Del Monte M (2017) Geomorphodiversity index: quantifying the diversity of landforms and physical landscape. Sci Total Environ 584–585:701–714. https://doi.org/10.1016/j.scitotenv.2017.01.101

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

Millennium Ecosystem Assessment (2005) Ecosystems and human well-being Island Press Washington, D.C.

Mitten D, Overholt JR, Haynes FI (2018) Hiking: a low-cost, accessible intervention to promote health benefits. Am J Lifestyle Med 12:302–310. https://doi.org/10.1177/1559827616658229

Muñoz L, Hausner VH, Runge C (2020) Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. People Nat 2:437–449. https://doi.org/10.1002/pan3.10083

Nielsen FÅ (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. CEUR Workshop Proc 718:93–98

NRW (2021) Contains Natural Resources Wales information © Natural Resources Wales and Database Right. All rights Reserved

Ordnance Survey (2020) Crown copyright and database rights 2020 Ordnance Survey

Oteros-Rozas E, Martín-López B, Fagerholm N (2018) Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. Ecol Indic 94:74–86. https://doi.org/10.1016/j.ecolind.2017.02.009

Parks KE, Mulligan M (2010) On the relationship between a resource based measure of geodiversity and broad scale biodiversity

patterns. Biodivers Conserv 19:2751–2766. https://doi.org/10.1007/s10531-010-9876-z

Pereira DI, Pereira P, Brilha J, Santos L (2013) Geodiversity assessment of Paraná State (Brazil): an innovative approach. Environ Manage 52:541–552. https://doi.org/10.1007/s00267-013-0100-2

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190:231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Prosser CD (2013) Our rich and varied geoconservation portfolio: the foundation for the future. Proc Geol Assoc 124(4):568–580

Prosser CD, Burek CV, Evans DH (2010) Conserving geodiversity sites in a changing climate: management challenges and responses. Geoheritage 2:123–136. https://doi.org/10.1007/s12371-010-0016-7

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing

Richards DR, Tunçer B (2018) Using image recognition to automate assessment of cultural ecosystem services from social media photographs. Ecosyst Serv 31:318–325. https://doi.org/10.1016/j.ecoser.2017.09.004

Roberts HV (2017) Using Twitter data in urban green space research: a case study and critical evaluation. Appl Geogr 81:13–20. https://doi.org/10.1016/j.apgeog.2017.02.008

Roux DJ, Smith MKS, Smit IPJ (2020) Cultural ecosystem services as complex outcomes of people–nature interactions in protected areas. Ecosyst Serv 43

Schirpke U, Meisch C, Marsoner T, Tappeiner U (2018) Revealing spatial and temporal patterns of outdoor recreation in the European Alps and their surroundings. Ecosyst Serv 31:336–350. https://doi.org/10.1016/j.ecoser.2017.11.017

Schwemmer C (2019) imgrec: An Interface for Image Recognition. R package version 0.1.1

Serrano E, Ruiz-Flaño P (2007) Geodiversity: a theoretical and applied concept. Geographica Helvetica 62(3):140–147

Sharples C (1993) A methodology for the identification of significant landforms and geological sites for geoconservation purposes

Tenerelli P, Demšar U, Luque S (2016) Crowdsourcing indicators for cultural ecosystem services: a geographically weighted approach for mountain landscapes. Ecol Indic 64:237–248. https://doi.org/10.1016/j.ecolind.2015.12.042

Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD - a platform for ensemble forecasting of species distributions. Ecography (cop) 32:369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Thuiller W, Araújo MB, Lavorel S (2003) Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. J Veg Sci 14:669–680. https://doi.org/10.1111/j.1654-1103.2003.tb02199.x

Thuiller W, Georges D, Engler R (2012) Biomod2: ensemble platform for species distribution modelling

Tieskens KF, Schulp CJE, Levers C (2017) Characterizing European cultural landscapes: accounting for structure, management intensity and value of agricultural and forest landscapes. Land Use Policy 62:29–39. https://doi.org/10.1016/j.landusepol.2016.12.001

UNEP-WCMC and IUCN (2021) Protected planet: the World Database on Protected Areas (WDPA)

Van Zanten BT, Van Berkel DB, Meentemeyer RK (2016) Continental-scale quantification of landscape values using social media data. Proc Natl Acad Sci U S A 113:12974–12979. https://doi.org/10.1073/pnas.1614158113

van Ree CCDF, van Beukering PJH, Boekestijn J (2017) Geosystem services: a hidden link in ecosystem management. Ecosyst Serv 26:58–69. https://doi.org/10.1016/j.ecoser.2017.05.013

Van Ree CCDF, van Beukering PJH (2016) Geosystem services: a concept in support of sustainable development of the subsurface. Ecosyst Serv 20:30–36. https://doi.org/10.1016/j.ecoser.2016.06.004

Van Berkel DB, Tabrizian P, Dorning MA (2018) Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. Ecosyst Serv 31:326–335. https://doi.org/10.1016/j.ecoser.2018.03.022

Walden-Schreiner C, Rossi SD, Barros A (2018) Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. Ambio 47:781–793. https://doi.org/10.1007/s13280-018-1020-4

Wilson T, Lovelace R, Evans AJ (2019) A path toward the use of trail users' tweets to assess effectiveness of the environmental stewardship scheme: an exploratory analysis of the Pennine Way National Trail. Appl Spat Anal Policy 12:71–99. https://doi.org/10.1007/s12061-016-9201-7

Wood SA, Guerry AD, Silver JM, Lacayo M (2013) Using social media to quantify nature-based tourism and recreation. Sci Rep 3. https://doi.org/10.1038/srep02976

Wu CC, Li CW, Wang WC (2021) Low-impact hiking in natural areas: a study of nature park hikers' negative impacts and on-site leave-no-trace educational program in Taiwan. Environ Impact Assess Rev 87

Viera AJ, Garrett JM (2005) Anthony J. Viera, MD; Joanne M. Garrett, PhD (2005). Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37(5):360–63. Fam Med 37:360–3

Wilcer SR, Larson LR, Hallo JC, Baldwin E (2019) Exploring the diverse motivations of day hikers: implications for hike marketing and management. J Park Recreat Admi. https://doi.org/10.18666/jpra-2019-9176

Yan Y, Schultz M, Zipf A (2019) An exploratory analysis of usability of Flickr tags for land use/land cover attribution. Geo-Spatial Inf Sci 22:12–22. https://doi.org/10.1080/10095020.2018.1560044

Zwoliński Z, Najwer A, Giardino M (2018) Methods for assessing geodiversity. In: Geoheritage. Elsevier, pp. 27–52