University of Southampton

Faculty of Environmental and Life Sciences

Geography and Environmental Science

Social Media, Geodiversity and the Provision of Cultural Ecosystem Services

by

Nathan Fox

ORCID ID 0000-0002-2816-9751

Thesis for the degree of Doctor of Philosophy

July 2021

University of Southampton

<u>Abstract</u>

Faculty of Environmental and Life Sciences

Geography and Environmental Science

<u>Doctor of Philosophy</u>

Social Media, Geodiversity and the Provision of Cultural Ecosystem Services

by

Nathan Fox

To provide a more holistic approach to the conservation of ecosystem services (ES) there is a need to further develop our understanding of how features of biotic and abiotic nature, as well as people and society, interact to provide them. However, the role of geodiversity – the diversity of geology, geomorphology, sediments and soils and hydrology – is overlooked in ES literature and frameworks. Furthermore, geosystem services (GS) – the services that geodiversity provides in isolation of interactions with biotic nature – are also currently underrepresented in ES science.

This thesis will focus on the role of geodiversity in providing cultural ecosystem services (CES), in particular how we interact with geodiversity when undertaking recreational actives. Here, social media datasets from the website Flickr and, for the first time in the field of CES, Reddit are used to assess human-nature interactions through a range of analytical methods including image content analysis, textual sentiment analysis and distribution modelling. The results of these methods contribute to our understanding of both the complex relationship between geodiversity and CES and to the applications of social media data to CES studies.

First, the empirical methods highlight that geodiversity is important at driving both the distribution of CES as well as the positive experience of the activity undertaken. It is demonstrated that geomorphological features, such as topography, and hydrological features, such as coastal waters and lakes, play an important role in determining the distribution and experience of the recreational activity of hiking. The results also highlight the complex relationships between geodiversity and biodiversity features, such as trees and plants, as well as between geodiversity and human-made features, such as trails and roads, in providing CES. The results of these studies can help inform future geoconservation management with the aim of promoting the sustainable use of geodiversity to ensure the future of the ES it provides.

Second, this work advances current uses of social media data by providing novel methods of obtaining data through an accessible R package, *photosearcher*, as well as informing on the best practice for enriching social media datasets. Furthermore, we investigate Reddit as a novel source of data for CES and demonstrate its usefulness in assessing a range of CES.

It is suggested that future work continues to investigate the role of geodiversity on ES, using both social media data and other analytical methods, to better inform the holistic conservation of ES for now and for future generations.

Table of Contents

Table of Tables

Table of Figures

List of Accompanying Materials

photosearcher R package ....The photosearcher R package was peer-reviewed by rOpenSci and is now hosted by them on their GitHub repository https://github.com/ropensci/photosearcher. The open peer review is also available on their GitHub repository https://github.com/ropensci/software-review/issues/325.

Research Thesis: Declaration of Authorship

Print name: NATHAN FOX

Title of thesis: Social Media, Geodiversity and the Provision of Cultural Ecosystem Services

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Fox, N., Graham, L.J., Eigenbrod, F., Bullock, J.M. and Parks, K.E., 2020. Incorporating geodiversity in ecosystem service decisions. Ecosystems and People, 16(1), pp.151-159.

Fox, N., August, T., Mancini, F., Parks, K.E., Eigenbrod, F., Bullock, J.M., Sutter, L. and Graham, L.J., 2020. “photosearcher” package in R: An accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX, 12, p.100624.

Signature: .... Date:...........................

Acknowledgements

I would like to thank my supervisory team Dr Katherine E Parks, Dr Laura J Graham, Professor Felix Eigenbrod and Professor James M Bullock for their guidance throughout the production of this thesis. This work would not have been possible without their combined knowledge and expertise and I am grateful that I have had this amazing opportunity to undertake research with such an amazing team. I would like to express my gratitude for their patience and motivation and for helping me pursue an area of research that continues to be exciting after the last four years of hard work. My entire supervisory team have fostered an incredible environment for me to develop both professionally and personally and I could not have hoped for a better experience.

I would also like to express thanks to my supervisors for providing a platform for me to engage and collaborate with other scientists. In particular, I would also like to thank my collaborators at the UK Centre for Ecology and Hydrology, Dr Tom August and Dr Francesca Mancini for their help and advice on developing the photosearcher R package. Their guidance has helped advance my technical skills and shaped the nature of my research interests.

I am also grateful for all my friends and family who have supported me throughout the last four years, their encouragement and love has greatly helped motivate me, particularly in the difficult times we have faced due to the COVID-19 pandemic. Finally, I would like to say thank you to my partner Victoria for constantly being there for me and providing me with love and support.

Definitions and Abbreviations

ANN ....Artificial Neural Network

API ....Application Planning Interface

AUC ....Area Under the Receiver Operating Characteristic Curve

CICES ....Common International Classification of Ecosystem Services

CTA ....Classification Tree Analysis

ES...Ecosystem Services

FDA......................................Flexible Discriminant Analysis

GAM ....Generalized Additive Model

GBM ....Generalized Boosting Model

GLM.....................................Generalized Linear Model

GS ....Geosystem Services

IPBES ....Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services

MARS...................................Multiple Adaptive Regression Splines

MaxEnt ....Maximum Entropy

MEA.....................................Millennium Ecosystem Assessment

SD ....Standard Deviation

SRE ....Surface Range Envelop

TSS.......................................True Skill Statistic

UK..United Kingdom

UN ....United Nations

UNCED.................................United Nations Conference on Environment and Development

UNESCO...............................United Nations Educational, Scientific and Cultural Organization

USA......................................United States of America

Chapter 1 Introduction

1.1 Overview

This thesis contributes both theoretically and empirically to the understanding of the relationship between geodiversity and cultural ecosystem services (CES), combining novel datasets from social media websites with sources of data for geodiversity. The work here advances the increasing use of social media datasets for assessing CES, first, by creating accessible and reproducible methods of accessing datasets from the website Flickr and second, by demonstrating how the social media website, Reddit can also be used to develop our understanding of CES. This thesis focuses on the role of geodiversity in underpinning recreational ES, assessing different activities in a range of locations through multiple analytical methods including assessing the spatial distribution of social media posts, image content analysis and textual metadata analysis. This chapter outlines the rationale and background of the research undertaken in this thesis and provides an overview of the thesis aims and objectives as well as the structure of the preceding chapters.

1.2 Background and rationale

1.2.1 The need for ecosystem services conservation

One of the most pressing challenges that humanity currently faces is the protection, and potential enhancement, of ecosystem services (ES), the material and non-material benefits gained from the natural environment (Díaz et al. 2018). ES have multiple benefits for human physical and mental wellbeing, as well as having a high economic value (Costanza et al. 1997; Díaz et al. 2015). However, global ES are under threat from both environmental change and anthropogenic activity including over-exploitation, land-use change, urbanisation, biodiversity loss and climate change (Millennium Ecosystem Assessment 2005; Haines-Young and Potschin 2010; Verburg et al. 2012; Kang et al. 2018). These pressures have meant that the supply of global ES is declining (Costanza et al. 2014) and that without intervention will continue to decline (Kubiszewski et al. 2017).

1.2.2 Protecting our abiotic environment

Geodiversity, the abiotic equivalent of biodiversity, is often defined as the range of geological, geomorphological, soil and hydrological features and processes (Gray 2013). As with biodiversity, geodiversity is under threat from a wide range of natural and anthropogenic pressures including urbanisation and climate change (Hjort et al. 2015). Therefore, the concept of geoconservation, the active management of geodiversity in relation to natural processes, was developed in the early 1990s (Sharples, 1993). Geoconservation techniques include the creation of protected areas such as the United Nations Educational, Scientific and Cultural Organization (UNESCO) Global Geoparks Programme (Prosser 2013), statutory legislation (Prosser 2008; Brown et al. 2012) and the ex situ conservation of vulnerable specimens (Gray 2008a). However, the concept of geoconservation is yet to be implemented ubiquitously (Brilha 2017). One potential limitation to the wider adoption of geodiversity and geoconservation is the ambiguity in the concept of geodiversity presented across publications (Erikstad 2013). Improving the clarity in geodiversity theory may help to better translate the science into effective policy and planning decisions for geoconservation (Matthews 2014).

1.2.3 Towards a holistic understanding of ecosystem services

The conservation of valuable ES is impeded by our limited understanding of the ecological, social and geological processes that contribute to them (Kremen 2005; Fisher et al. 2009). ES are delivered and maintained through complex interactions between biodiversity, geodiversity, people and society (Gray 2012; Gordon and Barron 2013; Alahuhta et al. 2018). To better understand this relationship, a plethora of frameworks and models aimed at understanding the role of ecosystems have been developed. However, the majority of these frameworks focus on the role of biodiversity in ES production (Díaz et al. 2018; Haines-Young and Potschin 2018), with few having any focus on geodiversity and ES, (but see Van Ree and van Beukering 2016; van Ree et al. 2017). The under-representation of geodiversity in ES theory and management undermines any attempts at holistic conservation and requires additional research. Increasing our understanding of the role of geodiversity in producing ES and demonstrating the value of geodiversity to society may help to facilitate the protection of geodiversity features and processes (Gordon et al. 2012, 2018).

1.2.4 Improving the assessment of ecosystem services

Biodiversity, geodiversity and social science datasets are key to understanding the relationship between nature, society and ES. However, traditional ecological and social assessments are limited by high financial costs and time-intensive methods (Wood et al. 2013). This means that primary datasets needed for ES assessments are challenging to obtain, even on smaller spatial and temporal scales (Hjort et al. 2012). Because of these limitations, studies often use proxies for ES, such as estimating ES based on land cover type (Eigenbrod et al. 2010). These ES proxies are not always suitable for assessing ES, including assessing multiple services and trade-offs, global scale assessments and unpacking fine-scale trends (Naidoo et al. 2008; Eigenbrod et al. 2010; Stephens et al. 2015). To overcome these challenges, novel sources of primary data, such as citizen science and social media websites, have started to be utilised as sources of primary data for social and environmental variables, in particular, to assess the non-material physical and mental benefits we receive from nature – or CES (Wood et al. 2013; Ghermandi and Sinclair 2019). Here, social media websites provide an inexpensive and quick to collect source of data, which can be applied to assessing CES over a range of spatial and temporal scales (Figueroa-Alfaro and Tang 2017). Previous implementations of social media for assessing CES have started to disentangle the relationship between geodiversity and CES, particularly geomorphology and hydrology (Van Zanten et al. 2016; Van Berkel et al. 2018). Social media data, therefore, present opportunities to assess the role of geodiversity on CES.

1.3 Research aims and objectives

1.3.1 Research aim

Aim: This thesis aims to harness social media datasets to better understand the role of geodiversity in underpinning the supply of cultural ecosystem services.

1.3.2 Research objectives

Objective 1: develop a holistic framework of the relationship between geodiversity and ecosystem services that provides the foundations for analytical assessments.

Objective 2: develop an accessible and reproducible method of returning datasets from social media sites useful for assessing CES.

Objective 3: utilise social media datasets to explore the relationship between geodiversity and a range of cultural ecosystem services.

1.4 Ethics

Once the final design for the methods used in this thesis was complete, they were discussed with the faculty Ethics Champion. As the methods used purely look at aggregated data from social media sites it was concluded that the work did not require a formal assessment using the University's Ethics and Research Governance Online System. The methods and data used within this thesis do however adhere to all ethical guidelines. In particular, we adhere to the UK Data Protection Act 2018 and all seven principles outlined in its General Data Protection Regulation (GDPR):

"(a) processed lawfully, fairly and in a transparent manner in relation to individuals ('lawfulness, fairness and transparency');

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes ('purpose limitation');

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes subject to implementation of the appropriate technical and organisational measures required by the GDPR in order to safeguard the rights and freedoms of individuals ('storage limitation');

(f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')."

1.5 Thesis structure

This thesis has been undertaken in the "paper approach" and is made up of five sequential papers. At the time of submitting this thesis, two have been published, one in "SoftwareX" (Fox et al. 2020a) and one in "Ecosystems and People" (Fox et al. 2020b); two have been accepted for publication in a special issue in "Ecosystem Services"; and one that is under review in "Geoheritage".

Chapter 1 has outlined the rationale and background for the main aims and objectives for the research undertaken in this thesis.

Chapter 2 reviews the current literature surrounding the main themes presented in this thesis, geodiversity and ecosystem services and social media datasets and ecosystem services. The theory and understanding developed in the chapter provide the basis for the research questions and method employed by the following chapters. The literature review is expanded within chapter 3, which explores the position of geodiversity within ecosystem science, policy and management.

Chapter 3 (Paper I). The complex relationship between geodiversity and ecosystem services are often omitted from ecosystem services frameworks, policy and management. This chapter provides a nested set of three frameworks which together give clarity to the definition of geodiversity, whilst operationalising it within the existing “ecosystem services cascade model” (Haines-Young and Potschin 2010). The work here underpins the rest of the chapters by providing the conceptual foundations for the quantitative assessment of the relationship between geodiversity and ecosystem services.

Chapter 4 (Paper II). Although social media websites are becoming increasingly popular amongst cultural ecosystem service researchers, methods used to obtain suitable datasets require relatively advanced coding skills. Coupled with the lack of transparency of data collection methods from current publications, the advanced skills needed to access social media datasets limits non-data scientists from undertaking these assessments. This chapter develops an accessible and reproducible method of obtaining data from the website Flickr in the form of a peer-reviewed package within the R coding environment (R Core Team 2020). This package allows for quick and simple data extraction from Flickr and provides the foundations for the proceeding chapter as the main source of data for assessing the role of geodiversity in ecosystem service delivery and maintenance.

Chapter 5 (Paper III). Standard methods for assessing cultural ecosystem services using social media datasets have started to become established within the literature, with many studies utilising either the geotagged location of posts, image content analysis or textual sentiment analysis. However, the full availability of data from these social media sites is not always suitable for the intended study. This chapter synthesises the various standard methods to demonstrate that a holistic approach utilising multiple methods of analysis can provide a more robust approach to assessing cultural ecosystem services and provides the potential to better understand the role that geodiversity plays.

Chapter 6 (Paper IV). The previous chapters all provide insights into the best methods for harnessing and preparing social media data for ecosystem service assessments. This chapter combines the findings of the previous chapters and applies this to assessing the relationship between geodiversity and recreational activities undertaken in Wales, UK. The paper harnesses data from the social media site Flickr following the methods from **chapter 4**, whilst the data cleansing is informed by the results of **chapter 5**. The data from these sources is then used to empirically assess the relationship between geodiversity and cultural ecosystem services using distribution modelling following the theory established in **chapters 2 and 3**. Finally, the results from this chapter are interpreted from a geoconservation focus to inform the future protection of geodiversity and the services it provides.

Chapter 7 (Paper V). Social media datasets used in cultural ecosystem service assessments have been derived from a range of sources including Flickr, Instagram and Twitter. However, recent changes to access rights for some social media sites, such as Instagram and Twitter, means that it is increasingly difficult to produce comparative research spanning multiple sources of data. This chapter presents the social media website Reddit as a novel source of data for cultural ecosystem services assessments and critically assesses the potential role for its use in future assessments. It is highlighted that the unique discussion-based structure of Reddit can help to untangle the relationship between geodiversity and cultural ecosystem services.

Chapter 8 synthesizes the findings of the previous chapters and discusses the importance of these papers for future research. The novel sources of social media data presented in this thesis have merit for understanding both cultural ecosystem services and their relationship with geodiversity. Future work will be undertaken to ensure that researchers have accessible and reproducible methods of harnessing datasets from sources presented in the thesis.

Chapter 2 Literature review

2.1 Ecosystem services

ES are generally defined as the benefits which individuals, groups and society gain from nature. These services are inherently produced by the interaction of the living and non-living aspects of ecosystems with people and society (Gray 2012; Díaz et al. 2015; Haines-Young and Potschin 2018). Since the inception of the concept, many ways of defining ES have been developed (Fisher et al. 2009). Commonly used definitions of ES include:

- *"the conditions and processes through which natural ecosystems, and the species that make them up, sustain and fulfill human life"* (Daily 1997);
- "*the benefits human populations derive, directly or indirectly, from ecosystem functions*" (Costanza et al. 1997);
- "*the benefits people obtain from ecosystems*" (Millennium Ecosystem Assessment 2005)

Here, ES are defined following the Millennium Ecosystem Assessment (2005) and Common International Classification of Ecosystem Services (CICES) (Haines-Young and Potschin 2012, 2018) and can be categorised into three final service groups: provisioning services, regulating services and cultural services. Previously this definition also included "supporting services" as one of the main categories of ES. Recent updates to ES definitions, recognise that supporting services are not final services with direct benefits to people and therefore are better defined as "intermediate services" (Fisher et al. 2009; Haines-Young and Potschin 2012).

Provisioning services can be defined as the material and energy outputs of ecosystems and include food, fibre, freshwater, construction materials and biofuels (Kandziora et al. 2013; Palacios-Agundez et al. 2015). Regulating services are those that regulate ecological processes, such as carbon storage, microclimate regulation and mitigating surface water runoff (Larondelle et al. 2014; Cortinovis and Geneletti 2019). Cultural services are defined as the non-material benefits and can be further categorised into other services including recreational, aesthetic and spiritual services (Milcu et al. 2013). Intermediate services are not classed as a final service, but are services that underpin ecosystem functions and provide the pathways for provisioning, regulating and cultural services to be delivered and maintained, e.g. the provision of a platform for other services to exist or nutrient cycling (Fisher et al. 2009; Gray 2011). Due to the complex interactions between these services and the benefits and values they provide, the sustainable management of ES could be improved by simultaneously assessing the drivers, relationships and trade-offs across multiple services (Bennett et al. 2009; Spake et al. 2017).

Though ES are generated through complex human-nature interactions, ES literature has generally focused on the role of nature in the delivery of ES, with the human aspects of co-production often being omitted from studies (Fischer and Eastwood 2016). It is therefore acknowledged that ES research, policy and planning should better recognise that CES are co-produced (through the making of things) and co-constructed (through the making of meanings) by people and therefore only arise from the interaction of people with the biophysical environment (Chan et al. 2011; Fischer and Eastwood 2016; Fish et al. 2016). A better understanding of the co-production of these complex human-nature interactions may help to improve sustainable approaches to ES conservation (Bennett et al. 2015; Palomo et al. 2016).

Final ES provide multiple benefits and values to individuals, groups and society as a whole (Small et al. 2017). It has been highlighted that the value, both monetary and non-monetary, of ES benefits differ among groups and individuals and therefore the distinction between benefits and values should be clarified, with benefits relating to any positive change in human physical and mental well-being and values relating to the worth of these benefits (Potschin and Haines-Young 2011; Haines-Young and Potschin 2012). The values most often associated with ES are purely their given economic value, such as the price of goods (Haines-Young and Potschin 2012; Reynaud and Lanzanova 2017). However, ES values can also be measured by their societal and cultural values. In some cases, there is no direct human use of a service (e.g. intermediary services), there is non-consumption of the service (e.g. some cultural services), or the ecosystem provides direct benefits to health and well-being (e.g. consumption of clean freshwater versus dirty freshwater) and therefore the value of these services may be non-monetary (Haines-Young and Potschin 2012; Small et al. 2017). Furthermore, valuing non-material CES is complicated as the perceived benefits and values vary between individuals due to differing cultural and societal beliefs (Daniel et al. 2012; Havinga et al. 2020). This makes the valuation of some ES particularly difficult if assessed solely from an economic viewpoint (Small et al. 2017). ES assessments should therefore also acknowledge other non-monetary values obtained from the service benefit. For example, the metaphysical values of ES (the value gained independently from the experience of an ES), such as, existence value (satisfaction from knowing certain ecosystems or aspects of within it exist) and bequest value (gratification from conserving nature for future generations), are important to consider in ES assessments (Chan et al. 2012).

Globally, ecosystems, along with the ES that they provide, are under threat from a range of anthropogenic and environmental pressures (Millennium Ecosystem Assessment 2005; Haines-Young and Potschin 2010; Verburg et al. 2012; Kang et al. 2018). These pressures have contributed to a global decline in ES supply (Costanza et al. 2014), with case studies of decline being found globally, across all ecosystems. For example, in Niger provisioning ES from wetlands have declined due to oil and gas exploration and water pollution (Adekola and Mitchell 2011); in the North-East Atlantic the provisioning and regulating services provided by marine kelp forests are under threat from climate and non-climate induced stressors (Smale et al. 2013); in Italy urbanisation and intensive agricultural are threatening recreational ES (Schirpke et al. 2018b); and globally services provided by tropical peatlands are threatened by agricultural expansion (Roucoux et al. 2017). Modelled scenarios for global ES demonstrate that policy interventions such as the United Nations' Sustainable Development Goals are necessary to halt and reverse the current trends and enhance ES and human wellbeing (Kubiszewski et al. 2017). It is, therefore, apparent that one of the most important environmental challenges that humanity now faces is to secure the long-term sustainability of ES, whilst protecting and enhancing them for future generations (Díaz et al. 2018).

Because of their high cultural, social and economic values, coupled with the current threats that they face, ES have the potential to be highly influential for decision-making processes (de Groot et al. 2010). Policy and management decisions can be effectively informed through the visualisation of spatio-temporal distribution of services across local, regional, national, continental and global scales (Maes et al. 2012; Kandziora et al. 2013). However, our ability to accurately map ES is limited, with studies usually relying on using proxies or models to map ES; in particular, there is still a large reliance on land-cover as a means of assessing the capacity of an ecosystem to deliver services (Lavorel et al. 2017; Mayer and Woltering 2018; Lautenbach et al. 2019). Using land cover as a proxy to map ES can provide poor estimates for a range of ES including provisioning and cultural services (Eigenbrod et al. 2010). Furthermore, other approaches to mapping ES, such as mapping ecosystem aspects (e.g. individual trees or soil mulch cover) can provide more accurate estimates of the actual measured ES over the estimate generated by land cover proxies (Zhao and Sander 2018; Kearney et al. 2019). Though proxies may show some large-scale ES, they may not be appropriate for influencing smaller local scales at which conservation decisions are more likely to take place (Naidoo et al. 2008). The limitations of mapping ES must be addressed as inaccurate estimations of ES distributions and values may impact the appropriateness of the policy and management decision they inform (Zhao and Sander 2018).

2.2 Geodiversity

One of the first definitions for geodiversity appeared in the literature in 1993 as a response to the 1992 United Nations Conference on Environment and Development (UNCED) and the need for an abiotic equivalent to biodiversity (Serrano and Ruiz-Flaño 2007; Gray 2008a, b). The first English definition for geodiversity was broad in scope, and described it as 'the diversity of Earth features and systems' (Sharples 1993; Gray 2008b). Over time the simple nature of the original definition has been interpreted differently. Variables considered as measures of geodiversity range from the diversity of highly specific measures of landforms (Bailey et al. 2017), to the inclusion of climatic diversity (Parks and Mulligan 2010). Geodiversity can therefore be viewed as an additional form of environmental heterogeneity, representing the abiotic diversity of geology, geomorphology, hydrology and soils (Gray 2004, 2013; Barthlott et al. 2007).

Though the concept of geodiversity is starting to be more widely accepted, particularly in Europe, Australia and the USA, it has yet to be accepted in many countries (Gray 2008a, b; Erikstad 2013; Larwood et al. 2013). A potential barrier to the successful adoption of these concepts globally is the lack of a standardised definition, with definitions for geodiversity varying greatly between publications (Hjort and Luoto 2010; Anderson et al. 2015; Ruban 2017). Giving clarity to the terminology will not only reduce ambiguity in the literature but increase clarity in communicating science to policymakers (Matthews 2014).

In the absence of clarity, geodiversity is used interchangeably with related terms that have different meanings, for instance, geoheritage, which is defined as specific areas with scientifically or culturally significant geology (Serrano and Ruiz-Flaño 2007). Confusion between terms may be due to geoheritage stemming from the 1940s term “geodiversities”, which was used to describe geographic diversities and the links between people, landscapes and culture (Serrano and Ruiz-Flaño 2007; Melelli et al. 2017), and thus it is not analogous to the currently applied concept of geodiversity. A previous review (Gray 2013), of geodiversity definitions present in the literature, from 1996-2008, proposed geodiversity be defined as:

“*Geodiversity: the natural range (diversity) of geological (rocks, minerals, fossils), geomorphological (landforms, topography, physical processes), soil and hydrological features. It included their assemblages, structures, systems and contributions to landscapes.*”

This definition succinctly captures the major components of geodiversity and has been widely viewed as an accepted definition for geodiversity (Erikstad 2013). However, despite the recognition of an accepted definition, since its publication, there remains a continued discrepancy in definitions used across publications (Anderson et al. 2015; Ruban 2017). Articles published after the scope of the original review have continued to confound the issue by continuing to use a variety of definitions for geodiversity (Table 1).

Table 1 Selected geodiversity definitions present in literature from January 2008 to January 2021.

Definition	Key features included	Reference
"A landscape characteristic related to the heterogeneity of the physical properties of the earth surface"	Geomorphology, geology and climate. The paper also highlighted the influence of anthropogenic and biotic factors influencing the physical system.	(Benito-Calvo et al. 2009)
"A measure of environmental resource availability, which includes climate, topography, soils and geology"	Inclusion of climate. Introduces the idea of resource availability. Geomorphology simplified to topography.	(Parks and Mulligan 2010)
"A broad range of geological phenomena constituting the geological heritage"	Restricts geodiversity purely to geological components. Introduces a link to geoheritage.	(Ruban 2010)
"Settings defined by soil and topography"	Restricts geodiversity into two components. Geomorphology simplified to topography.	(Comer et al. 2015)
"The diversity of abiotic terrestrial and hydrological nature, comprising earth surface materials and landforms"	A more specific acknowledgement of surface materials and landforms.	(Bailey et al. 2017)
"The set of abiotic processes and features of Earth's critical zone (lithosphere, atmosphere, hydrosphere and cryosphere)"	Inclusion of atmosphere and cryosphere. Does not specifically mention geomorphology.	(Zarnetske et al. 2019)
"The abiotic diversity of the Earth's surface and sub-surface"	Actively acknowledges sub-surface features. Does not specify what constitutes abiotic.	(Alahuhta et al. 2020)

Although the main components highlighted in the Gray (2013) definition are relevant to some studies, many studies still define geodiversity as only one or a combination of the geodiversity components (Anderson et al. 2015; Melelli et al. 2017). Furthermore, most definitions for geodiversity ignore external drivers of geodiversity, though some include climate diversity (Parks and Mulligan 2010). Geodiversity should not be considered as an isolated science but be viewed as a holistic approach (Gordon et al. 2012). It has further been suggested that geodiversity should be viewed in the context of both endogenous and exogenous systems (Serrano and Ruiz-Flaño 2007).

First, climatic features and processes, including climate-induced weathering and erosion, are crucial to shaping features of geodiversity such as geomorphology, hydrology and soils (Zwoliński 2009). As geodiversity is temporally variable, areas with identical existing geodiversity but different climates would have different geodiversity after a short timeframe. Several articles, therefore, include climate as a core geodiversity definition (Benito-Calvo et al. 2009; Parks and Mulligan 2010), though the place of climate in the definition is still being debated with suggestions that definitions that include it be termed abiotic diversity (Anderson et al. 2015). Here, geodiversity is viewed as physical components only, with climate diversity being viewed as an external driver.

Second, geodiversity is often viewed as the "natural range" (Gray 2013), though this omits important geodiversity features and processes produced through human activity and culture (Benito-Calvo et al. 2009). For example, open pit quarrying impacts all components of geodiversity, including changes to geology, geomorphology and soils through rock excavation and to hydrology through changes to surface and ground waters (Bétard 2013). Anthropogenic features such as abandoned quarries can also harbour interesting geodiversity features, including exposed rock walls, bare surfaces and water bodies (Bétard 2013). Geodiversity definitions should therefore not be limited to natural features and processes but be viewed in relation to anthropogenic activities (Serrano and Ruiz-Flaño 2007; Erikstad 2013).

Third, biotic interactions can influence all components of geodiversity (Benito-Calvo et al., 2009), from sand dune morphology (Hjort et al. 2015) to the production of entire ecosystems through biogenesis (Brocx and Semeniuk 2010). Therefore, biotic interactions are highly influential to the functional diversity of geodiversity components. For instance, soils and their associated ecosystem properties are reliant on interactions with biodiversity, including above-ground plant carbon inputs and soil microorganisms (Schmidt et al. 2011). Areas with shared soil type but differences in plant diversity would have soils with differing functional diversity and therefore different geodiversity. Here, these interactions with biodiversity are acknowledged in the same manner as the interactions of climate and geodiversity, as an external influence and not a key component.

Fourth, a significant factor in determining the inclusion of geodiversity components is the spatial and temporal extent at which it is assessed. For instance, in vascular plant species richness models, geodiversity variables are more important at smaller geographical extents and grain size (Bailey et al. 2017). The scope of geodiversity assessed is therefore often defined by geographic extent (Çetiner et al. 2018), though this coupled with specifically targeted geodiversity components will heavily influence the results and conclusions drawn (Serrano and Ruiz-Flaño 2007). Furthermore, the timescales at which geodiversity changes usually occur over 100s to 1000s of years (Gray et al. 2013), with the assumption that abiotic conditions remain persistent, even under environmental change (Lawler et al. 2015). However, geodiversity can rapidly change due to active processes, anthropogenic activities or environmental change (Hjort et al. 2015). The definition of geodiversity must therefore be fluid and applicable to a range of spatial and temporal scales. Overall, the definition by Gray (2013), provides a good basis for the further development of the concept of geodiversity (Crofts 2014).

2.3 Geodiversity and ecosystem services

Though it is generally acknowledged that ES are generated through the interactions of geodiversity and biodiversity (Gray 2012; Gordon and Barron 2013; Alahuhta et al. 2018), our understanding of these complex interactions is limited (Fisher et al. 2009). The role of geodiversity in the delivery and maintenance of ES can be direct or indirect (Gordon and Barron 2013). The direct role of geodiversity can be described as any interaction that leads to the delivery of a final service, while the indirect role is when geodiversity provides a supporting ES or underpins biodiversity.

First, geodiversity indirectly underpins ES through the generation of biodiversity through habitat diversity, resource availability and niche variety (Parks and Mulligan 2010; Bailey et al. 2017). Through providing the foundations for habitats to exist (Gray 2012), geodiversity features and processes provide the essential supporting services for the production and maintenance of terrestrial, aquatic and marine habitats (Hjort et al. 2015). The influence of geodiversity on habitat creation is not just limited to regional and landscape scales but is also influential at local scales, with smaller geodiversity features providing unique opportunities for biodiversity, often harbouring distinctive communities, for example, caves, waterfalls, talus and desert springs (Hjort et al. 2015). Furthermore, the interactions between the components of geodiversity produce the fundamental resources for biodiversity: energy, water, space and nutrients. The combination of habitat diversity and resource availability provides a range of ecological niches for species to occupy (Parks and Mulligan 2010). Through the provision of different niches, geodiversity, throughout evolutionary history, has influenced functional trait diversity (the functional differences among the species in a community) and phylogenetic diversity (the range of distinct evolutionary histories represented in a community) (Pepper et al. 2013; Cheesman et al. 2018). As functional traits and phylogenetic diversity are important determinants of ecosystem function and in turn ES (Petchey and Gaston 2002; Hooper et al. 2005; Xie et al. 2018; Grab et al. 2019), geodiversity, therefore, indirectly plays a key role in the delivery and maintenance of ES.

Second, though geodiversity is often only regarded as a static background for biodiversity, geodiversity is dynamic with a wide range of active geological, geomorphological, hydrological and soil processes continuously interacting with each other and with biodiversity (Hjort et al. 2015; Lawler et al. 2015). Geodiversity plays an active role in all categories of ES, for example for provisioning services geodiversity can provide marine calcium carbonate needed for coral aquaculture (Barton et al. 2017) and improve food and bioenergy crop yields through hydrological inputs (Brauman et al. 2007); for regulating services geomorphology can interact with vegetation to regulate soil erosion rates (Sun et al. 2014), control stormwater flow (Hallema et al. 2016) and improve drought resistance in vegetation (Dubinin et al. 2021); and for cultural services geodiversity features such as geomorphology and hydrology can contribute to aesthetic qualities, recreational value, cultural heritage and spirituality (Van Berkel and Verburg 2014; Van Zanten et al. 2016; de Almeida Rodrigues et al. 2018; Van Berkel et al. 2018).

As well as providing services through interactions with biodiversity, geodiversity can deliver benefits and values in the absence of any interactions with external factors (Table 2). These “geosystem services” (GS), as with ES, can be provisioning, regulating or cultural (Gray 2011). Provisioning GS includes the provision of fresh groundwater, precious metals and geothermal energy (van Ree et al. 2017). Regulating GS includes the regulation of thermal flows, storage of freshwater in glaciers and flood control such as barrier island (Gray 2011; Van Ree and van Beukering 2016). Cultural GS includes culturally significant sites such as Uluru, Australia or recreational activities such as caving (Van Ree and van Beukering 2016; Gray 2019).

Table 2 Selected examples of provisioning, regulating and cultural geosystem services (Gray 2011; Van Ree and van Beukering 2016; van Ree et al. 2017).

Service category	Geosystem service
Provisioning	Fresh groundwater
	Rare-earth metals
	Geothermal energy
	Construction materials
	Food
	Ornamental materials
Regulating	Geological carbon sequestration
	Marine carbon sequestration
	Thermal storage
	Flood control
	Water quality
Cultural	Spiritual and religious sites
	The cultural heritage of the built environment
	Tourism and leisure
	Artistic inspiration
	Knowledge and education

To understand better the role of nature in ES, a plethora of frameworks linking nature to ES have been created. Many of these frameworks aim to understand the role of biodiversity in delivering ES (Swift et al. 2004; Tallis et al. 2008; Haines-Young and Potschin 2010, 2018; Mace et al. 2012; Díaz et al. 2015, 2018). However, to date, there are relatively few relational frameworks that consciously acknowledge either GS or the role of geodiversity in ES production (Van Ree and van Beukering 2016; van Ree et al. 2017). Some frameworks (Díaz et al. 2015) actively acknowledge their exclusion of GS. Furthermore, some frameworks acknowledge the role of the abiotic environment, but confuse the message by introducing new terminology such as “abiotic services” (Haines-Young and Potschin 2018). The confused position of geodiversity and geosystem services within the ES literature needs greater clarification to better translate the science into robust and suitable policy and management decisions.

2.4 Cultural ecosystem services

As the benefits and values of CES are down to individual differences and varying cultural norms, they are harder to quantify and are therefore comparatively under-researched when compared to provisioning and regulating services (Daniel et al. 2012; Hernández-Morcillo et al. 2013; Dickinson and Hobbs 2017). Therefore, to address this gap, this thesis will primarily focus on the role of geodiversity in providing CES. Several definitions and classifications are used for CES depending on the context of the study (Fish et al. 2016). For example, though recreational activities are regarded as a service in most CES literature (Milcu et al. 2013; King et al. 2017), the framework provided by Common International Classification of Ecosystem Services (CICES) (Haines-Young and Potschin 2018) questions its position as a final CES. Although these differences exist, the common consensus is that CES are the intangible benefits, such as improvements to health and wellbeing, received from interactions between people and the biophysical environment (Fish et al. 2016; Dickinson and Hobbs 2017). Here, we classify CES following Milcu et al. (2013), who reviewed CES literature and provided 11 subcategories of service; recreation and tourism: aesthetic values; spiritual and religious values; educational values; cultural heritage values; bequest; intrinsic and existence; inspiration; sense of place; knowledge systems; social relations; and cultural diversity. Furthermore, it is acknowledged that the benefits from these CES are obtained through a variety of pathways (King et al., 2017). These pathways to benefits can be summaries into six broad categories: cognitive (benefits from the development of knowledge), creative (benefits from influences on aesthetic appreciation and artistic expression), intuitive (benefits from the influence on instincts and senses), retrospective (benefits from reflecting on past experiences), regenerative (benefits from opportunities for recreation, leisure and tourism) and communicative (benefits from social relations, cultural identity, and sense of place) (King et al., 2017). In this sense, a CES can deliver benefits through multiple pathways, for example, hiking can be a CES that directly provides benefits through the restorative pathway (e.g. providing physical benefits), however hiking can also be viewed as a pathway to other CES, such as through the creative pathway (e.g. through providing a means of access to aesthetically pleasing scenery), or through the communicative pathway (e.g. as part of a society or for religious reasons such as pilgrimage) (Wilcer et al. 2019).

2.5 Geodiversity and cultural ecosystem services

Geodiversity plays an important role in underpinning services across the different subcategories of CES suggested by Milcu et al. (2013), either as ES or GS (Gray 2011, 2019; Gordon and Barron 2013; Gray et al. 2013). Geodiversity features can be important for providing cultural and spiritual values. For example, Uluru, a famous landform in Australia, is culturally and spiritually important to local Indigenous Austrailian groups (Hueneke and Baker 2009). Furthermore, smaller geodiversity features such as limestone outcrops and quarries found around the city of Brno, Czech Republic contribute to providing a sense of place for locals (Kubalíková 2020). Geodiversity features can also contribute to local heritage through the built environment. For example, monuments and buildings in Lisbon, Portugal are traditionally built using a distinct whitish limestone, named locally as liós, which as well as providing cultural heritage has influenced artistic expression (da Silva 2019).

Furthermore, features and processes of geodiversity can contribute to educational values and knowledge capital (Gray 2011; Gordon and Barron 2013). For example, information on historic environmental and ecological trends can be extracted from paleosediments (Jeffers et al. 2015). The historic trends and baselines from geodiversity features can benefit environmental monitoring and forecasting, such as the assessment of sea-level change (Gray 2011). Geodiversity also provides a vehicle for institutional education such as geological field trips (Gray 2011).

Geodiversity has a strong relationship with recreational activities and tourism, with some activities requiring certain geodiversity features to enable them to be undertaken. For example, outdoor rock climbing inherently relies on geological and geomorphological landforms and is exemplified by famous natural landmarks such as Devils Tower and El Capitan in the USA which attract a large number of climbers annually (Gray 2012). Water-based sports such as outdoors swimming, kayaking and sailing are all reliant on hydrological systems, both freshwater and marine, to exist. The relationship between geodiversity and other recreational activities, such as hiking and camping, are harder to disentangle. For example, though the value of hiking can be related to geomorphological features such as elevation and slope, this relationship is complex, and the benefits and value of these geodiversity features may be site-specific or a result of variations in an individual's personal preference over time (Chhetri 2015). Furthermore, different user groups, such as novice hikers or those with mobility issues may prefer hiking in areas of flat land without variation in geomorphological features (Moore et al. 1996).

Geodiversity, and the CES it provides, helps to generate “geotourism”, a form of tourism analogous to ecotourism (Burek 2012). Both large scale and local geodiversity features can promote geotourism. For example, iconic geodiversity landscapes such as the Grand Canyon, USA, and international protected areas such as the UNESCO Global Geoparks attract millions of visitors annually (Gray 2008a; Ruban 2017), while local features such as caves and quarries can attract large numbers of both local and international tourists (Garofano and Govoni 2012; Kubalíková 2020). However, many CES including recreational activities and geotourism can cause damage to geodiversity and therefore require suitable conservation techniques to ensure that the CES, both ES and GS, provided by geodiversity can be continued to be used sustainably (Hjort et al. 2015).

2.6 Geoconservation

Geodiversity features and processes are under threat from a range of natural process and anthropogenic activities. Environmental change such as sea-level rise threatens to destroy or submerge geomorphological coastal features including stacks and dunes (Hjort et al. 2015). Highly populated urban areas can directly threaten hydrological systems, such as groundwater contamination from sewage systems leaks (Sorichetta et al. 2013) and geomorphological features such as the destruction of landforms (Hjort et al. 2015). Recreational and tourism activities can lead to in situ damage to geodiversity features including footfall induced erosion, damage to landforms or the removal of geodiversity features such as fossils and rocks (Gray 2008a). These activities not only have in situ impacts but are damaging to wider geodiversity, for example, urbanisation and land-use change can lead to changes in downstream hydrological and geological processes, while off-shore activities such as renewable energy and seabed trawling can alter wider sediment transport processes (Hjort et al. 2015). Furthermore, these threats to geodiversity, in turn, can reduce the capacity of ecosystems to provide ES. For example, contamination of lakes from agricultural run-off can impact the supply of clean drinking water and recreational activities (Reilly et al. 2021); while mining activities not only degrade the availability of geodiversity resources, they can directly impact the supply of ES in the surrounding area (Reverte et al. 2020). Therefore, alongside the development of the concept of geodiversity and in response to the 1992 UNCED and the growing global social and political agenda for environmental protection, the concept of geoconservation started to emerge (Burek and Prosser 2008).

Geoconservation was first introduced by Sharples (1993), who stated that “Geoconservation aims at conserving the diversity of Earth features and systems (‘Geodiversity’) and allowing their ongoing processes to continue to function and evolve in a natural fashion”. Though many variations of this definition evolved over time, the overarching principles of geoconservation remain the same; the active management of geodiversity features and processes (Burek and Prosser 2008; Gray 2018). Burek and Prosser (2008) highlight how the active management strategies associated with geoconservation means that geodiversity is not simply preserved, but instead, it is conserved in relation to natural processes, aligning with the original definition proposed by Sharples (1993). Geoconservation techniques can be undertaken over a range of spatial scales including the creation of physical barriers to prevent anthropogenic damage; the removal and storage of specimens such as fossils; and education, either through institutional learning or in situ notices (Gray 2008a; Henriques et al. 2011).

Though sometimes used interchangeably there is a distinction between the terms geoconservation and “geoheritage”. Geoheritage concerns the features and processes of geodiversity that have significant importance or value worthy of conservation, while geoconservation focuses on the active management of geodiversity and geoheritage sites (Gray 2018). Furthermore, a distinction should be made from the term “geosites” which represent the in situ feature of geodiversity with a high value (Brilha 2016).

The origins and development of geoconservation stem from the UK (Brocx and Semeniuk 2007; Burek and Prosser 2008). In the UK, geoconservation is undertaken through a wide range of instruments including statutory legislation and local conservation groups. First, statutory regulations such as the Wildlife and Countryside Act (1981), the Environmental Protection Act (1990) and the Countryside and Rights of Way Act (2000) set in law the precedent for the conservation of geodiversity (Prosser 2008; Brown et al. 2012). Second, the UK has developed Geodiversity Action Plans, which set out to conserve, enhance and promote geodiversity. These plans exist at a range of administration levels including Local Geodiversity Action Plans and a National Geodiversity Action Plan and are informed through geodiversity audits (Gray 2008b). Third, geodiversity is conserved through a network of different protected areas. At local scales geodiversity is protected by Local Nature Reserves designated for locally important geological features, and Sites of Special Scientific Interest which can be designated for the conservation of locally or regionally important geomorphology and geology and at a national level geodiversity is protected through both National Parks and Areas of Outstanding Natural Beauty, the latter of which aims at conserving all features of a landscape, as well as geological United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage Sites such as the Giant's Causeway (Prosser et al. 2010; Prosser 2013). Furthermore, the UK has created six Global Geoparks under the UNESCO Global Geoparks Programme: the English Riviera; Fforest Fawr; GeoMôn; North Pennines; North-West Highlands; and Shetland UNESCO Global Geoparks (Prosser 2013). These Global Geoparks are designated to promote the sustainable usage of the parks rich and diverse geodiversity through geotourism and education (Henriques and Brilha 2017). Fourth, there is a large number of individuals and groups in the UK who undertake voluntary geoconservation (Burek and Prosser 2008; Brown et al. 2012).

Good geoconservation practice is also being undertaken globally. Geodiversity features can be conserved ex situ, for example, the Natural History Museum of Lesvos Petrified Forest, Greece, a geological based museum that researches, manages and preserves the island's petrified forests and simultaneously provides education opportunities for locals and tourists (Larwood et al. 2013). In the USA, as well as government-led initiatives, charities are buying and protecting areas of high geoheritage value, such as The Nature Conservancy purchasing Egg Mountain, Montana, known for its dinosaur fossil deposits (Gray 2005). In Australia, geoconservation is supported by a varied approach to surveying and cataloguing the country's geodiversity. Starting in the 1960s the Geological Society of Australia began identifying geological monuments; over time, these assessments have expanded and there are now a variety of geoheritage committees undertaking geodiversity audits nationwide ranging from wider inventories and assessments of geoheritage sites to the assessments of individual geodiversity features such as caves and fossils (Cresswell 2019). Australia has also established protected areas such as the UNESCO World Heritage site Shark Bay which is protected for its geology and geomorphology and the UNESCO World Heritage site Uluru which holds cultural and spiritual value for indigenous peoples (Brocx and Semeniuk 2007; Gray 2019). Early examples of geoconservation are also present in Costa Rica where the 1955 Law of the Costa Rican Tourism Board made all areas in a 2km buffer from the centre of all volcano craters National Parks (Quesada-Román and Pérez-Umaña 2020).

Though the importance of geoconservation is starting to become more prominent on an international scale (e.g. through the UNESCO World Heritage Sites or the UNESCO Global Geoparks Programme), there are still barriers to its implementation. For example, the European Union does not have any geoconservation policy, instead, it focuses conservation efforts on biodiversity and ecosystems, which can limit efforts to establish geoconservation activities in European countries (Brilha 2017). Moving forward, geoconservation efforts globally could be improved through demonstrating and communicating the value of geodiversity to a wider audience (Gordon et al. 2012; Larwood et al. 2013). As CES have high social, cultural and economic value, improving our understanding of the role of geodiversity in these services could help promote future geoconservation efforts.

2.7 Social media and cultural ecosystem services

Due to the potential threats to the supply of CES, management decisions need to be better informed through methods that reliably identify, quantify and map CES (Tenerelli et al. 2016; Byczek et al. 2018). However, globally, datasets on ES are sparse, meaning services are often poorly mapped through proxies (Stephens et al. 2015). Furthermore, suitable indicators for valuing CES are underdeveloped (Tenerelli et al. 2016). Currently, the primary sources of data for

CES come from either monetary assessments or surveys of social preference (Tenerelli et al. 2016; Figueroa-Alfaro and Tang 2017). Due to labour-intensive methods and high financial costs, implementing these over large spatial and time scales is not always feasible (Wood et al. 2013; Kim et al. 2019). Moreover, methods need to be able to produce reliable maps as previous CES modelling studies have failed to do so (Byczek et al. 2018). As CES are dynamic, methods need to be able to investigate changes over time (Figueroa-Alfaro and Tang 2017). Therefore, to accurately assess the role of geodiversity in CES delivery there is a need for suitable primary datasets.

To overcome the issues of gathering large spatio-temporal datasets, crowdsourced datasets from social media are rapidly growing as a source of data in the field of environmental science, in particular CES (Ghermandi and Sinclair 2019). Types of social media sites include microblogging sites, such as Twitter and Weibo, and image and video sharing websites, such as Flickr, Panoramio and Instagram, both of which provide novel datasets for assessing CES (Table 3). Social media sites generally provide two main sources of data used for assessing CES, photographic posts and textual posts. Both types of post can contain additional metadata, such as the date and time the post was made or a georeferenced location of where a photograph was taken. This metadata allows researchers to assess who is interacting with CES (user profiles), where they are interacting with CES (georeferenced posts), when they were interacting with CES (post timestamps) and what they are interacting with and why (image and textual contents) (Heikinheimo et al. 2017).

Posts on Twitter have been used to examine human nature interaction in urban green spaces, providing textual information on a range of CES including recreational activities and religious events (Roberts 2017), as well as spatial information on the distribution of CES (Johnson et al. 2019). Photographs on Instagram have been used to compare CES across different ecosystems (Chen et al. 2020). Posts on Flickr have been used to assess CES over a range of different spatial extents including landscapes (Yoshimura and Hiura 2017; Oteros-Rozas et al. 2018), cities (Thiagarajah et al. 2015), regions (Figueroa-Alfaro and Tang 2017; Langemeyer et al. 2018) and continents (Van Zanten et al. 2016; Jeawak et al. 2017). However, due to changes in access to data from some social media sites (e.g. Paramio has ceased operation and Instagram has restricted data availability including blocking the return of location data), Flickr is now generally the most used social media site for valuing CES studies (Langemeyer et al. 2018; Retka et al. 2019).

Lee et al. (2019) identified three categories of methods used to assess social media photographs: those assessing the spatial and temporal distribution of photographs (Tieskens et al. 2017); those that look at the natural features at and surrounding the location of an image (Van Berkel et al. 2018; Hale et al. 2019); and those that examine the contents of the images themselves

(Thiagarajah et al. 2015; Clemente et al. 2019). Although Lee et al. (2019) classified these three methods of assessing photographs as distinct groups, many studies use multiple methods. For example, Richards and Tunçer (2018) assessed the image contents of posts on Flickr, categorised the images and then plotted the spatial distribution of the categories.

Table 3 Selected CES studies utilising data from social media sites. Studies have been selected to demonstrate a range of social media, study focus and study locations.

Source	**Study focus and location**	**Reference**
Flickr – photograph and video sharing website	Wildlife watching; Scotland, UK	(Mancini et al. 2019)
	Visitation rates to recreational sites; global study	(Wood et al. 2013)
	Recreational beneficiaries; Camargue region, France	(Gosal et al. 2019)
	Aesthetic values; Herakleia ad Latmos, Turkey	(Gülçin 2021)
	Change in mangrove CES; Singapore	(Thiagarajah et al. 2015)
	Marine protected area CES; Brazil	(Retka et al. 2019)
Weibo – microblogging website	Urban park visitation rates; Beijing, China	(Zhang and Zhou 2018)
Instagram – photograph and video sharing website	National park visitation rates; South Africa and Finland	(Tenkanen et al. 2017)
	Urban Park CES; New York City, USA	(Johnson et al. 2019)
	Dykeland and marsh CES; Cornwallis River, Canada	(Chen et al. 2020)
Panoramio – photograph sharing website	Human and natural features of high cultural value; Wales, UK	(Gliozzo et al. 2016)
	Aesthetic value; Nebraska, USA	(Figueroa-Alfaro and Tang 2017)
	CES hotspots; Southern Patagonia, Argentina	(Martínez Pastur et al. 2016)
Twitter – microblogging website	Urban park CES; Birmingham, UK	(Roberts 2017)
	The sentimental value of nature; Great Barrier Reef, Australia	(Becken et al. 2017)

Multiple methods of mapping the distribution of social media posts have been used in CES. For example, the spatial distribution of services can be obtained through hotspot and cluster analysis (Gliozzo et al. 2016; Figueroa-Alfaro and Tang 2017). Furthermore, the distributions of CES can be predicted using species distribution modelling methods such as MaxEnt models (Richards and Tunçer 2018; Walden-Schreiner et al. 2018; Arslan and Örücü 2020). Text-based posts on Twitter can also be geolocated and therefore mapped. For example, Johnson (et al. 2019) coded Twitter posts into different types of CES and mapped their distribution across urban greenspace. Mapping the spatial distributions of CES through social media data can help to inform future conservation planning and management (Gliozzo et al. 2016).

There are also multiple methods to assess the local environment where images are taken. For example, Sinclair et al. (2019) assessed how variables such as wetland attributes, human attributes and water quality are related to the visitation rates of wetlands in Kerala, India. Van Berkel et al. (2018) carried out a viewshed analysis of photograph locations and combined this with spatial layers to assess whether natural features such as coastlines of forests were visible from the location of where images on Flickr were taken. Zhang and Zhou (2018) categorised the types of parks that people were visiting into six categories based on the parks setting and its features: recreational parks, cultural relics parks, large urban parks, natural parks, community parks and neighbourhood parks, and used these classifications to assess the drivers of park visitation. Through assessing the natural features in the locations where images are taken, the relationship between biodiversity, geodiversity and CES can be examined.

Studies that classify the contents of images generally use one of two approaches, manual classification and automatic classification. Manual classification of images usually involves researchers examining the images and summarising the image contents as a category of CES (Richards and Friess 2015; Clemente et al. 2019). However, manual content analysis can be subjective, which is particularly problematic for services in which their value is ambiguous and more varied between individuals, such as aesthetic value (Zhang et al. 2020). Clemente et al. (2019) found that for the same images categorised by multiple researchers, services such as cultural heritage and inspiration were the most disagreed upon. Furthermore, manual classification of social media datasets is time-intensive and therefore is less applicable to larger datasets (Lee et al. 2019). Recently, studies have started using machine learning algorithms to categorise image contents including the Google Vision Cloud API (Richards and Tunçer 2018; Gosal et al. 2019) and Clarifai (Karasov et al. 2020).

The textual metadata from posts can be used to classify what CES people are undertaking or used to elicit CES values. For example, Ghermandi et al. (2020) classified CES based on words extracted from social media post titles and tags, while Jeawak et al. (2017) used tags from Flickr photographs to predict aesthetic value across Europe. Furthermore, sentiment analysis can be used on the textual metadata to extract opinions and feelings about nature and CES. For example, Becken et al. (2017) carried out sentiment analysis on Twitter posts to assess feelings to perceptions of the Great Barrier Reef; Wilson et al. (2019) used Twitter posts to analyse the sentiment of visitors to the Pennine Way National Trail, UK; and Do (2019) assessed to the sentiment of social media posts relating to wetlands in South Korea.

Another aspect of CES that can be elicited through social media posts is their temporal distribution. Tenkanen et al. (2017) compared social media posts to official visitation rates and demonstrated similarities in temporal visitation trend for national parks in South Africa and Finland; Walden-Schreiner et al. (2018) mapped the distribution of National Parks in Australia and Argentina; and Sinclair et al. (2020a) used social media posts to assess temporal trends including the hour of the day, the day of the week and the month of the year for visitation to German national parks. Furthermore, social media data can also be combined with sources of historic imagery to assess the changes in cultural values over time (Thiagarajah et al. 2015).

There are however some limitations with social media data. First, accessing data from social media sites requires relatively advanced skills in coding. To access data from websites such as Flickr and Twitter, researchers need to understand and interact with the website's Application Planning Interface (API), an interface that enables researchers to access the software via code. This limits the accessibility of API data to non-data scientists. To overcome this issue, tools such as the Natural Capital INVEST tool (Sharp et al. 2020) have been developed to make assessing social media data for CES studies more accessible. However, the INVEST tool limits the flexibility in searches and analysis, restricting users to specific methods such as only analysing the results through regression analysis.

Second, we need to further develop our understanding of the socio-demographic biases when using social media (Oteros-Rozas et al. 2018). Different social media sites have different communities and spatial distribution in uploads, therefore including data from multiple sources may provide data for a larger and more diverse number of individuals and help to better generalise human-CES interactions (Dunkel 2015; Gliozzo et al. 2016). However, not all social media sites are currently being used to assess CES. For example, although the social news aggregation site Reddit has been used as a source of data in other scientific disciplines (Baumgartner et al. 2020), recent reviews of CES and social media studies indicate that Reddit is currently not being used to assess CES (Ghermandi and Sinclair 2019; Zhang et al. 2020). Future work should therefore aim to increase the accessibility and reproducibility of the gathering of social media data as well as assess the viability of other social media sites such as Reddit as a source of data for CES studies.

2.8 Social media, geodiversity and cultural ecosystem services

Though to date, no study has had a specific focus on using social media datasets to assess the role of geodiversity in the provision and maintenance of CES, some studies have found links between geodiversity features and CES. For example, recreational activities can be influenced by geomorphological qualities. Continental-scale analysis using Flickr photographs found that geomorphological features such as mountainous terrain are related to recreational values (Van Zanten et al. 2016), while regional-scale analysis found a relationship between recreational value and the distance of Flickr photographs from the tops of mountains (Muñoz et al. 2020). Furthermore, studies have shown that hydrological features may contribute to aesthetic qualities. For example, Van Berkel et al. (2018) observed that almost half of the locations of photographs from social media in their study site contained water bodies within the locations viewshed and Tieskens et al. (2017) found that the contents of 40% of social media photographs in their study site contained water bodies. These studies highlight the potential applications for social media to be a powerful tool in disentangling the role of geodiversity in providing CES, particularly recreational activities and aesthetic qualities.

2.9 Conclusion

Geodiversity plays an integral role in the delivery and maintenance of ES, however, our knowledge of its relationship with biodiversity in providing these services is limited. The slow integration of the concept of geodiversity into ES science may be limited by the lack of clarity regarding terminology. **Chapter 3** aims to refine this terminology to enable the translation of the science into guiding practical policy and management decisions. **Chapters 5 and 6** apply these theoretical concepts to empirical studies to understand better the benefits and value of geodiversity to society. However, the mapping and assessment of this relationship is confounded due to the lack of suitable primary ES datasets and poor fitting proxies. In **Chapters 3 and 7** of this thesis, I demonstrate how social media can provide novel datasets for assessing ES, particularly CES. As a rapidly growing field of research, a multitude of methods for assessing CES through social media posts has emerged recently, however, no previous work has used social media posts to assess geodiversity in the context of CES and geoconservation. The results of some studies have begun to highlight the importance of geodiversity features to CES, particularly geomorphology and hydrology. While the results of some studies have highlighted the importance of geodiversity features to CES, particularly geomorphology and hydrology, and a multitude of methods for assessing CES via social media posts have emerged recently, no previous work has used social media posts to assess geodiversity in the context of CES and geoconservation. This thesis attempts to address this gap advancing upon these findings with focused research questions.

Chapter 3 Incorporating geodiversity in ecosystem service decisions

This chapter is presented as a reformatted version of the article published in Ecosystems and People: Fox, N., Graham, L.J., Eigenbrod, F., Bullock, J.M. and Parks, K.E., 2020. Incorporating geodiversity in ecosystem service decisions. Ecosystems and People, 16(1), pp.151-159.

3.1 Abstract

Holistic conservation of ecosystem services (ES) requires a greater understanding of how the interactions of biotic and abiotic aspects of nature provide them. Currently, geodiversity, the diversity of geology, geomorphology, sediments and soils and hydrology, as well as the services that they provide in isolation of interactions with biotic nature – geosystem services (GS) – are overlooked in ES literature and frameworks. Here, we provide a series of three nested frameworks which together help to provide clarity for both the theoretical role of geodiversity in service production as well as the basis for real-world management strategies. First, we present the 'Geodiversity Flower', a framework that can be operationalised to provide clarity in terminology to decision-makers. Second, we present the 'Geo-Eco Services Framework', which establishes the difference between ES and GS. The final framework presented is the 'Geo-Eco Services Cascade Model', which builds upon the widely used ES cascade model by demonstrating how geodiversity interacts with biotic nature to simultaneously provide ES and GS. Providing a holistic model that integrates both biotic and abiotic nature alongside ES and GS allows for a greater understanding of the roles of abiotic and biotic nature to services and their associated benefits and values to people.

3.2 Introduction

In the face of environmental change and the human exploitation of natural resources, there is a need to understand and manage the range of ES that benefit humankind, as well as to conserve the natural features and processes that produce them. Ecosystems are not only defined by a biological community of an area but through the interactions of both biotic and abiotic nature. It is these interactions between the living and non-living aspects of nature that contribute to the delivery and maintenance of ES (Gray 2012; Gordon and Barron 2013). It is increasingly recognised that geodiversity actively contributes to a wide range of ES across all service categories (Prosser 2013). Geodiversity is 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 (Gray 2013; Hjort et al. 2015). Geodiversity plays both direct and indirect roles in the delivery and maintenance of ES (Gordon and Barron 2013). The direct role of geodiversity in ES production occurs when interactions between abiotic and biotic elements of nature produce a final ES, such as the flow of rivers dispersing the seeds of hydrochorous plants (Table 4). Geodiversity also provides 'supporting or intermediate services' which indirectly regulate ES. It can be argued that these supporting services supplied by geodiversity underpin almost all ES, for example, the abiotic elements of soils provide key minerals, nutrients and water required to sustain living things – which illustrates the supporting role of geodiversity in providing the physical platform for the ecological functions that produce ES (Parks and Mulligan 2010; Hjort et al. 2015; van der Meulen et al. 2016). Furthermore, geodiversity, through the creation of a diversity of resources and niches (Parks and Mulligan 2010), passively underpins many aspects of biodiversity including species richness (Hjort et al. 2012; Bailey et al. 2017), functional trait diversity (Cheesman et al. 2018) and phylogenetic diversity (Pepper et al. 2013). These facets of biodiversity in turn underpin ecosystem functioning and services (Edwards et al. 2014). However, both the active and passive roles of geodiversity are generally excluded from ES assessments (Gray 2018).

Table 4 The active role of geodiversity in ES production

Service listed as 'biotic' in (Haines-Young and Potschin 2018)	**The active role of geodiversity in service production and maintenance**	**Reference**
Animals reared by in situ aquaculture for material purposes (Provisioning)	Provision of non-soil nutrients – e.g. marine calcium carbonate required for coral aquaculture	(Barton et al. 2017)
Seed dispersal (Regulating)	Transportation of seeds by geomorphological and hydrological factors – e.g. hydrochorous plant (plants that are dispersed by water i.e. coconuts)	(Araujo Calçada et al. 2015)
Bioremediation by microorganisms, algae, plants, and animals (Regulating)	Temperature regulation of chemical and biological reactions – e.g. heat storage capacity of different soil types	(Miri et al. 2019)
Disease control (Regulating)	Geological and hydrological influences on the epidemiology of diseases e.g. transmission of disease by hydrological systems	(Gordon and Barron 2013)
Characteristics of living systems that enable aesthetic experiences (Cultural)	Provision of different opportunities for aesthetic experiences e.g. wildlife watching facilitated by flat land vs panoramic views facilitated by higher altitudes and rough topography	(de Almeida Rodrigues et al. 2018)
Characteristics of living systems that enable scientific investigation or the creation of traditional ecological knowledge (Cultural)	Understanding historic trends in ES e.g. information on environmental and ecological changes contained in paleo-sediments	(Jeffers et al. 2015)

Geodiversity also contributes to a range of benefits to humans independent of interaction with biotic nature, which are often referred to as geosystem services (GS) or abiotic ES (Gray 2012; Van Ree and van Beukering 2016). There is a wide range of GS, including the provision of rare-earth metals, regulation of thermal flows and sites of cultural significance (Van Ree and van Beukering 2016). However, these benefits are not typically classed as ES due to the absence of any interaction with biotic nature. The absence of biotic interactions should be reflected in their terminology. The term ecosystem is misleading in this context and so we prefer the term GS to abiotic ES. Though the term GS has been used to define subsurface services (services occurring below the pedosphere) (Van Ree and van Beukering 2016), here we define it as all services associated with geodiversity independent of interactions with biotic nature (Gray 2011).

The current biocentric focus of ES literature potentially omits the value of services driven by geodiversity from any economic assessments and management strategies. This issue is exemplified by the plethora of frameworks that link biotic nature with ES (Swift et al. 2004; Tallis et al. 2008; Mace et al. 2012; Díaz et al. 2018), and the relatively few that directly acknowledge the role of geodiversity in ES delivery and maintenance (Van Ree and van Beukering 2016; van Ree et al. 2017; Gray 2018; Potschin-Young et al. 2018). Specifically, no framework yet exists that operationalises the linkages and interactions between geodiversity and biotic nature as well as the interactions of ES and GS within a single framework.

One framework that has included geodiversity (Van Ree and van Beukering 2016) identified the contribution of geodiversity to the provision of GS. However, the framework represents the flow from abiotic components to GS as a separate system, which fundamentally ignores the importance of the interactions between abiotic and biotic components in underpinning ES (e.g. nutrient and habitat provision or biological weathering) as well as trade-offs between ES and GS (e.g. competition for space). The exclusion of GS from ES literature and frameworks means that decision-makers may not be fully informed of the importance of geodiversity in the delivery of these valuable services, nor the trade-offs caused by management decisions (van der Meulen et al. 2016). Here, we introduce three nested concepts which enable the un-packing of the roles of biotic and abiotic nature in service provision and together provide a conceptual framework enabling a more holistic approach to management decisions.

3.3 The geodiversity flower

The first obstacle to including geodiversity into an ES framework is a lack of clarity in geodiversity terminology. Geodiversity is highly multifaceted (Serrano and Ruiz-Flaño 2007; Parks and Mulligan 2010; Gray 2013; Ruban 2014; Bailey et al. 2017), with studies defining geodiversity using different combinations of components. This means terminology surrounding geodiversity may not be transparent to policy-makers and resource managers. We address this issue here with what we call the ‘Geodiversity Flower’ – a framework designed to provide a clear and flexible description of the components of geodiversity (Fig. 1). We represent the structures and processes of geodiversity as petals in the Geodiversity Flower. Intersecting petals indicate the interactions and combinations of two or more geodiversity components. As previous studies have used different combinations of these components to define geodiversity, based on their specific geographic extent and aims (Serrano and Ruiz-Flaño 2007; Bailey et al. 2017), we do not aim to be prescriptive in our definition – our approach is flexible and allows for studies to contextually define geodiversity, whilst providing clarity. Our definitions for the main components of geodiversity are designed to be broader than other frameworks, for example, Hjort et al. (2015), includes topography as a standalone component. Here, we argue that geomorphology can be a broadly applied category that encompasses physical landscape features including topography. Furthermore, to better include other sediments, including marine sediments, we have updated the labelling of soils to sediments and soils. Though the regions of the Geodiversity Flower with the most overlapping segments indicate a more holistic definition of geodiversity, the term can be applied to any of the regions – including those representing single components. To define geodiversity clearly and consistently, the Geodiversity Flower can be operationalised by highlighting which intersections have been used to define geodiversity, thus providing better clarity to policy-makers and resource managers and allowing for better integration of the term into an ES framework.

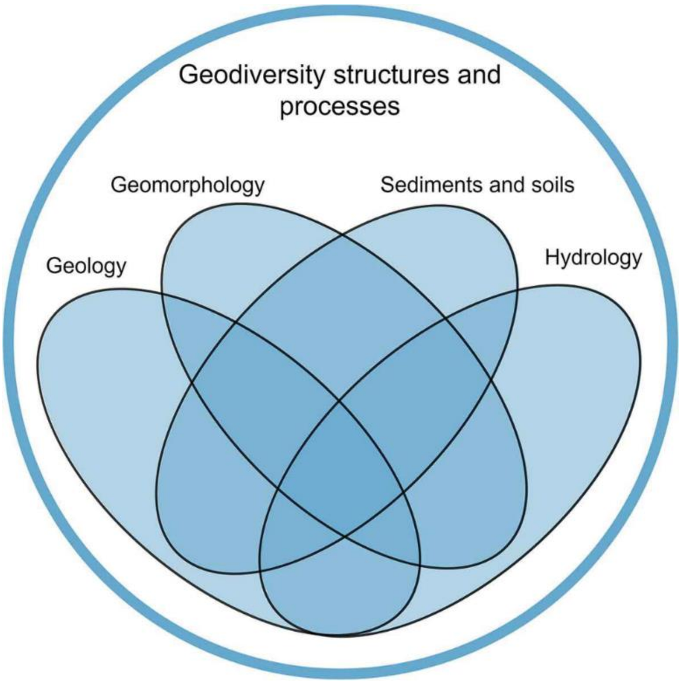


Figure 1 The Geodiversity Flower unpacks geodiversity into petals representing each of its major features, geology, geomorphology, sediments and soils and hydrology. The intersections between petals represent their combinations and interactions, whilst the term geodiversity can be applied to represent the diversity in any single petal or any variation of the intersections of petals.

3.4 The ES-GS framework

Another issue with mainstreaming geodiversity in ES assessment is that the position of geodiversity and GS within current ES science remains confused. The position of GS varies between different ES frameworks. Frameworks, for example, the Millennium Ecosystem, Assessment (Millennium Ecosystem Assessment 2005) and the Common International Classification of Ecosystem Services V5.1 (CICES) (Haines-Young and Potschin 2018)recognise services provided by hydrological features and processes, such as coastal and marine water used as an energy source, as an ES. Conversely, though the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (Díaz et al. 2015) acknowledges that services can be delivered without interaction with living-components of nature (i.e. aquifers and minerals), it does not include these services in its scope and only focuses on services that involve interactions with biotic components. This multifaceted approach to the inclusion of GS by frameworks presents a confusing picture to decision-makers as to whether these services are important or not. Moreover, IPBES and The Economics of Ecosystems and Biodiversity (Kumar 2010) both contain the word ‘biodiversity’ in their names, exemplifying the biocentric viewpoint that biodiversity is the only aspect of nature governing ES.

As well as a varying position between frameworks, GS position remains confused within frameworks. For example, CICES (Haines-Young and Potschin 2018) has started to acknowledge ‘abiotic services’, though only through additional material, and while still maintaining a focus on ‘living processes’. Though CICES now rightly labels services associated with water as ‘abiotic services’, recognising they are primarily driven by hydrology and geomorphology, it elevates services such as surface water for drinking and groundwater used as an energy source alongside ‘biotic services’ and separates them from others driven by geodiversity. For example, the maintenance of soil quality is still classed as a ‘biotic service’ – ignoring the fact that soil quality and quantity are also driven by geodiversity. This is partly due to the role of geodiversity in supporting or intermediary services not being included in the scope of CICES. Moreover, the labelling of ‘abiotic services’ is confusing and does not give clarity to whether these services are being viewed as ES or GS. While CICES highlights that the boundary between abiotic and biotic services is blurred and cannot be defined practically, it goes on to then create a dichotomy between ES and GS. We suggest that this classification of what counts as a service is artificially constructed and inconsistently applied. If services driven by water, an abiotic component, are considered in tandem with ES then why are other abiotic services such as mineral fuels and ornamental materials from geological features disregarded? To provide consistency across all services, GS should be clearly distinguished from ES that are primarily driven by abiotic features and processes.

To address the confusion between GS and ES driven by geodiversity, we present the Geo-Eco Services Framework – a conceptual framework aimed at providing clarity to the differentiation of ES and GS (Fig. 2). In Figure 2a the overlapping segments indicate the interactions between biotic nature and geodiversity where combinations of these interactions can give rise to ES. The non-overlapping geodiversity segment of the Geo-Eco Services Framework can be explicitly labelled as GS, representing services that can be delivered and maintained in the absence of biotic nature. Figure 2b demonstrates that ES can be primarily driven by either abiotic or biotic features and processes, with ES primarily driven by geodiversity being fundamentally different to GS as they require interactions with biotic nature to be delivered and maintained. Here, we note that there is no real distinction between ES that are primarily driven by geodiversity and those that are primarily driven by biotic nature and both should be equally viewed as ES. The value in highlighting this difference are the implications for better targeting management strategies – i.e. if an ES is mainly influenced by abiotic elements, management strategies should have a greater focus on these elements over biotic nature.

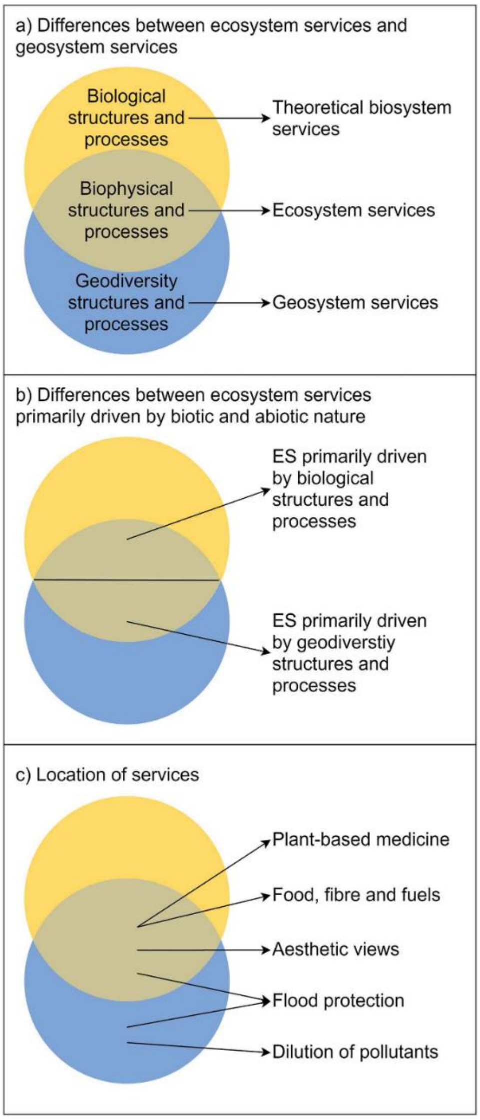


Figure 2 The Geo-Eco Services Framework, a) the different aspects of nature contributing to ES and GS, b) the difference between biotic driven ES and geodiversity driven ES, c) the hypothetical location of a range of services. While biosystem services presented in panel a are theoretically possible, the authors can think of no realised examples for panel c.

The Geo-Eco Services Framework further highlights that geodiversity (either directly or indirectly), is fundamental for ES delivery and maintenance. We include 'biosystem services' as a theoretical concept only – though these standalone biotic services could exist, we would argue that there is no real-world system in which geodiversity does not in some way directly or indirectly impact on biotic services. For example, ES such as medicinal materials from plants are mainly driven by biotic features and processes, however, geodiversity still plays an integral role in the delivery of such services by providing necessary supporting services such as water and nutrients. In contrast, GS such as construction materials and hydroelectric power can exist in the absence of interaction with biotic nature. However, because of the complexity in the interactions between abiotic and biotic nature in real-world systems, we acknowledge that most GS will be in some way influenced by biotic nature. Moreover, depending on the interactions that give rise to the service some services can be classified either as an ES or a GS (Figure 2c). For example, flood protection could be provided as a GS through coastal geology and geomorphology reducing wave action or flood protection could be an ES, delivered through the interactions of river geomorphology and riparian vegetation to slow the flow of a river.

3.5 The geo-eco services cascade model

The seminal ES 'cascade model' framework introduced by (Haines-Young and Potschin 2010)demonstrates the relationship between biophysical structures and processes, ecosystem function, ecosystem service, benefit and value acting as a production chain, stepwise linking them in a cascade (de Groot et al. 2010; Potschin and Haines-Young 2011; Maes et al. 2012). The cascade model attempts to acknowledge the role of geodiversity and places biophysical structures and processes as the drivers of ES, however, the overarching theme places emphasis on the role of biotic nature. This is confounded by a lack of clarity in the terminology of biophysical structures and process which is often only associated with biodiversity (La Notte et al. 2017). By incorporating biophysical structures, the original cascade model and the updated versions that start to incorporate geodiversity and GS (Van Ree and van Beukering 2016; Potschin-Young et al. 2018) provide a good foundation for the holistic integration of both geodiversity and GS within an ES framework.

Here, we include geodiversity in the cascade model by considering its interactions with biotic nature as the primary driver of the multiple cascades, both ES and GS (Fig. 3a). Our Geo-Eco Services Cascade Model provides an organising structure helping to clarify the role of geodiversity in ES and GS production (Potschin-Young et al. 2018). By acknowledging ES and GS in tandem, our integrated services framework can also be applied to assess the trade-offs between multiple services (Lin et al. 2018). Though here the basic framework is displayed by three parallel cascades, we acknowledge that the system is neither linear nor isolated and that ecosystem functions, services and benefits from separate cascades may all interact with each other (see Fig. 3b). Figure 3b does not include all possible interactions, but instead acts as an illustrative example of the potential application of the Geo-Eco Services Cascade Model rather than an exhaustive map of the pathways by which services are affected by upstream gravel extraction from a mangrove system – we unpack this further below (for further example applications see Appendix A). To allow for a completely holistic approach, in Figure 3a we have again included the theoretical 'biosystem services', however in the real-world example in Figure 3b this pathway does not exist.

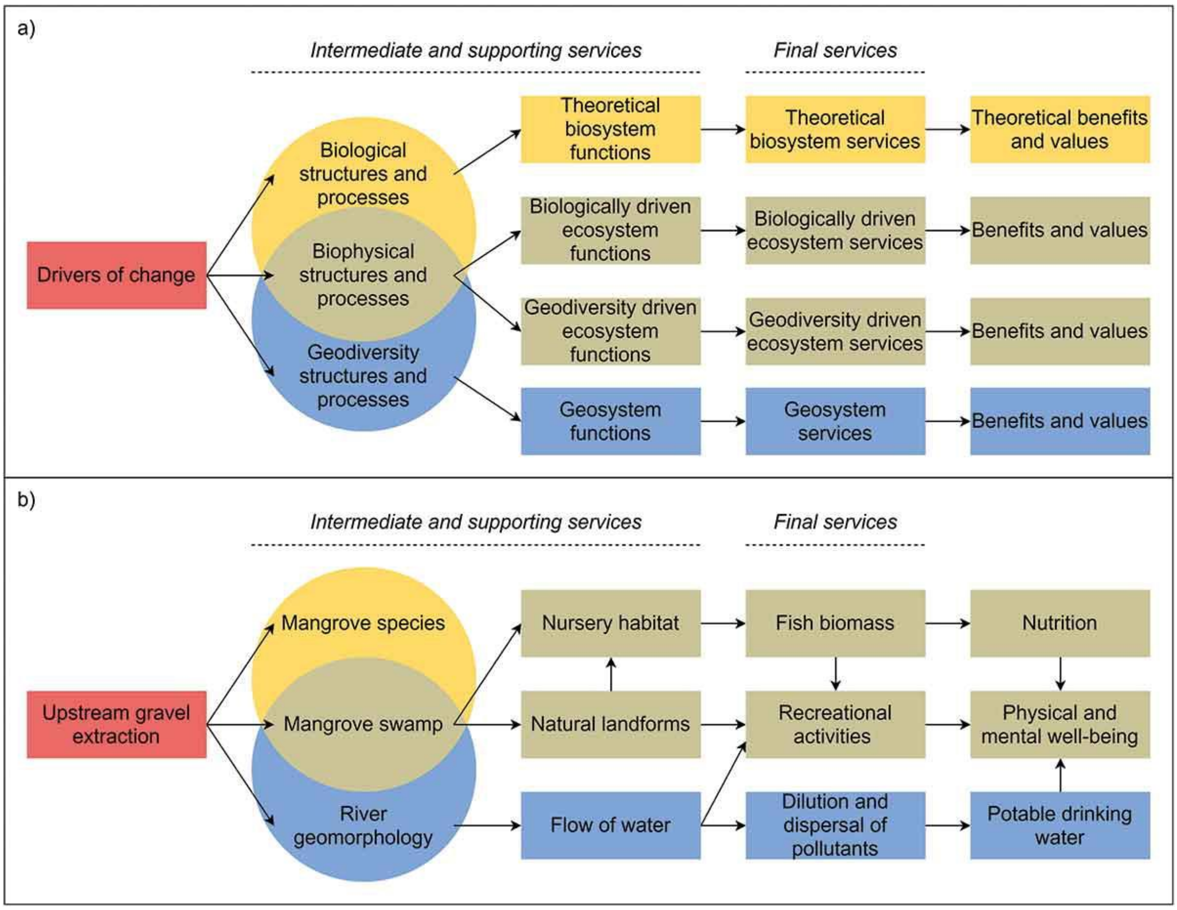


Figure 3 The Geo-Eco Services Cascade Model, (after Haines-Young and Potschin 2010), a) proposed updates to the cascade model, demonstrating the flow of both ES and GS services from the interactions of abiotic and biotic nature, b) application of the framework to a real-world ecosystem – mangrove swamps. We note that the applied framework does not represent the whole system, but instead represents some of the potential interactions and services

Our holistic framework, therefore, provides the conceptual foundations for hypothesis testing and quantitative assessment of the roles of abiotic and biotic components in service provision using methods such as generalized additive models (Alahuhta et al. 2018), mathematical equations (Maseyk et al. 2017), structural equation models (Deru et al. 2018) and Bayesian Belief Networks (Landuyt et al. 2013). Furthermore, as our framework allows for multiple interconnected cascades, we have also updated the previously labelled 'pressures' box (Haines-Young and Potschin 2010), to represent 'drivers of change', as a threat to one cascade may be beneficial to another. By updating the original cascade model to include drivers of change for multiple services the framework can be employed as an analytical framework aimed at guiding testable hypotheses about different management strategies (Potschin-Young et al. 2018; Spake et al. 2019).

3.6 Application of the framework for ES management

A concern with ES management is that some components of nature are unmanageable (Maseyk et al. 2017). Some abiotic aspects of soils, such as the physical processes governing the weathering of bedrock, are not practically manageable. However other aspects, such as soil mineral and water content, can be easily altered (Maseyk et al. 2017). Moreover, aspects of geomorphology can be manipulated for ES management, such as agricultural terracing to reduce soil erosion and water conservation (Ponette-González et al. 2015). The issue of unmanageable aspects is not unique to geodiversity and is shared by biotic elements such as trophic interactions. Therefore, like biotic nature, the trade-offs in both the manageable and unmanageable aspect of geodiversity need to be considered in ES management decisions (Maseyk et al. 2017; Spake et al. 2019).

Another challenge to managing GS and ES are the temporal differences between the replenishment of GS and ES. Generally, services that are primarily driven by abiotic nature occur over longer timescales than those that are primarily driven by biotic nature (Gray et al. 2013) (Fig. 4). This is because the formation of geodiversity components vs biological components that are then drawn on to form the service may take a long time due to difference in geological and biological timescales. However, the use of ES and GS occurs at the same or similar socio-economic timescales. When services fall under both the GS and ES definition (for example, ornamental materials may be an ES, e.g. corals, or a pure GS, e.g. semi-precious stones), and the GS is non-renewable on a human time scale (centuries, decades or less) they tend to be disregarded from ecological assessments (Gray et al. 2013; Brilha et al. 2018), arguably making their preservation even more important. However, the renewability of biotic features and processes providing services can be just as at risk as abiotic features and processes. For instance, the renewability of agricultural products could be diminished if we do not sustainably manage underlying supporting services, such as soil quality (Crenna et al. 2018). Both the economic and environmental trade-offs of prioritising short-scale services over relatively longer scale services should therefore be appropriately considered during ES decision making.

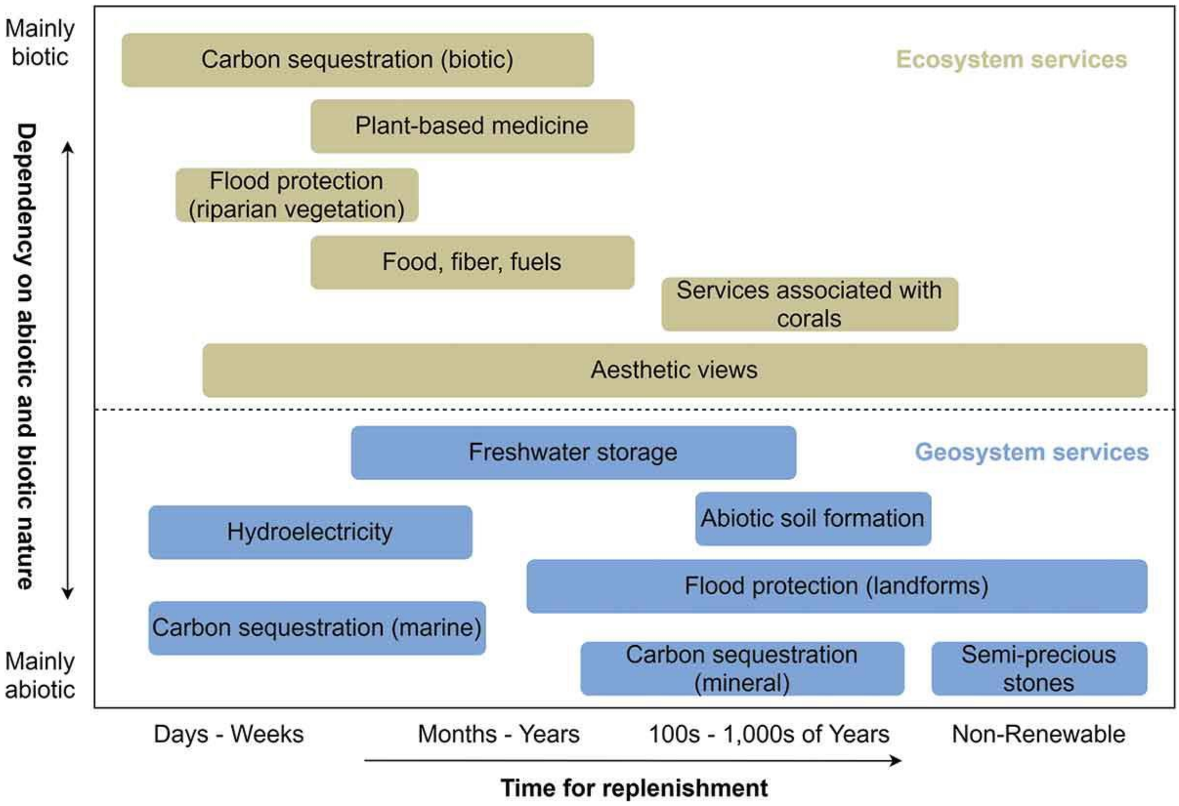


Figure 4 Time taken for the replenishment of a range of ES and GS and their dependencies on abiotic nature. Here, services are placed on a general timescale, but these timescales may vary

Figure 3b demonstrates how the application of the Geo-Eco Services Cascade Model for assessing a real-world ecosystem, mangrove swamps, can overcome both challenges. The mangrove swamp ecosystem is comprised of manageable and unmanageable aspects of nature as well as their associated services which occur over a range of temporal and spatial scales. In a real-world application, the cascades are not as linear as the blank framework template. The geosystem function ‘flow of water’ not only provides its own separate GS of ‘dilution and dispersal of pollutants’, by diluting the concentration of pollution, regardless of any biotic processes, but also contributes to an ES of recreational activities, i.e. boating and swimming. Furthermore, services can interact, with fish biomass (facilitated by fish population or diversity) interacting with boating (facilitated by navigable bodies of water) to present additional opportunities for recreational activities such as fishing. By including drivers of change, the framework can be utilised to assess the potential impacts of proposed anthropogenic activities on the provision of multiple services. For example, here the Geo-Eco Services Cascade Model allows for consideration of both the direct and indirect impacts of upstream gravel extraction on the entire system from biological, geodiversity and biophysical structures and processes to ES and GS and their benefits and values. With the addition of empirical data, the Geo-Eco Services Cascade Model could be implemented to assess how short-term increases in sediment transportation from gravel extraction may impact mangrove species growth rates and the ES they provide (Noor et al. 2015), or how long-term decreases in sediment transportation could alter the geomorphology and natural landforms and thus long-term ES and GS delivery and maintenance.

3.7 Moving forward: applying the framework

As geodiversity and GS are omitted from most ES literature and frameworks, ES policy and decisions often place focus on the management of living systems. This current fragmentation of management and policy impedes efforts to halt and reverse declines in ES. Without consideration of GS, there is the risk that ES management will generate conflicts between ES and GS, such as the destruction of natural landforms by human-constructed flood defences. Therefore, ES management decisions need to incorporate the role of geodiversity in the delivery and management of both ES and GS. By providing consistency and clarity to geodiversity terminology through the Geodiversity Flower, geodiversity can be better integrated into ES science, policy and management in a more transparent manner.

By taking account of abiotic and biotic nature, as interacting components, our work builds upon the successful cascade framework and provides a novel update that acts as an organising framework, providing clarity regarding the conceptual role of geodiversity in ES production. Through the provision of an organising structure, our updated cascade model also provides the foundations for empirical studies evaluating the relationship of abiotic and biotic nature to ecosystem function, service, benefit and value. By updating threats to include all drivers of change, both positive and negative, as well as providing scope for multiple cascades for different ES and GS, our Geo-Eco Services Cascade Model allows for a greater understanding of trade-offs. Utilising our framework allows for stakeholders to empirically test the impact of proposed service use and management strategies on multiple services. We recommend that future studies build upon this framework using empirical methods to advance our knowledge of the functional links between abiotic and biotic nature, and socio-economic and socio-cultural systems.

Chapter 4 "photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr

This chapter is presented as a reformatted version of the article published in SoftwareX: Fox, N., August, T., Mancini, F., Parks, K.E., Eigenbrod, F., Bullock, J.M., Sutter, L. and Graham, L.J., 2020. "photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX, 12, p.100624.

4.1 Abstract

The social media website Flickr contains a wealth of spatial and temporal metadata, which can play an important role in environmental research including cultural ecosystem service and ecological assessments. However, the uptake of Flickr is potentially limited by issues with accessibility to the Flickr Application Planning Interface (API), which limits results and restricts searches. Here, we introduce photosearcher, an R package aimed at overcoming these challenges. We provide examples of how photosearcher can be used as an accessible and reproducible method of accessing large spatio-temporal datasets from the Flickr API.

4.2 Motivation and significance

4.2.1 Scientific motivation

Biodiversity and social science datasets are key to many areas of environmental research, from understanding species distributions to guiding the management of CES. Social media sites such as Flickr, Facebook, Twitter and Instagram and other online sites such as Wikipedia are becoming recognized as potential sources of data for, not only for cultural ecosystem service assessments but also increasingly for ecological questions (Li et al. 2013; Wood et al. 2013; Byczek et al. 2018; Mittermeier et al. 2019). First, due to high financial costs, time-intensive methods and logistical difficulties, biological datasets are often limited or incomplete across even small spatial scales (Hjort et al. 2012; Wetzel et al. 2018). By overcoming many of the limitations of extensive large-scale surveys, social media sites can provide large spatio-temporal datasets (Van Zanten et al. 2016; Kim et al. 2019). Furthermore, as natural ecosystems and protected areas are at risk from overexploitation by people, understanding visitation rates can be useful for proactive conservation (Hadwen et al. 2007). Here, social and demographic data provided by social media sites can represent actual visitation rates (Wood et al. 2013), which present opportunities to understand how humans interact with nature and how best to inform management choices relating to conservation and ecotourism.

Here, we develop an approach to accessing data from Flickr (Flickr 2021), an image and video hosting site with a large database of photographs accompanied by accessible metadata. Flickr has advantages as a source of data as it has an active user base with up to 25 million new uploads a day (Ding and Fan 2019) and generally a wider demographic of users than other social media sites (Oteros-Rozas et al. 2018). Furthermore, the photograph's metadata can be obtained by making calls to the Application Planning Interface (API), an interface for accessing the Flickr server. This metadata usually contains a georeferenced location as well as the time and date the image was taken and has the potential to be used as a primary source of data for answering ecological questions. Flickr has already been successfully used in cultural services studies, such as wildlife watching (Mancini et al. 2019), recreational activities (Graham and Eigenbrod 2019), landscape aesthetic qualities (Figueroa-Alfaro and Tang 2017), and visitation rates in both protected areas (Kim et al. 2019) and national parks (Tenkanen et al. 2017). Additionally, Flickr has vast potential as a source of biodiversity data (Barve 2014). For example, it has been demonstrated as a successful tool for cross-validating Global Biodiversity Information Facility records (Wittich et al. 2018) and assessing ecological niches (Penã-Aguilera et al. 2019). It has been suggested that Flickr could be utilised to explore not just CES, but wider ecological questions at a large scale (Richards and Tunçer 2018). However, due to some limitations, the potential of Flickr as a source of data for a wider range of studies has yet to be fully explored.

4.2.2 Current limitations

Flickr has specific limitations that need to be addressed when using it as a data source. For example, searching for photographs for a given spatial location is restricted to searching via either a bounding box or a Flickr specific location identifier. This has meant researchers have added additional steps to data manipulation to download image metadata for specific search boundaries (Lee et al. 2019). Furthermore, searches for photographs through the Flickr API will only return 4,000 unique results per search criterion, limiting the ability to access data easily for spatially or temporally large searches. For searches that have more than 4,000 results, the API will appear to get metadata for all of them. However, the Flickr API only returns data for the first 4,000 images, after this the following pages of data are duplicates of the first 4,000. This means users can obtain what appears to be more than 4,000 results but end up having only the metadata for the first 4,000 unique images repeated multiple times. Some authors have limited their number of returns per query to fewer than 4,000 to get around this (Van Zanten et al. 2016). This workaround potentially omits the full range of data available and introduces biases, such as excluding early or new users of Flickr, or missing temporal patterns. Furthermore, the use of the API currently has limited accessibility and reproducibility. First, the API can only be accessed through a range of programming languages including Python, R and Java. To access datasets authors must be well versed in a programming language. Within R (R Core Team 2020)there is a set of generic packages that allow harvesting data through APIs. However, researchers who want to use these packages need to have an extensive understanding of the Flickr API as well as the numerous R packages needed to call to it. Second, authors rarely provide complete methodologies or their code, limiting the ability to replicate studies. To increase the uptake of Flickr as a source of data, there is a need for an application that makes API calls more reproducible and more accessible to all.

4.2.3 Related work

The use of an R package for making calls to the Flickr API improves the reproducibility of studies using this data as well as giving users control over what they search. The existing R package "FlickrAPI" (cran.r-project.org/web/packages/FlickrAPI/index.html) provides some limited functionality of the Flickr API within the R environment. Other tools such as the Natural Capital Projects INVEST Recreational Tool (Sharp et al. 2020) have also been developed to query the Flickr API. However, the FlickrAPI package only provides functions for obtaining information for a single known image and the INVEST tool only returns all images for an area. These tools do not provide functionality for searching based on criteria such as keywords or location. This, therefore, limits the functionality of these tools for ecological studies, which often require spatially explicit searches based on keywords, such as a target species. Furthermore, neither the FlickrAPI package nor the INVEST tool provides users with the functionality to download the raw images or return demographic data about Flickr users.

4.3 Software description

4.3.1 Software architecture

To overcome the challenges of using the Flickr API, we have developed the photosearcher R package (github.com/ropensci/photosearcher), aimed at facilitating reproducible requests to the Flickr API. The functions in this package make calls to the Flickr API and return both the raw photographs and their additional metadata in accessible formats, whilst overcoming the current limitations of larger spatial and temporal requests to the API.

4.3.2 Software functionalities

The photosearcher package provides a reproducible way of accessing geotagged photographs through search queries as well as several other functions that provide datasets useful for a range of ecological analysis. The photo_search function allows users to define a set of search criteria, which are then queried against the Flickr database. A data frame containing the metadata for the photographs matching the search criteria is then returned. To enable the use of Flickr across different disciplines, the photo_search argument text allows for searches to be defined by keywords. Searches for images will then only return photos that contain the keywords in their title, description or tags. Users can also limit the searches to find keywords in the photographs' tags only. As well as keywords, other search variables include minimum and maximum date the photograph was taken and a search location, provided as a bounding box, spatial layer or a Flickr specific location (where on earth identifier — woeid see: flickr.com/places/info/24865675). The ability to refine search parameters allows for a more focused approach to using Flickr's geotagged photographs by only returning those relevant to the study. The package also provides additional functionality for downloading images, getting user information and assessing related tags.

4.4 Illustrative examples

4.4.1 Spatial distribution and drivers of recreational cultural ecosystem services

The photo_search function returns a wealth of spatial, temporal and textural metadata. Here, we demonstrate the applications of this data by assessing recreational CES, by searching for photographs of hiking in the contiguous USA. We then used the results of this search with the user_info function to obtain social information on each of Flickr users. The general code is as follows (for a reproducible document Appendix B):

area_photos <- photo_search(mindate_taken = "2015-01-01", maxdaten_taken = "2020-01-01", maxdate_uploaded = "2020-01-01", sf_layer = USAboundaries:: us_states())

social_data <- user_info(user_id area_photos$owner)

The photo_search function returned 160,923 photographs for hiking in the USA between 2015 and 2020 in 61 minutes (Fig. 5). To return metadata for this large number of photographs the bare minimum number of necessary calls to the Flickr API would be 644 (250 photographs per search). The photo_search function, therefore, makes a minimum of 10.55 calls per minute to the API returning metadata for approximately 2,637 photographs (NB in order to minimize errors the photo_search makes more than the minimum number of calls).

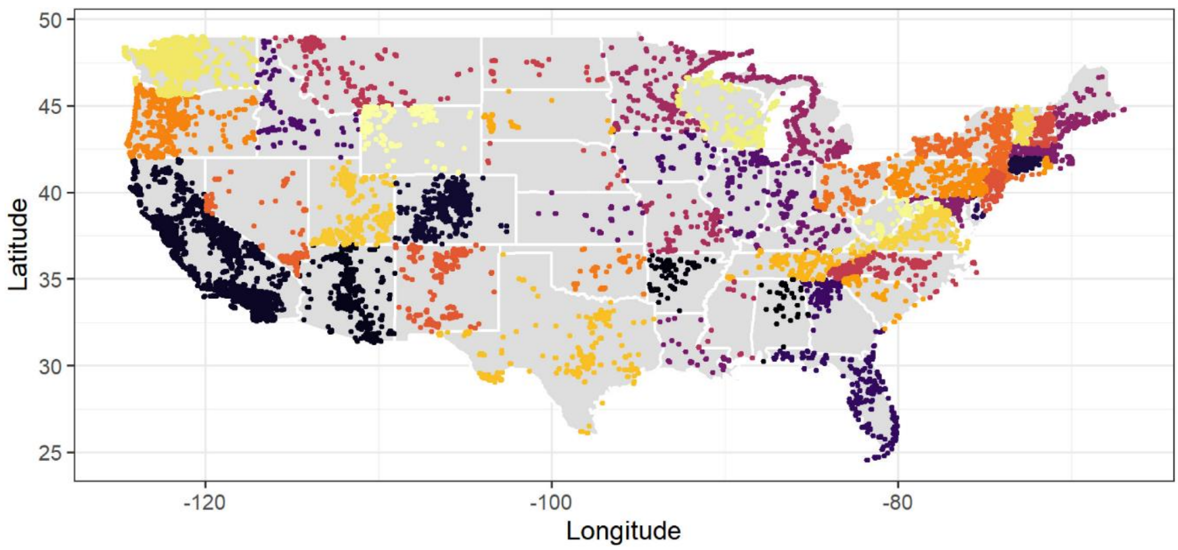


Figure 5 Flickr photographs containing the word hiking in its title description or tag, 2015–2020 (points are coloured by state).

Like the photo_search function, user_info typically returns large social datasets in short periods of time. Here, the user_info function took just under 24 minutes to return information on 6,514 individuals, about 271 users per minute. Normally, to get a users' information you have to make a new call to the API for each individual, however, the user_info function allows searches for multiple users at once, returning all available social data including hometown and occupation. The user_info function, therefore, provides an efficient method for obtaining large social datasets. Potential uses for the city datasets include network analysis to track travel route as well as to understand the social-economic drivers of supply and demand for CES. By being able to assess rapidly where visitors travel from, protected area managers can inform visitor management plans. The social datasets could also be combined with ecological datasets for studies such as understanding human-wildlife interactions or ecotourism management. The hometown information can be plotted by geocoding their location with functions such as geocode_OSM function in the tmap R package (cran.r-project.org/web/packages/tmap/) (Fig. 6).

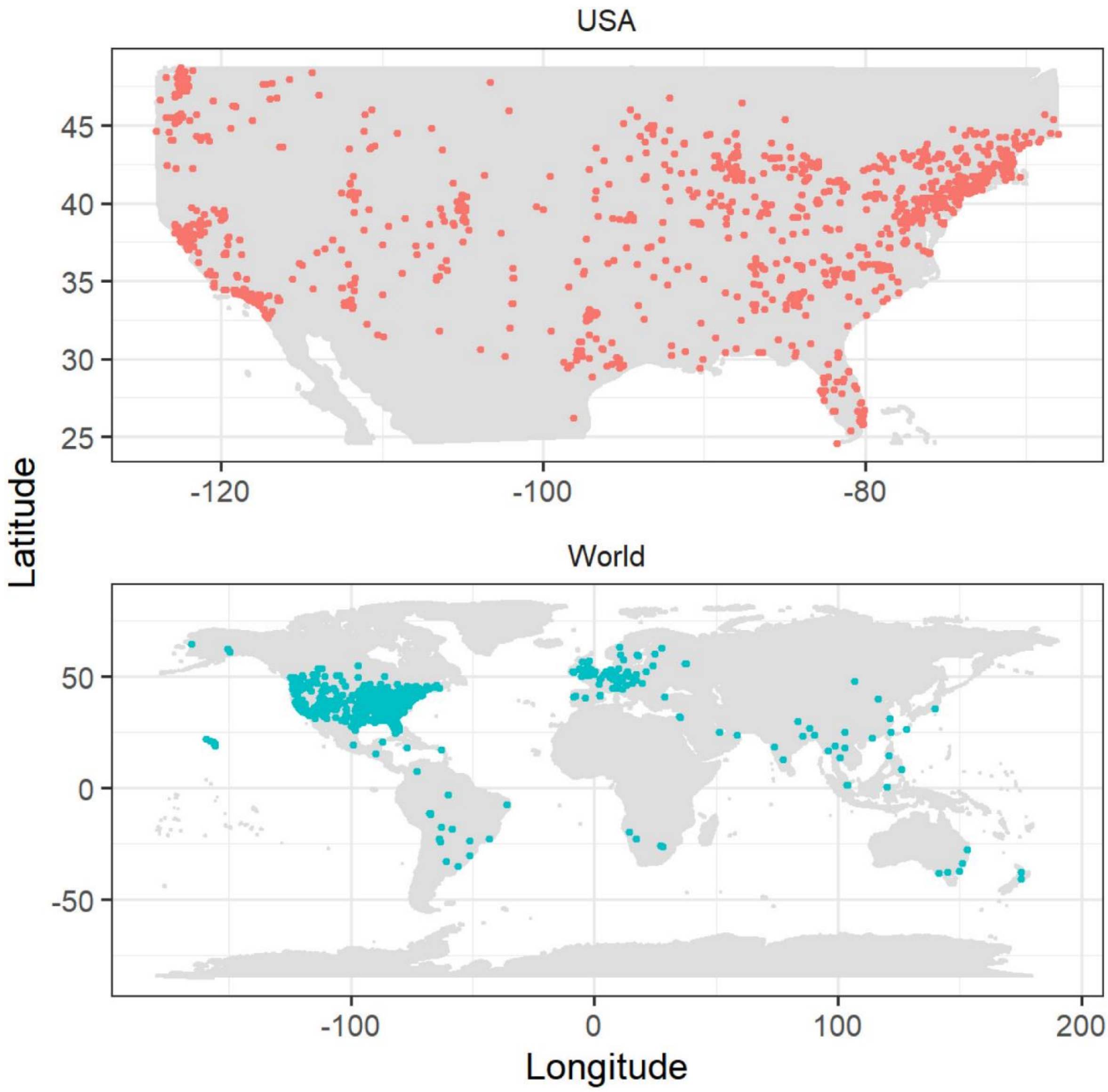


Figure 6 Geocoded hometowns of people undertaking hiking and posting to Flickr in the USA, 2015–2020.

4.4.2 Spatial and temporal distribution of species

To demonstrate the ease of using the photosearcher package for obtaining large ecological datasets, we utilise the photo_search function to find images metadata containing either the common or Latin names of a number of species (Table 5). The Flickr metadata can contain the complete date and time data, allowing for the investigation of temporal distributions such as migratory patterns, diurnal cycles and floral phenology. Flickr may be best suited to large charismatic species that are easily identifiable by the public, such as some birds (Barve 2014). The following piece of code outlines the basic search used (for a reproducible document see Appendix B).

species_name <- photo_search(mindate_taken "2000-01-01", maxdate_taken "2020-01-01", maxdate_uploaded "2020-01-01", text   species common or Latin name , has_geo TRUE)

Table 5 Search terms used, the number of results, time taken to search for the results, and number of API calls needed.

Species	**Text search**	**Results returned**	**Time**
Barn owl — *Tyto alba*	Common name	17,436	10.14 minutes
	Latin name	3,529	1.12 minutes
Red fox — *Vulpes Vulpes*	Common name	25,225	14.14 minutes
	Latin name	7,793	3.14 minutes
Brown bear — *Ursus arctos*	Common name	21,555	10.40 minutes
	Latin name	5,170	1.53 minutes

The photo_search function was able to return large datasets in short periods of time — i.e. returning 25,225 unique geotagged data points globally for the red fox in just over 14 minutes. These results reiterate that generally, this method does not result in exceptionally long search times. Furthermore, the results demonstrate that large spatial and temporal searches would require a large number of API calls, for example, a global study for barn owls would require 70 calls to the API and searches for brown bears would require 87. As *has_geo = TRUE*, the returned metadata contains latitude and longitude information, here we map the distributions of the photographs tagged with species names (Fig.7). Users should be aware that species distributions based on Flickr photographs may have erroneous points. First, Flickr users may misidentify species. To overcome the issue of mistagged images users should properly define their search criteria i.e. using the Latin name, or with a shapefile of its known distribution, or users can use classification techniques to confirm which photographs have positive sightings. Second, some distributions can be influenced by visitor attractions such as zoos and museums. These erroneous points can be removed using the CoordinateCleaner R package (cran.r-project.org/web/packages/CoordinateCleaner/index.html). Furthermore, the temporal metadata can be used to assess change in species over time (Fig. 8). Here we demonstrate that sightings of brown bears vary monthly, with fewer sightings occurring during periods of known hibernation. This temporal metadata could be combined with the spatial data to assess migratory patterns, or with photograph contents (accessible via the download_images function) to assess animal behaviour or plant phenology.

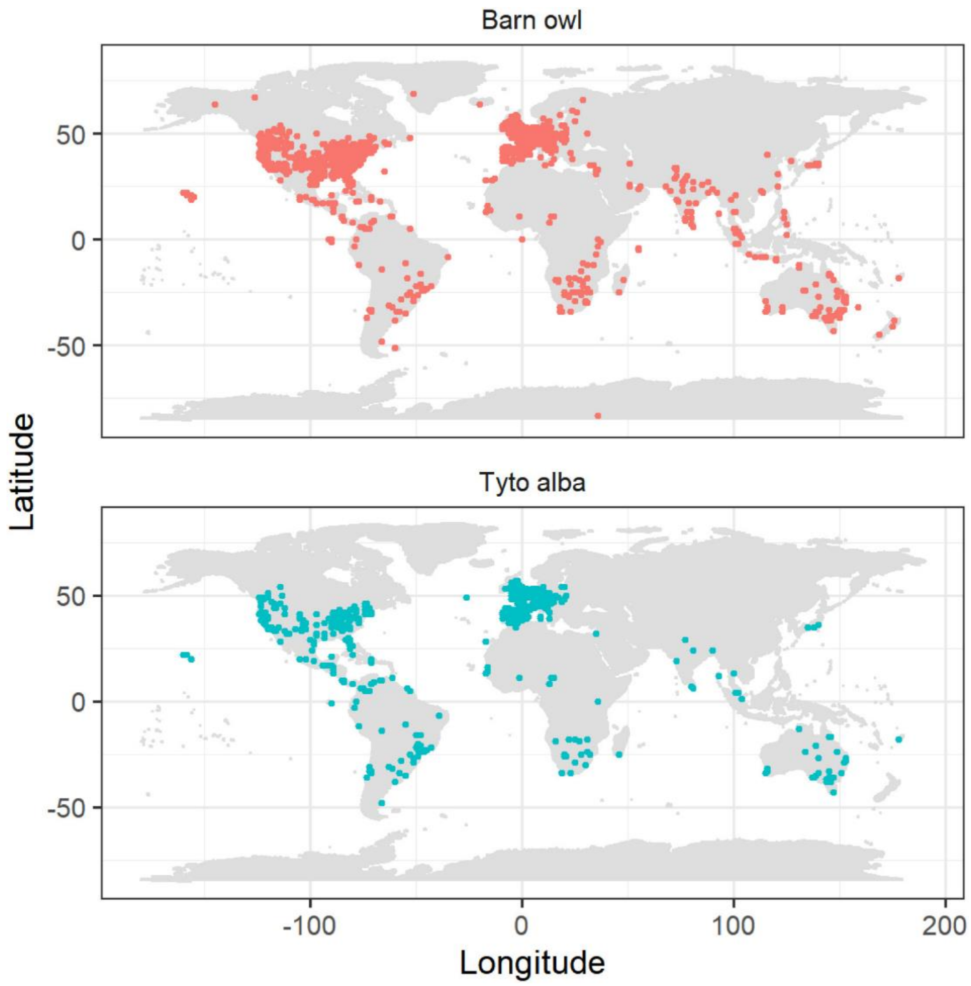


Figure 7 Spatial distribution of Flickr photographs of barn owls

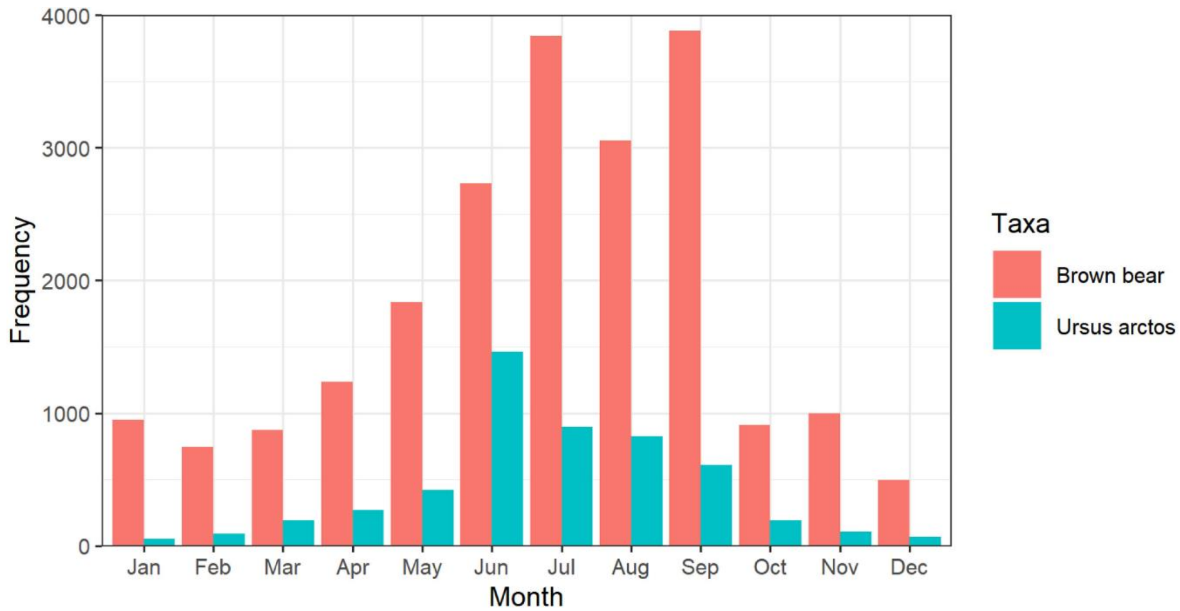


Figure 8 Temporal distribution of Flickr photographs of brown bears.

4.5 Impact

Photosearcher provides a more accessible and reproducible method of accessing the Flickr API, as well as overcoming limitations that prevent researchers from obtaining datasets. By creating photosearcher within the R environment it is freely available to all researchers. Furthermore, by consolidating the code into user-friendly functions the photosearcher package expands the accessibility of the Flickr dataset to non-data scientists. The simple functions also allow researchers to share their methods in a transparent and reproducible manner. However, we note that as people can add new uploads, edit metadata or delete their images, a search for the same criteria on two different occasions may return a different number of results. By providing arguments for limiting searches by the date they were uploaded the photo_search function helps to minimize any changes between repeated searches. This – combined with the ability to share the arguments used in the function calls or a full reproducible document (Appendix B) – makes photosearcher well suited to producing replicable results when working with Flickr data.

The photosearcher package allows researchers to obtain the full range of data available. To overcome the API limit of 4,000 results per query, photos_search requires the user to provide a minimum and maximum search date for when the photographs were taken. If the number of photographs matching the users defined criteria is less than 4,000, the metadata is returned. However, if the number of photographs is greater than 4,000 the metadata for the first 4,000 photographs are returned chronologically. The function then extracts the maximum date on which these images were taken and carries out a new search using this as the mindate_taken argument. The function does not assume that the new search contains fewer than 4,000 images and therefore checks whether the new search contains more than 4,000 results. In this way, the package will continue to dynamically split the initial search into new searches until it returns all available unique images from the initial search. The only time where all data may not be returned is if there were more than 4,000 images for a given second. As this process is automated it means users do not have to make additional calls manually to test which range of dates will return fewer than 4,000 results. Through using an automated method of splitting the searches, the photo_search function provides users with a time and cost-efficient method of data collection. Furthermore, unlike other software such as the INVEST tool, the photo_search function returns the full available metadata available for each photograph. This metadata can be useful for novel research by helping filter results to overcome some of the limitations of social media data. For example, by returning a Flickr-derived measure of spatial accuracy, users of the photosearcher package can quickly filter the returned results based on the accuracy of the spatial reference. Moreover, the anonymous user ID allows users to calculate visitation metrics such as photo–user–days (Wood et al. 2013), to overcome bias introduced by very active users. We have also provided an option to allow to supply a shapefile to search for a specific area. The photo_search function automatically transforms the provided shapefile to a bounding box which is then sent to the Flickr API to search for photographs. The function then extracts and returns only the responses from the original shapefile.

The other functions available in photosearcher are also designed to be useful in novel ecological assessments. For example, by returning the ID of the user uploading the images, additional analyses can be carried out using their publicly available data, returned by function user_info. Furthermore, the download_images function allows users to download the images themselves, which could be used for additional analysis or validation. The returned images could be classified by hand or through machine learning techniques to answer a range of ecological questions including the distribution of ES (Richards and Tunçer 2018) and identifying plant species (August et al. 2019). The plant species dataset (August et al. 2019) was derived from the outputs of the photo_search function.

4.6 Conclusions

The R package photosearcher provides an easily accessible and reproducible method for accessing large datasets from Flickr. The simple skill set needed to use the photosearcher package will increase opportunities for use of Flickr data by non-data scientists. By addressing the challenges and limitations associated with API access photosearcher provides the basis for a standardized method for API calls. The photosearcher package provides both a quick and inexpensive method of gathering large quantities of data, with the methods presented here demonstrating how the package can help provide extensive biological and social data. We hope that the package allows future studies to build upon the current use of Flickr in cultural ecosystem service research, whilst facilitating users to answer a wider array of ecosystem service and ecological questions

Chapter 5 Enriching social media data allows a more robust representation of cultural ecosystem services

This chapter is presented as a reformatted version of the manuscript accepted for publication in Ecosystem Services on the 16th of June 2021.

5.1 Abstract

Images and textual metadata from social media sites such as Flickr have been used to understand the drivers and distributions of cultural ecosystem services (CES). However, using all available data from social media sites may not provide an accurate representation of individual services. For example, an image of nature might be described negatively in the image's description. Here, we present a novel approach to refining social media data to represent CES better, including filtering by keywords, photograph content and enriching the data by including a measure of the sentiment expressed in the textual metadata. We demonstrate that using an enriched dataset of Flickr images representing hiking in the USA can lead to different results and conclusions when compared with results derived from the full dataset. Furthermore, we classified the contents of these hiking images and, using latent semantic analysis, clustered the images into ten groups based on the similarity of their content. The groups provide rich information, such as the importance of geodiversity and biodiversity in supporting a positive hiking experience. The application of this method can help to enrich social media data for CES studies, allowing researchers to further untangle the complex socio-ecological interactions that drive CES distributions, benefits and values.

5.2 Introduction

CES are the non-material benefits obtained from human-nature interactions including through recreation, cognitive development, aesthetic views and spiritual enrichment (Millennium Ecosystem Assessment 2005; Milcu et al. 2013). Though there are many different definitions and classifications used for CES, most literature focuses on the links between the biophysical environment and human wellbeing, while recognising that CES are intangible(Fish et al. 2016; Dickinson and Hobbs 2017) Here, we classify CES following Milcu et al. (2013), who divided CES into 11 subcategories; recreation and tourism, aesthetic values, spiritual and religious values, educational values, cultural heritage values, bequest, intrinsic and existence, inspiration, sense of place, knowledge systems, social relations, and cultural diversity. Furthermore, we note that CES benefits can be delivered through multiple pathways. (King et al. 2017) identified six pathways to CES benefits that reoccur across CES literature: cognitive (benefits from the development of knowledge), creative (benefits from influences on aesthetic appreciation and artistic expression), intuitive (benefits from the influence on instincts and senses), retrospective (benefits from reflecting on past experiences), regenerative (benefits from opportunities for recreation, leisure and tourism) and communicative (benefits from social relations, cultural identity, and sense of place). However, for these pathways to be actualised, there is a need for people to first recognise the potential benefits of biophysical features and then utilise these potential benefits (Spangenberg et al. 2014). Therefore, CES, as with other ES, are not provided by ecosystems independently of humans but are co-produced through our interactions with them (Fischer and Eastwood 2016).

Though the value of ES is generally provided by an economic metric, ES values can also be measured by its societal and cultural values and therefore the value of CES is often non-monetary (Reynaud and Lanzanova 2017; Small et al. 2017; Haines-Young and Potschin 2018). There are multiple methods of assessing CES value both, monetary (e.g. travel cost or willingness to pay) and non-monetary (photograph analysis or ranking methods) (Hirons et al. 2016). However, quantifying the benefits and values of CES are more difficult due to the perceptions of CES benefits being unique to individuals based on their social and cultural norms (Daniel et al. 2012; Havinga et al. 2020). Therefore, CES have been comparably under-researched compared to other ES and we, therefore, need to develop our understanding of the human-nature interactions that provide these services (Milcu et al. 2013; Dickinson and Hobbs 2017).

Previous literature assessing CES relationships, as well as wider ES relationships, tend to focus on the role of nature on CES production (Fischer and Eastwood 2016). Because people, culture and nature are so inherently interlinked, it can be difficult to disentangle what constitutes nature in the context of ES (Plumwood 2006; Hirons et al. 2016). Here, we view nature as the biophysical features of an ecosystem (Haines-Young and Potschin 2018), comprised of the interactions of biodiversity and geodiversity (Gray 2012; Gordon and Barron 2013; Potschin-Young et al. 2018; Fox et al. 2020b). There has been a particular focus on the role of biodiversity in CES production, with the role of geodiversity often omitted from studies (Fox et al. 2020b). Geodiversity can be viewed as the abiotic equivalent to biodiversity, representing 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 (Gray 2004; Hjort et al. 2015; Fox et al. 2020b). Geodiversity can also provide CES in the absence of biodiversity. These “geosystem services” include CES such as recreational activities (e.g. water-based sports, rock climbing and caving), spiritual sites (e.g. Uluru, Australia and the Torres del Paine, Chile), as well as providing opportunities for advancing scientific knowledge (e.g. the record of past climates and ecosystems contained in sediment, rock and ice cores) (Gray 2012; Kiernan 2015; Fox et al. 2020b). Often geodiversity is only assessed through landscapes types which are prescribed a general CES value, however, CES are not distributed randomly within a landscape, but are concentrated in hotspots that have specific features of biodiversity (e.g. forests and hedgerows) and geodiversity (waterbodies and geological formations) (Plieninger et al. 2013; Van Berkel and Verburg 2014). It is therefore important that the relationship of the individual features of biodiversity and geodiversity to CES be assessed.

There is also a need to recognise that CES are co-produced and co-created by people and therefore only arise from the interaction of people with the biophysical environment (Chan et al. 2011; Fish et al. 2016). Though CES are inherently co-produced through human-nature interactions, these relationships are often omitted from studies in favour of assessing the links between biophysical nature and ES (Fischer and Eastwood 2016). CES are co-produced through a variety of different pathways including human, financial and manufactured capital (Raymond et al. 2018). For example, CES can be co-produced through the influences of culturally important buildings (e.g. places of worships), managed landscape (e.g. agricultural land), organisations (e.g. museums, parks and gardens), or purpose-built infrastructure (e.g. hiking trails) (Plieninger et al. 2013; Van Berkel and Verburg 2014; Fischer and Eastwood 2016; Minkiewicz et al. 2016). There is a need to acknowledge the implications of the co-production for quantifying ES, and CES in particular, as a holistic approach to an understanding of these complex human-nature interactions can help to better shape their sustainable management (Bennett et al. 2015; Palomo et al. 2016).

To understand CES there is a need for suitable datasets that can assess the complex relationship between biodiversity, geodiversity and society. However, globally, datasets on ES are sparse, meaning services are often mapped through proxies (Stephens et al. 2015), with the primary sources of data for CES mostly coming from either monetary assessments, social surveys such as stated preferences, or onsite surveys (Tenerelli et al. 2016; Figueroa-Alfaro and Tang 2017; Mayer and Woltering 2018). Due to labour-intensive methods and high financial costs, implementing these over large spatial and time scales is not always feasible (Wood et al. 2013; Kim et al. 2019). Furthermore, management decisions need to be better informed through methods that reliably understand, identify, quantify and map CES (Tenerelli et al. 2016; Byczek et al. 2018). As CES are dynamic, methods also need to be able to reliably investigate changes over time (Figueroa-Alfaro and Tang 2017).

The potential of social media sites such as Flickr, Twitter and Facebook as a source of data for CES questions is starting to be realised (Kim et al. 2019). In contrast to social surveys, social media data is inexpensive, quick to gather and provides a means of mapping the distribution of CES and assessing changes over space and time (Fox et al. 2020a). Social media data has been used for CES studies, such as wildlife watching (Mancini et al. 2019), recreational services (Graham and Eigenbrod 2019; Sinclair et al. 2020b), aesthetic views (Van Berkel et al. 2018) and visitation rates in protected areas (Tenkanen et al. 2017; Kim et al. 2019; Sinclair et al. 2020a), providing key information for both tourism and conservation. However, social media data is often messy (Ghermandi and Sinclair 2019; Chen et al. 2020). Issues such as unknown or inaccurate spatial references (Figueroa-Alfaro and Tang 2017), unreliable image contents due to mistagged images or a mismatch between the content of a photograph and the location it was taken (Oteros-Rozas et al. 2018), and biases introduced by user groups (Langemeyer et al. 2018; Chen et al. 2020) therefore need to be accounted for.

To address the issues introduced by the vast volume of data on social media sites, CES studies tend to filter the returned results through several different approaches. For example, some studies filter results based on the geographic location the images were taken in - e.g. studies using the InVEST recreational model (Sharp et al. 2020). Searching for images within a given study site alone may return a large number of images not relevant to the specific ES of interest, or even to any ES at all. Other studies filter out images based on land cover types. For example, Tenerelli et al. (2016) excluded images of photographs found in urban areas. This method overlooks the fact that a photograph's location does not always represent the subject of the image (Yan et al. 2019). A photograph taken within an urban area may be a long-distance image of a CES such as an aesthetic natural view, whilst photographs taken in a natural land cover may not be of a CES - e.g. photographs of a car's interiors. Furthermore, CES are not confined to specific land cover types and so excluding on this basis may exclude relevant services such as those provided by urban green spaces and trees (Kondo et al. 2018)

Another approach to deciding which images to include is to search for photographs based on a set of criteria - e.g. a study looking for photographs of hiking may limit returned photographs to those containing the word "hiking" in the textual metadata (Graham and Eigenbrod 2019). However, limiting images based on text alone does not guarantee that the image itself represents an ES. For instance, a search for photographs of "biking" may return photographs of equipment such as bikes and helmets. It therefore cannot be assumed that an image containing textual metadata related to the use of an ES is relevant for assessing CES.

Some authors are starting to acknowledge this, for example, Havinga et al. (2020), who used the distribution of photographs from Flickr to assess the aesthetic quality of landscapes, suggest that not all photographs may have relevance to the study. They, therefore, recommended assessing photograph contents. To ensure that photographs are suitable for their studies, researchers have analysed the contents of images and refined them to those that meet relevant criteria. Image classification can be either manual (subjective) or automatic through machine learning techniques (objective). For example, Oteros-Rozas et al. (2018), manually classified the contents of images, labelling them with landscape features and CES. These labels were then used to identify bundles of landscape features and CES. One method of automatic tagging is the Google Cloud Vision API, a machine learning algorithm that can identify the contents of images. Richards and Tunçer (2018) used the Google Cloud Vision API to label the contents of Flickr images and subsequently used this information to map the distribution of plants and animals, whilst Gosal et al. (2019) labelled photograph contents using the Google Cloud Vision API to find groups of recreational beneficiaries.

However, we argue that a combination of textual metadata confirming the presence of a targeted service (for example, an image tagged “hiking”), and an image containing features of the natural environment (for example, an image of a mountain), still does not confirm that the service user experienced a positive benefit indicating a CES. Indeed, it may have been a negative experience - e.g. a user could caption the image with a complaint about a boring walk. Furthermore, as sharing photographs is influenced by societal pressures, photographs shared may not show the user’s preferred features of nature (Moreno-Llorca et al. 2020). The textual metadata can contain “text-private” information that can convey emotions and opinions which could not be elicited from the image contents. By contrast, the image contents will often contain “image-private” information, such as features of the image and colours not mentioned in the text (Huang et al. 2019). In particular, a dichotomy between the textual metadata and image contents can exist as the textual metadata tends to be more heterogeneous and contain non-descriptive terms and phrases, whereas the classification methods used to label the content of images provides more homogenous single descriptive terms (Yan et al. 2019). Here, we suggest that by assessing the sentiment expressed in the textual metadata we can get additional information about the quality of the experience, meaning one can enrich the CES data. As the different types of data from Flickr contain different information (image content, spatial, temporal and textual metadata), studies could combine all these data sources to obtain more information about the benefits received.

We suggest that social media data can be enriched to better understand user experience, elicited through sentiment analysis of the textual metadata, and may provide a more robust dataset for CES assessments. Lexicon-based sentiment analysis is a natural language processing technique used to calculate the semantics, opinions or emotions of words or phrases from text (Wilson et al. 2019). One form of analysis is polarity classification, which classifies text as either positive or negative and can be used to assess social media datasets (Koto and Adriani 2015). Sentiment analysis for ES assessment has been broadly applied to social media datasets from Twitter (Becken et al. 2017; Wilson et al. 2019) and Instagram (Do 2019). There has also been some limited application of sentiment analysis to Flickr textual metadata. For example, Brindley et al. (2019) used sentiment analysis on Flickr textual data to assess the perception of green space.

Though the position of recreation as a final service has been questioned (Haines-Young and Potschin 2018), recreational activities provide restorative benefits (e.g. increased physical wellbeing) and are therefore generally considered a CES (Millennium Ecosystem Assessment 2005; Milcu et al. 2013; Plieninger et al. 2013; King et al. 2017; Balzan and Debono 2018) In this article we will focus on hiking, a recreational activity that involves walking over an extended period, typically through natural or rural areas (Mitten et al. 2018). We consider hiking a CES here as it can directly provide restorative benefits, however, as the benefits from recreation may also be indirect (Balzan and Debono 2018), we recognize hiking could also be considered as a pathway to other CES (King et al. 2017) through co-production between human-nature interactions (Fischer and Eastwood 2016). For example, hiking may be undertaken for spiritual and religious motivations such as a pilgrimage, to experience aesthetic qualities of nature, or as a social activity with a sense of belonging (Collins-Kreiner and Kliot 2017; Wilcer et al. 2019). From this viewpoint, hiking can also provide multiple pathways to other CES benefits, for example, providing a means of access to aesthetic views (creative pathway) or providing a sense of place (communicative pathway) (King et al. 2017). Hiking is one of the most popular recreational activities, both in the USA and worldwide, with participants from all age categories (Wilcer et al. 2019). Despite its popularity, research on the drivers of hiking remains limited (Wilcer et al. 2019). Hiking, as with other ES, is driven by co-production between ecosystems and people, such as the influences of cultural landmarks and landscapes or through the provision of infrastructure such as signposted trails (Plieninger et al. 2013; Fischer and Eastwood 2016).

In this article, we present an analysis that refines social media data from Flickr using content analysis and enriches the data using a measure of sentiment value expressed in the textual metadata, particularly for mapping the distribution and understanding the drivers of CES. We then demonstrate the use of our method for a more differentiated analysis of hiking as a CES, focusing on understanding which of geodiversity, biodiversity and human features contribute to a positive hiking experience.

5.3 Methods

5.3.1 Data acquisition

A reproducible R file for the data collection methods has been included in the supplementary material (Appendix C). To comply with API terms and privacy policies all datasets were anonymised, stored with multiple layers of security and any unnecessary metadata was deleted.

We queried the Flickr API for photographs containing the text “hiking” in a photograph's title, description or tag metadata, in the contiguous 48 states of the USA, between 2015-01-01 and 2020-01-01. To ensure consistency and reproducibility in the study we used the photosearcher package (Fox et al. 2020a) within the R environment (R Core Team 2020). The photosearcher package allows searches of Flickr to be constrained by a shapefile. Here the search was limited to photographs taken and geotagged in the contiguous 48 states using a modified shapefile from the USAboundaries R package (Mullen and Bratt 2018). As the number of posts meeting these criteria can change over time (e.g. new photographs were taken during the study period but uploaded at a later date), we limited photographs to those uploaded before 2020-06-01 to increase the reproducibility of the search.

5.3.2 Content analysis

To ensure an image captures human-nature interactions, we categorised the contents of images to help filter out images that do not contain any biophysical features in their contents (e.g. indoor images). Here, the features within each image were automatically labelled using the Google Cloud Vision API (Google Cloud Vision API 2020), accessed through the imgrec R package (Schwemmer 2019). The Google Cloud Vision API is a pre-trained machine learning model that can detect image contents including objects, faces and text. For each image, we requested the API to identify and label the 10 most dominant features of the image. The generated labels had an associated confidence score between 0 and 1. Here we only kept labels with a confidence score of >0.6 (Gosal et al. 2019). We assessed each unique label identified by the Google Cloud Vision API and categorised them as either biophysical nature or not. Here, we identified non-biophysical nature words as synthetic objects (e.g. buildings and cars), relating to people (e.g. a person), and descriptive terms (e.g. black, text), as well as non-biophysical aspects of nature such as weather phenomenon (e.g. sky and sun). Biophysical nature labels were any label that was a feature of biodiversity (e.g. tree or bird of prey), geodiversity (e.g. lake or geology), or an ecosystem (e.g. rainforest or grassland). Furthermore, we included generic descriptions of landscapes (e.g. wilderness and natural landscape) as biophysical nature. Appendix D provides a full list of how each word was categorised. For each image, we calculated the percentage of labels categorised as a biophysical feature of nature.

5.3.3 Sentiment analysis

To ensure an image captures a positive human-nature interaction, and therefore a CES, we calculate the sentiment value expressed in the textual metadata. Here, we used the AFINN dictionary (Nielsen 2011) to summarise the sentiment of the textual metadata. The dictionary ranks words on a scale of -5 (the most negative words) to +5, (the most positive words), and can be an effective method for assessing social media datasets (Koto and Adriani 2015). For each image, the associated textual metadata was analysed using the AFINN dictionary and the overall sentiment value was calculated as the sum of all the positive and negative sentiment scores for that image. Each image was then categorised into two groups, (1) positive images - those with an overall positive sentiment score, (2) non-positive images - those with an overall negative sentiment score, an overall neutral sentiment score or where no sentiment was expressed in the textual metadata.

5.3.4 Impacts of refining the data

To understand the ramifications of refining data by contents and sentiment, we carried out two filtering processes: first selecting all images where the percentage of labels classified as biophysical nature exceeded a given threshold (e.g. 25% of the Google Vision Cloud API labels were of biophysical nature features), and second filtering only images with positive sentiment to ensure images were associated with a positive experience (Fig. 9). The final part of the analysis applied both filters, resulting in a final dataset representing images of human-natural interactions AND reflecting a positive experience. We extracted a random sample of 100 images from each category (*confirmedNature, nonNature, confirmedPositive, nonpositive*) to manually validate the automated process.

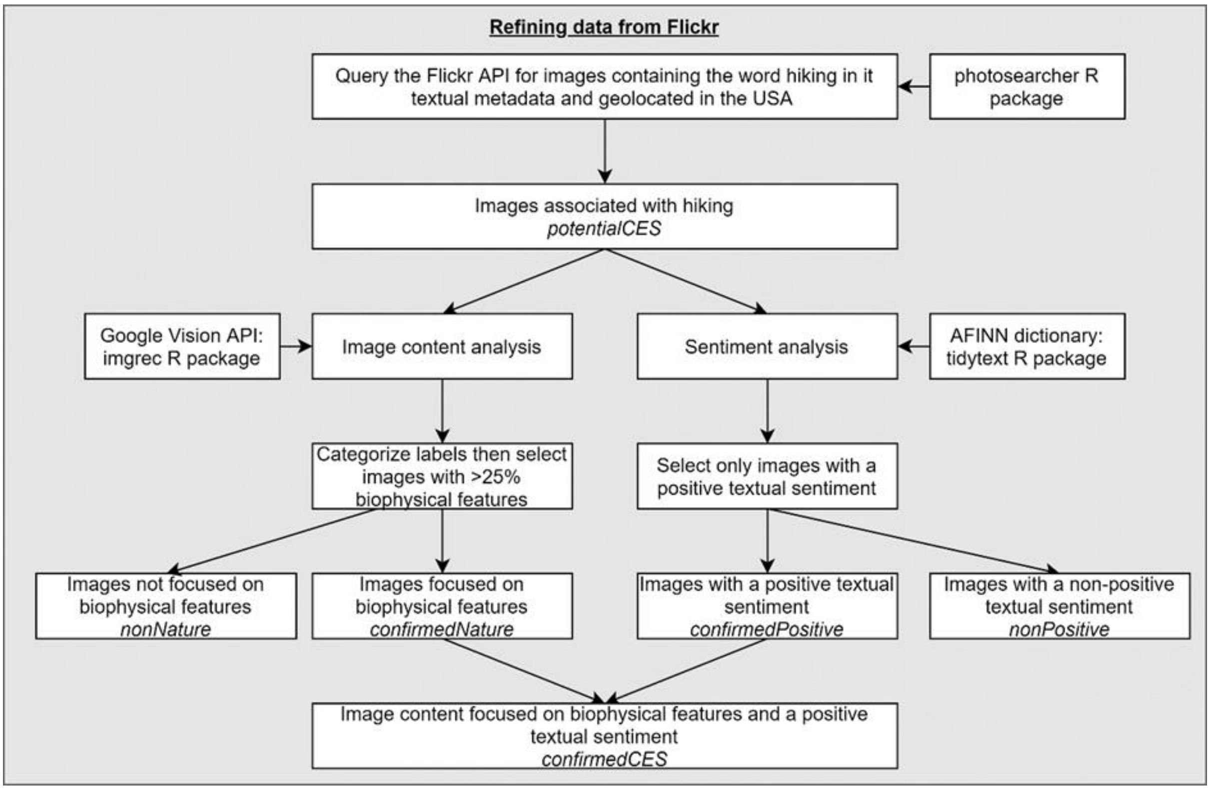


Figure 9 Processes applied to refining the Flickr dataset.

First, the full Flickr dataset was refined to images that were *confirmedNature* (thus removing any images that, whilst including the relevant activity in the associated textual data, were not taken within a natural or semi-natural setting and therefore did not represent an ES, for example, an image of somebody indoors with the caption "I wish I were hiking"). Two rasters were generated representing the number of uploads per pixel, one raster from the full dataset and one from the refined dataset. A moving window was used to assess localised differences in the two rasters (Fig. 10). The moving window assesses a square of pixels (3 x 3 pixels) and calculates the deviance of their values from a 1:1 line. To do this, the values of two rasters within the window are normalized between 0 and 1 and the absolute difference between the two pixels at the same location within the window was calculated. If the two sets of images have a perfect correspondence in the distribution within the window, the standardised upload values would fall along a 1:1 line. The deviance score is calculated as the mean of the absolute values of these differences (to account for negative deviations from the 1:1 line, following Willcock et al. (2019)). The moving window approach allowed us to represent spatial differences in deviance by calculating the deviance value for the number of uploads within the window and creating a new raster in which the deviance value was assigned to the central pixel of the window (e.g. the pixel in the middle of the 3 x 3 moving window). If the full dataset is a good proxy for images of nature, we expect the local deviance (defined by the window size) value to be less than 0.3 (Willcock et al. 2019); indicating that the refined and full datasets share similar distributions and that the filtering is not necessary. Where the deviance value is greater than 0.3 there is not a good fit between the two datasets (Willcock et al. 2019).

As landscape characteristics can drive recreational activities at a range of scales, and Flickr is a good proxy for recreation at a range of scales up to 50km (Graham and Eigenbrod 2019), here we map the difference in distribution when the number of uploads is aggregated to 25km^2. We also map the differences in distribution using a 3 x 3-pixel window (where a pixel is the size of one pixel of the underlying raster as determined by the spatial resolution), a standard size for aggregating fine-scale data (Graham et al. 2019). Furthermore, we map the differences based on refining by a threshold of 25% biophysical labels. This threshold was chosen to ensure that the refined set of images captured human-nature interactions, without completely excluding images containing human features as these could help to provide insight into the co-production of CES. However, as any changes in distribution may be a facet of the filtering method, or the method of mapping and calculating the deviance, we conducted a sensitivity analysis using all possible combinations of three thresholds for considering an image to be of nature (images with 25%, 50% and 75% labels classed as biophysical nature), three different spatial resolutions (5km^2, 10km^2 and 25km^2), and three sizes of moving window (3 x 3 pixels, 5 x 5 pixels and 7 x 7 pixels). This resulted in 27 datasets.

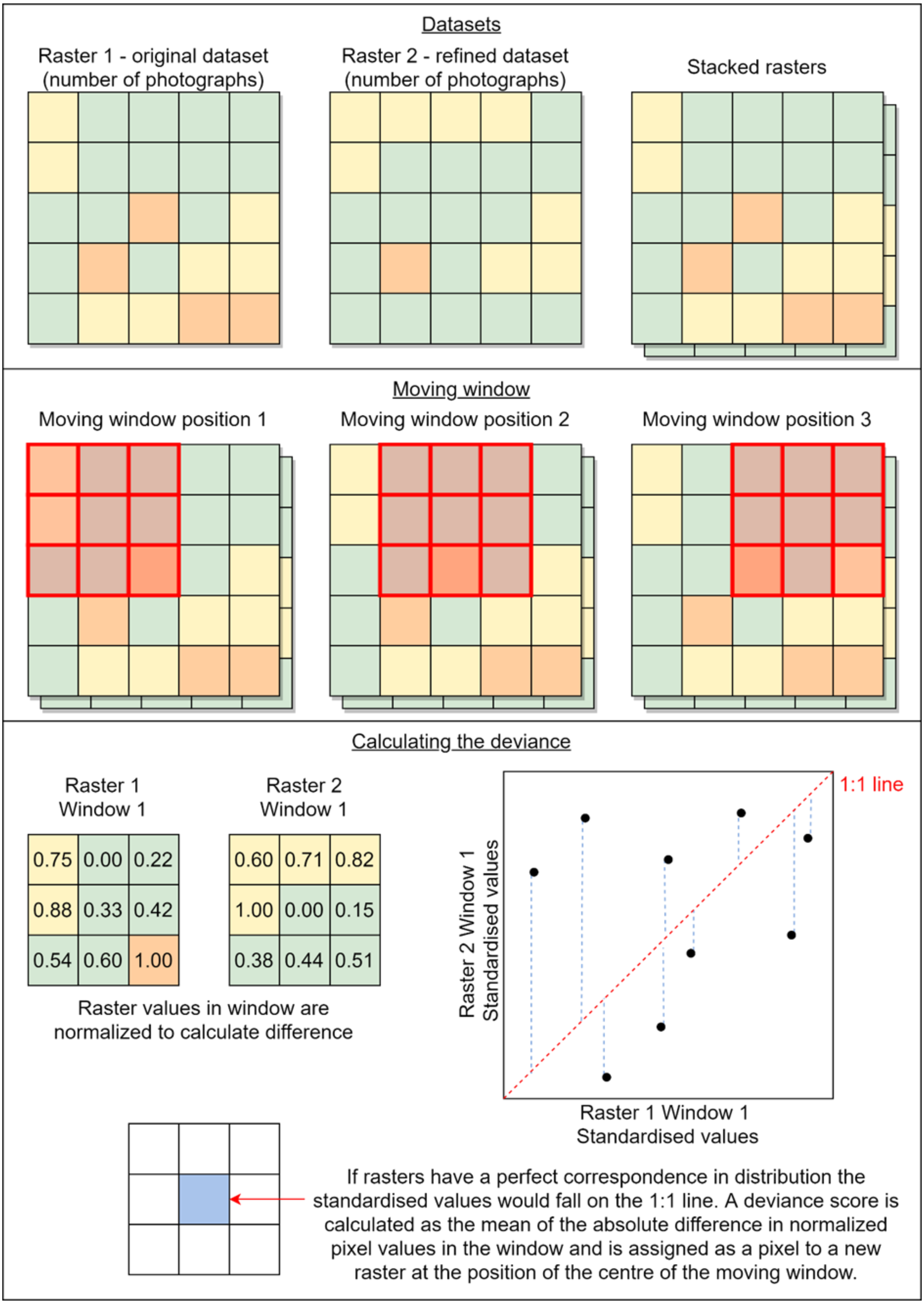


Figure 10 Methods for calculating the difference in distributions between the unrefined and refined datasets.

If there is a spatial pattern to the data (e.g. if all pixels that have high deviance values are spatially clustered such as in urban areas), the full data may be an applicable proxy if areas of anomalous data are accounted for in any analysis. However, if the distribution of deviance values is random the full dataset may not be a suitable proxy. We calculated the Moran's I for the local deviance maps. Moran's I can have a value of -1 to 1, with values closer to -1 showing a uniform distribution, values closer to 0 a random distribution and values closer to 1 a clustered distribution (Sankey 2017).

Second, we followed a similar process, filtering by the sentiment expressed in the textual metadata (thus removing any images that were not about a positive experience and therefore did not represent an ES, for example, an image with the caption "boring view"). In this case, we compared the number of images with positive sentiment, *confirmedPositive*, against the full dataset. In this case, we tested the sensitivity of the results to the chosen parameters by considering all combinations of the same three different pixel resolutions and three sizes of moving window (9 datasets in total). In this example, if the full dataset is a good representation of a positive experience, the local deviance values should be close to 0. We also tested for spatial patterns in local deviance using the Moran's I.

Finally, we refined the original dataset to images containing biophysical features and positive textual sentiment (thus capturing CES as images of positive experiences occurring within a natural, or semi-natural setting). Here, we mapped the distribution of these *confirmedCES* images versus the full number of images and calculated the local deviance values of the two datasets, again using all the 27 possible combinations of three thresholds for considering an image to be of nature (images with 25%, 50% and 75% labels classed as nature), three different spatial resolutions (5km^2, 10km^2 and 25km^2), and three sizes of moving window (3 x 3 pixels, 5 x 5 pixels and 7 x 7 pixels). Here, if the original full dataset was a good proxy for CES we would expect deviance values to be closer to 0 between the *confirmedCES* images versus all the images. We also tested the spatial uniformity of the local deviance values using Moran's I.

5.3.5 Mapping distributions and sentiment

The point locations of the *confirmedCES* images were aggregated to a raster layer using a pixel size of 25km^2, a suitable size for assessing spatial relationships based on Flickr data (Graham and Eigenbrod 2019). We also mapped the mean sentiment score of images falling within each pixel. To test the relationship between the number of uploads of *confirmedCES* images and the mean sentiment score of *confirmedCES* images in an area we calculated Pearson's correlation between the two raster maps. If a high number of uploads is related to a high sentiment value, we would expect a Pearson's correlation value to be closer to one.

5.3.6 Assessing human-nature interaction in images

To assess the relationship between biodiversity and geodiversity we carried out latent semantic analysis (LSA) on the content labels (Fig. 11). LSA was carried out in the R environment, primarily using the lsa R package (Wild 2015). LSA is a natural language processing technique that is used to assess the relationship between a collection of documents (in this case a user’s photographs from one day) and the terms used in them (in this case Google Vision Cloud API labels) as a term-document matrix (TDM) and can be used to examine how closely terms are related in use (Gefen et al. 2017). LSA on a TDM has been previously used to help describe recreational activities (Monkman et al. 2018; Gosal et al. 2019). Here, where the LSA shows terms are more closely related, this indicates that those Google Vision Cloud API labels are more commonly photographed together, for example, one might expect the labels “forest” and “tree” to be frequently photographed together.

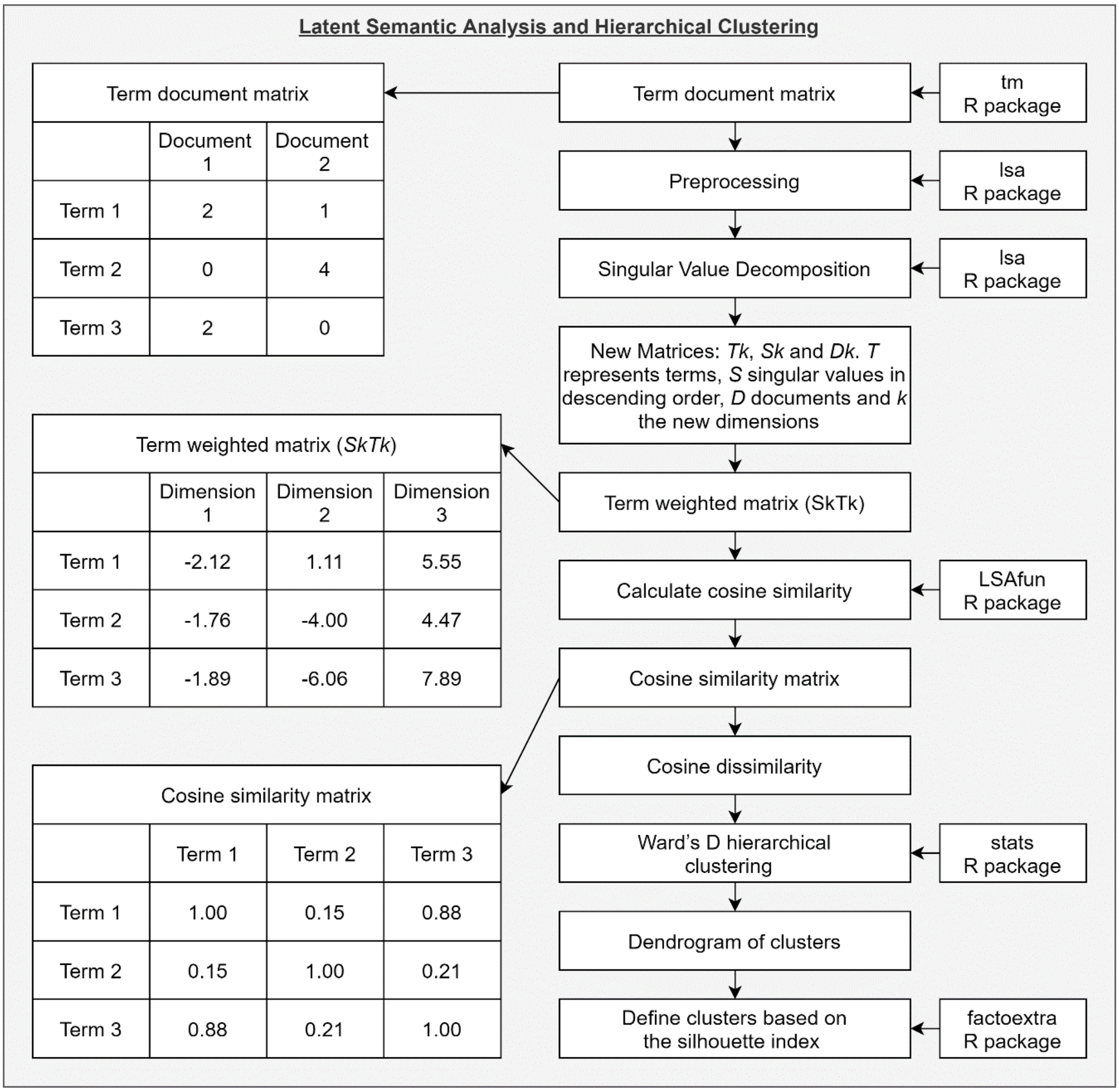


Figure 11 Example workflow for the methods of clustering photographs.

As LSA can be used to assess individual CES preference, we grouped all the labels from all images from a single user of a single day into a single document, building upon the photograph user-days (PUD) metric introduced by Wood et al. (2013). These new groupings were transformed into a term-document matrix (TDM) (Gosal et al. 2019). The TDM, *M*, contains the grouped photograph by a single user on a single day as the columns and the Google Vision Cloud API labels as the rows, with cells representing the frequency of the label appearing in that users photographs for that day. During the creation of the TDM we carried out several common LSA pre-processing procedures to ensure that only relevant words were kept (Evangelopoulos 2013; Gefen et al. 2017). First, we removed stop words, a list of around 400 common English words such as “the”, “of” and “them”. Second, to assess which features are most commonly photographed in association with hiking, we only selected labels that appear in at least 5% of the documents (the labels from each image, including non-nature labels). Third, we applied the “Term Frequency-Inverse Document Frequency” (TF-IDF) weighting to the TDM. The TF-IDF is one of the most commonly used weightings for LSA, where locally more weighting is given to terms that appear frequently in one document and globally less weighting is given to common terms and is necessary to control for the fact that some words appear far more frequently than others (Evangelopoulos 2013; Christian et al. 2016; Gefen et al. 2017).

After pre-processing the LSA was carried out on the weighted TDM. The LSA procedure carries out a singular value decomposition (SVD), a linear algebra method for the factorization of a matrix into a product of matrices. Here, the SVD takes our matrix M (an m x n matrix with m representing all images a user took in a day and n terms that the Google Vision Cloud API labelled in those images) and transforms this into three new matrices: T_k, S_k and D_k. T represents a term vector matrix, S represent a diagonal matrix containing singular values in descending order, D represents a document vector matrix and k the number of new dimensions (Gosal et al. 2019). The LSA process represents the data in k-dimensional semantic space, by reducing the original dimensions whilst preserving the most information. This method allows for the original space vector to be represented in the lower-dimensional term and document vectors. Here the dimensionality reduction (k) was automatically calculated using the standard "fraction of the sum of the selected singular values to the sum of all singular values" method (Gosal et al. 2019). This method selects the points on the diagonal matrix of descending singular values where the sum of the S singular values divided by the sum of all the S singular values are equal or greater than 0.5 (Gefen et al. 2017). From the new matrices, we calculated a term weighted matrix (S_kT_k). The term weighted matrix represents each term as a row and each dimension (or latent semantic factor) as a column (Evangelopoulos 2013).

By carrying out the SVD, the terms can now be projected in multidimensional space and the similarity between Google Vision Cloud API labels can be calculated using cosine similarity (Evangelopoulos 2013). Cosine similarity measures the angle between vectors in multi-dimensional space, with the resulting values ranging from 0 to 1, with 1 representing total similarity. The diagonal in the matrix will always be 1 as a word is always the same as itself. Here we calculated the cosine similarity between all terms. Cosine similarity of 1 means that the labels always appear together and 0 means they never appear together.

To understand what features (including biophysical and non-biophysical nature as well as human features) are most often taken in the same images, we used hierarchical clustering to group the Google Vision Cloud API labels based on their cosine similarity. As hierarchical clustering uses a distance measure, we calculated cosine *dissimilarity* (1 - cosine similarity). We then carried out hierarchical clustering using Ward's D method, which has previously been shown to create unambiguous clusters for labels generated by the Google Cloud Vision API (Gosal et al. 2019). The

Ward’s D method builds a dendrogram through a bottom-up approach to clustering. Each element in the tree starts as an individual cluster, two clusters are then merged so that variance within clusters is minimized. This process is repeated until all elements are clustered on the tree.

A dendrogram can then be cut at a chosen height to provide the final clusters, with the choice of height resulting in the selection of different final clusters. Though studies can choose an arbitrary height to cut the dendrogram and select the clusters, indices such as the silhouette index can be used to find the optimal number of clusters based on the given dendrogram (Wang and Xu 2019). Here, the dendrogram was cut into *x* clusters, based on the silhouette index, where each element is assigned a value between -1 and 1. Elements with higher numbers are closer to the other elements in their cluster than elements in other clusters (Pagnuco et al. 2016). The value of *x* with the highest silhouette index was chosen to be the number of clusters.

We summarised the types of images photographed when hiking based on the clusters. For each user's daily images, we calculated the number of Google Vision Cloud API labels belonging to each cluster then categorised the users daily images as which cluster was the most dominant (where labels of two or more clusters appear equally dominant in an image, that image was classified as a combination of those clusters). The mean sentiment of the images belonging to each cluster or combination of clusters was calculated.

5.4 Results

5.4.1 Full and enriched datasets

There were 179,700 geotagged photographs containing the text “hiking” in a photograph's title, description or tag metadata for the years 2015-2020 in the USA. The distribution of hiking images was largely concentrated in the west, the northeast, and some southern states (Fig. 12a). Here, the final refined dataset chosen to represent *confirmedCES* were images containing more than 25% of the image content labels being classified as nature and a positive sentiment expressed in textual metadata. There were 43,427 images that met these selection criteria, 24.17% of the full hiking dataset. The distribution of these *confirmedCES* images follows some similar patterns to the full dataset with the largest concentration of uploads along the west coast of the US in states such as Washington, southwest states, such as Arizona and Utah, northeastern states around the Great Lakes and along the Appalachians (Fig. 12b). Though there are similarities in the distribution of the *confirmedCES* images and the full dataset images, there are also areas where the distributions have different patterns for example, particularly with many regions having fewer uploads, such as southern California, Arkansas and the west coast of Florida.

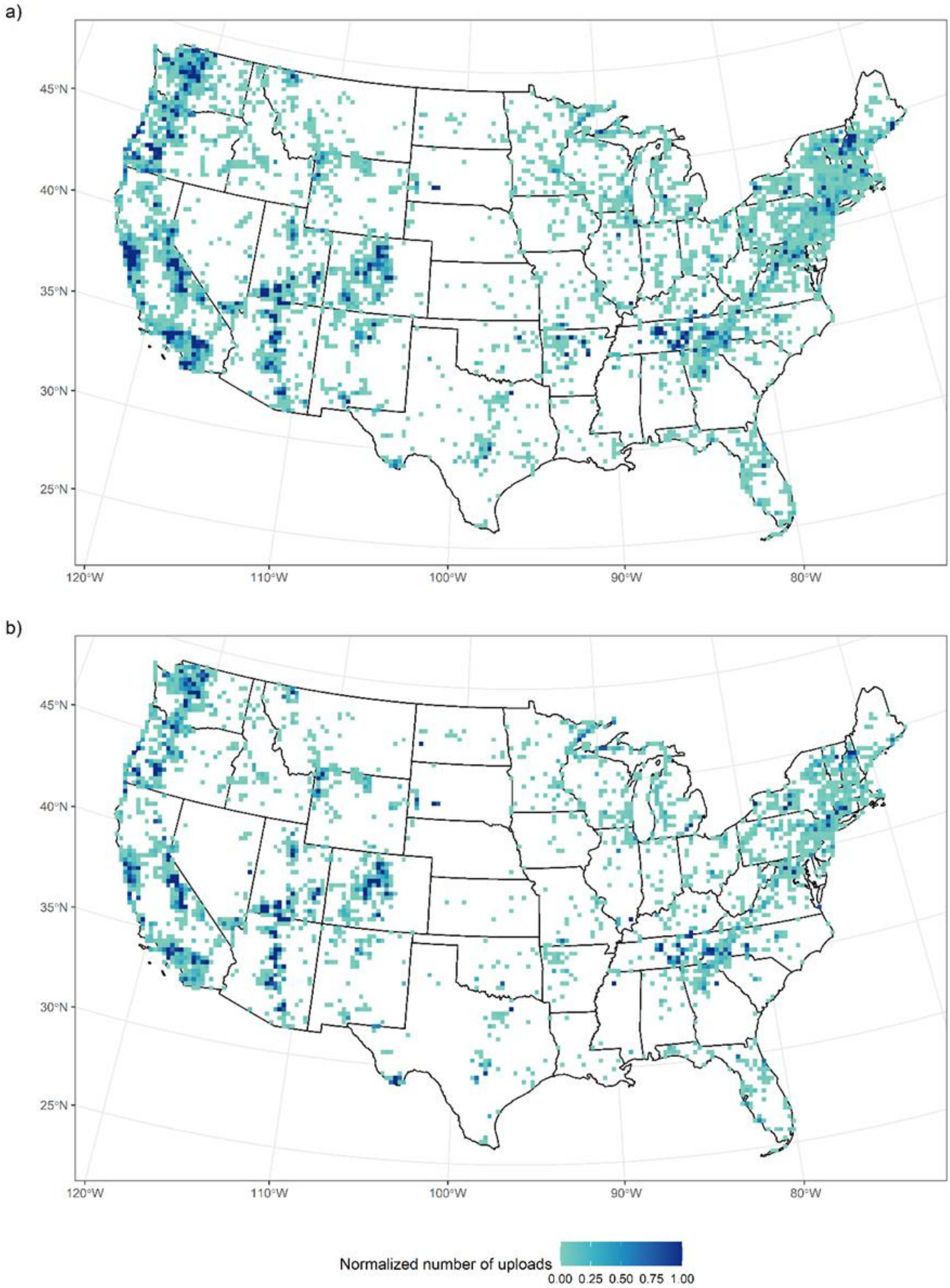


Figure 12 Number of images from Flickr in the USA between 2015-2020 with the term "hiking" in the images title, tag or description, results were normalized (spatial resolution 25km^2). a) the full Flickr dataset, b) images categorised as *confirmedCES* (images where the percentage of content labels classified as nature were >25% and a positive sentiment was expressed in textural metadata).

The similarities in some of the large-scale distributions may be misinterpreted to mean that overall the distribution of uploads from *confirmedCES* and the full dataset images are similar. However, the local variation in the deviance from a 1:1 line differs spatially, with high and low deviance distributed across the whole of the United States (Fig. 13). Here, 20.81% of pixels had a deviance value of > 0.3 indicating that these regions did not have a strong relationship between the number of uploads that were *confirmedCES* and from the full dataset images. As the Moran's I value for the local deviance distribution was 0.27, this indicates that the distribution is close to random, and therefore not spatially uniform. It may therefore not be possible to use the full dataset alone to select areas that represent a good proxy for a positive CES experience. The non-conformity between the full and filtered datasets is distributed across the US including in the northwestern states, around the Great Lakes, areas of California and throughout Florida.

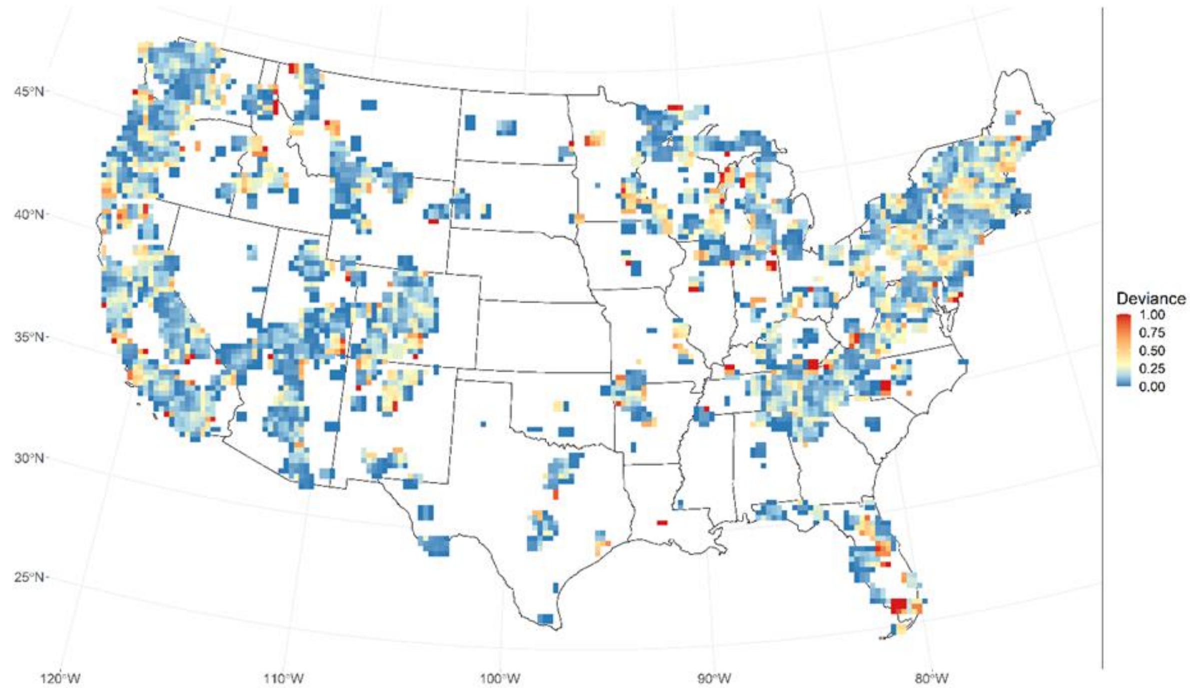


Figure 13 The local deviance in the scaled number of images from the full Flickr and *confirmedCES* images dataset (images where the percentage of content labels classified as nature were >25% and a positive sentiment was expressed in textural metadata) from the 1:1 line. Pixel size 25km^2; window size 3 x 3; a break point of 0.3 was applied to indicate areas of high deviance.

The differences in the distribution of these local deviance values do not appear to be a facet of the chosen mapping techniques as the deviance scores are similar across the full range of spatial resolution, window size and nature threshold combinations (Appendix E). When two variables are held the same (e.g. the same nature threshold and window size) and the other varies (e.g. differing pixel size), the deviance scores remain similar. We do note that there are some small changes based on the selection of each, for example, the larger the pixel size the smaller the overall deviance. However, this change is minimal and the overall deviance score remains similar, suggesting that the selection of mapping technique has had limited influence on the results drawn from the choices of aggregating to $25km^2$, refining based on nature threshold of 25% and calculating deviance using a window size of 3 x 3. Therefore, although the absolute values of the deviance vary based on the threshold, window size and pixel size values chosen, the conclusions which are drawn are qualitatively the same.

5.4.2 Individual influence of filtering by content only

When filtering photographs by the percentage of nature labels based on a threshold of 25%, 160,873 images of nature remained, compared with 179,700 images before filtering. The deviance values vary depending on the threshold, window size and spatial resolution, but again here the conclusions drawn are the same. From the moving window analysis, the distribution of the *confirmedNature* images does not deviate greatly from the full dataset (Fig. 14). The deviance values are again not spatially uniform with a Moran's I value of 0.26, indicating a random spatial pattern. Here, only 2.11% of the pixels are above a deviance value of 0.3 - indicating that images of hiking from Flickr may generally be a good proxy for images of human-nature interactions. The patterns of positive and negative local deviance values are similar overall for all combinations of the varied nature threshold, spatial resolution and window size, indicating that these choices did not influence the results when filtering by content alone (Appendix E).

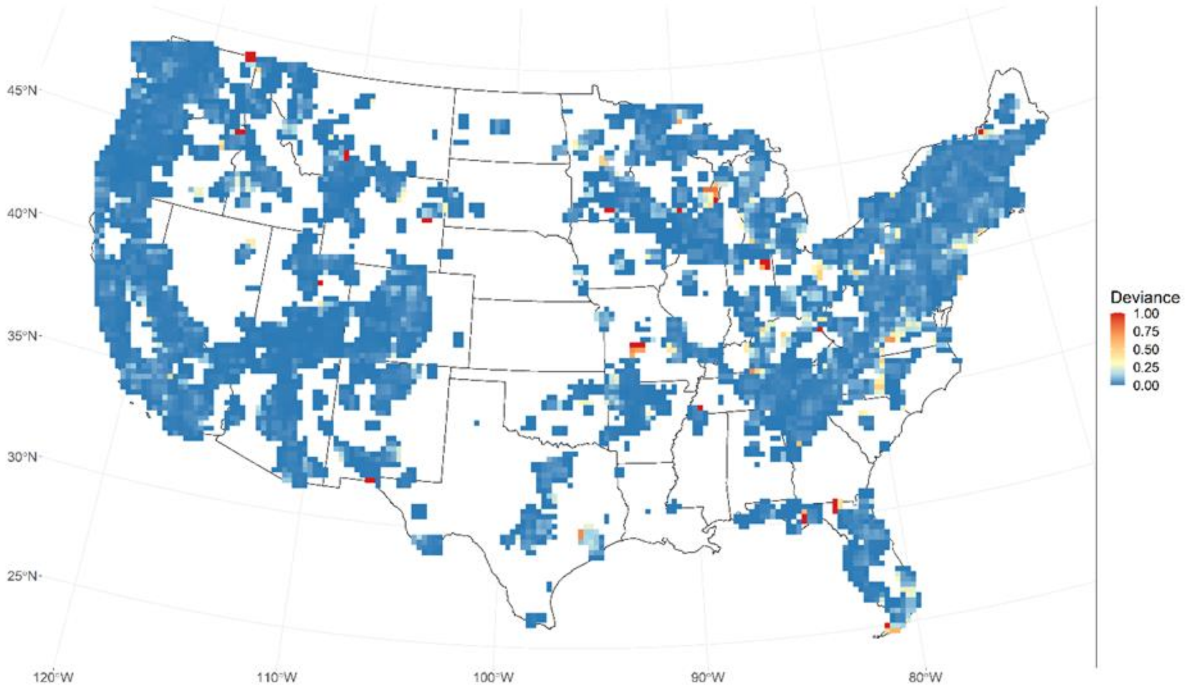


Figure 14 The local deviance in the scaled number of images from the full Flickr and *confirmedNature* images dataset (images where the percentage of content labels classified as nature were >25%) from the 1:1 line. Pixel size $25km^2$; window size 3 x 3; a break point of 0.3 was applied to indicate areas of high deviance.

5.4.3 Individual influence of filtering by sentiment only

Filtering out images with a non-positive sentiment removed a greater number of images than filtering by image contents. From the 179,700 images in the full dataset, 61,091 (34.00%) contained a non-null sentiment score, with 48,607 (27.05%) having a positive and 12,484 (6.95%) having a negative associated sentiment value. There was some difference in the local deviance values between the spatial distribution of the refined *confirmedPositive* and full dataset depending on the window size and pixel size.

Here, 19.08% of the pixels have a deviance value of greater than 0.3, indicating that in these locations, photographs of hiking on Flickr may not be a good proxy for positive images of hiking. Furthermore, the deviation of uploads from the *confirmedPositive* and full dataset is not uniform (Fig. 15), so it may not be possible to differentiate between areas where Flickr photographs are a good proxy for a positive experience and those where they are not. The random structure to the distribution (Moran's I = 0.27) suggests that there is no spatial structure to the distribution of the locations which deviate between the full and filtered dataset. As with the previous examples, the patterns of local deviance are similar across spatial resolution and window size and therefore the choices in mapping the distribution have not impacted the results (Appendix E). As there was little to no difference in spatial distribution when refining by nature threshold alone, but a larger difference when refining by sentiment alone, this indicates that differences observed in the analysis of the *confirmedCES* versus full dataset are primarily driven by the inclusion of refining by sentiment.

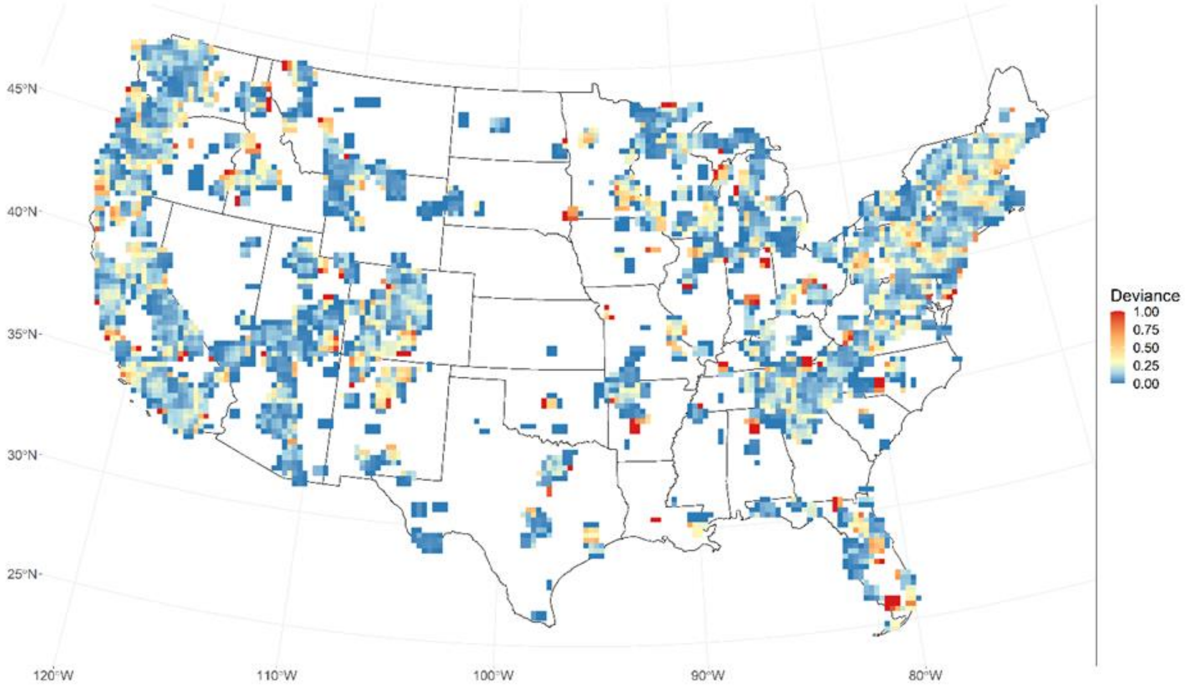


Figure 15 The local deviance in the scaled number of images from the full Flickr and *confirmedPositive* images dataset (images where a positive sentiment was expressed in textural metadata) from the 1:1 line. Pixel size 25km^2; window size 3 x 3; a break point of 0.3 was applied to indicate areas of high deviance.

5.4.4 Validation of methods

The manual validation of the image content and textual sentiment analysis indicates that the automated methods have high accuracy. When assessing images classified as presenting human-nature interactions we agreed with 98% of the images categorised this way at the 25% biophysical nature label thresholds (*confirmedNature)*: and 100% at 50% and 75% threshold (Table 6). However, when assessing images deemed not to be focused on human-nature interactions (*nonNature)* the method incorrectly included some images containing human-nature interactions, particularly when using a threshold of 50% and 75% biophysical labels. When using a 25% threshold, the images that were incorrectly labelled as *confirmedNature* were of artwork or indoors artificial water features. For the images that were incorrectly included as *nonNature,* the images tended to be where the biophysical nature features were out of focus (e.g. a photograph focused on a person in the foreground, with a small amount of scenery in the background) or where the image had been edited (e.g. in black and white or with text over the image). A threshold smaller than 25% would start to include more images, that though contain biophysical nature labels, do not relate to human-nature interactions as *confirmedNature* (e.g. images of pets, indoor plants or artificial water features). There was a range of correctly identified *nonNature* images including indoor images (e.g. furniture), non-hiking activities (e.g. indoor music events or an American football game inside a stadium), food and drink (e.g. packets of food), photographs of an object (e.g. a photograph of another photograph) or art (e.g. generated images), all having a biophysical nature label percentage of between 0 to 25%. In general, using a threshold of 25%, therefore, provides a good balance of excluding most images not related to human-nature interactions whilst not incorrectly excluding the large number of human-nature images that using thresholds of 50% and 75% did.

Table 6 Validation of automated filtering methods

	confirmedNature		**nonNature**		**Both**
Threshold	True-positive	False-positive	True-negative	False-negative	Overall accuracy
25% biophysical nature labels	98.00%	2.00%	70.00%	30.00%	84.00%
50% biophysical nature labels	100.00%	0.00%	39.00%	61.00%	69.50%
75% biophysical nature labels	100.00%	0.00%	26.00%	74.00%	63.00%

For the positive expressed sentiment there was a 96% agreement between the automated process and manual interpretation. Many posts contained positive sentiments (e.g. “beautiful”, “exciting” or “wonderful”). Where there were differences in our manual validation and the automated method, the posts tended to include a location name that was inherently expressing a positive sentiment but no indication of experience was given (e.g. “walk at Lucky Boy Vista”) or a mixed sentiment expressed in the text (positive and negative experience e.g. “nice view but overall the hike was terrible”). For the non-positive textual sentiments, the automated process was 92% accurate. Many posts contained no indication of sentiment, just simple descriptive phrases (e.g. “hiking along the river”), though some positive post were incorrectly included as the place or feature has an inherently non-positive connotation though the user expressed a positive view of it (e.g. “fun at Lost Creek Lake” or “beautiful poison oak”), or where users used a double negative (e.g. “this view is not bad”). Overall, the automated sentiment value provides a good indication of whether a positive sentiment expressed in the text.

5.4.5 Enriching spatial distribution

Through enriching the social media data with the inclusion of textual sentiments, we were not only able to plot the distribution of the number of uploads but assess which locations have the highest associated sentiment value (Fig. 16). The Pearson's correlation between the *confirmedCES* uploads and *confirmedCES* mean sentiment raster maps indicates a weak to no correlation (R = -0.004, t = -0.21, df = 2349, p-value = 0.83). This suggests that there is no correlation between areas with a high number of uploads and areas that have a high associated sentiment value.

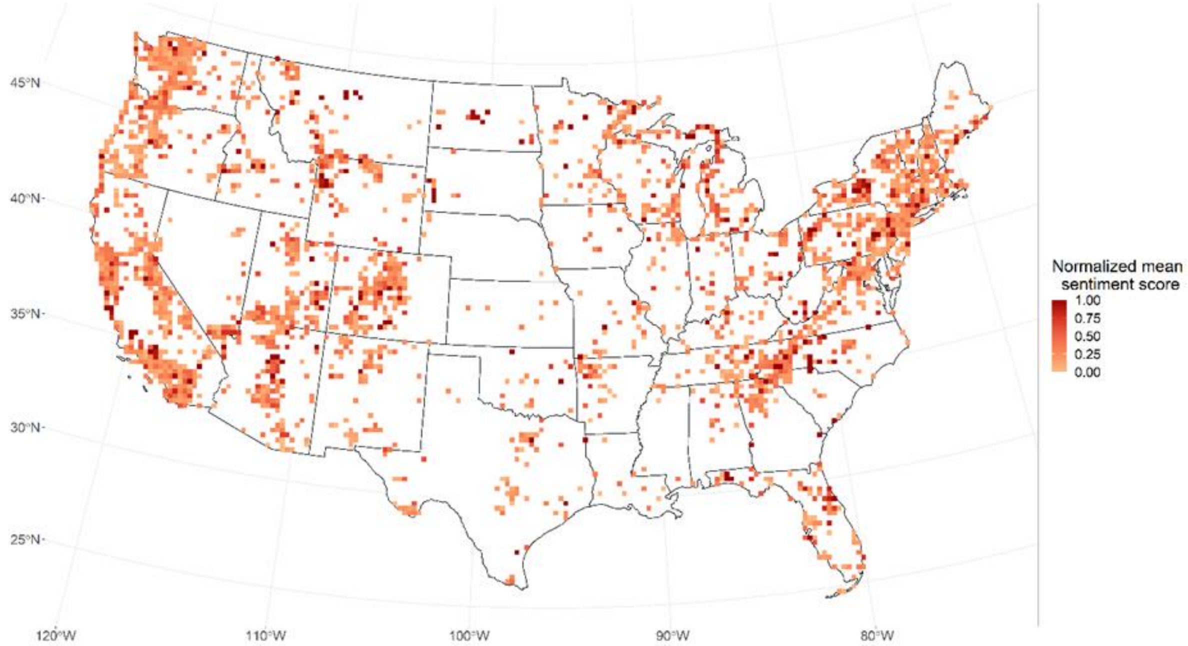


Figure 16 Mean sentiment value of *confirmedCES* images (images where the percentage of content labels classified as nature were >25% and a positive sentiment was expressed in textural metadata), results were normalized (pixel size 25km^2)

5.4.6 Enriching content analysis

When grouping images by a single user in a single day, 67 labels appeared in at least 5% of these *confirmedCES* images. From the cosine dissimilarity coefficient of the labels, the silhouette index indicated that when using Ward's D clustering, ten categories for the labels are suitable (Table 7). The ten classes can be summarized as: "mountains and hills", "forests", "views of lakes", "snow covered mountains", "vegetated trails and parks", "hydrological features", "geological or arid landscapes", "human-geology interactions", "grasses" and "flowers". The individual clusters tend to only contain either biodiversity labels (e.g. "flowers") or geodiversity labels (e.g. "mountains and hills"), indicating that the focus of images may only capture one aspect of biophysical nature.

Table 7 Google Vision Cloud API labels clustered into ten classes using Ward's D clustering method.

Cluster	**Google Vision Cloud API labels**
Mountains and hills	Cloud, highland, hill, hill station, mountain, mountainous landforms, sky, wilderness
Forests	Biome, branch, forest, green, leaf, nature reserve, natural environment, northern hardwood forest, old-growth forest, tree, tropical and subtropical coniferous forests, woodland, woody plant, trunk
Views of lakes	Lake, Canidae, natural landscape, nature, reflection
Snow covered mountains	Alps, mountain range, ridge, snow, winter
Vegetated trails and parks	Adaptation, botany, landscape, plant, plant community, shrubland, soil, state park, trail, vegetation, wildlife
Hydrological features	Body of water, river, stream, water, watercourse, waterfall, water resources
Geological or arid landscapes	Badlands, canyon, formation, geological phenomenon, geology, national park, rock wadi
Human-geology interactions	Adventure, atmospheric phenomenon, bedrock, outcrop, recreation
Grasses	Grass, grassland
Flowers	Flower, flowering plant

When categorising images based on which cluster's labels were most dominant in the photograph, there were 165 different combinations of the most dominant cluster (Fig. 17). The clusters that were most frequently photographed were "mountains and hills", "forests", "geological or arid landscapes" and "hydrological features". Many of the cluster combinations have large standard deviations, though some combination of clusters (e.g. "views of lake/human-geology interactions") have relatively low standard deviation values.

Though the most photographed cluster was “mountains and hills”, this cluster has a relatively low sentiment score when the sole dominant class in the photograph. However, there is a high relative sentiment associated with images containing both geomorphological and hydrological features (e.g. “mountains and hills/views of lakes” and “mountains and hills/hydrological features”), indicating that interactions between these features may be important at driving a positive hiking experience. The cluster of images containing “forests/vegetated trails and parks/geological or arid landscapes” had the highest mean sentiment, suggesting that a combination of geodiversity, biodiversity and human features can interact to provide a highly positive hiking experience. There is a relatively low sentiment value associated with “forests” when on their own, however, there is a higher sentiment in images containing both forests in combination with other clusters (e.g “mountains and hills/forests/hydrological features” and “mountains and hills/forests/vegetated trails or parks”), indicating that forests alone may not provide a relatively positive hiking experience, but forests in combination with a trail or park, or with geomorphological features (such as mountains) may help to increase a positive hiking experience. Other aspects of biophysical nature, such as “geological or arid landscapes” and “hydrological features”, are both highly photographed with a relatively high mean sentiment score. Many of the cluster combinations have large standard deviations, though some combination of clusters (e.g. “views of lakes/human-geology interactions”) have relatively low standard deviation values.

Figure 17 Count of images, their mean sentiment value and standard deviation of sentiment value for the 25 most frequently photographed clusters, where a user's images from a single day were categorised based on which cluster most of their assigned Google Vision Cloud API labels belonged to.

5.5 Discussion

Metadata from social media websites such as Flickr provide a source of big data for many CES applications. Being able to use the full dataset as a proxy for images of human-nature interactions may benefit researchers looking at general patterns by removing the need to manually or automatically tag image contents, which can be time-intensive or financially expensive (Richards and Tunçer 2018). However, this data may not accurately represent the CES and studies, therefore, need to refine data from social media and remove posts that are inappropriate for CES analysis (Oteros-Rozas et al. 2018). These studies assume that the content of the image is representative of a user's interactions with CES (Langemeyer et al. 2018). Therefore, content analysis alone provides incomplete information and does not allow one to fully untangle the human-nature relationship that drives CES. Furthermore, filtering based on textual analysis can be useful to subset CES images (Ghermandi et al. 2020a), but refining based on sentiment alone results in a dataset that includes non-CES images. Refining social media datasets by text alone also does not provide a complete and valid CES dataset (Chen et al. 2020). Through enriching the filtered images of nature with their associated textual sentiment value, we can start to provide a more robust understanding of CES distributions, and the aspects of nature that provide a positive experience.

For the content analysis, we aimed to refine our dataset to represent images of human-nature interactions, we, therefore, included images containing any aspects of biophysical nature. Studies with different aims should filter their images accordingly, for example, Oteros-Rozas et al. (2018) chose to include only images of landscapes in their study and removed images that featured animals. Some CES studies have stated a preference for using images from Panoramio as it prevents uploads of images of people and synthetic objects and therefore did not need to be refined (Casalegno et al. 2013; Martínez Pastur et al. 2016). However, as of 2016 Panoramio is no longer available. Here, we found little difference in the distribution of the *confirmedNature* and full datasets, with low deviance between the two groups for most of the US. This suggests that Flickr may therefore be a good proxy for images of biophysical features across a range of spatial scales. However, studies that use Flickr, or other social media sources, need to acknowledge that the returned dataset may contain images that are unsuitable for their specific CES assessment and should ensure these images are refined out accordingly. Overall, we found that a threshold of 25% of biophysical nature labels has provided a suitable set of images representing human-nature interactions, while generally excluding images that are not of CES. By not omitting all human features (e.g. Panoramio), a threshold of 25% biophysical nature labels can provide results that allow for more robust recommendations for improving the sustainable management of these complex human-nature interactions (Bennett et al. 2015; Palomo et al. 2016).

Overall, refining by sentiment value has a larger impact on the spatial distribution and size of the dataset than refining by image content. This suggests that the full Flickr dataset, at least for hiking, may be sufficient to show overall patterns of the distribution of images of human-nature interactions, but not necessarily a positive CES experience. Though there are some areas of low deviance between the number of uploads between the full Flickr and *confirmedCES* datasets, there are many areas of disagreement. Furthermore, as the spatial distribution of these areas is random it means that it may be difficult to account for the areas of high disagreement without refining the dataset. For example, if the distribution of high deviances were clustered, it may be appropriate to use the full dataset but filter out only the areas of high disagreement. The results for the *confirmedPositive* set of images were parallel to the *confirmedCES* images, with areas of high deviance randomly distributed across the US. The inclusion of sentiment, therefore, has implications for the number and distribution of images, (e.g. in the north western states where there was high deviance between the number of confirmedCES and the full number of images). To better understand positive human-nature interactions, we recommend that social media data be refined in a two-step process, filtering by content and enriching with sentiment value.

A more robust method of mapping CES occurrences, such as the one presented here, could help to inform policy and decision-makers (Clemente et al. 2019). For example, when the number of uploads is aggregated to a given location to understand visitation rates (e.g. national park, land cover or state level), it could be interpreted that these regions provide the largest supply of CES and are therefore the most important for future management strategies (Figueroa-Alfaro and Tang 2017). Here, we have highlighted a discrepancy between the number of images uploaded and the mean sentiment value expressed by textual metadata of the images in the area. Higher uploads of images taken in each area may indirectly be influenced by accessibility (Richards and Tunçer 2018), and therefore may not necessarily represent the sites with the highest sentiment value. By enriching the data with sentiment value, we can evaluate the distribution of CES in ways other than purely measuring the number of visitors. For example, as areas with a higher number of uploads are potentially more vulnerable to damage (Hausmann et al. 2019), mapping areas of high sentiment in combination with visitation rates means management decisions can be better targeted to alleviate pressure from overused locations (Clemente et al. 2019). Furthermore, the textual sentiment value could be used in conjunction with the temporal metadata to assess changes in visitor opinions over time (Becken et al. 2017).

Some of the clusters of image contents identified as being frequently photographed or having a high associated sentiment have already been widely explored in CES literature. For example, Van Zanten et al. (2016) found that some of the best predictors of recreational value at a landscape scale were geomorphological features such as hills and mountains and Oteros-Rozas et al. (2018) found that mountains are particularly associated with hiking. Furthermore, the impact of being close to water and vegetation on hiking has also been widely explored (Pastorella et al. 2017; Schirpke et al. 2018a; Aiba et al. 2019). Photographs that contain a combination of clusters can provide information on the interactions between different biophysical features which give rise to CES. For instance, some hikers may prefer natural mountainous areas with forest cover and others may prefer natural mountainous areas closer to water (Pastorella et al. 2017; Schirpke et al. 2018a).

By including a measure of sentiment we can start to understand how the interactions between biophysical and human features influence the hiking experience. For example, the high sentiment value associated with the cluster “forests/vegetated trails and parks/geological or arid landscapes" further demonstrates that a positive experience of CES may be enhanced through its co-production with people (Fischer and Eastwood 2016). The high standard deviations in sentiment value between images classified as different combination of clusters are unsurprising as CES experience is unique and varies between individuals (Daniel et al. 2012; Havinga et al. 2020). For example, the relationship between geomorphological features and hiking is complex, often with factors such as elevation, slope and landforms having site-specific influences on hiking experiences depending on people motivations for hiking (Chhetri 2015; Wilcer et al. 2019).

Some of the clusters identified here, particularly those relating to geodiversity features such as the geological features class, are not as well explored in CES literature, (Fox et al. 2020b). For example, the labels in the clusters associated with "geological or arid landscapes" resemble those of landscapes in southwestern states, such as Arizona and New Mexico, where there is a high number of *confirmedCES* images. For example, the Grand Canyon National Park in Arizona is dominated by iconic geodiversity landscapes and is one of the most visited tourist attractions, not just in the USA, but worldwide (Gray 2008a). These landscapes are often dominated by canyons with exposed strata, a variety of slope morphologies and talus and scree on the canyon floor and are popular hiking destinations (Gray 2008a). By enriching the social media data with the textual sentiment value we can start to understand the relative importance of these features to a positive hiking experience. Here, "geological or arid landscapes" were frequently photographed, they were less frequent than "forests". It does however provide a relatively more positive hiking experience than "forest" when either was the sole dominant cluster. This suggests that within the USA, though hiking in vegetated areas may be more common than through geological landscapes, they both can provide a positive recreational experience. In other countries where geological or arid landscapes are the prevailing ecosystem, the relationship between geodiversity and biodiversity on the hiking experience may be different. Future work should therefore aim to quantify the relationship of these underrepresented landscape features to CES experience across different study sites.

Here, our analysis has shown that the grouped images (based on a single user's photographs on a single day) can contain multiple different content label clusters. We note that here the individual clusters of labels focus on biodiversity and geodiversity features separately, though this may be a facet of the clustering procedure or the method in which the machine learning classifier labels images. As with other studies which clustered photograph content (Lee et al. 2019) photographs were frequently assigned multiple equally dominant clusters, indicating an interaction between geodiversity, biodiversity and human features. It may not always be possible to include features of the multiple clusters in a single image. For example, a hiker may be interested in photographing large scale vegetation such as a forest from the top of a mountain (Aiba et al. 2019), but these images may omit the mountain itself. We suggest that these relationships may not be properly captured by assessing single photographs and therefore, recommend that future studies assess each Flickr user's images from a single day, such as used here, or by grouping all photographs from a smaller study site as a collective.

Refining by different nature thresholds, spatial resolution and window size all had some impact on the deviance between the refined dataset and the remaining dataset. However, across all combinations of refined datasets, there were no major changes in the spatial distribution and range of values expressed by the local deviance values. The smaller deviance associated with decreasing resolution (increasing pixel size) is down to the aggregation of data – aggregation of data at the coarser resolution loses the spatial structure of the distribution and the pixel values become more homogenised (Bian and Butler 1999). This relationship is also true for larger moving windows producing lower deviances (Graham et al. 2019). We recommend that researchers be aware of the impacts of their mapping decision – if data aggregation is to be carried out researchers need to choose a suitable scale for the study (Graham and Eigenbrod 2019).

This study further demonstrates that social media data can easily provide large spatial-temporal dataset compared to traditional methods (Fox et al. 2020a), however, future work could implement this over small spatial scales to assess regional or local trends. One limitation of social media data is that not all demographics are well represented (Oteros-Rozas et al. 2018) and management and policy decisions should acknowledge this. Studies could therefore combine results of social media studies with other datasets such as survey data (Graham and Eigenbrod 2019; Moreno-Llorca et al. 2020; Sinclair et al. 2020b). A further potential limitation of this multi-faceted filtering process is that although we can be more confident that the images we are assessing represent CES, there are substantially fewer suitable images in the refined dataset compared to the original dataset downloaded from Flickr. However, we still had a large sample size of 43,427 images for one activity over 5 years. Other studies have found success when using smaller subsamples of the full dataset (Langemeyer et al. 2018; Chen et al. 2020). We recommend that social media data can best capture CES when researchers refine their dataset by the content of the images and enrich the data with a measure of sentiment, which can provide a reliable representation of CES.

5.6 Conclusion

The method tested combining the analysis of the contents of social media (Flickr) images and the sentiment value expressed in their textual metadata reduces the amount of available data but provides a robust measure of CES for large scale spatio-temporal studies. Through the application of the refined dataset, images of hiking in the USA were clustered into groups representing the interactions between people and biophysical nature. Hiking as a CES is driven and maintained by complex interactions between biodiversity and, geodiversity and people, and by refining data on two levels, our methods allow us to start to unpack which feature contribute to a positive experience. The results of this study encourage future social media and CES studies to acknowledge that the full social media dataset may not be suitable for their chosen study, as though it may be a good proxy for images of nature it may not be a good proxy for a positive experience. It is suggested that future work enrich their social media datasets with a measure of textual sentiments to provide a more robust representation of positive human-nature interactions.

Chapter 6 Using social media datasets to assess the relationship between geodiversity and recreational ecosystem services in Wales, UK

This chapter is presented as a reformatted version of the manuscript submitted to Geoheritage on the 2nd of May 2021, which is currently under review.

6.1 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 whilst 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 whilst hiking, this analysis can help to inform future geoconservation methods by focusing efforts on these important features.

6.2 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. 2020b). 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 management (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). Geoconservation includes: 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 ES – the benefits and values 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 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, ES and anthropogenic threats.

To develop our understanding of the relationship between geodiversity, ES, 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 just 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 ES (Van Zanten et al. 2016; Van Berkel et al. 2018) and recreational ES (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 now take an ensemble approach by combining 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.

6.3 Methods

6.3.1 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 of anthropogenic structures or people, whilst 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 classifications see Appendix D). 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 25% 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 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.

6.3.2 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: 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 8). 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 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 10km (Graham and Eigenbrod 2019), here, predictor variables were summarised into 10km^2 grid cells, e.g. elevation range within the 10km^2 grid cell. To transform the individual variables raster to a comparable scale, each of the raster maps were normalized to a 0-1 scale (Appendices G and H).

Table 8 Predictor variables chosen for species distribution modelling.

Category	Variable	Description
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)
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 differnt 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)

6.3.3 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 standardised 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 a 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.

6.3.4 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 indices 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. 18).

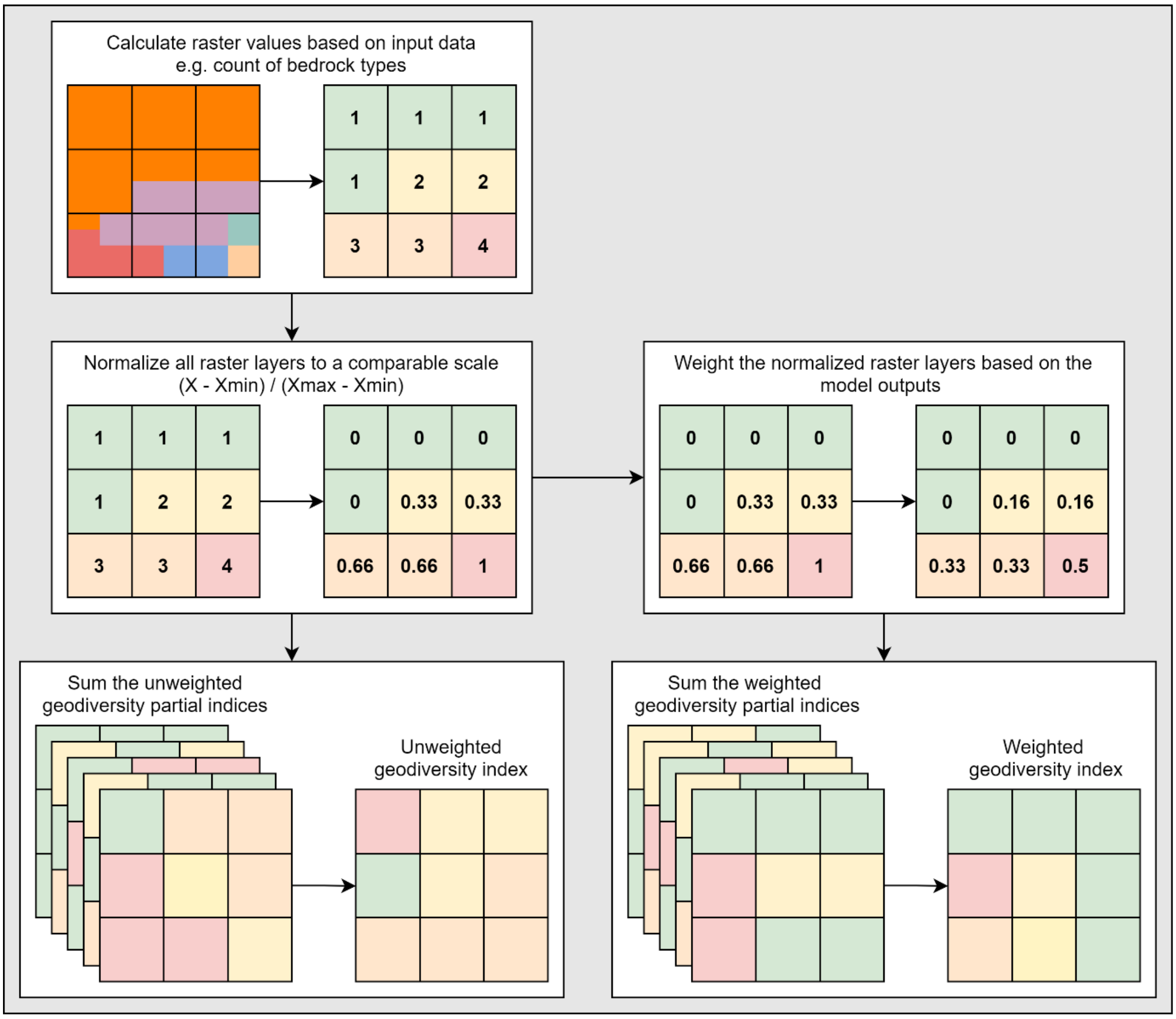


Figure 18 Methods for creating unweighted and weighted geodiversity indices.

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 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).

6.3.5 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 whilst 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 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.

6.4 Results

6.4.1 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. 19).

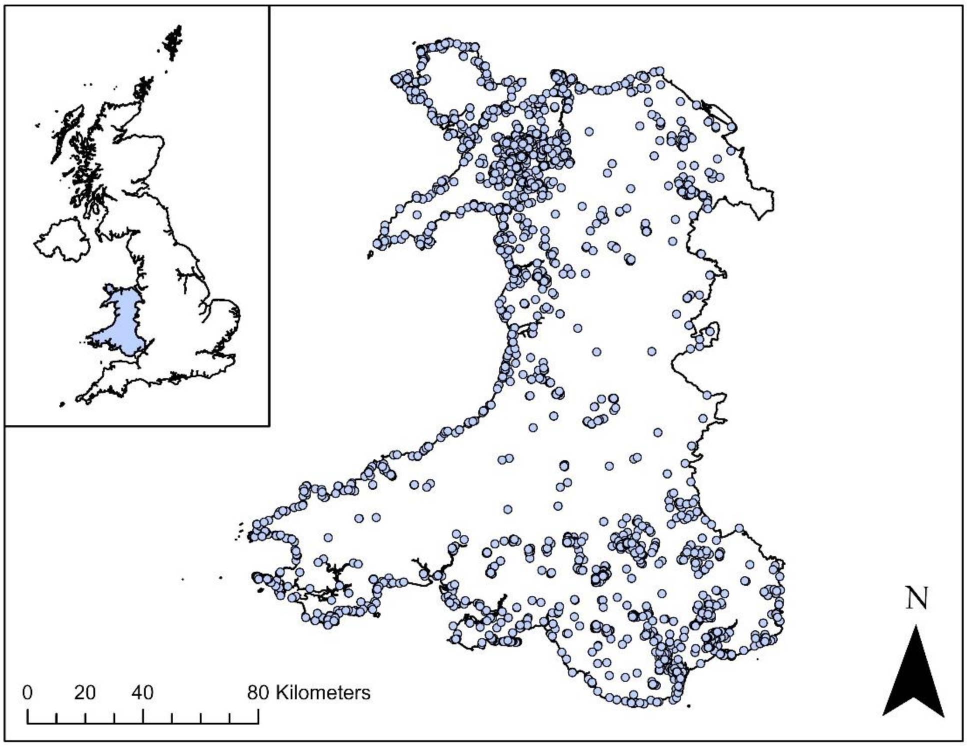


Figure 19 Distribution of hiking images in Wales; these images are classed as a CES as they have both a focus on biophysical features of nature and contain a positive textual sentiment expressed in the metadata. The inset map shows the location of Wales in comparison to the rest of the UK.

6.4.2 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 (Appendix I).

The order of the most important predictor variables varied with the model algorithm used, though overall the distance to the coast was consistently the most important predictor (Fig. 20; for individual models see Appendix J). 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.

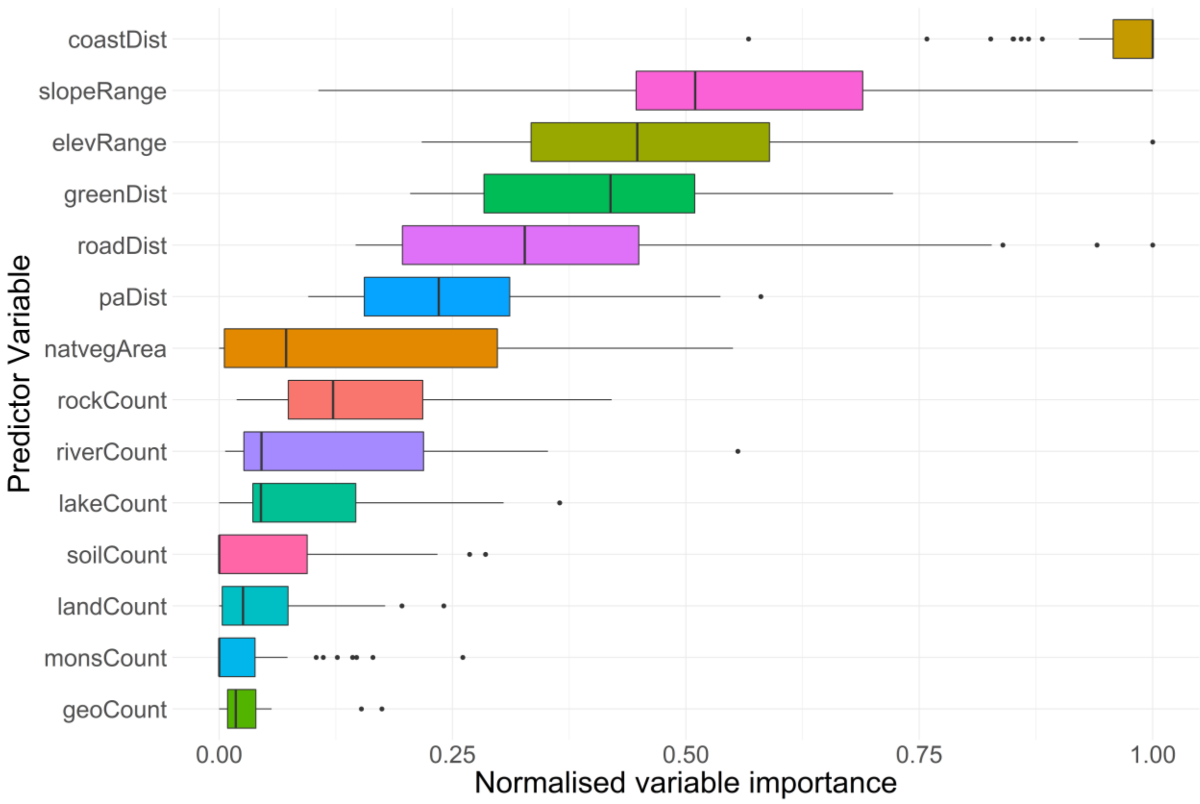


Figure 20 Normalised variable importance for ANN, CTA, GBM and RF (model algorithms with a mean kappa statistic and TSS value >0.6 and AUC value > 0.8).

Inspection of the response curves demonstrates that the probability of images of hiking decreases with distances further away from the coast (Fig. 21, for individual model responses, see Appendix K). 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.

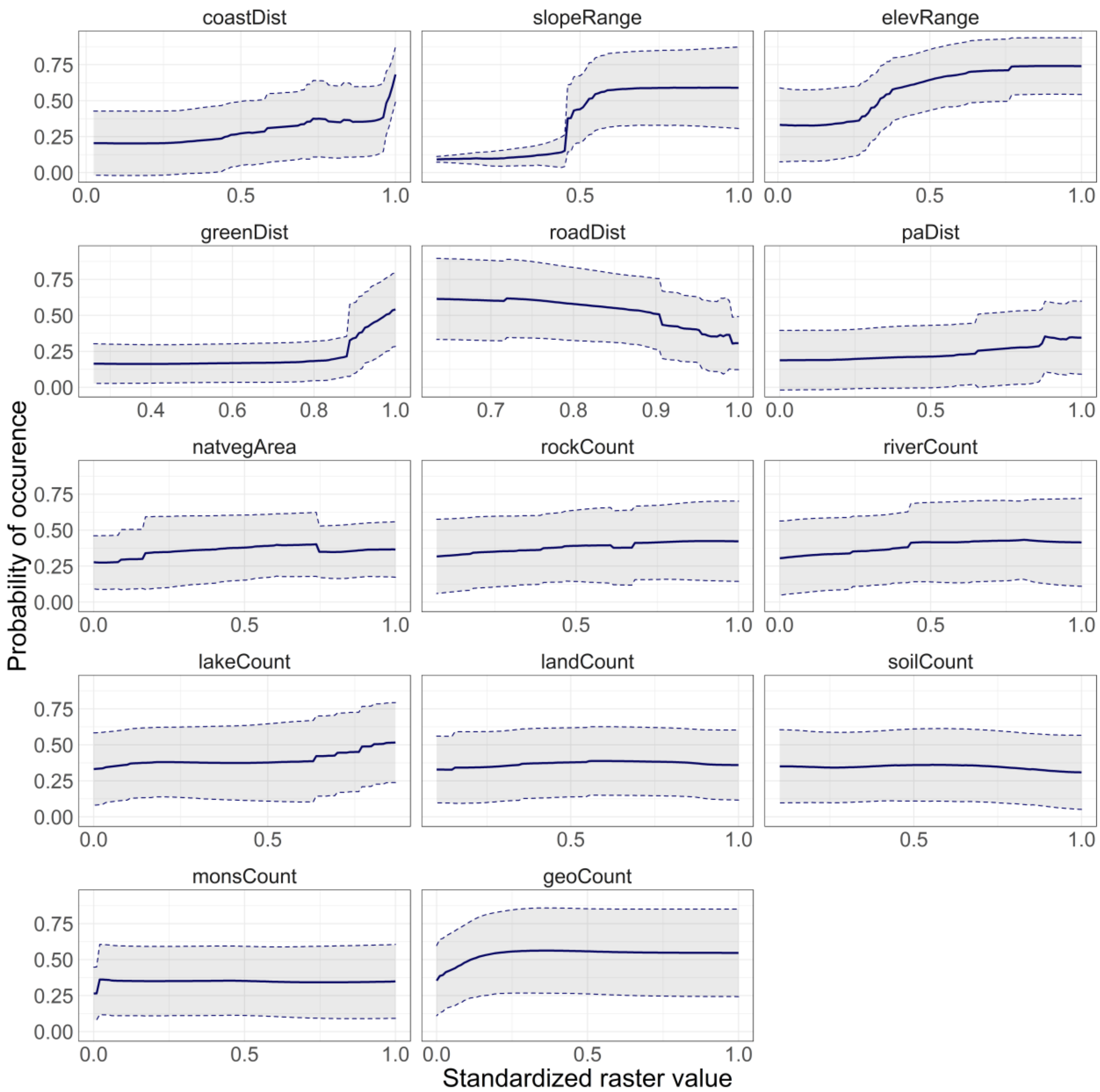


Figure 21 Mean, +/- 1 SD response curves for all variables from all runs of the ANN, CTA, GBM and RF models. The normalised raster values were created for comparable scale for each variable summarised within the 10km^2 square: for distance variables, the larger normalised values are, the closer to the 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.

6.4.3 Geodiversity map

The final geodiversity weightings vary based on the method chosen (Table 9). 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. 22). In general, the two indices agree (45.41% of pixels remained unchanged), with both indices showing very high geodiversity in Snowdonia National Park. 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.

Table 9 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

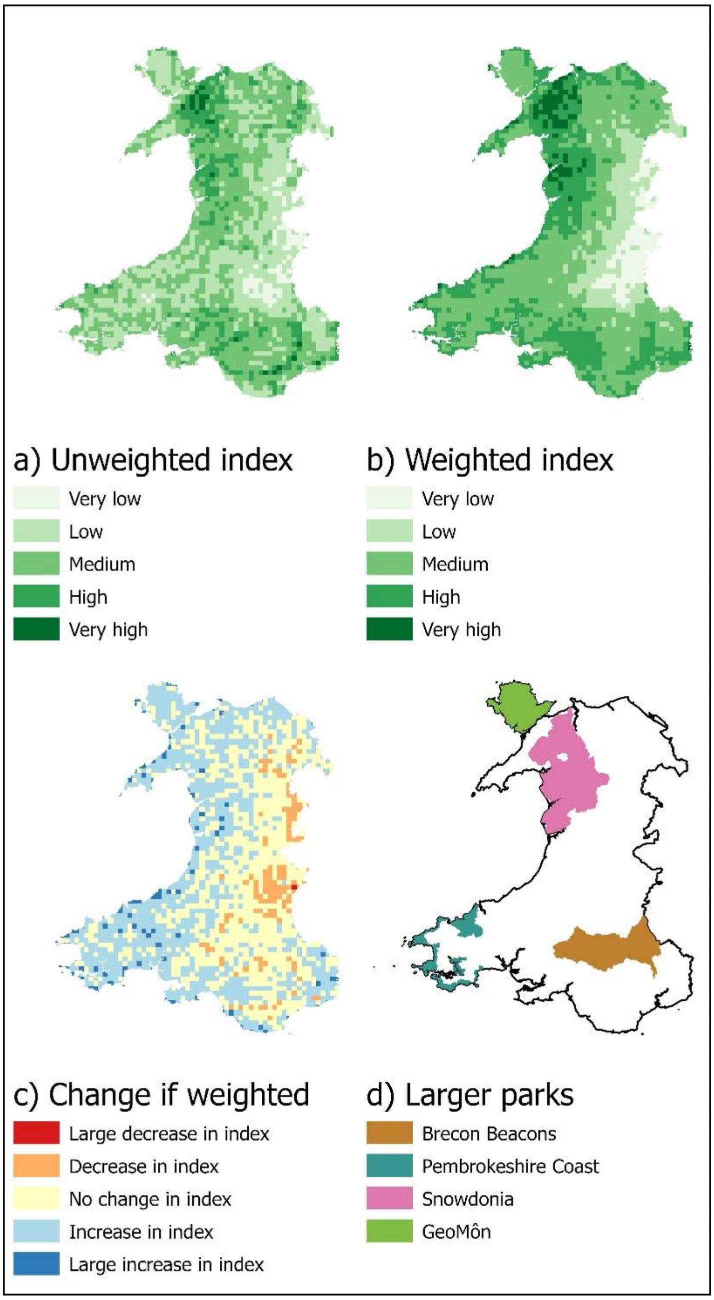


Figure 22 Normalised geodiversity index – partial indices: count of lakes; count of rivers; Euclidian distance to coast; range in elevation; range in slope; count of landscape types; count of bedrock types; count of geosites; and count of soil types. a) partial indices unweighted, b) partial indices 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).

6.4.4 Human nature interactions

In total there were 1,105 different image content labels returned by the Google Vision Cloud API. Whilst hiking, people take photographs of both geodiversity and biodiversity features (Fig. 26).

For geodiversity, people more frequently photograph elements of topography and water bodies. Furthermore, images containing river landforms (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. 23). Images associated with “building”, “nature” and “coastal and oceanic landforms” have the largest associated mean sentiment. Images of topographic features such as “highland” and “mountain”, as well as 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.

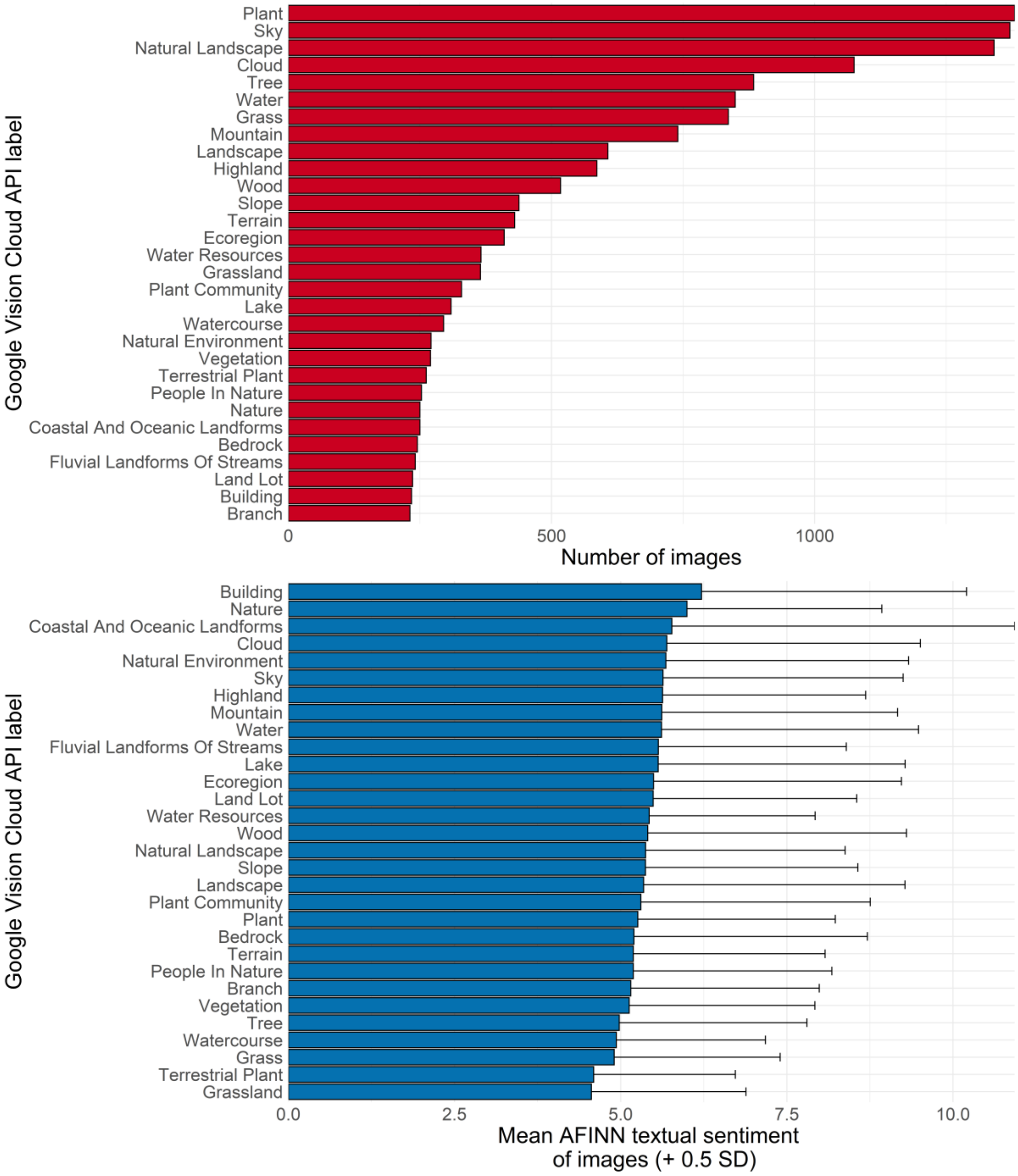


Figure 23 The 30 most frequently photographed features based on the Google Vision Cloud API labels, and the mean textual sentiment of all images containing those features.

6.5 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 geoconservation efforts should be focused and which features of geodiversity should be prioritised.

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 in a coastal-inland gradient that beaches are the most visited areas and Ghermandi et al. (2020b) found a higher positive sentiment expressed in coastal images when compared to other landscape types. 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 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. 2018a). Though here we found that number of lakes and rivers 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, 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), found observations 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 are relatively important in the distribution of CES in Wales. Furthermore, as with other studies, we also found that vegetation cover recreation is generally not the most important driver 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. 2018a; 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).

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 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).

6.6 Conclusions

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 recreation with both distance to roads and distance to greenspace access points contributing 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 whilst 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.

Chapter 7 Reddit: A novel data source for cultural ecosystem service studies

This chapter is presented as a reformatted version of the manuscript accepted for publication in Ecosystem Services on the 15th of June 2021.

7.1 Abstract

Social media sites have been gaining traction as a source of novel data for environmental research, particularly for cultural ecosystem service (CES) assessments. However, Reddit, a discussion-based site, has yet to establish itself as an important source of data for CES research, possibly due to researchers not being aware of its potential applications or because Reddit posts lack georeferencing information. Here, we demonstrate how researchers can search Reddit for CES datasets related to recreation and how specific pages on Reddit may provide data for other CES such as aesthetics. Using named-entity recognition, we developed an automated method of geocoding the approximate location of where images in Reddit posts were taken. Furthermore, we compare posts from Reddit and Flickr for a range of recreational activities and compare the content and textual metadata of images relating to hiking. Though there is potential for Reddit data to be used in spatial analysis, we highlight the limitations associated with georeferencing posts. We recommend that data from Reddit is best suited to assessing general trends in CES, either for a given service or place. By demonstrating the value of big data from Reddit we hope to encourage its inclusion in future CES and environmental research.

7.2 Introduction

Big data from social media sites has multiple benefits over conventional methods of data collection for environmental studies, providing access to large spatio-temporal scale datasets, through inexpensive and quick data collection methods (Barve 2014). The use of social media data is therefore becoming more prominent in environmental research, ranging from the use of Twitter to understand animal life cycles (Hart et al. 2018) and prepare for natural hazards (Wang et al. 2016; Mendoza et al. 2019), to Flickr being used to assess species niches (Penã-Aguilera et al. 2019) and map invasive species (Allain 2019). One of the biggest applications for social media data in environmental research has been the assessment of cultural CES (Ghermandi and Sinclair 2019).

CES are the non-material goods and benefits obtained through nature and are derived from the interaction of biodiversity (biotic nature) and geodiversity (abiotic nature) (Gray 2011, 2012; Fox et al. 2020b). CES include aesthetic value, recreational services and a sense of place and can enhance physical and mental well-being (Haines-Young and Potschin 2010). They can deliver multiple benefits for both residents and tourists, supporting local and regional economies (Schirpke et al. 2016; King et al. 2017). However, the exploitation, destruction and consumption of natural landscapes by humans for activities such as intensive agriculture, urban development and recreational activities can damage ecosystems and reduce their capacity to provide CES (Figueroa-Alfaro and Tang 2017). Furthermore, our understanding of CES is more limited than that of provisioning and regulating ES (Milcu et al. 2013; Gordon et al. 2018), particularly because our interactions with CES are subjective and vary between individuals, which makes obtaining practical measurements of their benefits and values difficult (Daniel et al. 2012; Havinga et al. 2020). By developing a better understanding of the natural and social drivers of CES we can help inform policy and management strategies to alleviate the threats to their sustainable use (Clemente et al. 2019). Researchers, therefore, need to understand better the supply and demand of these services over relevant temporal and spatial scales (Langemeyer et al. 2018). Here, social media datasets provide relatively quick and cost-effective data collection for assessing CES, versus traditional methods, and provide novel approaches to assessing how CES are generated as well as their perceived benefits and values over a range of spatial and temporal scales (Wood et al. 2013; Ghermandi and Sinclair 2019; Fox et al. 2020a).

Due to the vast quantity of data available on social media websites can be viewed as a source of big data and therefore benefit from the emergence of big data approaches to assessing human-nature relationships (Retka et al. 2019). Social media sites, including Twitter and Weibo (microblogging sites), Flickr, Instagram and Panoramio (image sharing sites), have already been widely used to assess a range of CES. Aesthetic value has been assessed through textual metadata from Twitter (Johnson et al. 2019), image and geographic distribution from Instagram (Guerrero et al. 2016; Chen et al. 2020), and image content and geographic distribution from Flickr (Figueroa-Alfaro and Tang 2017; Tieskens et al. 2018). Recreational preferences have been studied using Flickr (Van Zanten et al. 2016; Gosal et al. 2019; Graham and Eigenbrod 2019) and Weibo (Zhang and Zhou 2018). Furthermore, Flickr has also been used to assess changes in cultural values over time (Thiagarajah et al. 2015) and identify trade-offs between multiple CES (Allan et al. 2015). However, some social media sites have either ceased operating (e.g. Paramio) or introduced restrictions to accessing data (e.g. Instagram) and therefore Flickr is becoming the main source of data for CES studies (Langemeyer et al. 2018; Retka et al. 2019).

Metadata available from Reddit, the social news aggregation and discussion orientated social media website, has been used in a vast array of scientific studies across a range of disciplines (Baumgartner et al. 2020), including health and psychology (Jamnik and Lane 2017; Park et al. 2018), technological development (Derczynski et al. 2018; Völske et al. 2018) and political studies (Guimarães et al. 2019). Reddit, which is broken up into different forums or “subreddits'' themed around different topics, allows for a user to post a range of media such as images and text posts. These posts, along with their associated metadata, draws parallels to the types of data from other social media sites that are currently used in CES studies. However, there appears to have been little to no application of big data from Reddit to assess any ecosystem service. A systematic review of the applications of social media data in environmental research did not include any studies using Reddit as a source of data (Ghermandi and Sinclair 2019). A search of the titles abstracts and keywords on Web of Knowledge (https://wok.mimas.ac.uk) and Scopus (www.scopus.com) for “ecosystem servic*” (the * denotes any end to the term e.g. service or services) AND “Reddit”, carried out on 10th February 2021, returned no results.

As there have been few studies comparing social media sources for CES, there is a need for a greater understanding of the impacts of differences in data availability and biases among the various social media sites used as data sources (Oteros-Rozas et al. 2018). We, therefore, find it surprising that big data from Reddit has yet to be explored in the context of CES, though we postulate that this may be for two key reasons: researchers not being familiar with the website and its potential uses; and that posts on the website are not georeferenced. In this paper, we aim to provide an overview of Reddit and to compare data from the site with that from another social media site, Flickr. We provide examples of how data from Reddit can be used to assess recreational, aesthetic, spiritual and cultural CES and address how Reddit can be a novel source of data for commonly used CES methods such as assessing image contents and textual sentiment. We also provide an insight into the potential uses and limitations of Reddit for spatial assessments.

7.3 Methods

Here we present multiple methods for searching Reddit for data suitable for CES assessments via its Application Programming Interface (API), a computing interface that allows researchers to access a platform via code. First, we searched the site for all posts containing a specific keyword and compared these posts to those found using the same keyword search on Flickr. Second, we searched for posts on subreddits that are based around topics of interest to CES research. Third, we demonstrate a method for geocoding an estimated location for posts from Reddit as well as combining a place keyword search with another keyword, or within a subreddit to find posts linked to a particular location.

7.3.1 Data sources

7.3.1.1 Reddit

Reddit is a social media site consisting of over two million different communities called subreddits (Table 10). Subreddits are built around a topic, each with its own rules on posts and comments. The type of post is highly variable among subreddits. For example, the subreddit “r/EarthPorn” is limited to photographs of landscapes, accompanied by a text title and a comment section, whereas the subreddit “r/Culture” hosts primarily text-based posts with a title and a comment section.

Posts and comments from Reddit can be searched and returned through the Reddit API, with text and image posts, as well as their metadata including the title, comments, date posted, how many upvotes (the number of people that like a post) a post has, and the ratio of upvotes to downvotes (the number of people that dislike a post). These data types are similar to data already being used in CES and social media studies derived from Flickr, Instagram and Twitter.

Accessing data from Reddit has multiple benefits for researchers. First, data from Reddit is freely available. Second, the data is accessible across multiple software tools and programming languages. For example, the Pushshift tool (Baumgartner et al. 2020) provides researchers with an accessible method for querying and retrieving data. The tool also benefits researchers by providing built-in functionality which overcomes Reddit’s 100 object limit per search. For researchers familiar with writing scripts, functionality for searching the Reddit API is available in multiple programming languages: packages “RedditExtractoR” (Rivera 2019) and “rreddit” (Kearney 2020) for the R environment; “Python Reddit API Wrapper” (Boe 2020) within the Python environment; “jReddit” (jReddit 2020) within the Java environment.

Table 10 Selected examples of subreddits linked to cultural ecosystem services as of 10/02/21.

Service	**Subreddit**	**Extract of the group description**	**Number of members**	**Primary metadata type**
Aesthetic views	r/EarthPorn	“EarthPorn is your community of landscape photographers and those who appreciate the natural beauty of our home planet.”	20.9m	Images
	r/BotanicalPorn	“High quality images of plants (fungi are allowed!).”	167k	Images
Recreational activities	r/Outdoors	“Outdoors is for *all* outdoor experiences, not limited to any specific interest. Caving, mountain climbing, cycling, bushcraft, gardening, sailing, plants, birds, trees, going for a stroll -- it's all on topic here!”	2.7m	Images
	r/Hiking	“The hikers' subreddit.”	1.3m	Images
Tourism	r/Travel	“r/travel is a community about exploring the world. Your pictures, questions, stories, or any good content is welcome.”	5.7m	Images and text
Spirituality and sense of place	r/Spirituality	“Here, we discuss such things as personal transformation, the meaning of life, death, and moments of clarity.”	190k	Text
	r/Culture	“A subreddit dedicated to sharing and discussing the many aspects of culture”	6.3k	Text

7.3.1.2 Flickr

Flickr is a popular social media site that hosts images and videos with up to 25 million uploads a day (Ding and Fan 2019). Flickr has a broad user base, with a range of motivations for uploading photographs (Oteros-Rozas et al. 2018), and therefore has potential as a source of data for a wide range of CES. Posts on Flickr can have associated metadata that includes textual titles, description and tags; spatial location in the form of latitude and longitude of where the image was taken; and the time and date the image was taken. Flickr metadata is accessible through tools such as the “photosearcher” package in the R environment (Fox et al. 2020a), and stand-alone software such as the InVEST Recreational tool (Sharp et al. 2020).

7.3.2 Data collection and analysis

A reproducible R file for the data collection methods has been included in the supplementary material (Appendix L). To comply with API terms and privacy policies all datasets were anonymised, stored with multiple layers of security and any unnecessary metadata was deleted.

7.3.2.1 Keyword search

First, to find posts related to recreational activities, we searched the Pushshift tool (Baumgartner et al. 2020) for any posts on Reddit containing a single keyword for four different activities; “hiking”, “camping”, “skiing” and “kayaking”, found in any textual metadata uploaded by the user e.g. the title or description of the post. We also constrained the search to any posts that were uploaded between the 1st of January 2020 and the 1st of January 2021. We then repeated this query on Flickr, using the photosearcher R package (Fox et al. 2020a), ensuring that we made a comparable search using the same keywords, again found in any of the posts textual metadata, and within the same uploaded date range. We summarized the number of uploads per month as well as the mean character length of the title and text of the post, and the mean number of likes and comments on the images for each activity across platforms. Furthermore, as posts on Reddit can be in a range of formats other than images and text traditionally used in CES studies (e.g. links to other websites or videos), we calculate the percentage of posts that were images or text.

To compare the contents of the images posted on the two sites, we took a random sample of 1,000 images related to hiking from both sites (images listed as adult material were not included in the sample selection). The contents of the images were automatically tagged using the Google Cloud Vision API (Google Cloud Vision API 2020). The Google Cloud Vision API is a machine learning algorithm that labels the content of images. The algorithm is based on a large pre-trained dataset and can label image contents into millions of predefined categories including objects and expressions. Here, we used the “imgrec” R package (Schwemmer 2019) to label each image with the 10 objects the algorithm first detects. To ensure that the image contents were accurate without manual validation we only kept labels that had a confidence score of > 0.6 (Gosal et al. 2019).

To compare the hiking images from Reddit and Flickr we used a chi-square test to compare the two sources of data in terms of their image content (frequency of Google Cloud Vision API labels). As the dataset is relatively large, some statistical tests may indicate statistical significance ($p < 0.05$) irrespective of real-world significance in the data. Furthermore, statistical significance does not provide information on the size of the effect (Kim 2017). Here, we primarily focus on the individual contribution of features (${x^2}_i$ eq. 1) to the total effect size $x^2 = \sum {x^2}_i$, enabling us to understand better the difference between the two datasets (Oakes and Farrow 2007).

$${x^2}_i = \frac{(\mathrm{obs}_i - \mathrm{exp}_i)}{\mathrm{exp}_i}$$

Equation 1 Chi-square test, where obs_i and exp_i are the observed and expected values of feature i, respectively.

As textual metadata can be useful for understanding characteristics of CES or eliciting the emotion of CES beneficiaries (Brindley et al. 2019; Hale et al. 2019), we also returned textual metadata for analysis for the random sample of hiking images. As images uploaded to Reddit can only contain a title, with no description text, the most comparable source of textual data for images from Flickr and Reddit are the comment sections. We summarised the number of comments for these images as well as the number of unique users interacting with the posts. The sentiment expressed in each comment was calculated using the AFINN dictionary (Nielsen 2011), which has previously been used to assess the sentiment value expressed in social media text posts (Koto and Adriani 2015). This dictionary ranks words on a -5 (negative sentiment) to +5 (positive sentiment) scale. The sum sentiment of each post was calculated, and the mean sentiment score of the posts on each site calculated. We also filtered out automated messages, weblinks and commonly used words such as “the” and “is” and calculated the most frequently used words in comments on the two sites.

7.3.2.2 Subreddit search

A unique aspect of the Reddit API is the ability to search for individual subreddits. Here, we searched four subreddits that are themed around the aesthetic value of nature; “r/EarthPorn”, “r/BotanicalPorn”, “r/WaterPorn” and “r/DesertPorn”, as well as posts from two subreddit about two recreational activities (“r/Birding” and “r/Scuba”) and two subreddits that discuss spirituality and culture (“r/Spirituality'' and “r/Culture”). The results were limited to posts uploaded in the year 2020. The aesthetic views subreddits have a set of rules that mean all posts on the subreddit are of photographs of nature. Table 11 summarises the submission rules for the “r/EarthPorn” subreddit, these rules are similar across the other aesthetic subreddits assessed, though the subject of the photograph differs. The rules for the recreational and spiritual subreddits allow for both images and discussion-based posts. To compare the contents of the images posted in different subreddits, we took a random sample of 1,000 images posted on “r/EarthPorn” and “r/BotanicalPorn”. These images were then automatically tagged using the Google Vision Cloud API and the contents of the two sets of images were compared using a chi-square test.

Table 11 Selected rules for submissions to “r/EarthPorn” (as of the 10th February 2021).

Rule	**Description**
A photograph	“No Paintings, illustrations, gifs, videos, or interactive images.”
An image featuring a natural landscape	“Images must have visible land. Images with humans, machines, boats, roads, airplanes, farms, animals, buildings, or other man made objects in them will be removed.”
A photograph you took (OC)	“Or one which you can provide and post the original source for. Do not rehost non OC images to reddit or imgur.”
An unsilhouetted image	“Images where details in the landscape are not visible due to silhouetting will be removed.”
The location of the area in the photo	“When it comes to location, the more specific the better. If you wish to not disclose the location you should at the very least name the state/country. Rule of thumb for naming only the location (e.g. a lake, mountain): if one can find the place immediately by searching it in google it's fine. For possibly ambiguous locations add state/country for safety.”

7.3.2.3 Potential spatial uses for Reddit

As Reddit posts are not geolocated, it is not possible to directly map the distribution of the CES expressed in the posts. Instead, we developed an automated method for estimating the approximate location of images posted to Reddit, following a similar method to Harrigian (2018). The subreddit “r/EarthPorn” requires that posts must contain the image location in the title. To extract the location name, we used named-entity recognition – a technique that classifies words in a text into predefined categories, one of which is a named location (Alfred et al. 2014). Named-entity recognition was carried out using the “entity” R package (Rinker 2015). A subset of 10% of the name-entities was manually validated by comparing the returned name-entity with the post title. The extracted location names were then geolocated using the Google Maps API through the “ggmap” R package (Kahle and Wickham 2013). Based on the place name, the Google Maps API provides an estimated latitude and longitude. The global distribution of both sets of data was mapped and the percentage of uploads from each continent was calculated.

To assess whether Reddit posts can be used to assess general CES trends for a given location, we also searched Reddit for posts containing given place names. We carried out two types of search; first, we searched for posts containing a given place name as well as the term “hiking” and second, we searched for posts containing a given place name within the subreddit “r/EarthPorn”. The place names were chosen to represent a range of scales; national (“USA” and “UK”), regional (“Wyoming” and “Scotland”) and National Park (“Yellowstone” and “Cairngorms”). The searches were carried out for any post uploaded between the 1st of January 2010 and the 1st of January 2021. The total number of posts was calculated.

7.4 Results

7.4.1 Full datasets

For each activity, the number of posts varies across each site. For hiking, there were a similar number of posts uploaded to each site in 2020 with 145,036 hiking posts on Reddit and 148,535 on Flickr. There were also a similar number of posts relating to skiing across the two sites: 41,703 post on Reddit and 59,455 posts on Flickr. For camping, more posts were uploaded to Reddit (143,446) than Flickr (66,818); however, for kayaking, more posts were uploaded to Flickr (48,659) than Reddit (15,107). The number of uploads fluctuates across the year for both websites (Fig. 24). For hiking and skiing, even though there were a similar number of posts, Reddit had a greater quantity of unique users generating the posts. For hiking, Reddit had 88,075 unique users posting whilst Flickr had 9,392, while for skiing Reddit had 20,934 unique users whilst Flickr had 4,309.

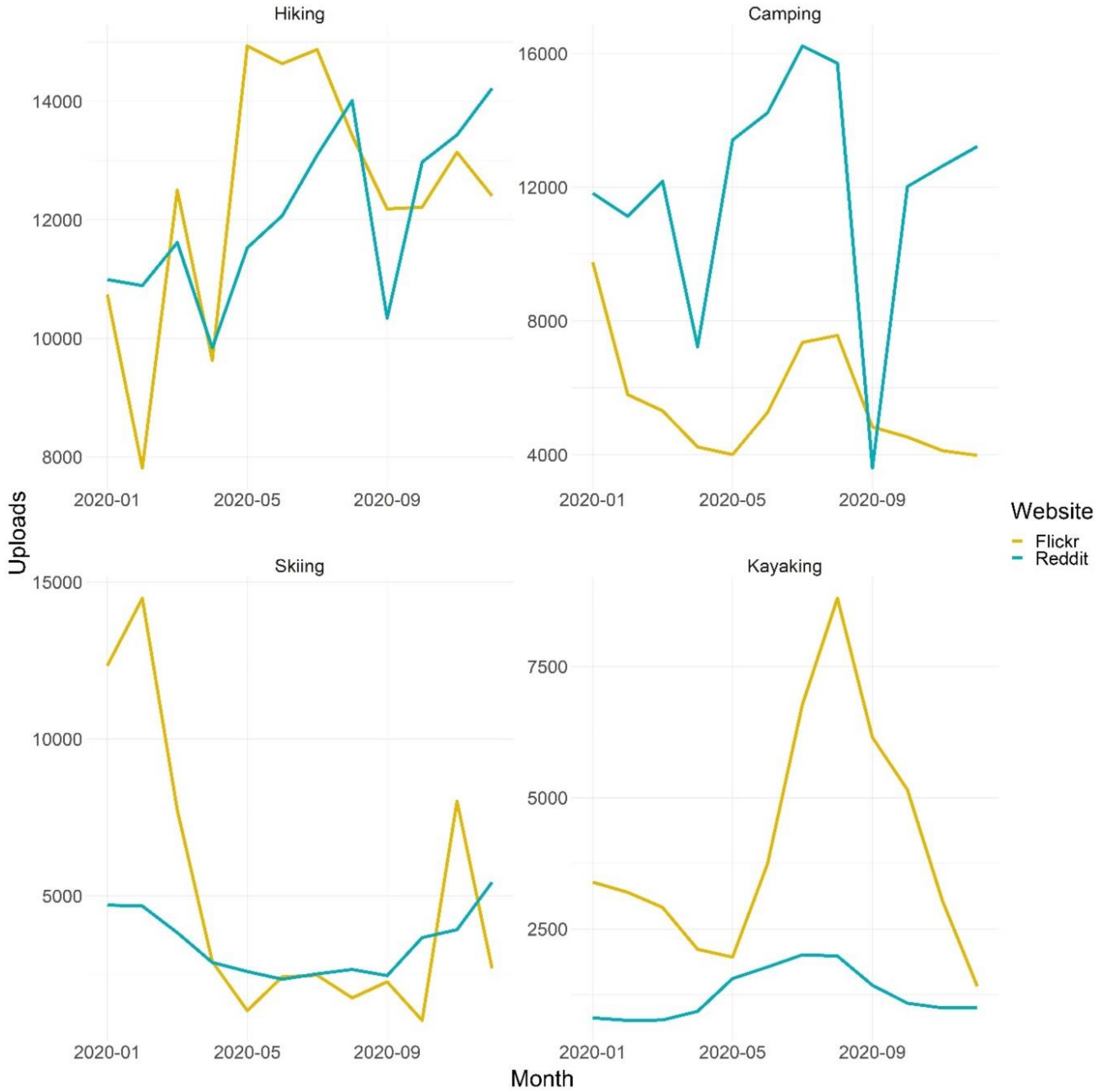


Figure 24 Uploads of posts including the words “hiking”, “camping”, “skiing” and “camping” to Reddit and Flickr between the 1st of January 2020 and the 1st of January 2021.

For each activity that we searched, many of the posts uploaded to Reddit were text-based (Fig. 25). Only around 15% of the posts returned via a keyword search from Reddit were images. Compared to posts uploaded to Flickr, posts on Reddit, in general, have longer titles and text descriptions as well as a higher number of comments (Fig. 26). Posts relating to hiking, camping and skiing on Flickr have, on average, more likes than posts on Reddit, though Kayaking posts on Reddit have a higher mean number of likes than those on Flickr.

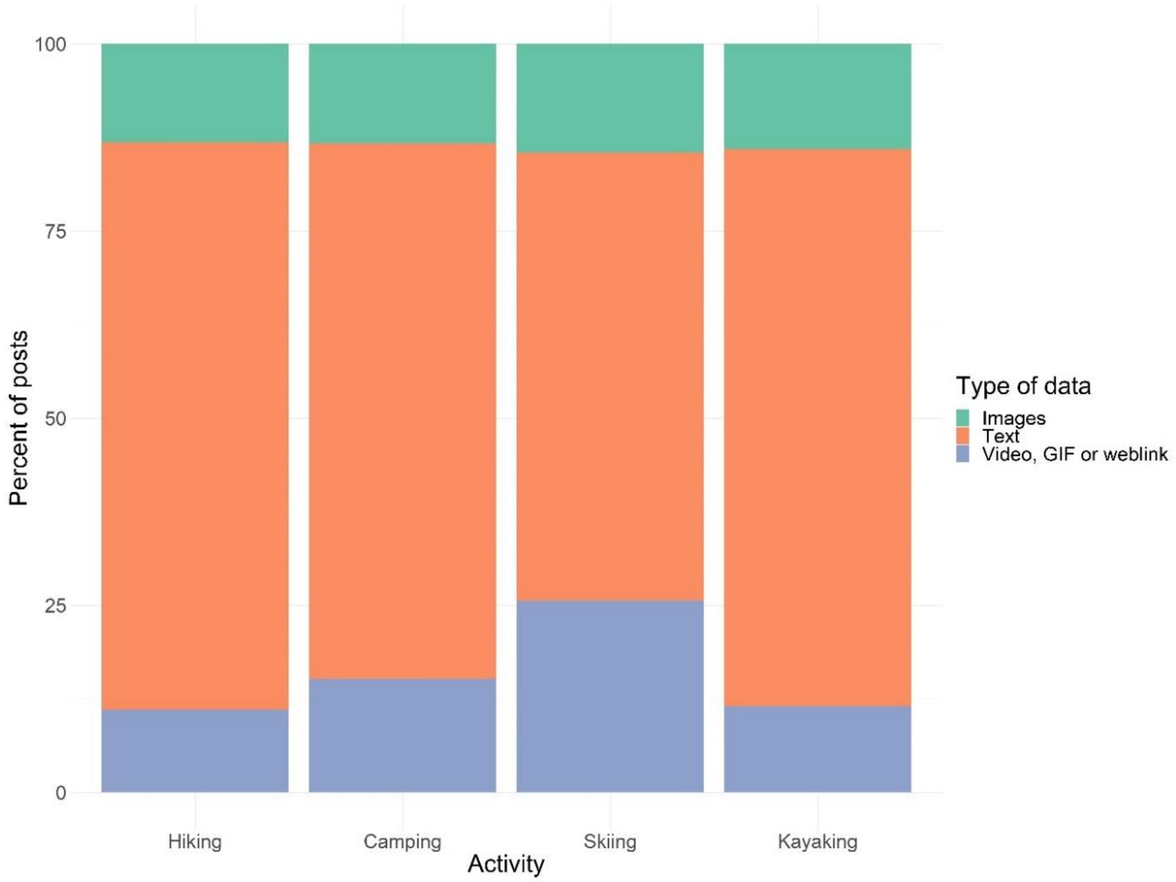


Figure 25 Types of posts uploaded to Reddit.

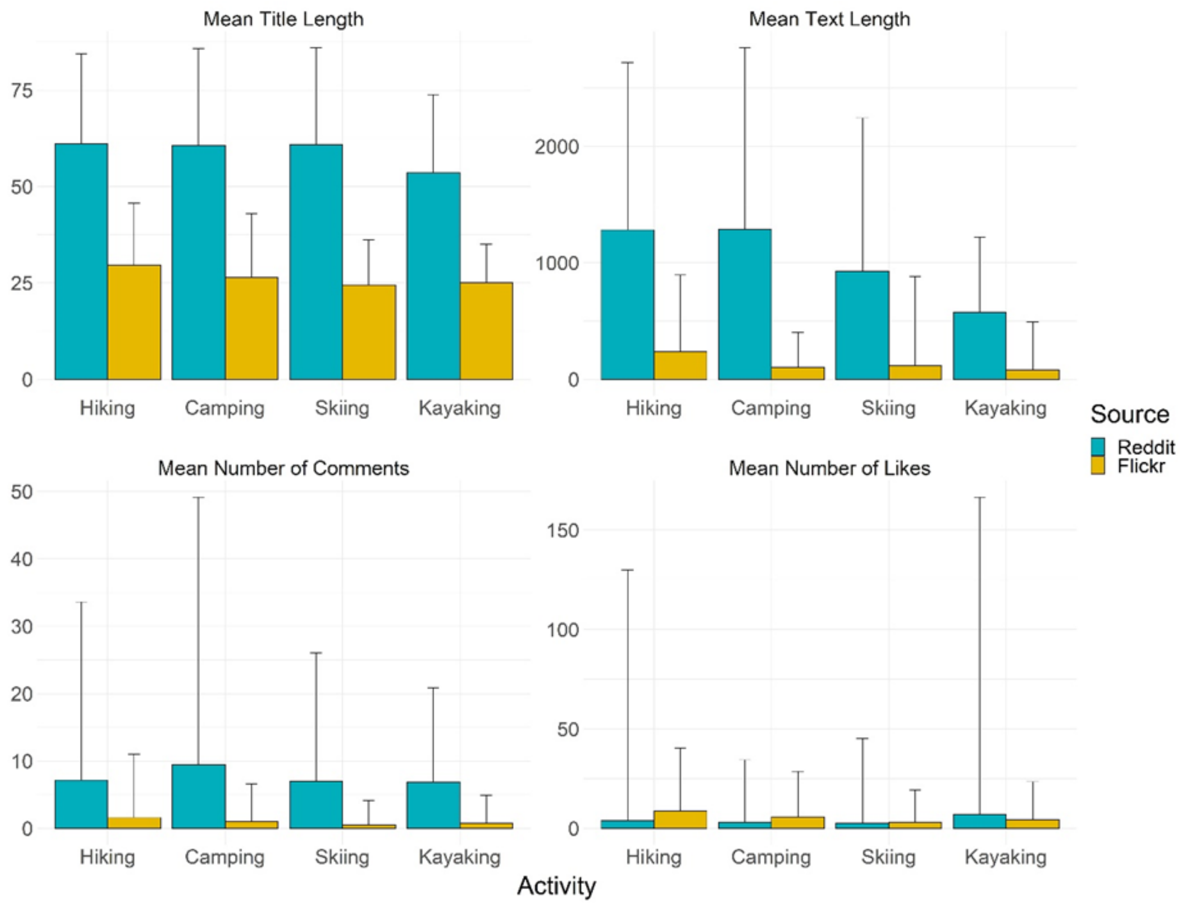


Figure 26 Summary of posts made on Reddit and Flickr (mean + 0.5 standard deviations).

While the majority of the most labelled objects were common between the two sets of images (e.g. tree and mountain), there was an overall significant difference in the contents of the two sets of photographs labelled by the Google Cloud Vision API, ($x^2 = 3{,}127.5, df = 1230, N = 13{,}582, p < 0.001$) The 15 Google Cloud Vision API labels (1.22% of the total number of unique labels) that had the highest contribution to the total x^2 effect size contributed 17.42% of the total x^2 value (Fig. 27a). Of these 15 labels, five ("plant community", "vegetation", "natural environment", "nature reserve" and "land lot") appeared more frequently in the images from Flickr (Fig. 27b). Though more frequent in Flickr images, the Google Cloud Vision API labels such as "plant community" and "natural environment" were present in 71 and 66 Reddit images, respectively. The other ten highest contributing labels, relating to dog walking, sports and people were more frequently photographed in Reddit images, with the labels such as "dog" and "dog breed" only being tagged in two of the Flickr images.

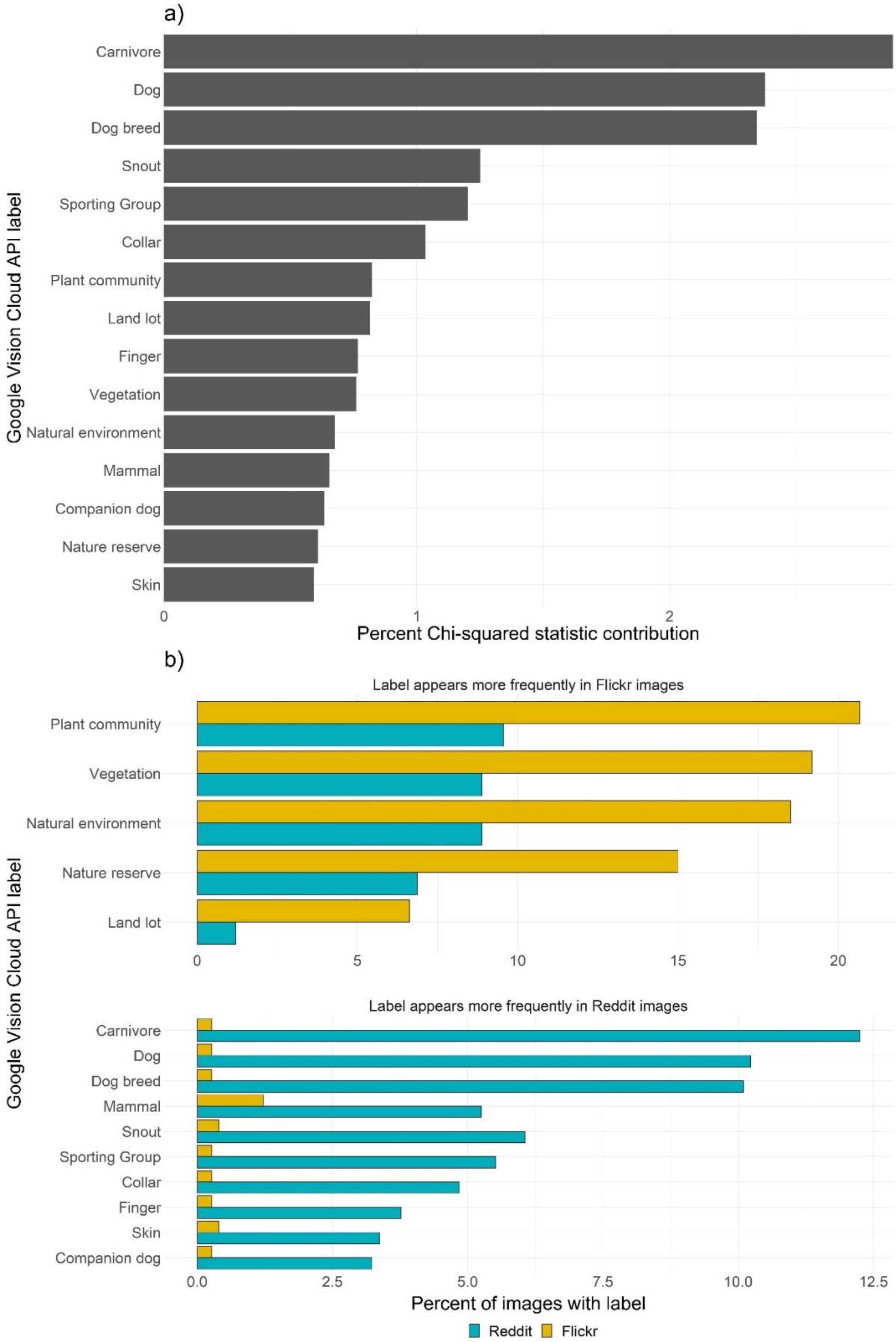


Figure 27 a) The 15 Google Cloud Vision API labels which had the greatest contribution to the overall Chi-squared statistic (larger values indicate a larger difference between Reddit and Flickr); b) The percentage of Reddit and Flickr images that the 15 labels appeared in.

For the 1,000 hiking images from Reddit, 702 posts had comments, while for the 1,000 Flickr images only 116 posts had comments. The 6,602 comments on the Reddit post were made by 4,142 unique users, while the 1,702 Flickr comments were made by 1,119 unique users. A sentiment score could be calculated for 642 Reddit comments and 108 Flickr comments, those where a score could not be calculated did not contain any words in the AFINN dictionary. In general, the sentiment expressed in Flickr comments was far higher than those on Reddit (Fig. 28). Only 1.90% of Flickr images expressed a negative or neutral sentiment, whilst 11.66% of Reddit comments expressed a negative or neutral sentiment. Many of the non-unique Flickr comments are “awards” while on Reddit they were automatically generated messages from moderators of the subreddit. After filtering, the most used words in Flickr and Reddit comments suggest that Flickr users more frequently comment general positive comments regarding the picture, such as “wonderful” and “excellent”, while Reddit users more frequently comment regarding features of the photograph, such as “trail”, “water” and “dog” (Fig. 29).

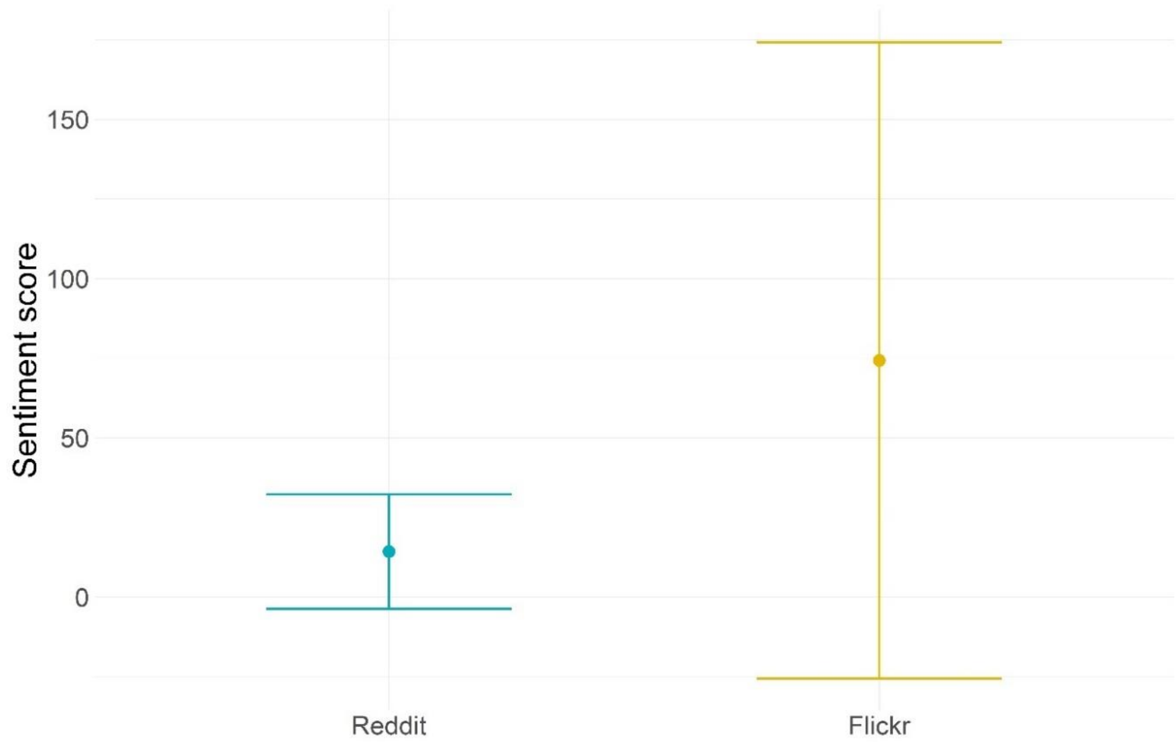


Figure 28 Mean +- 0.5 standard deviations for the AFINN sentiment score expressed in the comments of hiking images on Reddit and Flickr.

Figure 29 The 20 most frequently used word in Flickr and Reddit comments after filtering.

7.4.2 Subreddit search

Of the subreddits relating to aesthetic values, “r/EarthPorn” was the most popular of the four we searched, with 77,717 photographs uploaded in 2020. The subreddit “r/BotanicalPorn” had 5,289 uploads, “r/WaterPorn” 2,168 and “r/DesertPorn” 823. The number of uploads to each subreddit varies by month (Fig. 30). The subreddits “r/Spirituality” and “r/Culture” also had a relatively large number of uploads during the year 2020 with 30,528 and 4,579 uploads, respectively. Furthermore the recreational based subreddits “r/Birding” had 17,280 posts and “r/Scuba” had “6,064” posts.

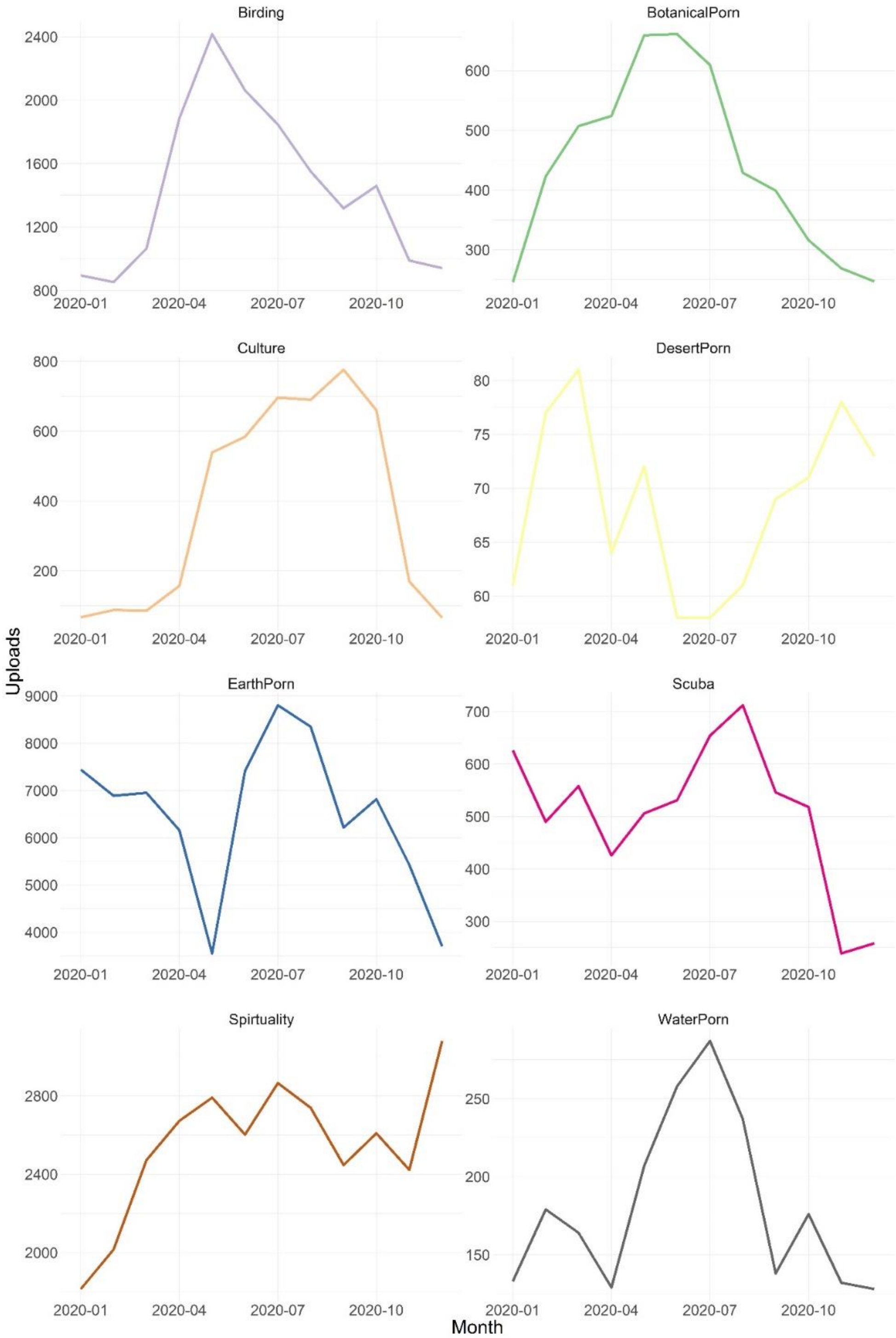


Figure 30 Uploads of posts to the subreddits “r/Birding”, “r/BotanicalPorn”, “r/Culture”, “r/DesertPorn”, “r/EarthPorn”, “r/Scuba”, “r/Spirituality” and “r/WaterPorn” between the 1st of January 2020 and the 1st of January 2021.

The was a large contrast between the labelled objects in images from the “r/EarthPorn” and “r/BotanicalPorn” subreddits, with an overall significant difference in the contents of the two sets of photographs labelled by the Google Cloud Vision API, ($x^2 = 10{,}205.5, df = 765, N = 13{,}196, p < 0.001$). The 15 Google Cloud Vision API labels (1.95% of the total number of unique labels) that had the highest contribution to the total x^2 effect size contributed 36.26% of the total

x^2 value (Fig. 31a). Of these 15 labels, seven, all relating to plants and flowers, appeared more frequently in the images from “r/BotanicalPorn” (Fig. 31b). The other highest contributing labels, relating to landscapes, were more frequently photographed in “r/EarthPorn” images.

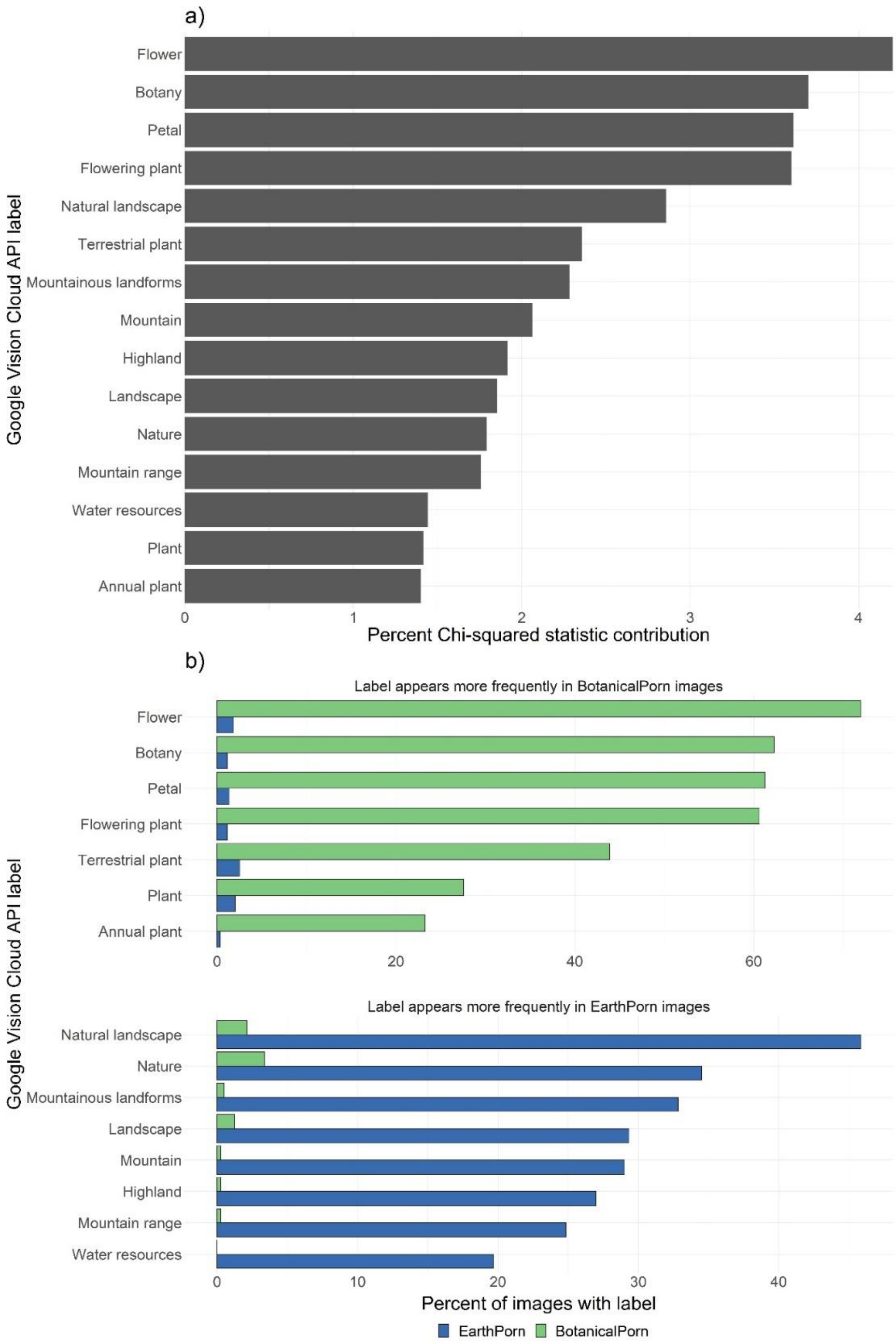


Figure 31 a) The 15 Google Cloud Vision API labels which had the greatest contribution to the overall Chi-squared statistic (larger values indicate a larger difference between Reddit and Flickr); b) The percentage of “r/EarthPorn” and “r/BotanicalPorn” subreddit images that the 15 labels appeared in

7.4.3 Potential spatial uses for Reddit

Our automated method for estimating image location returned latitude and longitude for 574 “r/EarthPorn” subreddit images (57.4%) (Fig. 32). The vast majority of images (65.26%) were distributed across North America. Overall, there were fewer images taken in the other continents, with Europe and Asia having relatively higher numbers of images than Oceania, South America and Africa.

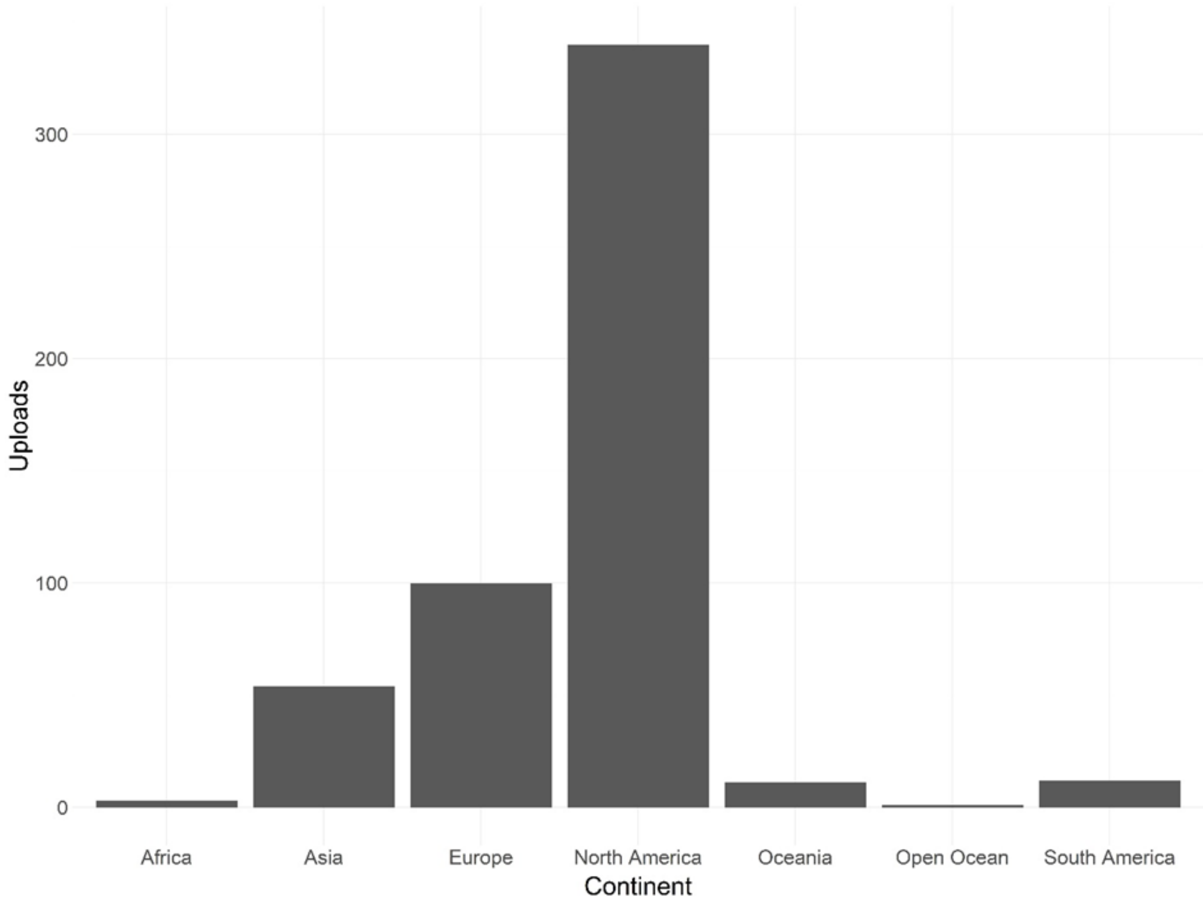


Figure 32 Estimated locations of a subset of photographs from the “r/EarthPorn” subreddit.

When searching the Reddit API for posts relating to a place name as a keyword the number of posts varies depending on the spatial scale and location (Table 12). For both searches containing a separate keyword (“hiking”) and those from a specific subreddit (“r/EarthPorn”) a large number of posts were returned.

Table 12 Number of posts, when searching Reddit with a location name as a criterion.

Scale	Search Criteria	Number of Posts
National	Text = “USA” AND “hiking”	13,148
	Text = “USA” AND Subreddit = “r/EarthPorn”	12,336
Regional	Text = “Wyoming” AND “hiking”	1,209
	Text = “Wyoming” AND Subreddit = “r/EarthPorn”	3,399
National park	Text = “Yellowstone” AND “hiking”	2,794
	Text = “Yellowstone” AND Subreddit = “r/EarthPorn”	4,334
National	Text = “UK” AND “hiking”	8,196
	Text = “UK” AND Subreddit = “r/EarthPorn”	5,539
Regional	Text = “Scotland” AND “hiking”	2,528
	Text = “Scotland” AND Subreddit = “r/EarthPorn”	5,539
National park	Text = “Cairngorms” AND “hiking”	87
	Text = “Cairngorms” AND Subreddit = “r/EarthPorn”	131

7.5 Discussion

The main aim of this paper was to understand the potential applications for Reddit as a complementary or alternative source of CES data from social media sites. Here, we explored two methods of searching the Reddit API: a keyword search and searching specific subreddits. In general, we were able to return a relatively large number of posts relating to a range of CES (recreation, aesthetic, spirituality and culture). Searches made via the keywords search showed that Reddit has a comparable number of available posts on recreational CES to Flickr. However, the posts returned via a keyword search on Reddit are primarily text-based, which is unsurprising given that Reddit is marketed as a discussion-based social media site. The two sites had similar numbers of posts for hiking and skiing, though Reddit had more posts about camping and Flickr had more posts about kayaking. This suggests that the choice of the site may depend on the activity of interest and thus the suitability for CES research is context-dependent. Furthermore, even when the posts had a similar number of uploads between sites, the posts on Reddit were contributed by a far greater quantity of unique users. This gives rise to the potential for posts to be generated by a more diverse user base than Flickr. There are however socio-demographic biases associated with social media sites (Duggan and Smith 2013; Retka et al. 2019), and these need to be explored fully before making generalisations about the wider population. For example, Reddit has a large user base with high socio-demographic diversity; (Duggan and Smith 2013) estimated that around 6% of internet users were active on Reddit. They found that there is a bias towards male users (8% of male internet users compared to 4% female) and that a higher percentage of internet users aged 18-49 use Reddit than those over 50.

The biggest limitation of Reddit is that the posts do not have geotagged locations. Our automatic method for estimating the approximate location of a photograph calculated latitude and longitude for 57.4% of the Reddit posts. From our analysis of landscape photographs, the distribution of images uploaded to the "r/EarthPorn" subreddit is primarily concentrated in North America, though many images were also from Europe and Asia. Harrigian (2018) estimated the distribution of the Reddit users base through geolocating statements in their comments and found that the demographic was primarily people living in North America, followed by Europe and Asia. Harrigian (2018) also provides a potential method of establishing user origins, a key feature in understanding CES interaction from Flickr (Wood et al. 2013; Sinclair et al. 2020a). The demographic of users and distribution of posts may have implications for studies that wish to assess CES across different continents, with previous studies assessing CES in North America potentially missing out on the wider range of photographs available from Reddit.

As the users of both Reddit and Flickr are concentrated in western, developed countries, this could also be a source of potential bias. Where studies are at a global or super-continental scale, data from Reddit and Flickr should therefore be used in combination with each other and with other sources of data that are popular in other areas of the world. For example, in China where Flickr is banned and Reddit is not a popular social media site, alternative social media sites such as Weibo (Zhang and Zhou 2018), or travel comment portals websites such as Tuniu Travel (Dai et al. 2019), should be used to bridge the gap in CES data. At local and regional scales other sources of data may also help to complement social media data such as on-site survey data (Sinclair et al. 2020a), online surveys (Moreno-Llorca et al. 2020) and national statistics (Graham and Eigenbrod 2019). Future work should begin to assess the respective biases in these alternative sources to ensure they are comparable. Furthermore, both Flickr and “r/EarthPorn” are related to images pertaining to high-end photography, which may restrict the demographics to only those with access to such technology (Chen et al. 2020). One possible source of data that we suggest needs exploring is other subreddits focused on natural landscapes, such as “r/AmatureEarthPorn”, which do not restrict uploads to high-quality images and therefore may have greater representation of landscapes from a wider demographic.

There are however several caveats to geocoding Reddit post locations. First, the extracted location name from the named-entity recognition may not be correct due to ambiguity in the text, spelling or language differences, or capitalizations (Goyal et al. 2018). Given “r/EarthPorn” is an English language forum, this may not have been a significant issue in our analyses. The issue with multi-part names being extracted to a single word place name means that the finer spatial scale of the location is lost. The rules of the subreddit specify that place names included in the post title should be as specific as possible. However, the named-entity recognition method often identified the location as the regional (i.e. state) or country part of the place name, losing the finer detail of the image’s location. Though the named-entity recognition method can correctly recognise and extract places names with multiple parts (e.g. "Ocean Beach", "San Francisco" was correctly identified), for many multi-part place names the finer location detail can be lost. For example, “Mt. St. Helens, Washington” was extracted as “Washington”. The automated extraction of the landscape image place name presented here may be best suited for generalising large-scale distributions. However, as the Reddit posts normally contain specific location details in their titles, studies that wish to assess spatial distribution on a finer scale may find success in manually extracting the place name.

Second, the high number of available geocoding algorithms, as well as the potential for ambiguity in the named entity locations extracted from the Reddit comments, can introduce errors in the geocoded results (McDonald et al. 2017). For example, there are multiple locations globally named “Portland”; without more context, the geocode algorithm may not correctly code the location. Third, though the geocoding method can provide a latitude and longitude with high spatial accuracy, when geocoding is based on a general location name, the location will be plotted to a single point within that region. For example, multiple photographs taken in completely different areas of the Badlands National Park, US, all containing “Badlands National Park” in their title, will all be aggregated to the same point location. Furthermore, though this method was successful on posts to “r/EarthPorn”, other subreddits may not stipulate that a location must be present in the text. We suggest that future studies using Reddit data for spatial analysis should consider methods for reducing geocoding inaccuracies (McDonald et al. 2017). Another possible source of geocoding the location of a post is the Google Cloud Vision API which can estimate the location of an image; however, this process is currently only capable of locating popular sites.

Due to the limitations of geocoding Reddit posts, we do not recommend using posts from Reddit to assess the spatial variation of CES similarly to those from Flickr, Twitter or Instagram (Graham and Eigenbrod 2019; Chen et al. 2020). Instead, one potential method for getting CES data for a location without the need for geocoding posts is searching for a given name place alongside other keywords or within a subreddit. This method has previously been used in CES studies from Flickr, for example, Thiagarajah et al. (2015) searched Flickr for photographs based on the place names of four mangrove sites in Singapore, while Roberts (2017) queried Twitter posts for any containing the names of urban green spaces in Birmingham, UK. Here, we showed that searches for Reddit posts with a relevant study site as a keyword provide a relatively large dataset across spatial scales and locations. Though we have demonstrated that Reddit data has the potential for spatial studies, we acknowledge these limitations do restrict the use of Reddit's data to assess spatial variations in CES and therefore suggest that Reddit posts are more suited to generalising CES interactions for a given search criteria e.g. a study area or specific activity. However, these limitations do not hinder the use of data for studies that assess CES through content analysis and textual analysis.

We have shown that photographs associated with hiking from both Reddit and Flickr can both be used in the same image content analysis techniques, thus illustrating their potential for CES studies that use content analysis of images, without additional spatial analysis (Thiagarajah et al. 2015). Oakes and Farrow (2007) demonstrated that words with the highest percentage contribution of the total x^2 value, relative to the other words in the set, best highlight the differences in two groups of words. Here, the small number of labels contributing to a high percentage of the total x^2 value indicates that, in general, many images contain similar scenes, but the difference between the two sites is driven by a small number of features identified with the Google Cloud Vision API. The differences between the two sites may be reflected in the user's motivations for undertaking hiking. As the reasons to undertake hiking are multifaceted (Wilcer et al. 2019), the difference in demographics between users of Reddit and Flickr suggests they may be undertaking hiking or uploading images to each site for different reasons. For example, results from our subset of images suggest that Reddit users are more likely to participate in hiking for physical activity and dog walking, whilst Flickr users are more likely to undertake hiking to access aesthetic views.

We have demonstrated that as the contents of images from Reddit and Flickr can provide essentially the same information about CES, Reddit may be a valuable additional source of data for assessing aesthetic landscape qualities (Oteros-Rozas et al. 2018) or recreational preferences (Gosal et al. 2019; Lee et al. 2019). The difference in contents may also be down to the motivations to upload to each platform. Kipp et al. (2017) found that Flickr users have multiple motivations for uploading photographs including wanting to get an opinion on their photographs and because they have an interest in a particular subject. However, as one of the main features of Reddit is the ranking of posts through user votes (Duggan and Smith 2013), further work should be undertaken to assess whether the relative motivations for uploading to Reddit are similar to other social media sites.

Comparison of image content from uploads to the subreddits “r/EarthPorn” and “r/BotanicalPorn”, which focus on photographs of different aspects of nature, demonstrated distinctions between the two - and therefore provide unique sources of data for assessing the role of different aspects of nature to CES. Building on this, “r/WaterPorn” and “r/DesertPorn” may help to provide a robust dataset for untangling the contributions of geodiversity to CES (Fox et al. 2020b). Furthermore, subreddits are not just useful for assessing aesthetic CES, but can also provide a large source of data for spirituality and recreation. There is a far larger range of subreddits available than accessed here, each with a unique theme that can help to understand CES, for example, “r/Travel” (a discussion board for travel) could be a useful source of data for understanding the links between tourism and CES and “r/CityPorn” (images of cityscapes and urban areas) may help to investigate urban ES, although this may require some content filtering to remove purely architectural images. As our keyword searches return significantly more text-based posts than images, researchers should familiarise themselves with the different subreddit as potential sources of images, for example, titles of posts in “r/EarthPorn” generally do not contain words like “landscape” or “view” and would therefore not be returned through a keyword search looking for images relating to aesthetics. The results presented here demonstrate that Reddit has the potential to be a significant source of image data and may be beneficial to CES studies that incorporate content analysis.

Studies can also use textual metadata to assess CES (Roberts 2017; Hale et al. 2019; Johnson et al. 2019). Flickr images tend to have description metadata that the uploader provides, which has been demonstrated to be useful in textual analysis such as sentiment analysis (Brindley et al. 2019) or eliciting information on CES from the text (Hale et al. 2019). A disadvantage of photographs uploaded to Reddit is that images do not have an equivalent description by the uploader, therefore we only compare the comment sections of the two websites. As many posts on Reddit have comments and because Reddit is a discussion-based platform, this large online database may help to understand the opinions of thousands of individuals. As the perception of the CES can only be drawn from those that comment (Dai et al. 2019), having a larger number of unique individuals interacting with CES related posts may enable the results to be generalised to the wider population and therefore better help to inform policy, planning and management (Dunkel 2015). Here, the text comments from the two sites vary regarding the sentiment expressed, with Flickr images having a more positive associated sentiment score, but also a large variability within the score. The subset analysed here also showed very few negative comments on Flickr, whilst on Reddit, negative sentiment was more frequently expressed. Moreover, the actual text contained within the comments differs between the two sources, with comments on Flickr tending to be more general appraisals of the photograph, while Reddit comments are more often a discussion around the image themselves, thus potentially providing richer information on the users' perspective of CES. Having access to a wider range of opinions, both positive and negative, may help to better generalise attitudes to CES.

As Reddit is designed to be a discussion-based forum it may contribute to richer information on the users' perspective of CES. For example, the "r/Spirituality" subreddit encourages users to contribute to the discussion of any aspects of spirituality regardless of religion or ideology, thus providing the potential to assess the opinions of people from a wide range of backgrounds. Furthermore, Reddit comments can be longer than most other social media sites (e.g. Twitter has a 280-character limit and Instagram has a 300-character comment limit) and therefore a user can discuss their opinions in greater detail (Gkotsis et al. 2017). The discursive nature of Reddit provides researchers with a unique opportunity to assess which aspects of a certain image or video people appreciate. There is also scope for this interactive and discussion-based platform to be used in experimental studies in which researchers post content and monitor feedback. Though as with all social media-based studies, we recommended that the ethics of these studies be discussed in further detail. We suggest that Reddit data is particularly useful for studies that wish to analyse users' comments in conjunction with the metadata available for each image for a more robust assessment of CES.

For studies carrying out image content or textual analysis we suggest that combining Reddit data alongside other sources of data, would be useful in CES because (1) images and text from Reddit can provide comparable data used to assess aspects of CES; (2) Reddit potentially contains additional data previously overlooked; (3) they have different geographical biases (e.g. Reddit to North America, Flickr to Europe and Weibo to Asia). We, therefore, suggest that a more holistic approach to assessing CES would be to include cross-platform analysis including multiple sources (Retka et al. 2019). However, we note that Reddit may not be suitable for integrating into studies assessing spatial variations in CES. Data integration, the bringing together of data from multiple sources, could be implemented to allow data from social media sites to be analysed as a complete unit. Data integration methods, which control for differing biases and sizes of datasets, have been successfully used in other scientific fields such as species distribution modelling (Isaac et al. 2020) and those using satellite imagery (Aires 2014). As accessing data from Reddit requires a similar skill level as accessing datasets from other social media websites, data integration of these multiple sources is feasible. The tools and software used in this manuscript make these datasets more accessible and reproducible for non-data scientists and enable us to start to bridge the gap in integrating multiple sources. We therefore recommend that CES and wider environmental science studies make use of these tools to include the vast amount of data from Reddit alongside other social media data sources in their future studies.

7.6 Conclusion

We have demonstrated that posts from Reddit can be used in commonly applied CES assessment methods, such as image content analysis and textual analysis, which leverage the power of big data from social media sites. The results of this study show that Reddit can provide a large source of data similar to Flickr. However, the posts available on Reddit are not geolocated and the geocoding of a post's location has several limitations meaning that Reddit is not as suited to assessing the spatial variation of CES as other social media sites. The large quantity of data available on Reddit is most appropriate for assessing general trends in CES through image content analysis and textual analysis. The discursive nature of Reddit provides a unique opportunity to assess a wide range of CES including recreational activities, aesthetic views, spirituality and culture. We argue that Reddit should be more widely considered as a useful source of data for CES studies and we hope that this paper sets a precedent for including big datasets from Reddit in future studies.

Chapter 8 Conclusions

This thesis aimed to harness social media datasets to better understand the role of geodiversity in CES, and by doing so has contributed to our understanding of both the complex relationship between geodiversity and CES and to the applications of social media data to CES studies. The thesis makes key contributions to our understanding of the relationship between geodiversity and CES (specifically recreation), whilst also developing novel and replicable methods for mining social media data.

First, this thesis has furthered our theoretical understanding of the relationship between geodiversity and ES. **Chapter 3** provides a novel framework that contributes to the understanding of the relationships between geodiversity and biodiversity, and the pathways from which their interactions lead to ES and can be used to target research questions around the relationship between ES and geodiversity. Furthermore, the ES-GS framework provides further clarity to the position of GS and provides a platform for these previously under-represented services to be assessed holistically alongside ES. This work has begun to be cited in a range of studies assessing geodiversity, ES and GS, as well as wider issues relating to abiotic nature (Finisdore et al. 2020; Kubalíková et al. 2021; Mckay 2021; Norrman et al. 2021; Manosso et al. 2021).

By providing a measure of textual sentiment to images with different combinations of biophysical features in the image content, **Chapters 5 and 6** begin to untangle the relationship between geodiversity and biodiversity and their roles in providing a positive CES experience for the recreational activity of hiking. The results from these two studies further reinforce that geomorphological features, in particular mountains and their associated landforms, are important in underpinning hiking. This work also demonstrates that hydrological features (such as coastlines and waterbodies) can contribute to a positive hiking experience, as can under-researched geodiversity features, such as geological landforms and arid landscapes. Further, this thesis highlights how the interactions of geodiversity and biodiversity features, such as trees and plants, can help to provide a more positive CES experience. The importance of the human aspect of ES was also found to be influential, further highlighting that ES are co-produced and that human infrastructure such as trails, roads and greenspaces can enrich the CES experience. By mapping areas of most importance to recreation, as well as understanding what people interact with whilst undertaking activities such as hiking, planning and decision-makers can use these results to better inform the creation and management of protected areas and also promote sustainable use of geodiversity for such activities – e.g. through the UNESCO Global Geopark Programme.

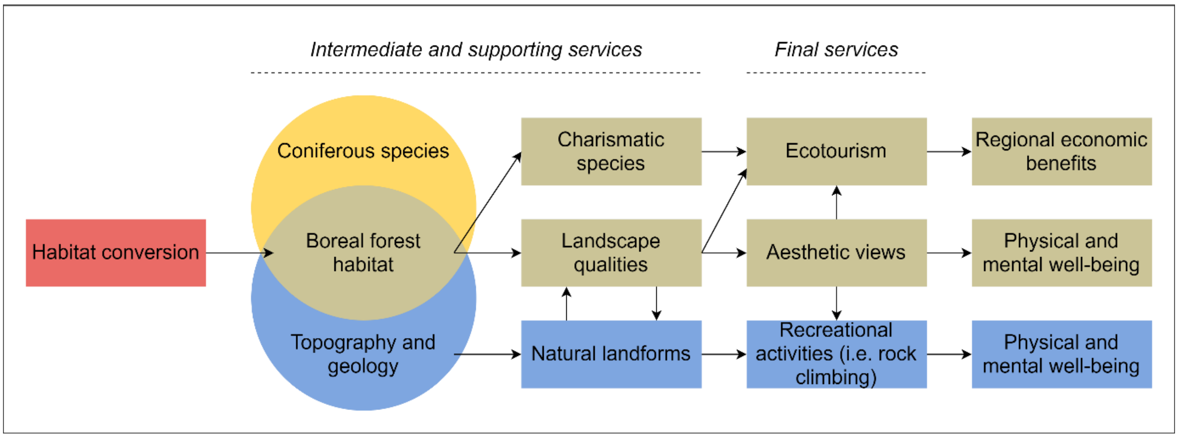
Second, this thesis has contributed to the development of social media datasets as a tool for environmental and CES assessments. The photosearcher R package presented in **Chapter 4** (Fox et al. 2020a) has already been used in the creation of published articles and datasets assessing human-landscape interactions (Gülçin 2020, 2021), forest recreation (Ciesielski and Stereńczak 2021), and plant species distributions (August et al. 2019, 2020). The package has also been discussed in relation to the use of social media data in CES studies (Dolan et al. 2021; Johnson et al. 2021). It is hoped that the photosearcher R package will be continued to be used to assess a wide range of CES across different spatial and temporal scales and in different study locations.

In **Chapter 7** it was demonstrated that the social media site Reddit can be a useful source of data for inclusion in CES studies and that it could be combined with data from Flickr to reduce biases introduced from assessing human-nature interactions based on a single social media site. Other studies have also started to utilise data from other novel social media sites, such as the Tencent messaging applications which can be used to assess real-time population densities (Liu et al. 2017) and social recommendation app Foursquare which has been used to assess users perceptions of green infrastructure (Martí et al. 2020). Furthermore, there are other social media sites, such as Tumblr, a microblogging site that shares similar available metadata to current sources which have yet to be fully explored as a source of CES data. There is also a range of other sources of citizen science data that can provide novel and insightful methods for understanding CES interactions (Havinga et al., 2020), including outdoor activity-sharing platforms such as Strava (a fitness tracking app), which can be used to assess recreational preferences (Sun et al. 2017); and geocaching apps (geolocated hidden containers), which can be a good indicator of CES (Rosário et al. 2019). Moving forward, the effectiveness of these sources of data should also be further investigated to ensure that a wider demographic of users is captured in CES studies. Therefore, starting with the addition of functionality to search Reddit, the photosearcher R package will be updated to include methods for accessible and reproducible data acquisition from other social media sites that are useful for CES and environmental studies, providing functionality for cross-platform analysis.

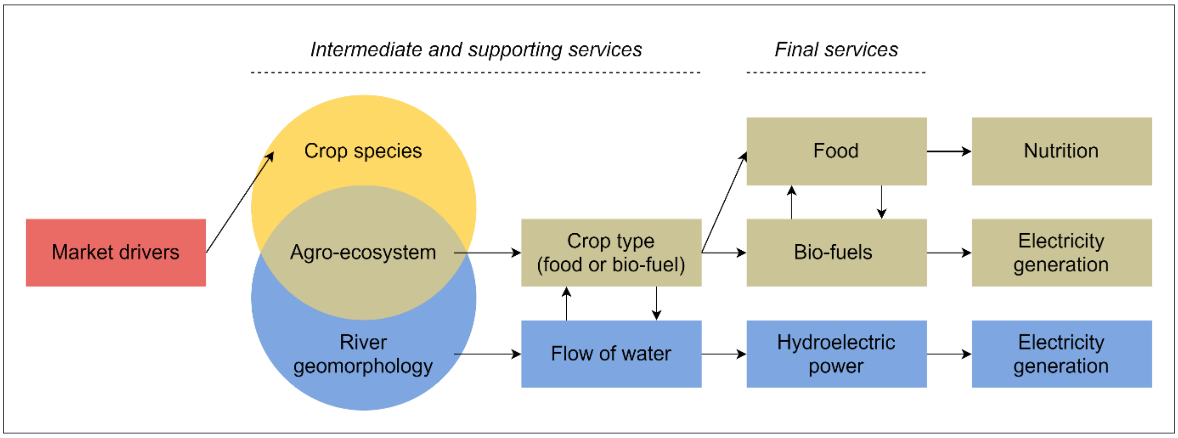
Future environmental science and CES research can continue to make advances by harnessing social media data and their associated metadata through novel approaches that have currently not yet been widely applied. For example, the Google Vision Cloud API can not only provide detailed labels on the contents of images but can also be applied to detecting landmarks and image properties. Through the detection of landmarks, the Google Vision Cloud API provides the possibility to assess the distribution of non-geotagged images in a novel approach that could complement the geotagging method presented in **Chapter 7**. Assessing the properties of images, such as the colour spectrum found within images, may help to provide novel insights into how we interact with the natural environment, such as how tree colours impact visual-sensory landscape qualities (Vaz et al. 2019). Furthermore, we have demonstrated the value of enriching social media data with textual sentiment analysis to better show positive human-nature interactions, rather than assuming all interactions are positive. Future work could also explore innovative measures of sentiment value available through social media data, such as facial expression analysis (Do, 2019), or new sentiment classifiers (Johnson et al. 2021), to further explore the range of benefits and values associated with human-nature interactions.

Overall, social media datasets are a powerful source of data that can be harnessed to explore a wide range of environmental questions including CES interactions. By using multiple analytical methods across datasets from different social media websites, this thesis contributes to our understanding of the role of geodiversity to CES, in particular recreational activities. Geodiversity features, such as geomorphology, including topography and elevation, and hydrology, including marine and freshwater systems, play an important role in providing recreational opportunities. Though geodiversity underpins CES, recreational and tourism activities can damage geodiversity features and process and pose a threat to its sustainable use. Harnessing social media data can not only provide information on the role of geodiversity in delivering and maintaining ES but it can help to inform the management strategies used to help ensure the future conservation of its features, processes and the ES these provide.

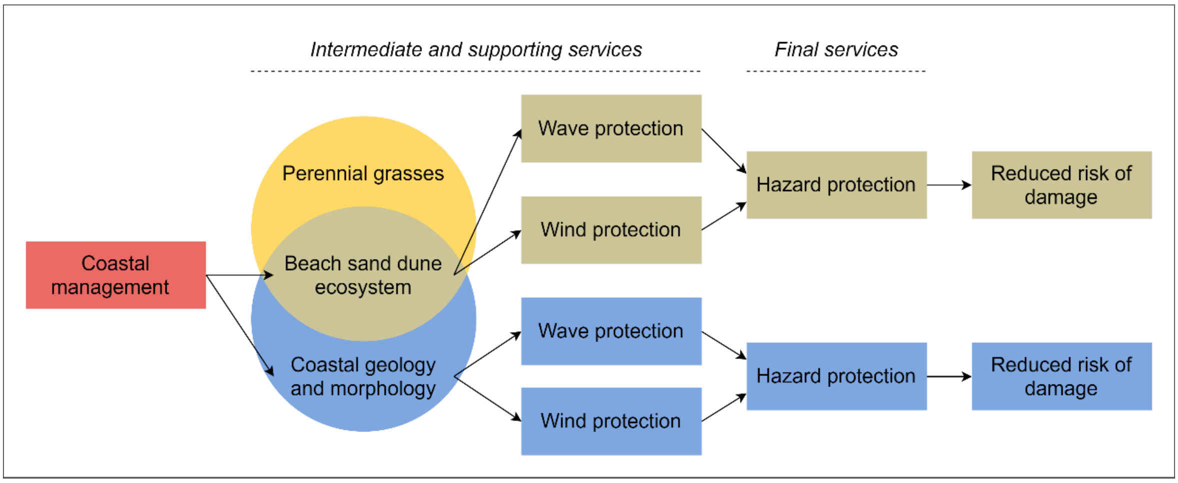
Appendix A Additional applications of the ES-GS cascade framework



A.1 Application of the framework to a real-world ecosystem – boreal forests. Here the framework provides a foundation for further research into the relative roles of biotic and abiotic nature to differing cultural ecosystem services. This example further highlights the interactive nature of services provision, highlighting that changes to the aesthetic quality of an area may impact on the recreational GS. By better understanding the interrelatedness of abiotic and biotic nature, as well as ES and GS, application of the framework could help better inform on the impacts of habitat conversion to a range of services.



A.2 Application of the framework to a real-world ecosystem – agricultural ecosystems. Here the framework can be implemented to assess the impact of crop choice on water quantity. As well as assessing the trade-offs between the value of food crops versus bio-fuel crops, empirical studies could utilise this framework to assess whether the irrigation needs of crop types will impact on local water levels and therefore any related ES or GS, i.e. hydroelectric power.



A.3 Application of the framework to a real-world ecosystem – beach sand dunes. Here the framework further highlights how the same ecosystem functions can be generated by both geodiversity alone, as well as its interactions with biotic nature. These functions in turn provide the same services and benefits and values. Understanding the relative value of each mechanisms would help to inform the economic viability of different management strategies.

Appendix B Repoducible data collection methods for Chapter 4

B.1 Installation

```
# pacman package allows for better loading of uninstalled packages
if(!"pacman" %in% installed.packages()) install.packages("pacman")
library(pacman)

# load required libraries and install if needed
p_load_gh("ropensci/photosearcher")
p_load(ggplot2,
       ggthemes,
       dplyr,
       ggmap,
       tmaptools,
       USAboundaries)
```

B.2 Cultural ecosystem service data

```
#Get photograph metadata for images of hiking in the USA

#Load shapefile
contiguous_us <- USAboundaries::us_states()
contiguous_us <- contiguous_us[!contiguous_us$name == "Alaska", ]
contiguous_us <- contiguous_us[!contiguous_us$name == "Hawaii", ]
contiguous_us <- contiguous_us[!contiguous_us$name == "Puerto Rico", ]

#add col for mapping points by state
contiguous_us$mapid <- 1:nrow(contiguous_us)

#search flickr for photographs
USA_hiking <- photosearcher::photo_search(mindate_taken = "2018-06-01",
                                          maxdate_taken = "2019-01-01",
                                          maxdate_uploaded = "2020-01-01
",
                                          text = "hiking",
                                          sf_layer = contiguous_us)
)

#add state name to the the hiking photographs to plot as colours
USA_hiking <- USA_hiking %>%
  rename(mapid = within)

USA_hiking <- merge(USA_hiking, contiguous_us, by = "mapid")

#plot map
ggplot(data = USA_hiking,
       aes(x = longitude, y = latitude, colour = state_abbr)) +
  borders("state", colour = "white", fill = "gray87") +
  geom_point(size = 0.5) +
```

```
  scale_x_continuous(name = "Longitude") +
  scale_y_continuous(name = "Latitude") +
  scale_colour_viridis_d(option = "B") +
  coord_fixed() +
  theme_bw(base_size = 9) +
  theme(strip.background = element_blank(),
        plot.margin=grid::unit(c(0.25,0.25,0.25,0.25), "mm"),
        legend.position = "none")

#save plot
ggsave(filename = "USA_hike.png",
       height = 7.365,
       width = 14.287,
       units = "cm")

#extract user ids
user_ids <- data.frame(USA_hiking$owner)

#extract unique users
user_ids <- distinct(user_ids)

#search for their information
user_info <- photosearcher::user_info(user_id = user_ids$USA_hiking.owne
r)

#only get users that have a city listed
user_city <- subset(user_info, city > 0)

#add country to end to increase geocode accuracy
user_city$addr <- paste(user_city$city, user_city$country, sep = " ")

#correct coding for geocoding
Encoding(user_city$addr) <- "UTF-8"
user_city$addr <- iconv(user_city$addr, "UTF-8", "UTF-8",sub='')
user_city$addr <- iconv(user_city$addr, 'utf-8', 'ascii', sub='')

#get geocoded location sample first 100 to speed up example
geo_city <- tmaptools::geocode_OSM(user_city$addr[1:100])
```

B.3 Species data

```
#Search for images
#common name
barn_owl <- photosearcher::photo_search(mindate_taken = "2000-01-01",
                                        maxdate_taken = "2020-01-01
",
                                        maxdate_uploaded = "2020-01-
01",
                                        text = "barn owl")
#Latin name
Tyto_alba <- photosearcher::photo_search(mindate_taken = "2000-01-01",
                                         maxdate_taken = "2020-01-01
",
                                         maxdate_uploaded = "2020-01-
01",
                                         text = "Tyto alba")
```

Appendix C Repoducible data collection methods for Chapter 5

C.1 Installation

```
# pacman package allows for better loading of uninstalled packages
if(!"pacman" %in% installed.packages()) install.packages("pacman")
library(pacman)

# Load required libraries and install if needed
p_load(USAboundaries,
       dplyr,
       tidytext,
       data.table,
       ggplot2,
       sf,
       sp,
       viridis,
       raster,
       rasterVis,
       ggwordcloud,
       patchwork,
       purrr,
       scales,
       tidyr
)

# Load from GitHub
p_load_gh("ropensci/photosearcher",
          "cschwem2er/imgrec")
```

C.2 Flickr data

```
#load shape file of contiguous states
# Get US48 data
us48 <- us_states()
us48 <- us48[!us48$name == "Alaska", ]
us48 <- us48[!us48$name == "Hawaii", ]
us48 <- us48[!us48$name == "Puerto Rico", ]
# search Flickr
flickr_results <- photosearcher::photo_search(mindate_taken = "2015-01-
01", maxdate_taken = "2020-01-01", maxdate_uploaded = "2020-06-01",
text = "hiking”, sf_layer = us48)
```

Appendix D Categorisation of Google Vision Cloud API labels

D.1 Biophysical nature labels

Biophysical nature labels
"Accipitridae, Accipitriformes, Acerola family, Acmon Blue, Acorn, Acorn Woodpecker, Acridotheres, Aegean cat, Aeolian landform, african daisy, African leopard, Africanis, Agaric, Agaricaceae, Agaricomycetes, Agaricus, Agati, Agave, Agave azul, Aglais, Aglais io, Airedale terrier, Akbash dog, Akita, Akita inu, Alaskan malamute, Alaunt, Albatross, Alfalfa, Algae, algerian iris, Alismatales, Alligator, Alligator gar, alligator lizard, Allium, Alluvial fan, Aloe, Alpaca, alpine aster, alpine forget-me-not, alpine sea holly, Alpine strawberry, Alps, Alstroemeriaceae, Amaranth, Amaranth family, Amaranth grain, Amaryllis belladonna, Amaryllis family, American alligator, American aspen, American Bittern, American black bear, American Black Duck, American bulldog, American cocker spaniel, American coot, American crocodile, American crow, American curl, American Goldfinch, American hairless terrier, American larch, American lobster, American Mourning Dove, American painted lady, American pit bull terrier, American pokeweed, American Redstart, American rosefinches, American staffordshire terrier, American Toad, American Tree Creeper, american witch hazel, Amphibian, Anaxyrus, Anchovy, and melon family, and prickles, Andean condor, Anemone, angel's trumpets, Angelica, Anguidae, Animal migration, Annual plant, Anole, Ant, Antarctic flora, Antelope, Antelope jackrabbit, antelope squirrels, Anthriscus, Anthurium, Antler, Apalone, Apatura, Apatura iris, Aphids, Appenzeller sennenhund, Apple, Aquatic plant, Arabian camel, Arachnid, Araneus, Araneus cavaticus, Arapaima, Arch, Archidendron pauciflorum, Archipelago, Arctic char, Arctic ocean, Arctostaphylos, Arctostaphylos uva-ursi, Arecales, Argali, Argiope, Argynnis, Aricia, Aristea, Aristotelia chilensis, Arizona Black tailed Prairie Dog, Arizona Cypress, aromatic aster, Arpophyllum, Arrowgrass, Arrowroot, Arrowroot family, Arroyo, Artemisia, Arthropod, Artichoke thistle, Artificial fly, Arum, Arum family, Asclepiadoideae, Ash, Aspin, Aster, Asterales, Atlantic canary, Atlantic puffin, Attalea speciosa, Aubretia, Audubon's Cottontail, Auk, Auricularia, Australian cattle dog, Australian kelpie, Australian shepherd, avalanche lily, Baby carrot, Badlands, Bait fish, Bald eagle, balloon flower, balsam fir, Baltic clam, Baltic gray seal, Bamboo, Banana, Banana family, band winged grasshoppers, Banded water snake, Bandog, Banksia, Barbary fig, Barbary sheep, Barberry family, barberton daisy, Barbet, Barking Tree Frog, Barlia, Barn owl, Barren ground Caribou, Basset fauve de bretagne, Bath White, Batholith, Bavarian

mountain hound, Bay, Bayou, Beach, Beach moonflower, Beagador, Beagle, Beak, beaked hazelnut, Bear, Bearded collie, Bearded Seal, beardtongue, Beauceron, Beaver, Bedrock, Bedug, Bee, Bee balm, bee eater, Bee pollen, Beech, Beetle, Begonia, Bell pepper, Bell peppers and chili peppers, Bellflower, Bellflower family, Belostomatidae, Belted Kingfisher, Bengal tiger, Berberis, Berger blanc suisse, Bernese mountain dog, Berry, Bewick s Wren, Bichon, Big cats, bighorn, Bight, Bilberry, Biome, Birch, Birch family, Birch sap, Bird, Bird-of-paradise, Bird migration, Bird nest, bird of paradise, Bird of prey, Bison, Bittern, Bivalve, black-eyed susan, Black and white Warbler, Black capped Chickadee, Black cat, Black grouse, Black hawk, Black mamba, Black maple, Black mouth cur, Black mustard, Black norwegian elkhound, Black oak, Black Skimmer, Black stork, Black swallowtail, Black swan, Black tailed jackrabbit, Black throated Green Warbler, Blackberry, Blackbird, Blanket flowers, Blister beetles, Bloodhound, BloodrootSanguinaria canadensis, Blossom, Blowfish, blowflies, Blue-tongued skink, Blue cardinal flower, Blue eyed grass, Blue headed Vireo, Blue jay, Blue lacy, Blue sow thistle, Blue winged Teal, blue wood aster, blue woodland phlox, Blueberry, Bluebird, Bluebonnet, Bluegill, Bluetick coonhound, Boa, Boa constrictor, Boar, Boat tailed Grackle, Body of water, Boerboel, Bog, Bolboschoenus, Bolete, Bombycidae, Bombyliidae, Bombyx mori, Bonsai, Bony-fish, Borador, Borage family, Borassus flabellifer, Border terrier, Boreal Toad, Borzoi, Bosnian coarse-haired hound, Boston terrier, Botanical garden, Botany, bottlebush, Bottlenose dolphin, Bougainvillea, Boulder, Bovine, Box jellyfish, Box turtle, Boykin spaniel, Boysenberry, Braided river, Bramble, Brambling, Branch, Branched asphodel, Braque d'auvergne, Braque du bourbonnais, Braque francais, Braque saint-germain, Brassica rapa, Breckland thyme, Brewer's Blackbird, Broadleaf arrowhead, broadleaf pond lily, Broholmer, Bromelia, Bromeliaceae, Bronze hammerhead shark, Broomrape, Broomrape family, Brown bear, brown hare, Brown Pelican, Brown snake, Brown trout, Brush-footed butterfly, Buckthorn family, Bud, buddleia, Budgie, Buffaloberries, Bufflehead, Bufo, Bug, Bulbul, Bull, Bull and terrier, Bull shark, Bull terrier, Bull terrier (miniature), Bulldog, Bullfrog, Bullmastiff, Bullsnake, bulrush, Bumblebee, Burdock, Burmese python, Burnet rose, Burro, Bustard, Butomus, Butte, buttercup, Butterfly, butterfly milkweed, Butterflyfish, Buzzard, Cabbage, Cabbage butterfly, Cactus, Caesalpinia, Cairn terrier, Calabash, Calabaza, Calamondin, Caldera, Calendula, Calf, Calidrid, California condor, california lilac, California live oak, California newt, California sea lion, California slender salamander, California wild rose, Californian white oak, Calochortus, Camas, Camberwell Beauty, Camel, Camelid, Camellia, camomile, Canaan dog, Canada columbine, Canada goose, Canadian eskimo dog, Canadian fir, Canal, Canary, Candytuft, Cane corso, Cane toad, Canidae, Canis, canis lupus tundrarum, Canna family, Canna lily, Cannon, Canoe birch, Canopy, Cantaloupe, canvasback duck, Canyon, Caper family, Capybara, Caraway, Carcharhiniformes, Cardigan welsh corgi, Cardinal, Caridean shrimp, Carnivore, Carnivorous plant, Carolina Chickadee, Carolina dog, Carolina rose, Carolina wren, Carp, Carpathian shepherd dog, Carpenter ant, Carpenter bee, Carrot, Cartilaginous fish, Caryophyllales, Cash crop, Castilleja, Cat, Catahoula bulldog, Catalan sheepdog, Catchfly, Caterpillar, Catfish, cattle egret, Cattleya, Cattleya labiata, Cavachon, Cavalier king charles spaniel, Cavapoo, Cave, Cay, Cecropia Moth, Cedar Waxwing, Celastrina, Centaurium, Centella, Centipede, Central asian shepherd dog, Cephalopod, Cereal, Cestrum, Cetacea, Cetoniidae, Chaga mushroom, Chalk, Chalkhill blue, Chamaemelum nobile, Chameleon, Chamois, chamomile, Champignon mushroom, Channel, Chaparral, Charadriiformes, Chasmanthe, Chelidonium, Chelonoidis, Chelydridae, Cherimoya, Cherry, Cherry blossom, Chestnut-backed chickadee, Chestnut sided Warbler, Chickadee, Chicken, Chicory, Chihuahua, Chilopsis, china aster, Chinchilla, Chinese hawthorn, Chinese hibiscus, Chipmunk, Chloraea, Chocolate Daisy, Chokeberry, Chokecherry, Chorus frog, Christmas Orchid, Christmas tree, Chrysanthemum coronarium, Chrysanths, Chrysopogon zizanioides, Chrysops, Cicada, Ciconiiformes, Cinder cone, Cinquefoil, Cirque, Citrullus, Citrus, Clam, Clay, Claytonia, Clematis, Clementine, Cliff, Cliff Swallow, Climbing salamander, closed blue gentian, Cloudless Sulphur, Clover, cluster-lilies, Cnidaria, Coast, Coast horned lizard, Coastal and oceanic landforms, Coastal cutthroat trout, Coca, Coccoloba uvifera, Cockapoo, Cocker spaniel, Cockle, Cockroach, Coconut, Cod, Coenagrion, Coenonympha, coho, Colias, Colias croceus, Colias hyale, Colias sareptensis, Collard greens, Collared lizard, Collie, colobus, Colorado blue columbine, Colorado spruce, coltsfoot, Colubridae, Columbian spruce, Columbine, Combretaceae, Comfrey, Common bottlenose dolphin, Common Buckeye, Common chameleon, Common chimpanzee, Common dolphins, Common evening primrose, Common Gallinule, Common Garter Snake, Common Map Turtle, common milkweed, Common persimmon, Common rue, Common sage, common shepherd's purse, Common snapping turtle, Common tern, Common tormentil, Common Wood nymph, Common yabby, common yellow violet, common zinnia, Companion dog, Compost, Conch, Condor, Coneflower, Conger eel, Conifer, Conifer cone, Coonhound, Cooper's Hawk, Copepod, Coquelicot, Coraciiformes, Coral, coral aloe, Coral fungus, coral honeysuckle, Coral reef, Coral reef fish, Coregonus lavaretus, Corgi-chihuahua, Cormorant, Corn, Corn kernels, corn poppy, Cornales, Corona, Corset, Cosmos, Costus family, Coton de tulear, Coucal, Cougar, Cove, Cow-goat family, Cow parsley, Cowslip, Coyote, Crab, Cranberry, Crane-like bird, crape myrtle, Crassocephalum, Crater lake, Crayfish, Creek, creeping thistle, creeping wood sorrel, Crenate orchid cactus, Crepis paludosa, cretan crocus, Cricket, Cricket-like insect, Crimson columbine, Crinum, Crocodile, Crocodilia, Crocosmia, Crocus, Crow, Crow-like bird, crown of thorns, Cruciferous vegetables, Crustacean, Cuckoo, Cuckoo Wasps, Cuculiformes, Cucumber, Cucurbita, Cupido (butterfly), Curly coated retriever, Currant, Custard-apple, Cut flowers, Cutthroat trout, cutworms, Cynara, Cynorkis, Cynthia (subgenus), Cyprinidae, Cypripedium, Cytinus, Czechoslovakian wolfdog, Dachshund, Dactylorhiza praetermissa, Dagger, Dahlia, Dairy cow, Daisy, Daisy family, Dall's sheep, Dalmatian, Damselfly, dandelion, Dandelion, Danish swedish farmdog, Daphne, Dark eyed Junco, Dark green fritillary, Darkling beetles, Date palm, Datura, Datura inoxia, Davidson's Plum, Dayflower, Dayflower family, Daylily, Decapoda, Deciduous, Deep sea fish, Deer, Delphinium, Dendrobium, Desert, Desert horned lizard, Desert iguana, Desert Palm, desert rose, Desert tortoise, Dewberry, Dhole, Dianthus, Dicotyledon, Digitalis, Dike, Dill, Dingo, Dingy skipper, Distaff thistles, Ditch, Dive computer, Dobermann, Dog, Dog breed, Dogbane family, Dogo guatemalteco, Dogue de bordeaux, Dogwood family, Dolichopodidae, Dolphin, Domestic long-haired cat, Domestic pig, Domestic rabbit, Domestic short-haired cat, Domesticated turkey, Dorotheanthus bellidiformis, Dossinia, Double crested Cormorant, Douglas' squirrel, Dowitcher, Downhill, Downy Woodpecker, Dragon's mouth orchid, Dragon lizard, Dragonflies and damseflies, Dragonfly, Dragonfruit, Drainage basin, Drentse patrijshond, Drosophila melanogaster, Dry lake, Duck, Ducks, Dugong, Dune, Dung beetle, Dutch clover, Dutch smoushond, Dutchman's pipe, Eagle, Eagleray, Earl grey tea, Earless seal, Earth, earthenware, Earthstar, Earthworm, Earwigs, East-european shepherd, East siberian laika, Eastern Bluebird, Eastern box turtle, Eastern chipmunk, Eastern Indigo Snake, Eastern Meadowlark, Eastern newt, Eastern prickly pear, Eastern Screech owl, eastern skunk cabbage, Eastern Tailed blue, Eastern tent caterpillar, Eastern Tent Caterpillar, Eastern Tiger Swallowtail, Eastern Towhee, Eastern Wood pewee, Ebony trees and persimmons, Echeveria, Echinoderm, Ecoregion, Edible mushroom, Eggplant, Egret, Elaeis, Elapidae, Elder, Elderberry, Electric guitar, Electric ray, Electronics accessory, Elephant, Elephants and Mammoths, Eleutherodactylus, Elk, Elm, Elymus repens, Emberizidae, emperor moths, Emperor penguin, Emu, English cocker spaniel, English lavender, english marigold, English shepherd, English springer spaniel, English toy terrier, Enokitake, Entlebucher mountain dog, Epidendrum, Epiphyllum, Erosion, Erythronium 'pagoda', Escarpment, Eschscholzia californica, Estuary, Euanthe sanderiana, Eucalyptus, Eumenidae, Euphrasia, Euphydryas, Eurasian golden oriole, Eurasian magpie, Eurasian Red Squirrel, Eurasier, european gallinule, European garden spider, European green lizard, European herring gull, European marsh thistle, european michaelmas daisy, European plum, European robin, European shorthair, European Swallow, Euryops pectinatus, Evening primrose, Evening primrose family, Evening stock, Evergreen, Evergreen candytuft, Evergreen rose, Everlasting sweet pea, Extinct volcano, Faboideae, Falcon, Falconiformes, Fall webworm, Fault, Fawn, Fawn lily, Feather, Feeder fish, Felidae, Fell, Felty germander, Fen, Fence Lizard, Fennec fox, Fennel flower, Feral goat, Fern, Fernleaf lavender, Ferns and horsetails, Fiddlehead fern, Field, Fig, Figwort, Fin, Finch, Finnish hound, Finnish lapphund, Finnish spitz, Fir, fire cherry, Fish, Fish Crow, Fishbone cactus, Fissure vent, Fjord, Flamingo, Flat-coated retriever, Flatweed, Flax, Flightless bird, Flock, Flood, Floodplain, Floribunda, Florida Redbelly Turtle, Flower, flowering dogwood, Flowering plant, Fluvial landforms of streams, Fly, flying snake, Foal, Forage fish, Forb, Forest, Forget-me-not, Formation, Formosan mountain dog, Fossil, Fouquieria, Four o'clock family, Four o'clock flower, Fowl, Fox, Fox Sparrow, Fox squirrel, Foxtail lily, fragrant white water lily, Franklin s Gull, French bulldog, French lavender, Freshwater crab, Freshwater marsh, Fringe-toed lizard, Fritillaria, Frog, Fruit, Fruit tree, Fumaria, Fungus, Fur, Fur seal, Gagea, Gal<e1>pagos tortoise, Galanthus, Galliformes, Gannet, Gar, Garden, garden loosestrife, Garden roses, Garlic chives, Garter snake, Gaura, Gazania, Gazelle, Geastrales, Gecko, geese and swans, Genipa americana, Gentian family, Gentiana, Geoemydidae, Geological phenomenon, Geology, Georgia pine, Geraniaceae, Geraniales, geranium, Geranium, Geranium cinereum, Gerbera, Gerbil, German longhaired pointer, German pinscher, German shepherd dog, German shorthaired pointer, German spitz, German spitz mittel, Germanders, Gesneriad family, Geyser, Giant freshwater stingray, giant goldenrod, Giant granadilla, Giant otter, Giant Swallowtail, Gila monster, Giraffe, Giraffidae, Glacial lake, Glacial landform, Glacier, Glacier cave, Glanville fritillary, Glass lizard, Glechoma hederacea, Glen of imaal terrier, globe thistle, Goat, Goat-antelope, Goats, Golden eagle, Golden retriever, Golden samphire, Goldendoodle, Goldenrod, goldfinch, Goldfish, goldmoss stonecrop, Gonepteryx rhamni, Goosander, Goose, Gooseberry, Gopher, Gopher tortoise, Gopher Tortoise, Gordon setter, gourd, Gourd, Granite, Granite dome, Grape, Grapevine family, Grass, Grass family, Grass snake, Grasshopper, Grassland, Gravel, Gray Catbird, Grayling (butterfly), Great black-backed gull, Great blue heron, Great dane, Great egret, great grey owl, Great heron, Great horned owl, great masterwort, Great pyrenees, Great Spangled Fritillary, Great white shark, Greater burdock, greater scaup, Greater short-horned lizard, Greater swiss mountain dog, greater yellowlegs, Green algae, Green heron, Green iguana, Green pufferfish, Green sea turtle, Green winged Teal (American), Greenland dog, Grevillea, grey alder, grey fox, Grey squirrel, grey whale, Greyhound, Grizzly bear, Ground beetle, ground squirrels, Groundcover, Groundhog, Grouse, Grove, Guanaco, Guard dog, Guernsey lily, Gulf Fritillary, Gull, Gun dog, Gypsy moth, Gypsywort, Gyromitra, Hackberry Emperor, hackmatack, Hairfinned silverfish, Hamster, Harbor, Harbor seal, hard-leaved pocket orchid, Hare, Harebell, Harvestman, Harvestmen, Havanese, Hawaiian hibiscus, Hawk, hawk moths, Hawker dragonflies, Hawksbill sea turtle, hawkweed, Hawthorn, Hazard, Headland, Health shake, heath aster, Hedge, Hedgehog cactus, Heliconia, Hellebore, Heloderma, Hemp family, Hen-of-the-wood, Heracleum (plant), Herb, Herbaceous plant, Herd, Herding dog, Hericium erinaceus, Heron, Herring, Herring family, Hesperia (butterfly), Hesperia comma, Heteromeles, Hibiscus, High brown fritillary, Highland, Hill, Hippeastrum, Hippophae, Hobomok Skipper, Hofmannophila pseudospretella, Hognose snake, Holacanthus, Holly, Holly blue, Hollyhocks, Homarus, Homarus gammarus, Honeybee, Honeysuckle, Honeysuckle family, hooded merganser, Hooded Warbler, Hordeum, Horehound, Horn, Hornbill, Horned lizard, Hornet, Hornwort, Horse, Horsehair crab, Horsetail, Horsetail family, Hortaya borzaya, Hose, Hot spring, Hound, House finch, house fly, House sparrow, House Wren, Houseplant, Hovawart, Hoverfly, Huckleberry, Hummingbird, Humulus lupulus, Huntaway, Hunting dog, Hyacinth bean, hybrid clover, Hydrangea, Hydrangea serrata, Hydrangeaceae, Hyla, Hymenocallis, Hymenocallis littoralis, Hypericum, Hyssopus, Ibis, Ibizan hound, Ice, Ice cap, Ice cave, Ice plant, Ice plant family, Icicle, Igneous rock, Iguana, Iguania, Iguanidae, Ilex verticillataAmerican Winterberry, Impact crater, Impala, Indian cobra, Indian elephant, Indian spitz, Indigo bunting, Inlet, Inonotus, Insect, Intrusion, Invertebrate, Io moth, Iris, Iris family, Iris japonica, Iris reticulata, Iris versicolor, Irish soft-coated wheaten terrier, Irish terrier, Irish wolfhound, Ironweed, Islet, Isopod, Italian greyhound, Ivy, Ivy family, Ixia, Jack pine, Jackal, Jagdterrier, Jaguar, Japanese beetle, Japanese Camellia, Jasmine, Jay, Jelly fungus, Jellyfish, Jerusalem artichoke, jewel beetles, Jewel bugs, joe pye weed, Jumping Cholla, Junco, Jungle, Juniper, Juniper berry, Kaempferia rotunda, Kalua, Kangal dog, kangaroo, Kangaroo, Karelian bear dog, Karkalla, Keeshond, Kelp, Kemp's ridley sea turtle, Killdeer, Killer whale, King charles spaniel, King crab, King shepherd, kingcup, Kingsnake, Kinosternidae, Kintamani, Kitten, Klippe, Koala, Kodiak bear, Koi, Koolie, Korat, Korean jindo dog, Kudu, Kunming wolfdog, Kuvasz, Labradoodle, Labrador husky, Labrador retriever, Lacerta, Lacustrine plain, lady tulip, Ladybug, Lagoon, Lagotto romagnolo, Lake, Lake district, Lake Erie Water Snake, Lamnidae, Lamniformes, Lampranthus, Lancashire heeler, Landscape, Lantana, Lapponian herder, Larch, Large-flowered cactus, Large-flowered evening primrose, Large skipper, Large tortoiseshell, Large White, Lari, Larix lyalliiSubalpine Larch, Lark, Larva, Laughing Gull, Lava, Lava cave, Lava dome, Lava plain, Lava tube, Lavandula dentata, Lavender, Lazuli Bunting, Leaf, Leaf beetle, Leaf Footed Bugs, Leaf vegetable, Leafhopper, Least Sandpiper, Least Skipper, Lecythidaceae, Ledum, Legume, Legume family, Lemon balm, Lemon beebalm, Lemur, Leopard, Leptidea, lesser burdock, Lesser celandine, Lesser Scaup, Lesser skullcap, Lettuce, Leucaena, Leuconotopicus, Levee, lilac, Lily, Lily family, Lima bean, Lime, Limenitis, Limestone, Limpkin, Lingonberry, Lingzhi mushroom, Lion, lionfish, Little blue heron, Little egret, Liverwort, Livestock, Livestock guardian dog, Lizard, Llama, Lobelia, loblolly pine, Lobster, Loch, Locust, lodgepole pine, Loggerhead sea turtle, Loggerhead Shrike, Longhaired whippet, Longhorn beetle, longstem marsh violet, loon, Loon, Loosestrife and pomegranate family, Lotus, Lotus family, Louisiana catahoula leopard dog, Lovebird, Lower Keys Marsh Rabbit, Ludisia, Luffa, Lulworth skipper, Luna Moth, Lungless salamander, Lupinus mutabilis, Lurcher, Lycaena, Lycaenid, Lycaon pictus, Lychee, Lymantria dispar dispar, Lymnaeidae, Macaque, Macaw, Macrocystis, Macrocystis pyrifera, Macropodidae, Madagascar hissing cockroach, Magnolia, Magnolia family, Magnolia Warbler, Magpie, Mahi Mahi, Mahonia, Maidenhair tree, Maine coon, Makhtesh, Mallard, Mallow family, Maltepoo, Maltese, Malus, Malvales, Mamba, Mammal, Manatee, Manchester terrier, Mandarin orange, Manta ray, Mantidae, Mantis, Mantis shrimp, Maple leaf, Marabou stork, Mare, Maremma sheepdog, Marguerite daisy, Marine biology, Marine iguana, Marine invertebrates, Marine mammal, Marjoram, Marmot, Marsh, Marsh fritillary, Marsh labrador tea, Marsh pea, Marsupial, Masai lion, Masked lapwing, Massif, Matsutake, mayflies, mayweed, Meadow, meadow jumping mouse, Meadow Vole, Meadowsweet, Medicinal mushroom, Meerkat, Megachilidae, Megalith, Melastome family, Melitaea, Melon, Membrane-winged insect, Menispermaceae, Mergus, Mertensia, Mexican hairless dog, mexican petunia, Milkweed, millipedes, miner's lettuce, Mineral, Mineral spring, Miniature australian shepherd, Miniature pinscher, Miniature Poodle, Mink, Miridae, mock orange, Mold, Mole salamander, Molluscs, Monarch butterfly, Mongoose, monkshood, Moonflower, Moonlight cactus, Moose, Moraine, Morkie, Morning glory family, Moschatel family, Mosquito, Moss, Moth, moth orchid, Moth Orchid, Moths and butterflies, Mound, Mound-building termites, Mount scenery, Mountain, mountain alder, Mountain Bluebird, Mountain Cottontail, Mountain cur, Mountain goat, mountain laurel, Mountain pass, Mountain range, Mountain river, Mountainous landforms, Mouse, Mud, Mudflat, Mudhol hound, Mulberry family, Mulch, Multiflora rose, Muridae, Muroidea, Mushroom, Musk deer, Muskrat, Mustang horse, Mustard and cabbage family, Mustard plant, Mustelidae, Mustelinae, Myna, Myrtle family, Nannyberry, Naranjilla, Narcissus, Narrow-leaved sundrops, Narrows, National park, Native american indian dog, Native raspberry, Native Sowthistle, Natural arch, Natural environment, Natural landscape, Natural material, Nature, Nature reserve, Nectar, Neotinea ustulata, Nepenthes, Nepeta, Nest, Net-winged insects, Nettle family, New caledonian crow, new england aster, New guinea singing dog, New Mexico maple, new york aster, Newfoundland, Newt, night heron, Nightingale, Nightshade family, Nile crocodile, Non-vascular land plant, North American newt, North american river otter, Northern Alligator Lizard, Northern Cardinal, Northern Copperhead, Northern flicker, Northern Grey Shrike, Northern hardwood forest, Northern Harrier, Northern inuit dog, Northern largemouth bass, Northern leopard frog, northern mockingbird, Northern pike, Northern shoveler, Northern two-lined salamander, Norwegian elkhound, Norwegian forest cat, Norwegian lundehund, Norwich terrier, Nova scotia duck tolling retriever, Noxious weed, Nunatak, Nurse shark, Nut, nutria, Nuts & seeds, Nymphalis, Nymphalis xanthomelas, Oak, Oat, Ocean, Ochlodes, Ocicat, Octomeria, Octopus, Oecanthidae, Oily fish, Olallieberry, Old-growth forest, Old english terrier, old field clover, Old german shepherd dog, Old World flycatcher, Old world monkey, Old World oriole, Olde english bulldogge, Olive ridley sea turtle, Oncorhynchus, Oniscidea, Ononis, orange hawkweed, Orange lily, Orange Sulphur, Orangutan, Orb-weaver spider, Orchid, Orchids of the philippines, Oregon grape, oregon pine, Organ Mountains Chipmunk, Organism, oriental poppy, oriole, Ormosia, Orris root, Ortolan bunting, Osprey, Ostrich, Ostrich fern, Otter, Outcrop, Owl, Ox, Oxeye daisy, Oyster, Oyster mushroom, Oystercatcher, Pacific newt, Pacific rhododendron, Pacific Treefrog, Pack animal, Paddy field, Painted Bunting, Painted turtle, Palamedes Swallowtail, Palm tree, Pansy, Papilio, Papilio machaon, Parakeet, Pararge, Parasite, parlour maple, Parrot, Parrotfish, Parsley family, Partridge, pasqueflower, Passion flower, Passion flower family, Passion fruit, Pasture, Patterdale terrier, Paw, Pea, Peach, Peafowl, Peanut, Pearl-bordered fritillar, Pearl Crescent, Pebble, Peccary, Peck s Skipper, Pedicel, Pekingese, Pelecaniformes, Pelican, Pembroke welsh corgi, Penguin, Peninsula, Peniocereus, Penny bun, Peony, peppered moth, Peppermint, Perching bird, Peregrine falcon, Perennial plant, perennial sowthistle, perforate st john's wort, Pericallis, Perilla, periwinkle, Perro de presa canario, Peruvian hairless dog, Peruvian lily, Pest, Petal, Petunia, Pezizales, Phalaenopsis sanderiana, Phallales, Phasianidae, Pheasant, Phengaris, Phragmites, Phyllanthus family, Phyllobates, Phytolaccaceae, Piciformes, Pied billed Grebe, Pieridae, Pigeons and doves, Pike, Pileated woodpecker, Pilotfish, Pine, Pine family, Pine nut, Pine Siskin, Pineapple, pink moccasin flower, Pinscher, Pintail, Piper auritum, Pipevine Swallowtail, Pistachio, Pistacia lentiscus, Piste, Pit bull, Pit cave, Pitcher plant, Plain, Plains Gartersnake, Plane-tree family, Plankton, Plant, Plant community, Plant pathology, Plant stem, Plantago, Plantation, Plateau, Plebejus, Pleurotus eryngii, Plott hound, Poales, Podophyllum peltatum, Pointing breed, Poison dart frog, Polar bear, Polar ice cap, Polder, Police dog, Polish greyhound, Polish hound, Polish tatra sheepdog, Pollen, Pollinator, Polygonia, Polyommatus, Polyphemus moth, Polyporales, Polystachya, Pomacanthidae, Pomacentridae, Pomeranian, Pond, Pond turtle, Pontia, Pony, Poodle, Poodle crossbreed, Pool, Poppy, Poppy family, Porcupine fishes, Porpoise, Portuguese man o' war, Portuguese water dog, Posavac hound, Possum, Potcake dog, Poultry, Prairie, Prairie dog, prairie pasqueflower, prairie vole, prickly pear, Prickly pear, Prickly rose, pride of madeira, Primate, Primula, Procyonidae, Promethea Silkmoth, Pronghorn, Protea family, Prunus spinosa, Prussian asparagus, Pseudemys concinna concinna, Psyllium seed husks, Puffin, Puggle, Pulasan, Puma, Pumpkin, Pupa, Puppy, Purple coneflower, Purple Gallinule, Purple loosestrife, Purple Martin, purple milkweed, Purple passionflower, Purple salsify, Pyrenean mastiff, Pyrrharctia isabella, Python, Python family, Quail, Quarry, Quartz, queen's lady's-slipper, Rabbit, Rabbits and Hares, Radish, Rafeiro do alentejo, Rainforest, Raised beach, Rallidae, Ram, Rangpur, Rapeseed, Rapid, Rare breed (dog), Raspberry, Rat, Rat terrier, Ratite, Ratonero bodeguero andaluz, Rattlesnake, raven, Raven, Ravine, Ray-finned fish, Rays and skates, red-backed sandpiper, Red-breasted merganser, Red-tailed hawk, Red back salamander, Red bellied Woodpecker, Red bud, Red bugs, Red clover, Red eared slider, Red fox, Red headed Woodpecker, Red juniper, Red leaf lettuce, Red mulberry, red pine, Red salamander, Red shouldered Hawk, Red snapper, Red spotted Toad, red trillium, Red winged Blackbird, Red wolf, Redwood, redwood sorrel, Reed, Reef, Reindeer, Relief, Reptile, Requiem shark, Reservoir, Retriever, Rhodesian ridgeback, Rhododendron, Rhododendron catawbiense, Ribbon snake, Ridge, Ring-necked pheasant, Ring billed Gull, Ringed-worm, Riodinidae, Riparian forest, Riparian zone, River, River Birch, River cooter, River delta, River island, River monitor, roadrunner, robber flies, robin, Rock, rock rose, Rock samphire, Rodent, Roe deer, Rook, Rooster, Root, Root vegetable, Rosa canina, Rosa dumalis, Rosa nutkana, Rosa omeiensis, Rosa rubiginosa, Rosa rugosa, Rosa sericea, Rosa wichuraiana, Rose, Rose breasted Grosbeak, Rose family, Rose hip, Rose order, Rosemary, Rosy garlic, Rottweiler, Rough collie, round leaved liverleaf, Rowan, Roystonea, Rubber Boa, Rubble, Rubus, Ruby-throated hummingbird, Ruddy Duck, Ruffed grouse, Rufous Hummingbird, Russo-european laika, Russula integra, Rye, Saarloos wolfdog, Sabal palmetto, Saccharina japonica, Sacred lotus, Sage, Sagebrush lizard, Sahara, Sakhalin husky, Salamander, Salamandra, Salmon, Salmon-like fish, Salmonberry, Salt evaporation pond, Salt lake, Salt marsh, Saltbush, Saltwater crocodile, Saluki, Samoyed, San Pedro cactus, Sand, Sand tiger shark, Sandhill crane, sandpiper, Sandpiper, Sapodilla, Sapodilla family, Sapsali, Sardine, Savanna, Savannah Sparrow, Sawfly, Scaled reptile, Scaphosepalum, Scarabs, Scarlet gourd, Scarlet oak, Scarlet Tanager, Scentless Plant Bugs, Schapendoes, Schipperke, Schisandrap, Schnoodle, Scilla, Scorpion, Scorpionfish, Scotch collie, Screech owl, Scrub Jay, Sea, Sea anemone, sea aster, Sea cave, Sea cows, Sea eagle, Sea ice, Sea lettuce, Sea otter, Sea snail, Sea turtle, Seabird, Seaduck, Seahorse, Seal, Seamount, Seaweed, Sedge family, Seed, Seedless fruit, Sego lily, Semipalmated Plover, Seppala siberian sleddog, Serbian tricolour hound, Sewellel, Shadbush, Shamrock, Shark, Sharp shinned Hawk, Sheep, Sheep's sorrel, shellbark hickory, Shellfish, Shetland pony, Shetland sheepdog, Shiba inu, shield bugs, Shield volcano, Shih-poo, Shih tzu, Shiitake, Shiloh shepherd dog, Shoal, Shore, Shorebird, Short-beaked common dolphin, Short-tailed blue, shortleaf black spruce, shortstraw pine, Shrike, Shrimp, Shrub, Shrub frog, Shrubland, Siberian fawn lily, Siberian husky, Siberian tiger, Side-blotched lizards, Sidewinder, Sighthound, Silene noctiflora, Silene nutans, Silk tree, Silken windhound, Silver-studded blue, Silver-washed fritillary, Silver spotted Skipper, silvertip fir, Silvery Checkerspot, Silybum, Singing sand, Sinkhole, sitka spruce, Skerry, Skink, Skipper (butterfly), Sled dog, Slope, Sloth bear, Sloughi, Slovak cuvac, Slug, Small skipper, Small to medium-sized cats, small white aster, smartweed-buckwheat family, smooth aster, Smooth collie, Smooth fox terrier, Smooth newt, smooth Solomon's seal, smooth sumac, Snail, Snails and slugs, Snake, Snake's head, Snapdragon, Snapper, Snout, Snout moths, Snow, Snow Bunting, Snow goose, Snow leopard, Snowdrop, Snowshoe hare, Snowy Egret, Snowy owl, Soapberry family, Soil, Solanales, Solanum, Solanum lycocarpum, Soldier beetle, Solomon's plume, Solomon's seal, Song Sparrow, Songbird, Sorbus, Sorrel, Southern Dogface, Southern Leopard Frog, southern magnolia, Sow thistles, Spaniel, spanish daisy, Spanish water dog, Sparrow, Spear thistle, Spearmint, Speckled wood (butterfly, Spectacled bear, Speleothem, Sperm whale, Spicebush Swallowtail, Spider, Spider web, Spinach, spines, Spinner dolphin, Spinone italiano, Spiny lobster, spirea, Spoonbill, Sporting lucas terrier, Spotted burclover, Spotted hyena, spotted jewelweed, Spotted Joe pye Weed, Spotted knapweed, Spotted salamander, Spotted Sandpiper, spring crocus, Spring Salamander, Springbok, Springtail, Spruce, Spruce-fir forest, Spurge family, Squaliformes, Squash, Squirrel, Squirrel tree frog, stable fly, Stabyhoun, Stachys affinis, Stack, Staffordshire bull terrier, Stag beetles, Stalactite, Stalagmite, Stallion, Stamp seal, Standard Poodle, Starfish, Starling, State park, Steller's sea eagle, Steller s Jay, Steller sea lion, Stenella, Steppe, Stevia rebaudiana, Stilt, Stingray, Stinkhorn mushroom, Stitchwort, Stock dove, Stonecrop family, Stony coral, Stork, Stratovolcano, Strawberries, Strawberry, Strawberry guava, Stream, Stream bed, Street dog, Striated Heron, Striped dolphin, striped squill, Subshrub, Succulent plant, sugar pine, Suidae, Sulfur Cosmos, Suliformes, Sumac, Summer squash, Summit, sunflower, Sunflower, Sussex spaniel, Sutherlandia frutescens, Swallow, Swallowtail butterfly, Swamp, swamp birch, Swamp Rose mallow, Swamp Sparrow, Swan, Swedish lapphund, Swedish vallhund, sweet birch, Sweet corn, Sweet granadilla, Sweet grass, Sweet gum, Sweet pea, Sweet peas, sweet pepperbush, Sweet scabious, Sweetscented bedstraw, Swift fox, Tabby cat, Tachinidae, Tadpole, Tagetes, Tagetes patula, Tall cinquefoil, Tambora, Tangelo, Tangerine, Tangle-web spider, Tansy, Tarantula, Tarn, Teasel, Teddy roosevelt terrier, Teff, Temperate broadleaf and mixed forest, Temperate coniferous forest, Tench, Termite, Tern, Terrain, Terrapin, Terrestrial animal, Terrestrial plant, Terrier, texas bluebonnet, Texas longhorn, Theaceae, Thermokarst, Thistle, Thorns, thuya, Thymelicus, Tibetan terrier, Tickseed, Tidal marsh, Tide, Tide pool, Tidy tips, Tiger, Tiger beetle, tiger lily, Tiger salamander, Tiger shark, Toad, toad lily, tommie crocus, Torch lily, Tornjak, Tortoise, Tosa, Toucan, Toy fox terrier, Toy Poodle, Trachemys, Trachyspermum ammi, Tree, Tree frog, tree poppy, Tree stump, Treehopper, Treeing feist, Tremella, Tributary, Triggerfish, Trillium, Tropical and subtropical coniferous forests, Trout, True frog, True salamanders and newts, True toad, trumpet creeper, Trumpeter swan, Trunk, Tuberose, Tuberous pea, Tucuxi, Tulip, Tulip poplar, Tulipa humilis, Tundra, Tundra swan, Turkey, Turkey Vulture, Turtle, Twig, Types of volcanic eruptions, Underground lake, Underwater, Underwing moths, upright yellow sorrel, Urtica, Vaccinium arboreum, Valdivian temperate rain forest, Valencia orange, Valerian, Valley, Vanessa (butterfly), Vanessa atalanta, Vanessa cardui, Variegated Fritillary, Vascular plant, Vegetable,

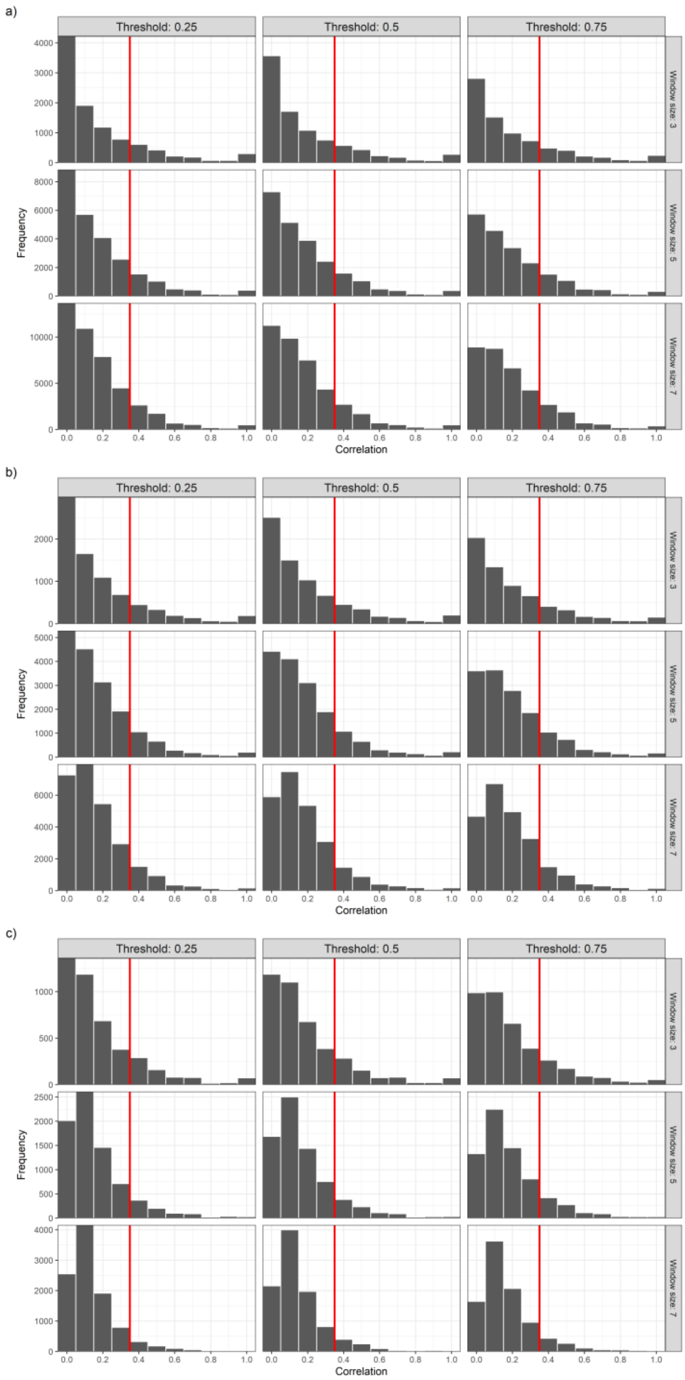
Vegetation, Veratrum, Verbascum, Verbena, Verbena family, Vertebrate, Viburnum, Viceroy (butterfly), Vine, Vineyard, Violet family, Violet woodsorrel, Viper, Virginia lungwort, Virginia Rose, Vitis, Viverridae, Vizsla, Volcanic crater, Volcanic field, Volcanic landform, Volcanic plug, Volcanic rock, Volcano, Vulture, Wadi, walking stick insect, Wall lizard, Wallflower, Walrus, Warthog, Warty newt, wasp, Wasp, Water, Water bird, Water dropwort, Water feature, Water forget me not, water lily, Water resources, water smartweed, Water snake, Watercourse, Waterfall, Waterfowl, Watermelon, Waterway, Wave, Waxwing, Waxworm, webbing clothes moth, Weed, Weevil, Weimaraner, Welsh Corgi, Welsh terrier, West highland white terrier, West indian gherkin, West Indian raspberry, West siberian laika, Western alligator lizard, Western conifer seed bug, Western Fence Lizard, Western Grebe, Western Gull, Western Kingbird, Western Meadowlark, Western Screech owl, Western Tiger Swallowtail, Western Whiptail, western yellow pine, Wetland, Whale, Whale shark, Wheat, Whippet, Whiptail, Whiskers, White-tailed deer, White Admiral or Red spotted Purple, White breasted Nuthatch, white cockle, White coffee, White crowned Sparrow, White horehound, White mulberry, White Pelican, White pine, White shepherd, White stork, white throated sparrow, white trillium, Wholphin, Whooping crane, wild carrot, Wild cat, wild cranesbill, Wild ginger, wild pansy, wild sweet potato, Wild turkey, Wilderness, Wildflower, Wildlife, Willet, Wind wave, windflower, Wine raspberry, winter aconite, Winter squash, Wire hair fox terrier, Wirehaired pointing griffon, Wisteria, Wolf, Wolf spider, Wolfdog, wombat, Wombat, Wood, wood duck, Wood ear, Wood Frog, wood rabbit, Wood sorrel family, Wood stork, Woodland, Woodland salamander, woodland sunflower, Woodpecker, Woodpecker finch, Woods' rose, Woody plant, Woolly sunflower, Working animal, Working dog, Worm, Wren, Xanthorrhoeaceae, Yak, Yarrow, yellow avalanche lily, Yellow breasted Chat, yellow Canada lily, Yellow fir, Yellow garden spider, Yellow headed Blackbird, yellow iris, yellow lady's slipper, Yellow nutsedge, Yellow rumped Warbler, Yellow Warbler, Yellowtail amberjack, Yucca, Zebra, Zebra Swallowtail, Zebu, zigzag clover, Zingiberales, Zinnia, Zinnia angustifolia, Zipper, Zooplankton, Zophobas morio"

D.2 Non-nature labels

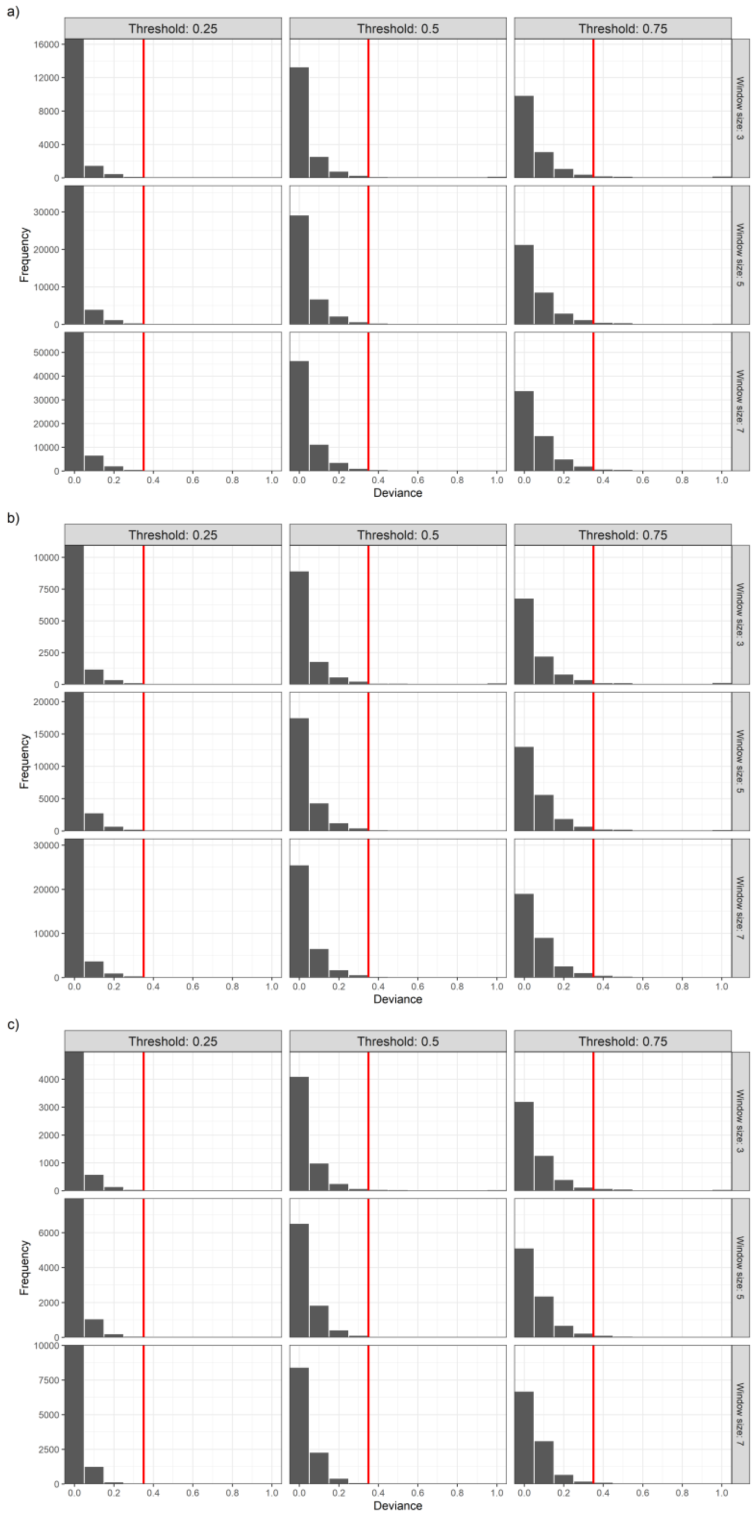
Non-nature words
<c0> la carte food, <c0> la carte food, <c9>p<e9>e, 100 metres hurdles, 110 metres hurdles, 1955 ford, 3d modeling, 3x3 (basketball), Abbey, Abdomen, Abrasive saw, Abseiling, Academic certificate, Academic conference, Academic dress, Acanthocereus tetragonus, Accordion, Accordionist, Acoustic-electric guitar, Acoustic guitar, Acrobatics, Acrylic paint, Acting, Action-adventure game, Action figure, Action film, Active pants, Active shirt, Acura, Acura mdx, Acura

rdx, Acura tsx, Adaptation, Address sign, Admiral, Adonis, Adventure, Adventure game, Adventure racing, Advertising, Aerial photography, Aerobatics, Aerobic exercise, Aerobics, Aerospace engineering, Aerospace manufacturer, Aerostat, Affogato, Afro, Afterglow, Agama, Agave nectar, Agricultural machinery, Agriculture, Aguas frescas, Air-raid shelter, Air force, Air gun, Air hockey, Air racing, Air show, Air sports, Air travel, Airboat, Aircraft, Aircraft cabin, Aircraft carrier, Aircraft engine, Airline, Airliner, Airman, Airplane, Airport, Airport apron, Airport terminal, Airship, Airsoft, Airsoft gun, Aisle, Album cover, Alcohol, Alcoholic beverage, Ale, Alfa romeo, Alfa romeo 8c competizione, Alfa romeo mito, All-terrain vehicle, Alley, Alloy wheel, Almond milk, Almshouse, Alpine skiing, Altar, Alto horn, Aluminium, Aluminium foil, Aluminum can, Alyssum, Amber, Ambulance, American food, American football, American frontier, Ammunition, Ammunition belt, Ammunition box, Amphitheatre, Amusement park, Amusement ride, Analog watch, Ancient history, Ancient roman architecture, Andouille, Angel<U+0092>s Tear, Angle grinder, Angling, Angry birds, Animal cracker, Animal fat, Animal feed, Animal figure, Animal shelter, Animal sports, Animal training, Animated cartoon, Animation, Anime, Ankle, Antenna, Antique, Antique car, Antique tool, Antonov an-2, Apartment, Apiary, appetizer, Apple cider, Apple crisp, Aqua, Aquanaut, Aquarium, Aquarium decor, Aquarium lighting, Aqueduct, Ar<ea>te, Ar<ea>te, Arbor day, Arborist, Arcade, Arcade game, Arch bridge, Archaeological site, Archaeology, Archery, Architecture, Arctic, Arena, Arena football, Arm, Arm wrestling, Armored car, Armored cruiser, Armour, Armrest, Army, Army men, Arrow, Art, Art exhibition, Art gallery, Art model, Art paper, Artifact, Artificial flower, Artificial island, Artificial turf, Artillery tractor, Artisan, Artist, Artwork, Asian, Asphalt, Astronomer, Astronomical object, Astronomy, Athlete, Athletic dance move, Athletic shoe, Athletics, Atlas, Atmosphere, Atmospheric phenomenon, Attic, Audi, Audi a7, Audi r8, Audience, Audio equipment, Auditorium, Aurora, Australian collie, Auto mechanic, Auto part, Auto race, Auto racing, Auto show, Autocross, Autograph, Autoharp, Automobile repair shop, Automotive bicycle rack, Automotive carrying rack, Automotive design, Automotive engine part, Automotive exhaust, Automotive exterior, Automotive ignition part, Automotive lighting, Automotive luggage rack, Automotive mirror, Automotive navigation system, Automotive parking light, Automotive side-view mirror, Automotive super charger part, Automotive tail & brake light, Automotive tire, Automotive wheel system, Automotive window part, Autumn, Aviation, Award, Award ceremony, Awning, Axe, Azure, Baby, Baby & toddler clothing, Baby bathing, baby blue eyes, Baby bottle, Baby carriage, Baby carrier, Baby float, Baby Products, Baby sleeping, Baby toys, Bachelorette party, Back, Back bacon, Backlighting, Backpack, Backpacking, Backstroke, Backyard, Bacon, Bacon sandwich, Badge, Badminton, Bag, Bagel, Baggage, Bagpipes, Baguazhang, Bailey bridge, Bait, Bake sale, Baked beans, Baked goods, Baker, Bakery, Baking, Balance, Balcony, Balinese, Ball, Ball game, Ball pit, Ballet tutu, Balloon, Ballroom, Baluster, Bamboo flute, Band plays, Bandage, Bandrek, Bandy, Bangle, Bangs, Banjo guitar, Bank, Banknote, Banner, Banquet, Bansuri, Baptistery, Bar, Bar stool, Barbecue, Barbecue grill, Barbecue sauce, Barbed wire, Barbell, Barber, Barbie, Barechested, Barefoot, Baritone saxophone, Barley, Barley water, Barn, Barque, Barquentine, Barrel, Barricade, Bartender, Barware, Baseball, Baseball bat, Baseball cap, Baseball equipment, Baseball field, Baseball park, Baseball player, Baseball uniform, Basement, Basic pump, Basilica, Basket, Basketball, Basketball court, Basketball hoop, Basketball moves, Basketball official, Basketball player, Basketball shoe, Bass, Bass boat, Bass drum, Bass guitar, Bassist, Bat-and-ball games, Bath toy, Bathing, Bathroom, Bathroom accessory, Bathroom cabinet, Bathroom sink, Bathtub, Batman, Batting glove, Battle, Battle gaming, Battleship, Bay breeze, Bazaar, Beach soccer, Beach volleyball, Beacon, Bead, Beam, Beam bridge, Beanie, Beard, Bearing, Bearskin, Beauty, Bed, Bed frame, Bed sheet, Bed skirt, Bedding, Bedroom, Bedtime, Beef, Beehive, Beer, Beer bottle, Beer cocktail, Beer glass, Beer pong, Beer stein, Beer tap, Beige, Belay device, Belgian waffle, Bell, Bell 206, Bell 212, Bell 214, Bell 412, Bell boeing v-22 osprey, Bell tower, Bell uh-1 iroquois, Belly dance, Belt, Belt buckle, Bench, Benchrest shooting, Bentley, Benz patent-motorwagen, Beret, Beverage can, Bia hoi, Biathlon, Bicycle, Bicycle accessory, Bicycle basket, Bicycle chain, Bicycle clothing, Bicycle drivetrain part, Bicycle fork, Bicycle frame, Bicycle handlebar, Bicycle helmet, Bicycle mechanic, Bicycle motocross, Bicycle part, Bicycle pedal, Bicycle racing, Bicycle saddle, Bicycle seatpost, Bicycle stem, Bicycle tire, Bicycle trailer, Bicycle trainer, Bicycle wheel, Bicycle wheel rim, Bicycles--Equipment and supplies, Bigtree, Bikejoring, Bikini, Billboard, Bin bag, Biniou, Binoculars, Bioluminescence, Biplane, Bird's-eye, Bird's-eye view, Bird bath, Bird feeder, Bird food, Bird supply, Birdhouse, Birth, Birthday, Birthday cake, Biscuit, Bishop, Bit, Black, Black-and-white, Black hair, Blackberry pie, Blackboard, Blacksmith, Blade, Blanket, Blazer, Blended whiskey, Blessing, Blizzard, Block party, Blond, Blouse, Blowing horn, Blowout, Blt, Blue, Blue-collar worker, Blue and white porcelain, blue violet, Blueberry pie, bluejacket, Bmw, Bmw 2002tii, Bmw new class, Bmw x1, Bmw z1, Bmx bike, Bmx racing, Board game, board short, Board track racing, Boardsport, Boardwalk, Boat, Boat trailer, Boathouse, Boating, Boats and boating--Equipment and supplies, Bob cut, Boba fett, Bobsleigh, Bocce, Body jewelry, Bodyboarding, Bodybuilder, Bodybuilding, Bodypump, Boeing b-17 flying fortress, Boeing c-97 stratofreighter, Boeing ch-47 chinook, Boeing e-3 sentry, Boeing vertol ch-46 sea knight, Boiling, Bolognese sauce, Bombard, Bone, Bonfire, Bongo, Bonnet, Book, Book cover, Bookcase, Bookselling, Boombox, Boot, Border collie, Bottle, Bottle cap, Bottled water, Boudin, Bouldering, Bouleuterion, bounce house, Bouquet, Boutique, Bow, Bow and arrow, Bowed string instrument, Bowie knife, Bowl, Bowl cut, Bowler, Bowler hat, Bowling, Bowling ball, Bowling equipment, Bowling pin, Bowls, Box, Box girder bridge, Boxer, Boxing, Boxing ring, Boy scouts of america, Bracelet, Braising, Brake, Brand, Brass, Brass instrument, Brassiere, Bratwurst, Bread, Breakfast, Breakfast cereal, Breakfast sandwich, Breakfast sausage, Breakwater, Breaststroke, Brewery, Brick, Bricklayer, Brickwork, Bridal accessory, Bridal clothing, Bridal party dress, Bridal shoe, Bride, Bridge, Bridle, Briefcase, Briefs, Brimstones, Briquet griffon vend<e9>en, Brisket, Brittany, Broadaxe, Broadcasting, Brochette, Brochure, Bronze, Bronze sculpture, Brown, Brown hair, Brug, Brunch, Bruschetta, Brutalist architecture, Bucket, Buckle, Buffalo burger, Buffalo wing, Buffet, Bugatti, Bugatti type 35, Bugle, Buick century, Buick invicta, Building, Building insulation, Building sand castles, Bulk carrier, Bull riding, Bulldozer, Bullet, Bulletin board, Bumper, Bumper sticker, Bungee jumping, Bunk bed, Bunker, Burr truss, Bus, Bus stop, Business, Business bag, Businessperson, Butorides, Buttercream, Butterfly stroke, Buttermilk, Button, Button accordion, Buzz cut, Byzantine architecture, C<e3>o da serra de aires, C<e3>o da serra de aires, Caber toss, Cabinetry, Cable, Cable-stayed bridge, Cable car, Cable management, Cadillac, Cadillac sts, Caf<e9>, Caf<e9>, Caf<e9> au lait, Cafeteria, Caff<e8> americano, Caff<e8> macchiato, Caffeine, Cage, Caipirinha, Cake, Cake decorating, Cake decorating supply, California-style pizza, Calligraphy, Calm, Camera, Camera accessory, Camera lens, Camera operator, Cameras & optics, Camouflage, Camp, Campfire, Camping, Campus, Canadian football, Canal tunnel, Candle, Candy, Cani cross, Canning, Canoe, Canoe slalom, Canoeing, Canola, Canopy bed, Canopy walkway, Cantilever bridge, Canyoning, Cap, Cape, Caper, Cappuccino, Caprese salad, Captain america, Car, Car dealership, Car seat, Car seat cover, Carabiner, Caramel color, Caravan, Caravanserai, Caravel, Carbon, Carbonated soft drinks, Carburetor, Card game, Cardboard, Cargo, Cargo ship, Caribbean, Carmine, Carnitas, Carnival, Carousel, Carpenter, Carpet, Carrack, Carriage, Cart, Carton, Cartoon, Carving, Cash, Casino, Cassette deck, Casting (fishing), Castle, Cathedral, Cauldron, Cauliflower cheese, Caving, Ceiling, Ceiling fan, Ceiling fixture, Celestial event, Cello, Cellular network, Cement, Cemetery, Center console, Centrepiece, Ceramic, Cereal germ, Ceremony, Cervelat, Cessna 206, Cg artwork, Ch<e2>teau, Ch<e2>teau, Chain, Chain-link fencing, Chainsaw, Chainsaw carving, Chair, Chaise longue, Chambered cairn, Champagne, Champagne cocktail, Champagne stemware, Championship, Chandelier, Changing room, Chapel, Chapulines, Charcoal, Chariot, Charoset, Chassis, Cheek, Cheering, Cheerleading, Cheerleading uniform, Cheese, Cheesesteak, Chef, Chemical compound, Cheque, Chess, Chessboard, Chest, Chest hair, Chest of drawers, Chevrolet, Chevrolet advance design, Chevrolet bel air, Chevrolet camaro, Chevrolet cavalier, Chevrolet colorado, Chevrolet corvette, Chevrolet corvette c6 zr1, Chevrolet hhr, Chevrolet malibu, Chevrolet nomad, Chevrolet superior, Chevrolet venture, Chiavari chair, Chicken coop, Chicken feet, Chicken fried steak, Chicken meat, Chicken tikka, Chiffonier, Child, Child art, Child model, Chime, Chimney, Chin, Chinese architecture, Chinese chicken salad, Chinese food, Chinese new year, Chocolate, Chocolate bar, Chocolate brownie, Chocolate cake, Chocolate ice cream, Chocolate truffle, Choir, Chopper, Chopsticks, Choreography, Christmas, Christmas decoration, Christmas eve, Christmas lights, Christmas ornament, Christmas stocking, Chrysler, Chrysler 300, Chrysler 300c, Chrysler concorde, Chrysler neon, Chrysler pt cruiser, Chrysler voyager, Church, Church bell, Churchill tank, Churrasco food, Chute, Cider, Cimarr<f3>n uruguayo, Cimbalom, Cinco de mayo, Cinematographer, Circle, Circuit component, Circuit training, Circular saw, Circus, City, City car, Cityscape, Clamp, Clarinet, Clarinet family, Clarinetist, Clary, Class, Classic, Classic car, Classic cocktail, Classical architecture, Classical music, Classical sculpture, Classroom, Claw, Clay animation, Clay pigeon shooting, Cleanliness, Cleat, Clergy, Cliff dwelling, Climbing, Climbing harness, Climbing hold, Clinic, Clip art, Clipboard, Clipper, Cloak, Clock, Clock tower, Close-up, Closet, Clothes dryer, Clothes hanger, Clothing, Clotted cream, Cloud, Clown, Club, Clutch, Clyde steamer, Coach, Coal, Coast guard, Coasteering, Coat, Cob, Cobalt blue, Cobblestone, Coca-cola, Cocido madrile<f1>o, Cockpit, Cocktail, Cocktail dress, Cocktail garnish, Coffee, Coffee cup, Coffee cup sleeve, Coffee milk, Coffee substitute, Coffee table, Coffeehouse, Coffeemaker, Cog, Coin, Coin purse, Cola, Colcannon, Cold weapon, Collaboration, Collage, Collar, Collection, College, College baseball, College ice hockey, Collegiate wrestling, Collision, Colonel, Color guard (flag spinning), Colorfulness, Coloring book, Colt, Column, Comb, Combat, Combat medic, Combat sport, Combat vehicle, Combination machine, Comfort, Comfort food, Comic book, Comics, Comma, Commemorative plaque, Commercial building, Commercial vehicle, Common blue, Communication Device, Community, Community centre, Compact car, Compact cassette, Compact mpv, Compact sport utility vehicle, Compact van, Company, Compass, Competition, Competition event, Competitive trail riding, Composite material, Compound bow, Computer, Computer accessory, Computer case, Computer cluster, Computer cooling, Computer desk, Computer hardware, Computer keyboard, Computer monitor, Computer monitor accessory, Computer network, Computer program, Computer speaker, Computer terminal, Concept car, Concert, Concert hall, Concrete, Concrete bridge, Concrete mixer, Concrete ship, Condiment, Condominium, Cone, Confectionery, Conference hall, Confetti, Conformation show, Connecting rod, Constellation, Construction, Construction equipment, Construction paper, Construction worker, Contact sport, Control panel, Control tower, Convenience food, Convenience store, Convent, Convention, Convention center, Conversation, Convertible, Convoy rescue ship, Cook, Cookie, Cookies and crackers, Cooking, Cooking show, Cookware and bakeware, Cool, Cooling tower, Cope, Copper, Corn on the cob, Cornet, Cornhole, Corporate headquarters, Cortado, Corvette stingray, Cosmetic dentistry, Cosmetics, Cosplay, Cossacks, Costume, Costume accessory, Costume design, Costume hat, Cottage, Cottage pie, Cotton candy, Couch, Countertop, Country-western dance, Coup<e9>, Coup<e9>, Coupe utility, Course, Courthouse, Courtyard, Cousin, Cowboy action shooting, Cowboy boot, Cowboy hat, Cracker, Cradle, Craft, Crane, Crane vessel (floating), Crankset, Crash, Crayon, Cream, Creative arts, Creeking, Crescent, Crest, Crew, Crochet, Crocosmia <d7> crocosmiiflora, Croft, Crop, Crop top, Cross, Cross-country cycling, Cross-country skier, Cross-country skiing, Cross-stitch, Cross country running, Cross training shoe, Crosscut saw, Crossfit, Crossover suv, Crowd, Crown, Crucifix, Cruise ship, Cruiseferry, Cruiser, Crumble, Crutch, Crypt, Crystal, Cuba libre, Cuisine, Culinary art, Cumberland sausage, Cumulus, Cup, Cupboard, Cupcake, Curb, Currency, Curry, Curtain, Cushion, Custom car, Customer, Cutlery, Cutting tool, Cycle polo, Cycle sport, Cycling, Cycling shorts, Cyclo-cross, Cyclo-cross bicycle, Cyclocomputer, Cylinder, Cylindrical grinder, Cymbal, Dai pai dong, Dairy, Dam, Dame<U+0092>s rocket, Dame<U+0092>s rocket, Dampfnudel, Dance, Dancer, Dandelion coffee, Dane axe, Darkness, Darts, Davul, Dawn, Day dress, Daylighting, Daytime, Deacon, Deck, Decoration, Decorative nutcracker, Decorative rubber stamp, Deep frying, Deer hunting, Defenseman, Delicacy, Delicatessen, Demolition, Demon, Demonstration, Denim, Depot ship, Derailleur gears, Design, Desk, Desktop computer, Dessert, Destroyer, Devil's bridge, Dew, Dhak, Dhol, Dhow, Diagram, Diamond, Diaper bag, Diary, Diatonic button accordion, Digital camera, Digital clock, Digital compositing, Digital photo frame, Digital piano, Diner, Dinghy, Dinghy sailing, Dining room, Dinner, Dinnerware set, Dinosaur, Diot, Dip, Diploma, Dirt jumping, Dirt road, Dirt track racing, Disabled sports, Disc brake, Disc dog, Disco, Dish, Dishware, Dispatcher, Display advertising, Display board, Display case, Display device, Display window, Distilled beverage, Divemaster, Diving, Diving equipment, Diving mask, Diwali, Dobos torte, Dock, Dock jumping, Dock landing ship, Document, Dodge, Dodge caravan, Dodge journey, Dodge magnum, Dodge neon srt-4, Dodge ram srt-10, Dodge ram van, Dodgeball, Dog agility, Dog bed, Dog clothes, Dog collar, Dog crate, Dog hiking, Dog sled, Dog sports, Dog supply, Dog toy, Dog walking, Dogo argentino, Doily, Doll, Dollar, Dollhouse, Dollhouse accessory, Dome, Dondurma, Doodle, Door, Door handle, Doorbell, Dormitory, Double bass, Double reed, Doughnut, Douglas ac-47 spooky, Douglas dc-2, Douglas dc-3, Douglas sbd dauntless, Downhill mountain biking, Downtown, Drag boat racing, Drag racing, Dragon, Drain, Drainage, Drama, Drawbridge, Drawer, Drawing, Dreadlocks, Dress, Dress shirt, Dress shoe, Drifting, Driftwood, Drilling rig, Drink, Drinking, Drinking establishment, Drinking water, Drinkware, Driveway, Driving, Drizzle, Drop, Drought, Drum, Drumhead, Drummer, Drums, Dry cleaning, Dry suit, Duathlon, Duckpin bowling, Duel, Duet, Dungeon, Durango boot, Dusk, Dust, Dutch oven, Duvet, Duvet cover, Dvd, Dye, Ear, Earthquake, Easel, East indiaman, Easter, Easter bunny, Easter egg, Eastern Cottontail, Eating, Eclipse, Eco hotel, Edsel pacer, Edsel ranger, Education, Educational toy, Egg, Egg salad, Egg sandwich, Eight-man football, Elbow, Elderflower cordial, Electric blue, Electric car, Electric locomotive, Electric vehicle, Electrical network, Electrical supply, Electrical wiring, Electricity, Electronic component, Electronic device, Electronic engineering, Electronic instrument, Electronic musical instrument, Electronic signage, Electronics, Elektroboot, Embellishment, Emblem, Embroidery, Emergency, Emergency service, Emergency vehicle, Emoticon, Employment, Enclosure, Endurance racing (motorsport), Endurance riding, Endurance sports, Enduro, Endurocross, Energy bar, Energy drink, Engagement ring, Engine, Engineer, Engineering, English draughts, English riding, Entertainment, Environmental art, Equestrian helmet, Equestrian sport, Equestrianism, Eraser, Erg, Escalator, Escargot, Espresso, Essex skipper, Estate, Euphonium, European Starling, Eurovans, Evening, Event, Eventing, Everyday carry, Executive car, Exercise, Exercise equipment, Exercise machine, Exhaust manifold, Exhaust system, Exhibition, Experimental aircraft, Explosion, Extinction, Extradosed bridge, Extreme sport, Eye, Eye glass accessory, Eye liner, Eyebrow, Eyelash, Eyewear, F<ea>te, F<ea>te, Facade, Face, Face mask, Facial expression, Facial hair, Factory, Factory ship, Fair, Fairchild republic a-10 thunderbolt ii, False morel, Family, Family car, Family pictures, Family reunion, Family taking photos together, Fan, Fan convention, Farm, Farmer, Farmhouse, Farmworker, Farrier, Fashion, Fashion accessory, Fashion design, Fashion illustration, Fashion model, Fashion show, Fast food, Fast food restaurant, Fastener, Feathered hair, Feature phone, Fedora, Feist, Fence, Fender, Ferrari 328, Ferrari california, Ferris wheel, Ferry, Festival, Fetish model, Fettuccine, Fiat 500, Fiction, Fictional character, Fiddle, Field archery, Field house, Field trial, Fife, Fighter aircraft, Figure drawing, Figurine, Filling station, Film camera, Film crew, Film noir, Film producer, Film studio, Filmmaking, Finger, Finger food, Finial, Fire, Fire apparatus, Fire department, Fire extinguisher, Fire hydrant, Fire marshal, Fire screen, Fire station, Firearm, Fireboat, Firefighter, Fireplace, Fireworks, First-class cricket, First aid, First generation chevrolet aveo, First generation ford mustang, Fish hook, Fish pond, Fish products, Fisherman, Fisheye lens, Fishing, Fishing bait, Fishing lure, Fishing rod, Fishing trawler, Fishing vessel, Fitness and figure competition, Fitness professional, Fixed link, Fixture, Flag, Flag Day (USA), Flag football, Flag of the united states, Flagship, Flagstone, Flame, Flange, Flap, Flare, Flash photography, Flashlight, Flat iron steak, Flat panel display, Flat white, Flatbread, Flatland bmx, Flautist, Flea market, Flesh, Flight, Flight attendant, Flip-flops, Flip (acrobatic), Floats, Floodlight, Floor, Floor hockey, Floor plan, Floorball, Flooring, Floral design, florist gayfeather, Floristry, Flower Arranging, Flowerpot, Flugelhorn, Fluid, Fluorescent lamp, Flute, Fluyt, Flxible metro, Fly fishing, Flyer, Flying disc freestyle, Foam, Fodder, Fog, Foil, Fokker 50, Folding chair, Folding table, Folk dance, Folk instrument, Folkrace, Fondant, Font, Food, Food additive, Food coloring, Food court, Food craving, Food grain, Food processing, Food storage, Food storage containers, Food truck, Foot, Football, Football equipment, Football gear, Football helmet, Football player, Footprint, Footwear, Ford, Ford crown victoria, Ford crown victoria police interceptor, Ford escape, Ford excursion, Ford expedition, Ford explorer, Ford f-350, Ford f-series, Ford fairlane crown victoria skyliner, Ford gt, Ford gt40, Ford model a, Ford motor company, Ford super duty, Ford transit, Ford windstar, Forehead, Forge, Fork, Forklift truck, Formal wear, Formula libre, Formula one, Formula one car, Fortepiano, Fortification, Foundation, Fountain, Fountain pen, Fourball, Foursome (golf), Fourth generation lexus ls, Fractal art, Framing hammer, Frankfurter w<fc>rstchen, Frapp<e9> coffee, Free climbing, Free solo climbing, Free weight bar, Freebord, Freediving, Freeride, Freestyle bmx, Freestyle motocross, Freestyle skiing, Freestyle swimming, Freestyle walking, Freeway, Freezing, freight car, Freight transport, French fries, French toast, Freshwater aquarium, Frico, Fried chicken, Fried clams, Fried egg, Fried food, Friendship, Friendship sloop, Frigate, Friterie, Frock coat, Frost, Frozen dessert, Frozen food, Frozen yogurt, Fruit salad, Frula, Frutti di bosco, Frying, Fuel, Fuel line, Fuel pump, Fuel tank, Full-rigged ship, Full-size car, Full breakfast, Full moon, Fun, Function hall, Funicular, Fur clothing, Furniture, Futon, Futsal, Fuzzy navel, Gadget, Galaxy, Galeas, Galgo espa<f1>ol, Galgo espa<f1>ol, Galician gaita, Galiot, Galium, Galleon, Gallery rifle shooting, Galley, Gambling, Game boy accessories, Gamekeeper, Gamer, Games, Garage, Garbage truck, Garden buildings, Garden cosmos, Garden gnome, Garden salad, Gardener, Gardening, Garnish, Gas, Gas mask, Gas pump, Gas stove, Gasoline, Gate, Gauge, Gazebo, Gear, Gear shaper, Gear shift, Gelato, Gemstone, General aviation, Geocaching, Geologist, Gesture, Ghanta, Gift basket, Gin and tonic, Ginger family, Gingerbread house, Girder bridge, Giri choco, Girl scouts of the usa, Gladiator, Gladiolus, Glass, Glass bottle, Glasses, Glider, Glitter, Globe, Glove, Gluten, Gmc, Gmc envoy, Go-go dancing, Go-kart, Goal, Goalkeeper, Goaltender, Goaltender mask, Goat cheese, Goggles, Gold, Gold medal, Goldeneye, Golf, Golf bag, Golf ball, Golf cart, Golf club, Golf course, Golf equipment, Golfer, Gomashio, Gong, Gong bass drum, Gordita, Goth subculture, Gothic architecture, Gothic fashion, Goulash, Government, Gown, Gps navigation device, Graduation, Graffiti, Graham cracker, Grain, Grain milk, Gramophone record, Grand prix motorcycle racing, Grandparent, Granola, Grape leaves, Graphic design, Graphics, Grave, Gravure idol, Gravy, Grazing, Green, Green-veined white, Greenhouse, Grenadier, Grey, Gridiron football, Grillades, Grille, Grilling, Grind rail, Gristmill, Grocer, Grocery store, Grog, Ground attack aircraft, Groundhog day, Groupset, Grumman f8f bearcat, Guard rail, Gugelhupf, Guilinggao, Guitar, Guitarist, Gumbo, Gun, Gun accessory, Gun barrel, Gun turret, Gunfighter, Gungdo, Guru, Gusli, Gym, Gymnastics, Gyro, Hachis parmentier, Hacienda, Hail, Hair, Hair accessory, Hair care, Hair coloring, Hair iron, Hair tie, Hairstyle, Half marathon, half track, Hall, Hall of fame, Halter, Hamburger, Hammer drill, Hammered dulcimer, Hammock, Hamper, Hand, Hand drum, Hand fan, Hand luggage, Hand tool, Handbag, Handball, Handgun holster, Handheld device accessory, Handrail, Handwriting, Handymax, Hangar, Happy, Hard hat, Hardcourt bike polo, hardhack, Hardtack, Hardtop, Hardware accessory, Hardwood, Harley quinn, Harrier, Harvest, Harvester, Hat, Hatchback, Hatchet, Haute couture, Hawker, Hay, Haze, Hazmat suit, Head, Head restraint, Headband, Headgear, Headlamp, Headphones, Headpiece, Headquarters, Headset, Headstone, Health care, Health care provider, Heart, Hearth, Heat, heater, Helicon, Helicopter, Helicopter pilot, Helicopter rotor, Helmet, Hemp milk, Heptathlon, Herder, Herding, Hero, Hi-hat, Hickory golf, High-speed rail, High-visibility clothing, High heels, High jump, High school, Highball, Highball glass, Highland games, Highway, Hiking, Hiking boot, Hiking equipment, Hiking shoe, Hill station, Hindu temple, Hip, Hip-hop dance, Hipparchia, Hippie, Historic house, Historic site, History, Hockey, Hockey pants, Hockey protective equipment, Hoe, Holding hands, Holiday, Holiday ornament, Holy places, Home, Home accessories, Home appliance, Home door, Home fencing, Homemaker, Homework, Hominy, Honda, Honda accord, Honda city, Honda civic, Honda cr-v, Honda element, Honda fcx clarity, Honda fit, Honda odyssey, Honda pilot, Honda prelude, Honda ridgeline, Honeycomb, Honeymoon, Hong kong-style milk tea, Honmei choco, Hood, Hoodie, Hoop (rhythmic gymnastics), hoopskirt, Hops, Horchata, Horizon, Horse and buggy, Horse harness, Horse racing, Horse supplies, Horse tack, Horse trailer, Horse trainer, Horseshoe, Horseshoes, Hospital, Hostel, Hot air balloon, Hot air ballooning, Hot chocolate, Hot dog, Hot dog stand, Hot hatch, Hot pot, Hot rod, Hot sauce, Hot toddy, Hotel, Hotteok, Hound trailing, House, House numbering, Household appliance accessory, Housewarming party, Hub gear, Hubcap, Hudson commodore, Hudson hornet, Hug, Huggies pull-ups, Hula, Hula hoop, Hulk, Human, Human anatomy, Human body, Human leg, Human settlement, Humpback bridge, Hunting, Hunting decoy, Hunting knife, Hurdle, Hurdling, Hurricane, Hut, Hutch, Hybrid bicycle, Hybrid vehicle, Hypericaceae, Hyundai, Hyundai sonata, Ice climbing, Ice cream, Ice cream cone, Ice cream maker, Ice Cream Sodas, Ice hockey, Ice hockey equipment, Ice hockey position, Ice hotel, Ice pop, Ice rink, Ice skate, Ice skating, Iceberg, Iced coffee, Icing, Icon, Idiophone, Igloo, Illustration, Impatiens, Inca rope bridge, Incandescent light bulb, Independence day, indian blanket, Indian cuisine, Individual sports, Indoor cycling, Indoor games and sports, Industry, Infant bed, Infantry, Infiniti g, Inflatable, Inflatable boat, Infrastructure, Ingredient, Ink, Inline skates, Inline skating, Inline speed skating, Inn, Interaction, Interactive kiosk, Interior design, International xt, Intersection, Intramural softball, Inventory, Invitation, Ipad, Iphone, Irish cream, Irish setter, Iron, Ironclad warship, Ironworker, Island, Isle of man tt, Issoria, Italian food, Italian sausage, Ivory, J<e4>mthund, J<e4>mthund, jack-in-the-pulpit, Jack-o'-lantern, Jacket, Jackhammer, Jacuzzi, Jamun, Japanese architecture, Jarana jarocha, Javelin, Jaw, Jazz, Jazz club, Jeans, Jeep, Jeep cj, Jeep dj, Jeep grand cherokee, Jeep patriot, Jeep wrangler, Jersey, Jester, Jet aircraft, Jet engine, Jet ski, Jewellery, Jheri curl, Jigging, Jin deui, Job, Jockey, Jogging, Joint, Journalist, Jug, Juice, Jukebox, Jumping, Jumping into the pool, Junction, Junk food, Junkers ju 52, Justice league, Karaage, Karahi, Karedok, Kart racing, Kayak, Kayaking, Keirin, Kendang, Kennel, Kennel club, Kettle, Kettlebell, Key, Keyboard, Keyboard player, Keychain, Khaki, Kia motors, Kia sorento, Kia soul, Kia sportage, Kick scooter, Kickflip, Kids' meal, Kielbasa, Kilt, Kimono, Kindergarten, Kiosk, Kir royale, Kiss, Kitchen, Kitchen & dining room table, Kitchen appliance, Kitchen stove, Kite, Kite sports, Kitesurfing, Knackwurst, Knee, Knife, Knight, Knit cap, Knitting, Knot, Kobe beef, Kofta, Kombucha, Korean taco, Kuchen, Label, Labyrinth, Lace, Lacrosse, Ladder, Ladder golf, Lady, Lager, Laksa, Lama, Lamb's quarters, Lamb and mutton, Lamborghini, Lamborghini hurac<e1>n, Lamborghini murci<e9>lago, Laminate flooring, Lamp, Lampshade, Lance, Lancia beta, Land lot, Land rover defender, Land rover discovery, Land rover series, Land vehicle, Landing, Landmark, Landscape lighting, Landscaping, Lane, Lantern, Lap, Laptop, Large m<fc>nsterl<e4>nder, Laser, Laser tag, Latex, Latex clothing, Lathe, Latte, Laugh, Launch, Laundry, Laundry room, Lavochkin la-5, Law enforcement, Lawn, Lawn game, Lawn mower, Lawn ornament, Layered hair, Lcd tv, Learning, Leash, Leather, Leather jacket, Lectern, Lecture, Led-backlit lcd display, Led display, Leg, Leggings, Lego, Lei, Leisure, Leisure centre, Lemonade, Lens, Lens flare, Leotard, Letter, Lexus, Lexus is, Library, Lid, Lifebuoy, Lifeguard, Lifejacket, Liftback, Light, Light-emitting diode, Light aircraft, Light commercial vehicle, Light fixture, Lighthouse, Lighting, Lighting accessory, Lightning, Lilac, Limousine, Line, Line art, Line dance, Linens, Lingerie, Lip, Lip gloss, Liqueur, Liquid, Liquid bubble, Liquor store, Listed building, Litter, Little black dress, Liver, Living room, Lobby, Local food, Lock, Locket, Lockheed c-130 hercules, Lockheed martin f-22 raptor, Lockheed xh-51, Locomotive, Loft, Log bridge, Log cabin, Logging, Logo, Long-distance running, Long-sleeved t-shirt, Long hair, Longboard, Longboarding, Longbow, Longcase clock, Loudspeaker, Loukaniko, Love, Loveseat, Luau, Lucha libre, Luge, Luggage and bags, Lugger, Lumber, Lumberjack, Lunch, Lupin, Lure coursing, Luxury vehicle, Luxury yacht, Lye roll, M113 armored personnel carrier, M35 2<U+00BD>-ton cargo truck, Macaroni and cheese, Macedonia, Machaca, Machine, Machine gun, Machine tool, Macro photography, Magazine, Magenta, Magic, Magyar ag<e1>r, Mai tai, Mail, Mailbox, Maillot, Major appliance, Majorelle blue, Majorette (dancer), Malt, Mammoth, Management, Mane, Manhattan, Manhole, Manhole cover, Manicure, Mannequin, Manor house, Mansion, Map, Maple, Maple syrup, Marathon, Marble, Marching, Marching band, Mardi Gras, Margarita, Marina, Marine protector-class coastal patrol boat, Marines, Maritime museum, Market, Marketplace, Maroon, Marriage, Marshmallow creme, Masala chai, Mascara, Mascot, Maserati 26m, Mashed potato, Mask, Mason jar, Masonry oven, Masque, Massage table, Massively multiplayer online role-playing game, Mast, Mat, Match play, Match rifle shooting, Material property, Mattress, Mattress pad, Mazda, Mazda cx-7, Mazda cx-9, Mcdonnell douglas f-15e strike eagle, Mcdonnell douglas f-4 phantom ii, Mcintosh, Mcnab, Meal, Measuring instrument, Meat, Meatball, Mechanic, Mechanical fan, Medal, Media, Medical, Medical assistant, Medical equipment, Medical glove, Medieval architecture, Mediterranean food, Medium tactical vehicle replacement, Medley swimming, Mee siam, Meeting, Megaphone, Melee weapon, Mellophone, Melting, Membranophone, Memorial, Memorial day, Menu, Mercedes-benz, Mercedes-benz 500e, Mercedes-benz c-class, Mercedes-benz cls-class, Mercedes-benz glk-class, Mercedes-benz m-class, Mercedes-benz r-class, Mercedes-benz sls amg, Mercedes-benz sprinter, Mercedes-benz viano, Mercedes-benz vito, Mercedes-benz w123, Mercedes-benz w124, Mercedes-benz w126, Mercedes-benz w221, Mercedes-benz w31, Mercury eight, Mercury mountaineer, Mesh, Messenger bag, Metal, Metal lathe, Metalsmith, Metalworking, Metalworking hand tool, Meteorological phenomenon, Meter, Metro, Metro station, Metropolis, Metropolitan area, Metropolitan bishop, Mexican food, Microcassette, Microphone, Microphone stand, Microvan, Microwave oven, Mid-autumn festival, Mid-size car, Middle-distance running, Middle ages, Midnight, Migas, Mil mi-24, Military, Military aircraft, Military camouflage, Military helicopter, Military officer, Military organization, Military person, Military rank, Military transport aircraft, Military uniform, Military vehicle, Milk, Milkshake, Milky way, Mill, Milling, Mime artist, Mimosa, Mincemeat, Miner, Mineral water, Mini SUV, Miniature, Miniature golf, Minibus, Minidisc, Mining, Minivan, Mirror, Missile, Mission, Mist, Mitsubishi, Mitsubishi pajero, Mixed-use, Mixture, Moat, Mobile device, Mobile home, Mobile phone, Mobile phone accessories, Mobile phone case, Mode of transport, Model, Model aircraft, Model car, Modern art, Modern pentathlon, Moisture, Mojito, Molding, Mole sauce, Mollete, Monarch, Monastery, Money, Mongolian food, Monk, Monochrome, Monochrome photography, Monocular, Monokini, Monolith, Monoplane, Monorail, Monoski, Monster truck, Monument, Moon, Moonlight, Moped, Morin khuur, Morning, morning glory, Morning glory, Mortar, Mortuary temple, Mosaic, Moscow mule, Mosque, Motel, Motif, Motocross, Motor oil, Motor ship, Motor vehicle, Motorcycle, Motorcycle accessories, Motorcycle boot, Motorcycle drag racing, Motorcycle fairing, Motorcycle helmet, Motorcycle racer, Motorcycle racing, Motorcycle speedway, Motorcycling, Motorized scooter, Motorized wheelchair, Motorsport, Moulder, Mountain bike, Mountain bike racing, Mountain biking, Mountain guide, Mountain village, Mountaineer, Mountaineering, Mousepad, Moustache, Mouth, Moveable bridge, Movie, Moving, Mozzarella, Muay thai, Muffin, Muffler, Mug, Mulberry, Multi-sport event, Multi-tool, Multimedia, Mural, Muscle, Muscle car, Muscovado, Museum, Mushing, Music, Music artist, Music venue, Music workstation, Musical, Musical box, Musical ensemble, Musical instrument, Musical instrument accessory, Musical keyboard, Musical theatre, Musician, Mustard, Mustard greens, Mythical creature, Mythology, Nail, Nail care, Nail polish, Nameplate, Nap, Nap mat, Napoleon iii style, Narrow-body aircraft, Nasi kandar, National historic landmark, National monument, Nativity scene, Natural foods, Natural rubber, Naval architecture, Naval officer, Naval ship, Naval trawler, Navel, Navy, Nebula, Neck, Necklace, Needlework, Negroni, Neighbourhood, Neo-burlesque, Neon, Neon sign, Net, Net sports, Netball, Netbook, New year, New year's eve, New Years Day, News, News conference, Newspaper, Newsprint, Night, Nightclub, Nightlight, Nike free, Nissan, Nissan 370z, Nissan cube, Nissan navara, Nissan patrol, Nissan rogue, Nissan skyline gt-r, Nissan x-trail, Non-alcoholic beverage, Non-commissioned officer, Non-Sporting Group, Nonbuilding structure, Nopal, Nordic combined, Nordic skiing, Nordic walking, North american a-36 apache, North american b-25 mitchell, North american fraternity and sorority housing, North american p-51 mustang, North american t-6 texan, Nose, Notebook, Novel, Number, Nuncio, Nurse, Nutcracker, Oar, Oasis, Oat bran, Obedience training, Obelisk, Oboe, Observation tower, Observatory, Obstacle race, Ocean liner, Odometer, Off-road racing, Off-road vehicle, Off-roading, Office, Office equipment, Office supplies, Official, Official residence, Oil lamp, Oil rig, Oil tanker, Oktoberfest celebrations, Old fashioned glass, Oldsmobile 442, Oliang, One-piece swimsuit, Onion ring, Open-wheel car, Open water swimming, Opening presents, Opera house, Operating theater, Optical instrument, Optoelectronics, Orange, Orange drink, Orange juice, Orange soft drink, Orator, Orchestra, Orchestra pit, Organ, Organization, Orienteering, Origami, Origami paper, Ornament, Outdoor bench, Outdoor furniture, Outdoor grill, Outdoor grill rack & topper, Outdoor play equipment, Outdoor power equipment, Outdoor recreation, Outdoor shoe, Outdoor structure, Outdoor table, Outer space, Outerwear, Outhouse, Outlet store, Oven, Overcoat, Overhead power line, Overpass, Oxygen mask, Oyaki, P<e2>tisserie, P<e9>tanque, Pachycephalosaurus, Packaging and labeling, Paddle, Padlock, Pagoda, Pain, Paint, Paintball, Paintball equipment, Paintball marker, Painter, Painting, Pajamas, Pakora, Palace, Pan frying, Pancetta, Pancit, Panel saw, Panorama, Pantry, Pantsuit, Pantyhose, Paper, Paper bag, Paper product, Parachute, Parachuting, Parade, Paragliding, Parallel, Paramedic, Parasailing, Parfait, Parish, Park, Park ranger, Parka, Parking, Parking lot, Parsley, Party, Party bike, Party favor, Party hat, Party supply, Passenger, Passenger car, Passenger ship, Pasteles, Pastrami, Pastry, Patchwork, Path, Patient, Patio, Patrol boat, Patrol boat, Pattern, Patty melt, Pavilion, Payphone, Pc game, Peaked cap, Pearl, Peddler, Pedestrian, Pedestrian crossing, Pedometer, Peeps, Pen, Pencil, Pencil case, Pendant, Pendulum, People, People in nature, People on beach, Perch, Percussion, Percussionist, Performance, Performance art, Performance car, Performing arts, Performing arts center, Perfume, Pergola, Persian, Persimmon, Personal care, Personal computer, Personal luxury car, Personal protective equipment, Personal water craft, Pes<e4>pallo, Pet supply, Petit basset griffon vend<e9>en, Petit basset griffon vend<e9>en, Petroleum, Pez, Phd, Photo caption, Photo shoot, Photobombing, Photograph, Photographer, Photographic film, Photographic paper, Photography, Photomontage, Physical fitness, Physical therapy, Physician, Physicist, Pianist, Piano, Picadillo, Picket fence, Pickling, Pickup truck, Picnic, Picnic boat, Picnic table, Picture frame, Pier, Piggy bank, Pilates, Pilgrimage, Pillow, Pilot boat, Pin-back button, Pinball, Ping pong, Pink, Pink family, Pink peppercorn, Pint, Pint glass, Pipe, Pipe insulation, Pipeline transport, Pitch and putt, Pizza, Pizza cheese, Pizza cutter, Place card, Place of worship, Placemat, Plaid, Plan, Plane, Planet, Plank, Plaster, Plastic, Plastic bag, Plastic bottle, Plastic wrap, Plate, Plate girder bridge, Platform supply vessel, Platter, Play, Play-doh, Player, Player piano, Playground, Playground slide, Playing in the snow, Playset, Plaza, Pliers, Plimsoll shoe, Plot, Plucked string instruments, Plumbing, Plumbing fixture, Pluot, Plush, Plymouth duster, Plymouth road runner, Plywood, Poached egg, Pocket, Podium, Poi, Point-and-shoot camera, Pointer, Pok<e9>mon, Poker, Poker table, Pole, Pole climbing (gymnastic), Pole vault, Police, Police car, Police officer, Polka dot, Pollution, Pom-pom, Pond hockey, Pontederia, Pontiac gto, Pony car, Ponytail, Pop music, Popcorn, Porcelain, Porch, Porchetta, Pork, Porsche, Porsche 911 classic, Porsche 912, Port, Portable communications device, Portable electronic game, Portable stove, Portable toilet, Portrait, Portrait photography, Post-it note, Post box, Poster, Poster session, Pot rack, Pot roast, Potato, Potato chip, Potato salad, Pottery, Poutine, Power inverter, Power station, Power supply, Powerboating, Powered paragliding, Powerlifting, Pra<U+009E>sk<fd> krysar<ed>k, Practical shooting, Praline, Pray, Pre-dreadnought battleship, Precipitation, Precision sports, Premiere, Prepackaged meal, Presbyter, Present, Presentation, Preserved food, Presidential palace, Pretzel, Priesthood, Princess Leia, Printing, Printmaking, Private school, Procyon, Produce, Product, Professional boxer, Professional boxing, Professional golfer, Professional wrestling, Professor, Project, Projection screen, Projector accessory, Prom, Promontory, Propeller, Propeller-driven aircraft, Property, Prophet, Proteales, Protest, Psychedelic art, Pub, Public event, Public library, Public space, Public speaking, Public transport, Public utility, Publication, Puddle, Pueblo, Pulled pork, Pump, Pumping station, Punch, Puppet, Puppy love, Pur<e9>e, Purple, purplehead, Putter, Pyramid, Quadrathlon, Quartz clock, Queen, Quill, Quilt, Quilting, Quincea<f1>era, Rabbi, Race car, Race track, Racer, Racing, Racing bicycle, Racing video game, Racket, Racquet sport, Radar, Radiator, Radio-controlled aircraft, Radio-controlled car, Radio-controlled helicopter, Radio-controlled toy, Radio telescope, Raft, Rafting, Railroad car, Railway, Rain, Rain and snow mixed, Rain boot, Rain gauge, Rainbow, Raincoat, Rally obedience, Rallycross, Rallying, Ramen, Ramsons, Ranch, Range rover, Ranged weapon, Rapper, Rapping, Raw milk, Reading, Real estate, Rear-view mirror, Rebellion, Receipt, Receptionist, Recipe, Recital, Recliner, Recording, Recording studio, Recreation, Recreation room, Recreational fishing, Rectangle, Recumbent bicycle, Recycling, Recycling bin, Red, Red carpet, Red flag, Red hair, Red meat, Red sky at morning, Red wine, redbud, Redhead, redshank, Reed instrument, Reflecting pool, Reflection, Reflex camera, Refried beans, Refrigerator, Regularity rally, Rehearsal dinner, Rein, Reinforced concrete, Religious institute, Religious item, Repinique, Republic p-47 thunderbolt, Rescue, Rescuer, Research vessel, Residential area, Resort, Resort town, Rest area, Restaurant, Restroom, Retail, Retirement home, Retro style, Revolver, Rhythmic gymnastics, Ribbon, Ribbon (rhythmic gymnastics), Rickshaw, Riding boot, Riding instructor, Riding mower, Riding toy, Rifle, Rigid-hulled inflatable boat, Rim, Ring, Ringlet, Rink bandy, Ristretto, Rite, Ritual, river, River juniper, Road, Road bicycle, Road bicycle racing, Road cycling, Road racing, Road surface, Road trip, Roadster, Roar, Roasted barley tea, Roasting, Robe, Robot, Rock-climbing equipment, Rock climbing, Rock concert, Rock dove, Rock fishing, Rocker, Rocker cover, Rocket, Rocket-powered aircraft, Rocking chair, Rodeo, Rogaining, Roller, Roller coaster, Roller derby, Roller skates, Roller skating, Roller sport, Rolling, Rolling stock, Rolls-royce, Roman temple, Romance, Roof, Roof lantern, Roof rack, Roofer, Room, Root beer, Rope, Rope bridge, Rotor, Rotorcraft, Rou jia mo, Rowing, Royal icing, rubber ducky, Ruddy turnstone, Ruffle, Rugby, Rugby player, Ruins, Ruler, Rum ball, Rum swizzle, Running, Running shoe, Runway, Rural area, Rust, RV, Ryuteki, Sachertorte, Sackbut, Saddle, Safari, Safety pin, Saguaro, Sail, Sailboat, Sailing, Sailing ship, Sailor, Saimin food, Saint patrick's day, Salad, Salsa, Salt-cured meat, Salt and pepper shakers, Samba, Samgyeopsal, Samurai, Sandal, Sandboarding, Sandpit, Sandwich, Sanitary sewer, Santa claus, Santoor, Sash window, Satay, Satchel, Satellite phone, Satin, Saucer, Sauces, Sauerkraut, Sausage, Saw, Saw chain, Sawhorse, Saxhorn, Saxophone, Saxophonist, Scabbard, Scaffolding, Scale model, Scar, Scarecrow, Scarf, Scene, Schematic, Schnecken, School, School bus, School uniform, Schooner, Schwenker, Science, Scissors, Sconce, Scone, Scooter, Scoreboard, Scotch egg, Scotch whisky, Scottish smallpipes, Scout, Scrap, Scrapbooking, Screen, Screen-printing, Screenshot, Screw, Scrubs, Scuba diving, Sculptor, Sculpture, Sea breeze, Sea kayak, Seafood, Seaplane, Seaside resort, Seasoning, Seat belt, Secondary school, Security, Security guard, Security lighting, Sedan, Segmental bridge, Segway, Sel roti, Self-help book, Self-portrait, Self-propelled artillery, Selfie, Selling, Seminar, Senna, Serpent, Serveware, Service, Sewing, Sewing machine, Sewing machine needle, Shack, Shade, Shadow, Shank, Sharing, Shashlik, Shawl, Shed, Sheet pan, Sheftalia, Shelby mustang, Shelf, Shell, Shelving, Shikoku, Ship, Ship of the line, Ship replica, Shipping container, Shipwreck, Shirt, Shoe, Shoe store, Shooting, Shooting range, Shooting sport, Shopkeeper, Shopping, Shopping bag, Shopping cart, Shopping mall, Short track motor racing, Shorts, Shot tower, Shotgun, Shoulder, Shovel, Shower, Shrine, Sibling, Side cap, Side dish, Sidecar, Sidewalk, Siding, Sign, Sign language, Signage, signaling device, Signature, Sikorsky hh-52 seaguard, Sikorsky s-61, Sikorsky s-61r, Sikorsky s-64 skycrane, Sikorsky s-70, Sikorsky s-92, Sikorsky sh-3 sea king, Silhouette, Silk, Sill, Silo, Silver, Silver medal, Simit, Singer, Singing, Single-lens reflex camera, Single-origin coffee, Sink, Siren, Sitting, Six-wheel drive, Skate, Skateboard, Skateboard deck, Skateboarder, Skateboarding, Skateboarding Equipment, Skatepark, Skater hockey, Skating, Skeet shooting, Skeleton, Sketch, Skewer, Ski, Ski binding, Ski boot, Ski cross, Ski Equipment, Ski helmet, Ski jumping, Ski mountaineering, Ski pole, Ski resort, Ski touring, Skier, Skiff, Skiing, Skijoring, Skimboarding, Skin, Skin care, Skirmish, Skordalia, Skull, Skullcap, Sky, Skyline, Skyscraper, Skyway, Slalom skiing, Slam dunk, Slate, Sled, Sled dog racing, Sledding, Sleep, Sleeping bag, Sleeping pad, Sleeve, Sleeveless shirt, Slider, Slipcover, Slipper, Sloop-of-war, Slopestyle, Slot machine, Slush, Small appliance, Smartphone, Smile, Smiley, Smoke, Smoking, Smoothie, Snack, Snapshot, Snare drum, Sneakers, Snifter, Snipe, Sniper rifle, Snips, Snorkeling, Snow angel, Snow blower, Snow boot, Snow bridge, Snow removal, Snowboard, Snowboarding, Snowkiting, Snowman, Snowmobile, Snowplow, Snowshoe, Soccer, Soccer-specific stadium, Soccer ball, Social group, Social work, Sock, Soda straw, Sofa bed, Sofa tables, Soft drink, Soft flag, Soft Serve Ice Creams, Soft tennis, Softball, Software, Software engineering, Solar energy, Solar panel, Soldier, Sole, Solution, Solvent, Sombrero, Song, Sorbet, Sorbetes, Sorghum, Sound, Sound stage, Soup, Sour cream, Sousaphone, Souvenir, Souvlaki, Sowing, Spa town, Space, Space bar, space shuttle, Spacecraft, Spaceplane, Spaghetti, Spandex, Spanish missions in california, Spark plug, Sparkler, Speaker, Spear, Special agent, Spectacle, Speech, Speedboat, Speedometer, Sphere, Spice, spiderwort, Spinach salad, Spinning, Spiral, Spire, Split-rail fence, Spoil tip, Spoiler, Spoke, Spokesperson, Sponge, Spoon, Spoon lure, Sport aerobics, Sport climbing, Sport utility vehicle, Sport venue, Sporting clays, Sporting Group, Sports, Sports bra, Sports car, Sports car racing, Sports drink, Sports equipment, Sports fan accessory, Sports gear, Sports sedan, Sports training, Sports uniform, Sportswear, Spotting scope, Spray, Spreader, Spring, Spring break, Spring greens, Springboard, Springerle, Sprint, Sprint car racing, Sprint football, Spritzer, Spur, Square, Squeezebox, Stable, Stadium, Stage, Stage equipment, Stain, Stained glass, Stairs, Stall, Stampe sv.4, Stamppot, Stand up paddle surfing, Standing, Staple food, Star, State school, Stately home, Stationary bicycle, Stationery, Statue, Steak, Steam, Steam engine, Steamboat, Steamed rice, Steel, Steel-toe boot, Steel casing pipe, Steeple, Steering part, Steering wheel, Stele, Stemware, Stencil, Stereophonic sound, Stew, Stick and Ball Games, Stick and Ball Sports, Sticker, Still life, Still life photography, Stinson reliant, Stock car racing, Stock photography, Stock trader, Stocking, Stomach, Stone carving, Stone wall, Stool, Stop sign, Stopwatch, Storage basket, Storage tank, Storm, Stout, Stove, Stovetop kettle, Strap, Strategy video game, Straw, Street, Street art, Street artist, Street dance, Street fashion, Street food, Street football, Street light, Street performance, Street sign, Street sports, Street stunts, Streetball, Strength athletics, Strength training, Stretching, Striking combat sports, String instrument, String instrument accessory, Strings, Studebaker gran turismo hawk, Student, Studio, studio couch, Stuffed toy, Stuffing, Stunt, Stunt performer, Style, Subaru, Subaru impreza wrx sti, Subcompact car, Suburb, Subway, Sugar-apple, Sugar cake, Sugar house, Sugar paste, Suit, Suit actor, Suitcase, Suite, Summer, Sun, Sun hat, Sun tanning, Sunbeam alpine, Sunbeam tiger, Sundae, Sunglasses, Sunlight, Sunlounger, Sunrise, Sunset, Super bowl, Superbike racing, Supercar, Superfood, Superfruit, Superhero, Superman, Supermarket, Supermini, Supermoto, Supervillain, Supper, Surf fishing, Surface lure, Surface water sports, Surfboard, Surfer hair, Surfing, Surfing Equipment, Surgeon, Suspension, Suspension bridge, Suspension part, Suzuki, Swan boat, Swat, sweatpant, Sweet cicely, Sweetness, Swim brief, Swim cap, Swimfin, Swimmer, Swimming, Swimming machine, Swimming pool, Swimsuit bottom, Swimsuit top, Swimwear, Swing, Swing bridge, Switch, Switchel, Sword, Symbol, Symmetry, Synagogue, Synchronized swimming, Synthetic rubber, T'ai chi ch'uan, T-shirt, T<e1>rogat<f3>, Tabla, Table, Table tennis racket, Tablecloth, Tablet computer, Tabletop game, Tableware, tacamahac, Tachometer, Tackle, Taco, Tag rugby, Tagliatelle, Tail, Take-out food, Takeoff, Talent show, Tall ship, Tan, Tanacetum balsamita, Tandem bicycle, Tandem skydiving, Tango, Tank, Tank ship, Tap, Tapestry, Taquito, Tar, Taramasalata, Target archery, Tarpaulin, Tartan, Tarte flamb<e9>e, Tarte flamb<e9>e, Tartiflette, Taste, Tattoo, Tattoo artist, Tavern, tawhana, Taxi, Tea, Teacher, Teacup, Teal, Team, Team sport, Technical drawing, Technology, Teddy bear, Teenage mutant ninja turtles, Teh tarik, Telecommunications engineering, Telemark skiing, Telephone, Telephone booth, Telephony, Telescope, Television, Television crew, Television presenter, Television program, Television set, Television studio, Television transmitter, Temperature, Temple, Temporary tattoo, Ten-pin bowling, Tennis, Tennis ball, Tennis court, Tennis Equipment, Tennis player, Tennis racket, Tennis racket accessory, Tennis shoe, Tent, Terrace, Text, Textile, Tgv, Thai food, Thatching, Theater curtain, Theatre, Theatrical property, Theatrical scenery, Theodolite, Therapy, Thermae, Thermal bath, Thermostat, Thigh, Thimbleberry, Thinking, Thoroughfare, Thread, Throat, Throw pillow, Throwing a ball, Throwing axe, Throwing knife, Thrush, Thumb, Thunder, Thunderstorm, Thuringian sausage, Ti plant, Tick, Ticket, Tie, Tied-arch bridge, Tieguanyin, Tights, Tiki, Tile, Tiltrotor, Timbales, Timer, Tin, Tin can, Tinto de verano, Tints and shades, Tiple, Tire, Tire care, Titanium, Toast, Tobacco products, Toddler, Toe, Toilet, Toilet seat, Tololoche, Tom-tom drum, Tomahawk, Tongue, Tool, Tool accessory, Tool and cutter grinder, Toolroom, Tooth, Top, Torch, Torii, Torte, Totem, Totem pole, Touch football (American), Touch rugby, Tour bus service, Touring car racing, Tourism, Tourist attraction, Tournament, Tow truck, Towed water sport, Tower, Tower block, Town, Town crier, Town square, Toy, Toy airplane, Toy block, Toy dog, Toy manchester terrier, Toy vehicle, Toyota, Toyota 4runner, Toyota camry, Toyota corolla, Toyota corolla e80, Toyota fj cruiser, Toyota prius, Toyota sienna, Toyota tacoma, Toyota venza, Toyota yaris, Track, Track and field athletics, Track cycling, Track racing, Track spikes, Tractor, Trade, Trademark, Trader, Tradesman, Tradition, Traditional sport, Traffic, Traffic congestion, Traffic light, Traffic sign, Trail, Trail mix, Trail riding, Trailer, trailer truck, Train, Train station, Trainer, Training, Training ship, Training wheels, Tram, Trampoline, Trampolining--Equipment and supplies, Transformers, Transmission tower, Transmitter station, Transparency, Transparent material, Transport, Transporter bridge, Trap shooting, Travel, Travel trailer, Tread, Treadmill, Tree house, Tree stand, Trekking pole, Tres leches cake, Trestle, Triangle, Triathlon, Tribe, trick-or-treat, Tricking, Tricycle, Trigger, Trikiti, Trilobite, Trip computer, Tripod, Triumphal arch, Trolleybus, Trombone, Troop, Trophy, Tropics, Trousers, Truck, Truck bed part, Truck driver, Trucker hat, Truggy, Trumpet, Trumpeter, Trunks, Truss bridge, Tteokguk, Tuba, Tubing, Tug of war, Tugboat, Tumbler, Tumbling (gymnastics), Tunnel, Turban, Turbine, Turkish angora, Turkish coffee, Turkish van, Turquoise, Turret, Tursu, Tusk, Tuxedo, Twine, Two-man saw, Two-way radio, Types of trombone, Typewriter, Tyrannosaurus, Uilleann pipes, Ukulele, Ultramarathon, Umbrella, Unadon, Undergarment, Underpants, Underwater diving, Underwater sports, Unesco world heritage site, Unicorn, Unicycle, Unidentified flying object, Uniform, Universe, University, Urban area, Urban design, Urinal, Urn, Utility knife, Vacation, Vacuum flask, Valentine's day, Van, Vanilla ice cream, Vanillekipferl, Varnish, Vase, Vault, Veal, Vegan nutrition, Vegetable juice, Vegetarian food, Veggie burger, Vehicle, Vehicle audio, Vehicle brake, Vehicle cover, Vehicle door, Vehicle registration plate, Veil, Velociraptor, Vending machine, Venison, Vernissage, Vespa, Vest, Vestment, Veterans day, Viaduct, Victorian fashion, Video camera, Video game arcade cabinet, Video game software, Videographer, Vienna sausage, Vietnamese iced coffee, Vigil, Viking, Villa, Village, Vintage advertisement, Vintage car, Vintage clothing, Viol, Viola, Violet, Violin, Violin family, Violinist, Violone, Vision care, Visor, Visual arts, Visual effect lighting, Vodka, Volkswagen, Volkswagen 181, Volkswagen beetle, Volkswagen crafter, Volkswagen golf mk1, Volkswagen golf mk5, Volkswagen lupo, Volkswagen tiguan, Volkswagen type 14a, Volkswagen type 2, Volkswagen type 2 (t3), Volleyball, Volleyball net, Volleyball player, Volvo amazon, Volvo c30, Volvo cars, Volvo pv444/544, Volvo v70, Volvo xc60, Volvo xc90, Vortex, Wafer, Waffle, Wagon, Waist, Waiting room, Waiting staff, Wakeboarding, Wakesurfing, Walking, Walking shoe, Walking stick, Walkway, Wall, Wall clock, Wall sticker, Wallball, Wallet, Wallpaper, Walt disney world, Wardrobe, Warehouse, Warehouseman, Warship, Washing, Washing machine, Waste, Waste collector, Waste container, Waste containment, Wat, Watch, Watch accessory, Water bottle, Water park, Water polo, Water polo ball, Water polo cap, Water sport, Water tank, Water tower, Water transportation, Water volleyball, Water well, Watercolor paint, Watercraft, Watercraft rowing, Waterskiing, Wax, Web page, Webbing, Website, Wedding, Wedding cake, Wedding dress, Wedding favors, Wedding reception, Wedding ring, Wedge, Weeder, Weight training, Weightlifter, Weightlifting, Weightlifting machine, Weights, Weisswurst, Welder, Welsh sheepdog, Western concert flute, Western riding, Western Tanager, Wetsuit, Whaler, Wheat beer, Wheel, Wheelbarrow, Wheelchair, Wheelchair racing, Wheelchair sports, Wheelchair tennis, Wheelie, Whipped cream, Whisky, White, White-collar worker, White coat, Whiteboard, Whitewater kayaking, Whole food, Whole grain, Wicker, Wide-body aircraft, Wiener melange, Wig, Wildfire, Wildlife biologist, Wind, Wind chime, Wind farm, Wind instrument, Wind machine, Wind turbine, Windjammer, Windmill, Window, Window blind, Window covering, Window film, Window screen, Window treatment, Windscreen wiper, Windshield, Windsports, Windsurfing, Wine, Wine bottle, Wine cellar, Wine glass, Winemaker, Winery, Wing, Wing chun, Winter, Winter sport, Winter storm, Wire, Wire fencing, Wok, Women's basketball, Women's football, Women's lacrosse, Wonder Woman, Wonders of the world, Wonton, Woo woo, Wood-burning stove, Wood chopping, Wood flooring, Wood shaper, Wood stain, Woodsman, Woodwind instrument, woodworking, Wool, Woolen, Work boots, Workbench, Workhouse, Workshop, Workwear, World, World rally championship, Woven fabric, Wrangler, Wreath, Wrestler, Wrestling, Wrinkle, Wrist, Wristband, Writing, Writing implement, Writing instrument accessory, Xing yi quan, Yacht, Yakiniku, Yard, Yawn, Yellow, Yoga, Yoga mat, yoga pant, York boat, Youth, Yuanyang, Yurt, Zabumba, Zebra crossing, Zither, Zombie, Zoo, Zumba

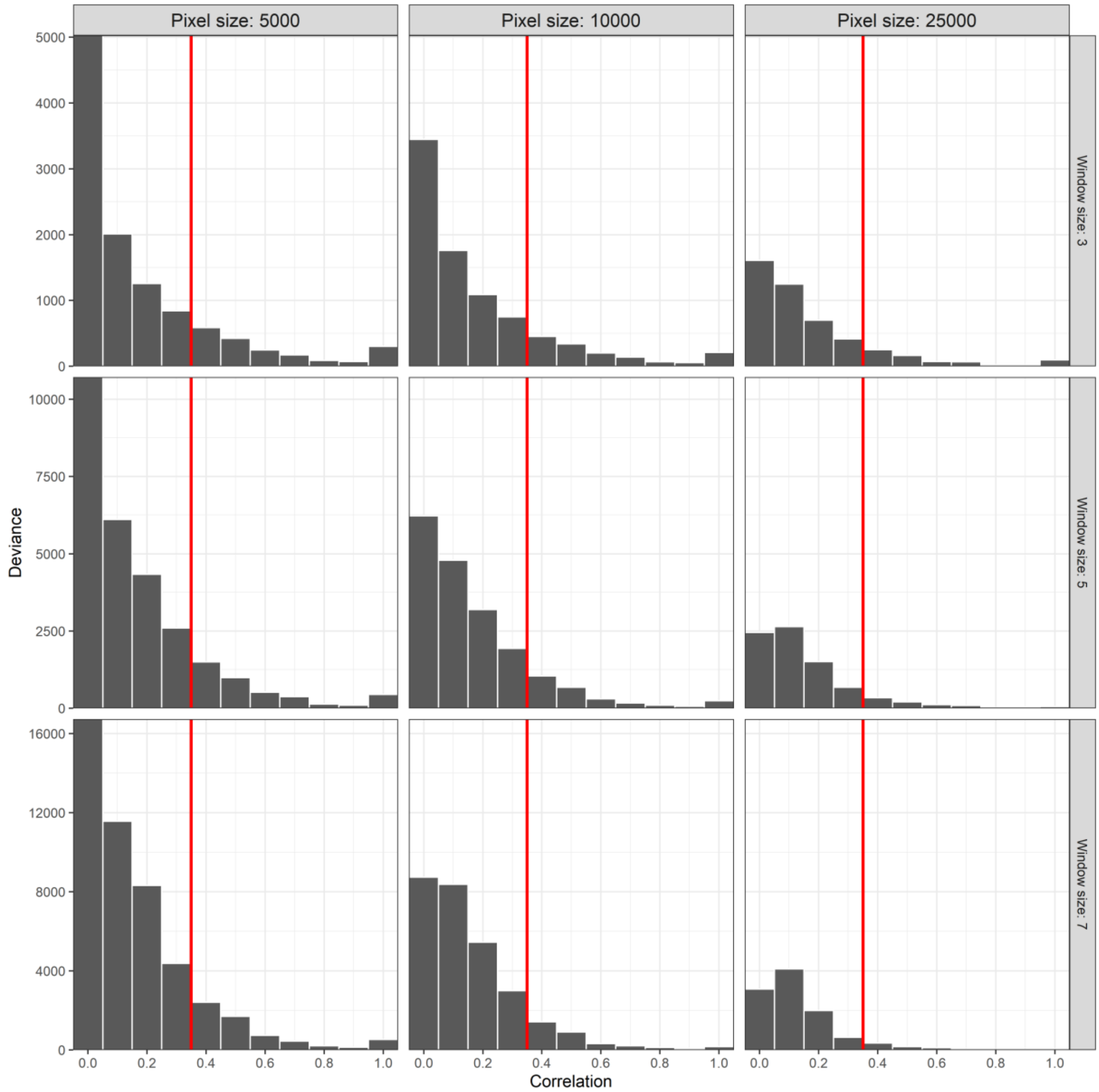
Appendix E Frequency of local deviance values for all runs



E.1 Frequency of local deviance values in the scaled number of images from the full Flickr and confirmedCES images dataset from the 1:1 line when varying the threshold for the number of features to be accepted as an image of nature, pixel resolution and window size. a) pixel size 5000, b) pixel size 10000, c) pixel size 25000.



E.2 Frequency of local deviance values in the scaled number of images from the full Flickr and confirmedNature images dataset from the 1:1 line when varying the threshold for the number of features to be accepted as an image of nature, pixel resolution and window size. a) pixel size 5000, b) pixel size 10000, c) pixel size 25000.



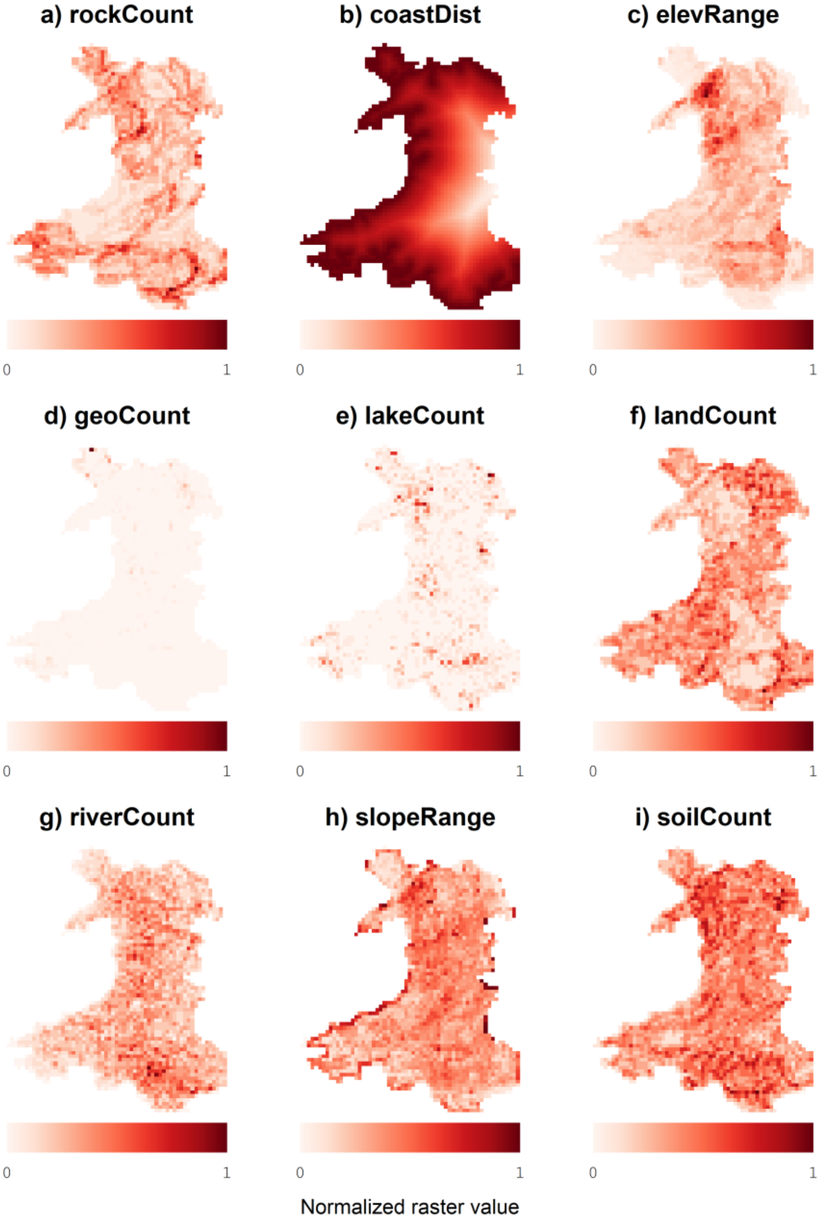
E.3 Frequency of local deviance values in the scaled number of images from the full Flickr and confirmedPositive images dataset from the 1:1 line when varying the threshold for the pixel resolution and window size.

Appendix F Reproducible data collection methods for chapter 6

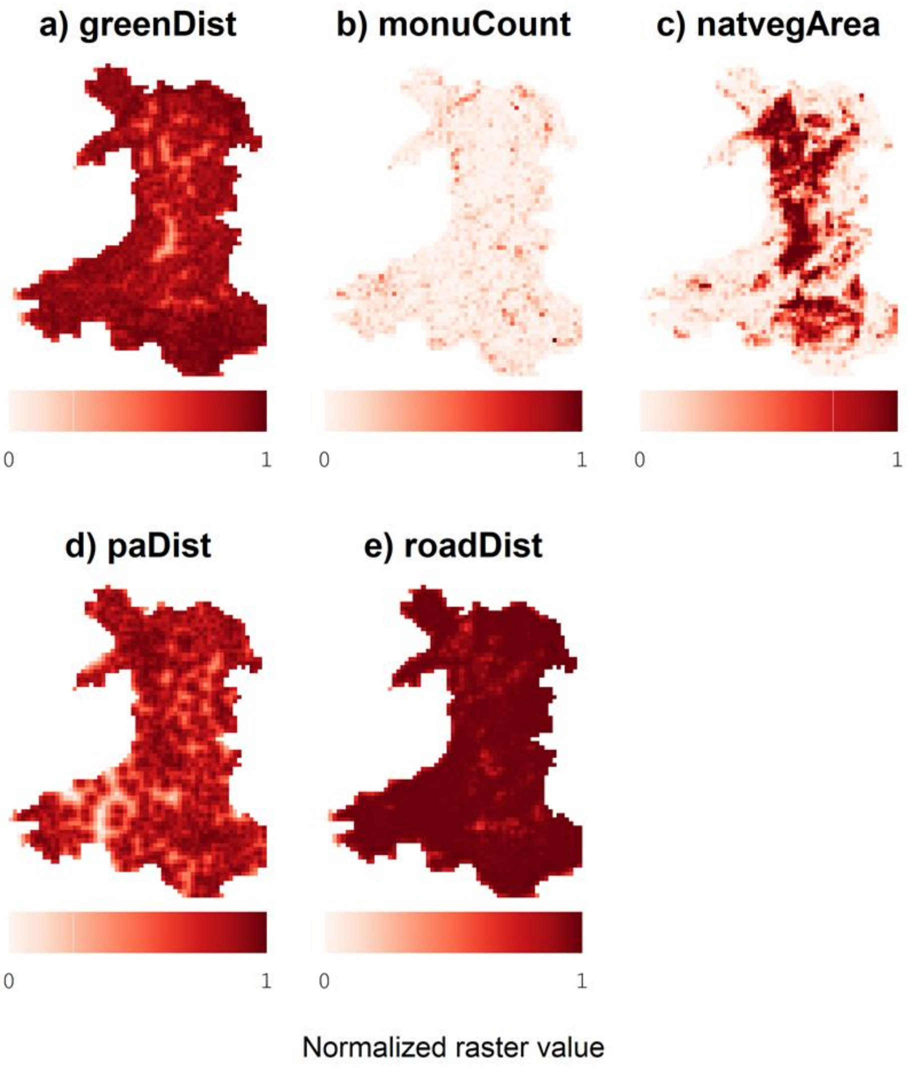
```
# pacman package allows for better loading of uninstalled packages
if(!"pacman" %in% installed.packages()) install.packages("pacman")
library(pacman)

# Load required libraries and install if needed
p_load(dplyr,
       sf,
       sp
)

# Load from GitHub
p_load_gh("ropensci/photosearcher")

#upload shape file
wales_sf <- sf::read_sf(".\\UK_shape\\UK_shape.shp") #directory to your shape
file
crs_reproject <- "+proj=longlat +datum=WGS84 +no_defs" #new crs to project to
wales_sf <- sf::st_transform(wales_sf, crs_reproject) # project shapefile

#add varients of terms
search_term <- c("hike", "hiking","walk","walking", "trek",
                 "trekking","ramble","rambling")

#set null data frames
out_df <- NULL
tmp_df <- NULL
flickr_out <- NULL
count <- NULL

#search through all the search terms and join together
for(i in 1:length(search_term)){

  print(search_term[i])
```

```
  flickr_tmp <- photosearcher::photo_search(mindate_taken = "2001-01-01",
                                            maxdate_taken = "2021-01-01",
                                            maxdate_uploaded = "2021-02-01",
                                            text = paste(search_term[i]),
                                            sf_layer = wales_sf)

  flickr_out <- dplyr::bind_rows(flickr_out, flickr_tmp)
  flickr_out <- dplyr::distinct(flickr_out)
```

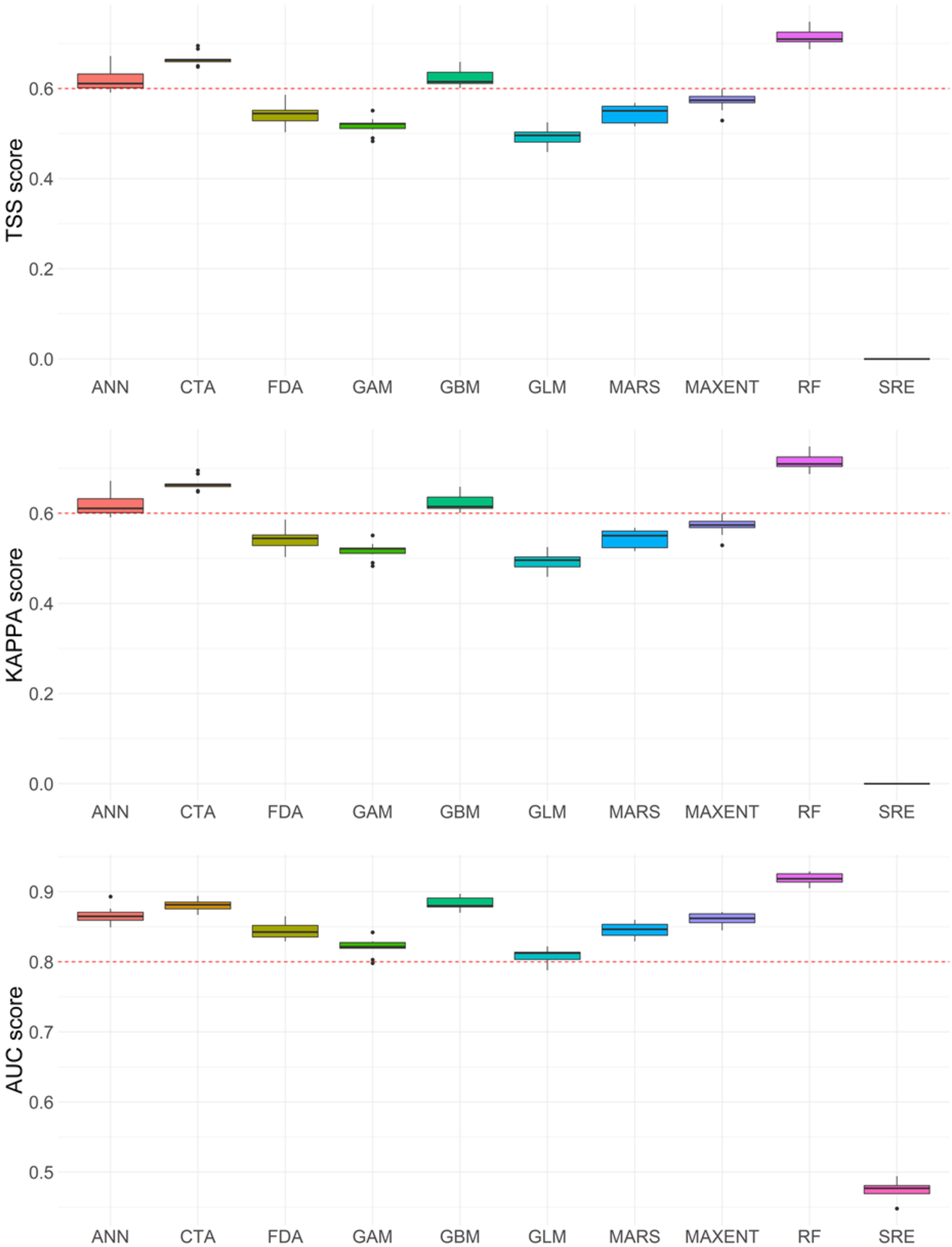
Appendix G Geodiversity variables used in the species distribution model, a) count of bedrock types, b) distance to coast, c) range in elevation, d) count of geosites, e) count of lakes, f) count of landscape types, g) count of rivers, h) range in slope, i) count of soil types.



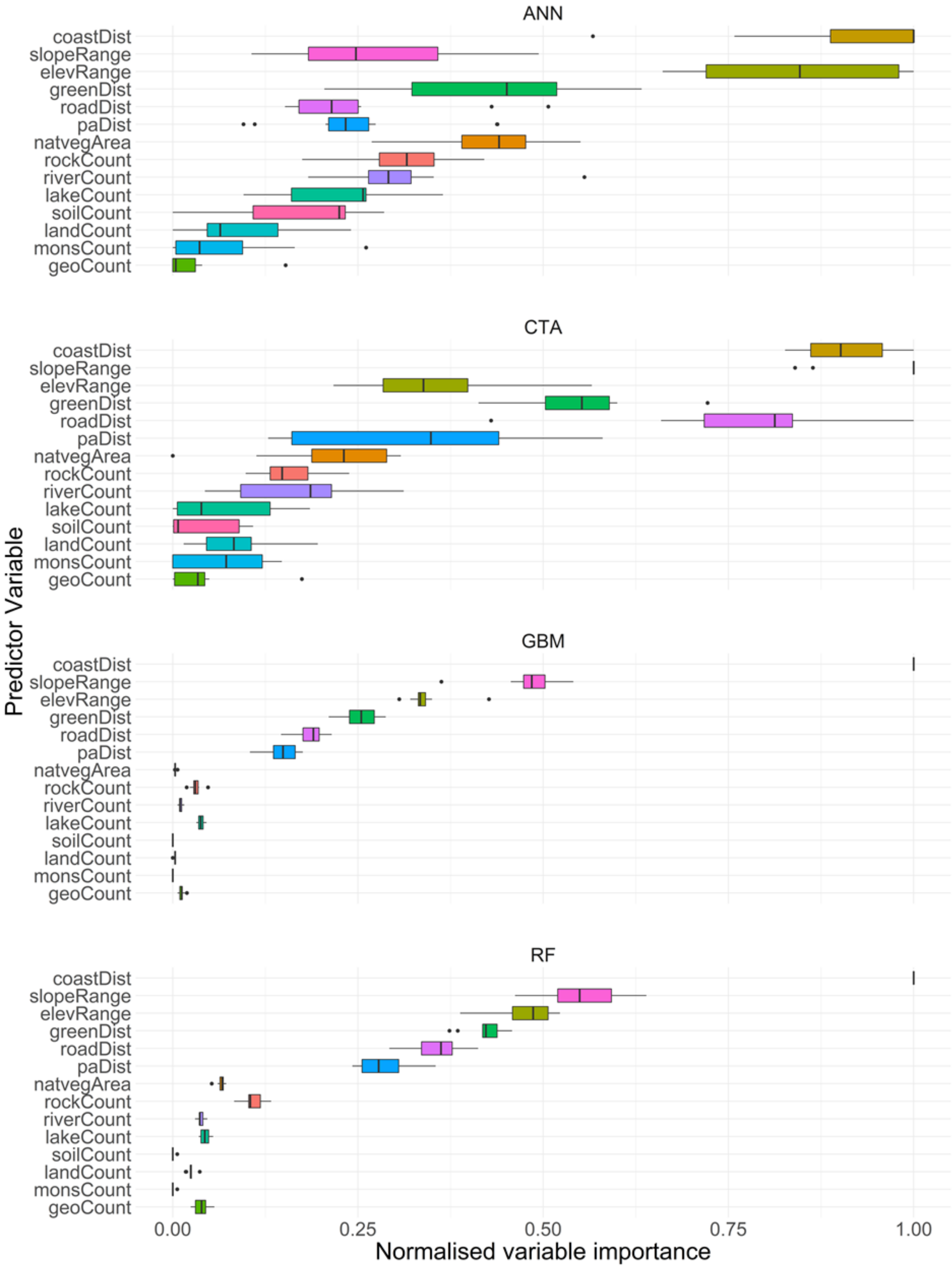
Appendix H Non-geodiversity variables used in the species distribution model, a) distance to greenspace access point, b) count of scheduled monuments, c) area of natural or semi-natural vegetation, d) distance to protected area, e) distance to road



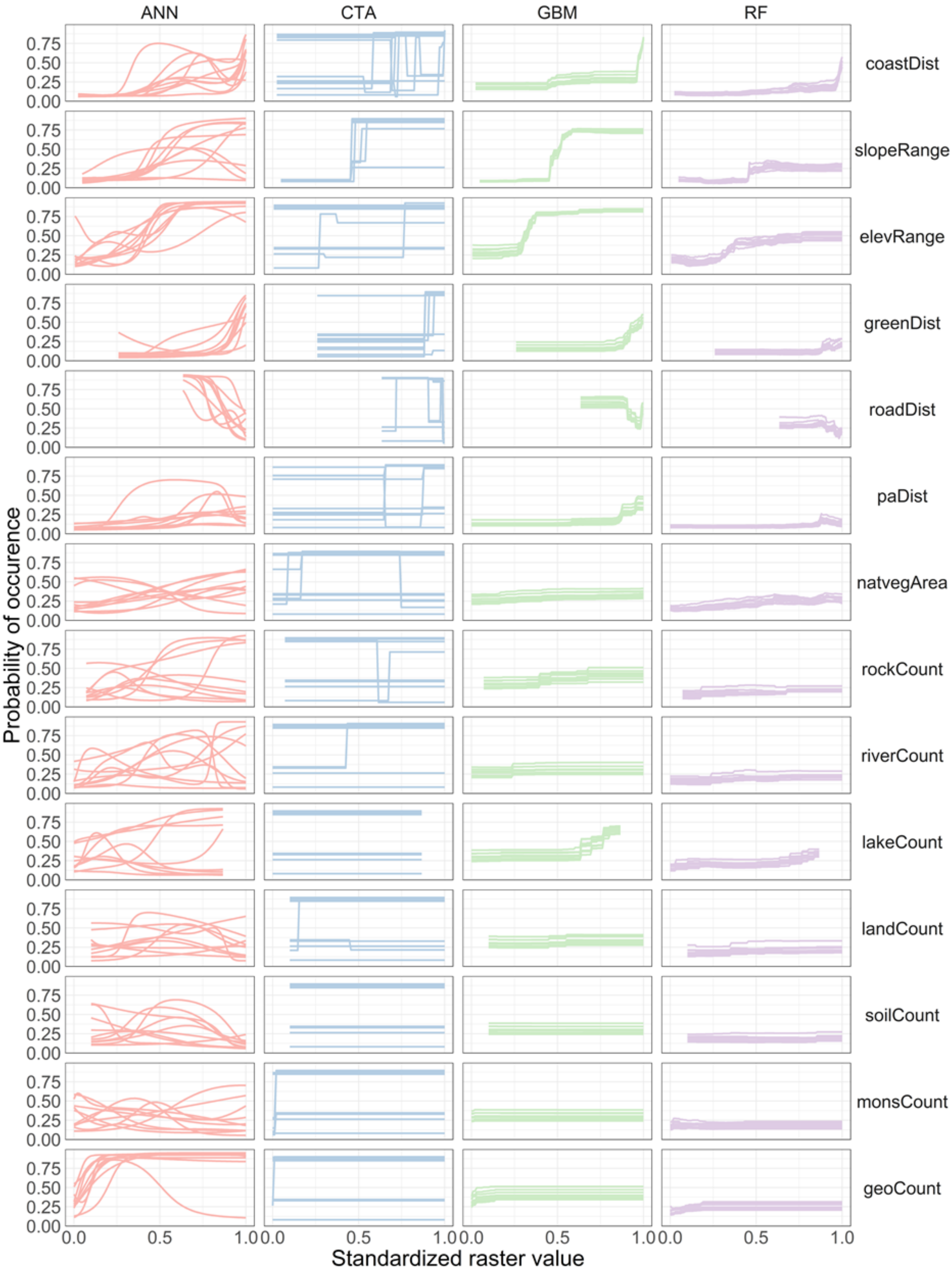
Appendix I Model performance



Appendix J Individual model variable importance



Appendix K Individual model response curves



Appendix L Reproducible data collection methods for chapter 7

```
# pacman package allows for better loading of uninstalled packages
if(!"pacman" %in% installed.packages()) install.packages("pacman")
library(pacman)

# Load required libraries and install if needed
p_load(dplyr,
       RJSONIO,
       devtools,
       tidytext,
       data.table,
       ggplot2,
       magrittr,
       translateR,
       rnaturalearth,
       sf,
       tidyr,
       rreddit)

# Load from GitHub
p_load_gh("ropensci/photosearcher",
          "cschwem2er/imgrec",
          "mkearney/rreddit",
          "trinker/entity")
```

L.1 2. Search Reddit

The following is a basic search of Reddit for the all posts with the term “hiking” in its textual metadata posted in the year 2020. We recommend that data be anonymized and additional irrelevant metadata be deleted before the outputs are saved.

```
#define search term and subreddit
search_term <- "hiking"
subreddit <- NULL

#define dates
start_date <- "2020-01-01"
end_date <- "2021-01-01"

#carry out search
for(i in 1){

  #combine search terms if multiple
  if(length(search_term) > 1){
    search_term <-  paste(search_term, collapse='+')
  }

  #create base url with dates and search terms
  base_url <- paste("https://api.pushshift.io/reddit/submission/search/
```

```
",
                    "?&after=",
                    start_date,
                    "&before=",
                    end_date,
                    "&sort=asc",
                    "&limit=100",
                    ifelse(!(is.null(subreddit)), paste0(
                      "&subreddit=", subreddit), ""),
                    ifelse(!(is.null(search_term)), paste0(
                      "&q=", search_term), ""),
                    sep = "")

  #parse api data
  jsondata <- jsonlite::fromJSON(base_url, flatten = TRUE)
  pushshift <- data.table::rbindlist(jsondata, fill= FALSE)

  #find new date to search from
  new_date <- pushshift$created_utc[nrow(pushshift)]
  new_date <- format(as.POSIXct(new_date, origin= '1970-01-01'))
  new_date <- gsub(" ", "%20", new_date, fixed = TRUE) #%20 is just spac
e for search

  print(new_date) #print to console to check how long to go

  #check if new search is needed
  check <- nrow(pushshift)

  while(check > 0) {

    #create base url with dates and search terms
    base_url <- paste("https://api.pushshift.io/reddit/submission/search
/",
                      "?&after=",
                      new_date,
                      "&before=",
                      end_date,
                      "&sort=asc",
                      "&limit=100",
                      ifelse(!(is.null(subreddit)), paste0(
                        "&subreddit=", subreddit), ""),
                      ifelse(!(is.null(search_term)), paste0(
                        "&q=", search_term), ""),
                      sep = "")

    #parse api data
    jsondata <- jsonlite::fromJSON(base_url, flatten = TRUE)
    tmp_pushshift <- data.table::rbindlist(jsondata, fill= FALSE)

    #merge to all data
    pushshift <- dplyr::bind_rows(pushshift, tmp_pushshift)

    #which columns to keep
    keeps <- c("author",
               "created_utc",
               "title",
```

```
                "selftext",
                "url",
                "permalink",
                "score",
                "num_comments",
                "subreddit",
                "over_18")

    #select only the keep cols
    pushshift <- pushshift %>%
      select(all_of(keeps))

    #find new date to search from
    new_date <- pushshift$created_utc[nrow(pushshift)]
    new_date <- format(as.POSIXct(new_date, origin= '1970-01-01'))
    new_date <- gsub(" ", "%20", new_date, fixed = TRUE) #%20 is just sp
ace for search

    print(new_date) #print to console to check how long to go

    #check if new search is needed
    check <- nrow(tmp_pushshift)

  }

}

#function to parse api results
set_lists_to_chars <- function(x) {
  if(class(x) == 'list') {
    y <- paste(unlist(x[1]), sep='', collapse=', ')
  } else {
    y <- x
  }
  return(y)
}


#turn results into a df
reddit_data <- data.frame(lapply(pushshift , set_lists_to_chars),
                        stringsAsFactors = F)

#write clean csv
readr::write_csv(new_frame, ".\\reddit_data")
```

L.2 3. Search Flickr for data

The following is a basic search of Flickr for the all posts with the term “hiking” in its textual metadata posted in the year 2020. We recommend that data be anonymized and additional irrelevant metadata be deleted before the outputs are saved.

```
flickr_data <- photosearcher::photo_search(mindate_taken = "2000-01-01",
                                           maxdate_taken = "2020-01-01",
                                           mindate_uploaded = "2020-01-0
1",
```

```
                                                    maxdate_uploaded = "2021-01-0
1",
                                                    tags = c("hiking"),
                                                    tags_any = FALSE)


#remove non-unique images
flickr_data <- unique(setDT(flickr_full), by = "id")

#save outputs
write.csv(flickr_data, ".\\flickr_full.csv")
```

List of References

Adekola O, Mitchell G (2011) The Niger Delta wetlands: Threats to ecosystem services, their importance to dependent communities and possible management measures. Int J Biodivers Sci Ecosyst Serv Manag 7:50–68. https://doi.org/10.1080/21513732.2011.603138

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

Aires F (2014) Combining Datasets of Satellite-Retrieved Products. Part I: Methodology and Water Budget Closure. J Hydrometeorol 15:1677–1691. https://doi.org/10.1175/jhm-d-13-0148.1

Alahuhta J, Ala-Hulkko T, Tukiainen H, et al (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

Alahuhta J, Toivanen M, Hjort J (2020) Geodiversity–biodiversity relationship needs more empirical evidence. Nat Ecol Evol 4:2–3. https://doi.org/10.1038/s41559-019-1051-7

Alfred R, Leong LC, On CK, Anthony P (2014) Malay Named Entity Recognition Based on Rule-Based Approach. Int J Mach Learn Comput 4:300–306. https://doi.org/10.7763/ijmlc.2014.v4.428

Allain S (2019) Mining Flickr: a method for expanding the known distribution of invasive species. Herpetol Bull 148:11–14. https://doi.org/10.33256/hb148.1114

Allan JD, Smith SDP, McIntyre PB, et al (2015) Using cultural ecosystem services to inform restoration priorities in the Laurentian Great Lakes. Front Ecol Environ 13:418–424. https://doi.org/10.1890/140328

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, et al (2015) Case studies of conservation plans that incorporate geodiversity. Conserv Biol 29:680–691. https://doi.org/10.1111/cobi.12503

Araujo Calçada E, Lenoir J, Plue J, et al (2015) Spatial patterns of water-deposited seeds control plant species richness and composition in riparian forest landscapes. Landsc Ecol 30:2133–2146. https://doi.org/10.1007/s10980-015-0236-y

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

August TA, Affouard A, Bystriakova N, et al (2019) AI validated plant observations from social media: Flickr images from central London 2011-2019

August TA, Pescott OL, Joly A, Bonnet P (2020) AI Naturalists Might Hold the Key to Unlocking Biodiversity Data in Social Media Imagery. Patterns 1:100116. https://doi.org/10.1016/j.patter.2020.100116

Bailey JJ, Boyd DS, Hjort J, et al (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

Balzan M V., Debono I (2018) Assessing urban recreation ecosystem services through the use of geocache visitation and preference data: A case-study from an urbanised island environment. One Ecosyst 3:. https://doi.org/10.3897/oneeco.3.e24490

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

Barthlott W, Hostert A, Kier G, et al (2007) Geographic patterns of vascular plant diversity at continental to global scales. Erdkunde 61:305–315

Barton JA, Willis BL, Hutson KS (2017) Coral propagation: a review of techniques for ornamental trade and reef restoration. Rev Aquac 9:238–256. https://doi.org/10.1111/raq.12135

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

Baumgartner J, Zannettou S, Keegan B, et al (2020) The pushshift reddit dataset. Proc 14th Int AAAI Conf Web Soc Media, ICWSM 2020 830–839

Becken S, Stantic B, Chen J, et al (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 surface processes and landforms. Earth Surf Process Landforms

34:1433–1445. https://doi.org/https://doi.org/10.1002/esp.1840

Bennett EM, Cramer W, Begossi A, et al (2015) Linking biodiversity, ecosystem services, and human well-being: three challenges for designing research for sustainability. Curr Opin Environ Sustain 14:76–85. https://doi.org/10.1016/j.cosust.2015.03.007

Bennett EM, Peterson GD, Gordon LJ (2009) Understanding relationships among multiple ecosystem services. Ecol Lett 12:1394–1404. https://doi.org/10.1111/j.1461-0248.2009.01387.x

Bétard F (2013) Patch-Scale Relationships Between Geodiversity and Biodiversity in Hard Rock Quarries: Case Study from a Disused Quartzite Quarry in NW France. Geoheritage 5:59–71. https://doi.org/10.1007/s12371-013-0078-4

Bian L, Butler R (1999) Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. Photogramm Eng Remote Sensing 65:73–84

Boe B (2020) PRAW: The Python Reddit API Wrapper.

Brauman KA, Daily GC, Duarte TK eo, Mooney HA (2007) The nature and value of ecosystem services: An overview highlighting hydrologic services. Annu Rev Environ Resour 32:67–98. https://doi.org/10.1146/annurev.energy.32.031306.102758

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

Brilha J (2017) Trends and challenges for geoconservation. In: EGU General Assembly Conference Abstracts. p 8626

Brilha J (2016) Inventory and Quantitative Assessment of Geosites and Geodiversity Sites: a Review. Geoheritage 8:119–134. https://doi.org/10.1007/s12371-014-0139-3

Brilha J, Gray M, Pereira DI, Pereira P (2018) Geodiversity: An integrative review as a contribution to the sustainable management of the whole of nature. Environ Sci Policy 86:19–28. https://doi.org/10.1016/j.envsci.2018.05.001

Brindley P, Cameron RW, Ersoy E, et al (2019) Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health. Urban For Urban Green 39:45–54. https://doi.org/10.1016/j.ufug.2019.01.015

Brocx M, Semeniuk V (2010) Coastal geoheritage: A hierarchical approach to classifying coastal types as a basis for identifying geodiversity and sites of significance in Western Australia. J R Soc West Aust 93:81–113

Brocx M, Semeniuk V (2007) Geoheritage and geoconservation - History, definition, scope and scale. J R Soc West Aust 90:53–87

Brown EJ, Prosser CD, Stevenson NM (2012) Geodiversity, Conservation and Climate Change: Key Principles for Adaptation. Scottish Geogr J 128:234–239. https://doi.org/10.1080/14702541.2012.725859

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

Burek C V., Prosser CD (2008) The history of geoconservation: An introduction. Geol Soc Spec Publ 300:1–5. https://doi.org/10.1144/SP300.1

Byczek C, Longaretti PY, Renaud J, Lavorel S (2018) Benefits of crowd-sourced GPS information for modelling the recreation ecosystem service. PLoS One 13:1–23. https://doi.org/10.1371/journal.pone.0202645

Casalegno S, Inger R, DeSilvey C, Gaston KJ (2013) Spatial Covariance between Aesthetic Value & Other Ecosystem Services. PLoS One 8:6–10. https://doi.org/10.1371/journal.pone.0068437

Čengić M, Rost J, Remenska D, et al (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

Çetiner ZS, Ertekin C, Yiğitbaş E (2018) Evaluating Scientific Value of Geodiversity for Natural Protected Sites: the Biga Peninsula, Northwestern Turkey. Geoheritage 10:49–65. https://doi.org/10.1007/s12371-017-0218-3

Chan KMA, Goldstein J, Satterfield T, et al (2011) Cultural services and non-use values. In: Natural Capital. Oxford University Press, pp 206–228

Chan KMA, Satterfield T, Goldstein J (2012) Rethinking ecosystem services to better address and navigate cultural values. Ecol Econ 74:8–18. https://doi.org/10.1016/j.ecolecon.2011.11.011

Cheesman AW, Preece ND, van Oosterzee P, et al (2018) The role of topography and plant functional traits in determining tropical reforestation success. J Appl Ecol 55:1029–1039. https://doi.org/10.1111/1365-2664.12980

Chen Y, Caesemaecker C, Rahman HT, Sherren K (2020) Comparing cultural ecosystem service delivery in dykelands and marshes using Instagram: A case of the Cornwallis (Jijuktu'kwejk) River, Nova Scotia, Canada. Ocean Coast Manag 193:105254. https://doi.org/10.1016/j.ocecoaman.2020.105254

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

Christian H, Agus MP, Suhartono D (2016) Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). ComTech Comput Math Eng Appl 7:285. https://doi.org/10.21512/comtech.v7i4.3746

Ciesielski M, Stereńczak K (2021) Using Flickr data and selected environmental characteristics to analyse the temporal and spatial distribution of activities in forest areas. For Policy Econ 129:. https://doi.org/10.1016/j.forpol.2021.102509

Clemente P, Calvache M, Antunes P, et al (2019) Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a Natural Park in Portugal. Ecol Indic 96:59–68. https://doi.org/10.1016/j.ecolind.2018.08.043

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

Comer PJ, Pressey RL, Hunter ML, et al (2015) Incorporating geodiversity into conservation decisions. Conserv Biol 29:692–701. https://doi.org/10.1111/cobi.12508

Copernicus (2021) Copernicus land portal

Cortinovis C, Geneletti D (2019) A framework to explore the effects of urban planning decisions on regulating ecosystem services in cities. Ecosyst Serv 38:100946. https://doi.org/10.1016/j.ecoser.2019.100946

Costanza R, D'Arge R, de Groot R, et al (1997) The value of the world's ecosystem services and natural capital. Nature 387:253–260. https://doi.org/https://doi.org/10.1038/387253a0

Costanza R, de Groot R, Sutton P, et al (2014) Changes in the global value of ecosystem services. Glob Environ Chang 26:152–158. https://doi.org/10.1016/j.gloenvcha.2014.04.002

Cranfield University (2021) The Soils Guide

Crenna E, Sozzo S, Sala S (2018) Natural biotic resources in LCA: Towards an impact assessment model for sustainable supply chain management. J Clean Prod 172:3669–3684.

https://doi.org/10.1016/j.jclepro.2017.07.208

Cresswell ID (2019) Geoheritage and Geoconservation in Australia. Aust J Earth Sci 66:753–766. https://doi.org/10.1080/08120099.2018.1516698

Crofts R (2014) Promoting geodiversity: Learning lessons from biodiversity. Proc Geol Assoc 125:263–266. https://doi.org/10.1016/j.pgeola.2014.03.002

da Silva CM (2019) Urban Geodiversity and Decorative Arts: the Curious Case of the “Rudist Tiles” of Lisbon (Portugal). Geoheritage 11:151–163. https://doi.org/10.1007/s12371-017-0253-0

Dai P, Zhang S, Chen Z, et al (2019) Perceptions of cultural ecosystem services in urban parks based on social network data. Sustain 11:. https://doi.org/10.3390/su11195386

Daily GC (1997) Nature’s Services: Societal Dependence on Natural Ecosystems. Island Press, Washington, DC

Daniel TC, Muhar A, Arnberger A, et al (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

de Almeida Rodrigues A, da Cunha Bustamante MM, Sano EE (2018) As far as the eye can see: Scenic view of Cerrado National Parks. Perspect Ecol Conserv 16:31–37. https://doi.org/10.1016/j.pecon.2017.11.004

de Groot RS, Alkemade R, Braat L, et al (2010) Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. Ecol Complex 7:260–272. https://doi.org/10.1016/j.ecocom.2009.10.006

Derczynski L, Nichols E, van Erp M, Limsopatham N (2018) Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition. 140–147. https://doi.org/10.18653/v1/w17-4418

Deru JGC, Bloem J, de Goede R, et al (2018) Soil ecology and ecosystem services of dairy and semi-natural grasslands on peat. Appl Soil Ecol 125:26–34. https://doi.org/10.1016/j.apsoil.2017.12.011

Díaz S, Demissew S, Carabias J, et al (2015) The IPBES Conceptual Framework - connecting nature and people. Curr Opin Environ Sustain 14:1–16. https://doi.org/10.1016/j.cosust.2014.11.002

Díaz S, Pascual U, Stenseke M, et al (2018) Assessing nature’s contributions to people. Science (80-) 359:270–272. https://doi.org/10.1126/science.aap8826

Dickinson DC, Hobbs RJ (2017) Cultural ecosystem services : Characteristics , challenges and lessons for urban green space research. Ecosyst Serv 25:179–194. https://doi.org/10.1016/j.ecoser.2017.04.014

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

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

Dolan R, Bullock JM, Jones JPG, et al (2021) The Flows of Nature to People , and of People to Nature : Applying Movement Concepts to Ecosystem Services. 1–18

dos Santos FM, de La Corte Bacci D, 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

Dubinin V, Stavi I, Svoray T, et al (2021) Hillslope geodiversity improves the resistance of shrubs to prolonged droughts in semiarid ecosystems. J Arid Environ 188:104462. https://doi.org/10.1016/j.jaridenv.2021.104462

Duggan M, Smith A (2013) 6% of Online Adults are reddit Users

Dunkel A (2015) Visualizing the perceived environment using crowdsourced photo geodata. Landsc Urban Plan 142:173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

Edwards DP, Tobias JA, Sheil D, et al (2014) Maintaining ecosystem function and services in logged tropical forests. Trends Ecol Evol 29:511–520. https://doi.org/10.1016/j.tree.2014.07.003

Eigenbrod F, Armsworth PR, Anderson BJ, et al (2010) The impact of proxy-based methods on mapping the distribution of ecosystem services. J Appl Ecol 47:377–385. https://doi.org/10.1111/j.1365-2664.2010.01777.x

Erikstad L (2013) Geoheritage and geodiversity management - the questions for tomorrow. Proc Geol Assoc 124:713–719. https://doi.org/10.1016/j.pgeola.2012.07.003

Evangelopoulos NE (2013) Latent semantic analysis. Wiley Interdiscip Rev Cogn Sci 4:683–692. https://doi.org/10.1002/wcs.1254

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

Finisdore J, Waltham N, Obst C, et al (2020) Valuation of disaster risk reduction ecosystem services of Australia’s coastal wetlands: review and recommendations.

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

Fish R, Church A, Winter M (2016) Conceptualising cultural ecosystem services : A novel framework for research and critical engagement. Ecosyst Serv 21:208–217. https://doi.org/10.1016/j.ecoser.2016.09.002

Fisher B, Turner RK, Morling P (2009) Defining and classifying ecosystem services for decision making. Ecol Econ 68:643–653. https://doi.org/10.1016/j.ecolecon.2008.09.014

Flickr (2021) Flickr. https://www.flickr.com/explore

Fox N, August T, Mancini F, et al (2020a) “photosearcher” package in R: An accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX 12:100624. https://doi.org/10.1016/j.softx.2020.100624

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

Garofano M, Govoni D (2012) Underground Geotourism: A Historic and Economic Overview of Show Caves and Show Mines in Italy. Geoheritage 4:79–92. https://doi.org/10.1007/s12371-012-0055-3

Gefen D, Endicott JE, Fresneda JE, et al (2017) A Guide to Text Analysis with Latent Semantic Analysis in R with Annotated Code Studying Online Reviews and the Stack Exchange Community. Commun Assoc Inf Syst 41:450–496. https://doi.org/10.17705/1CAIS.04121

Ghermandi A, Camacho-Valdez V, Trejo-Espinosa H (2020a) Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tour Manag 77:104002. https://doi.org/10.1016/j.tourman.2019.104002

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, Sinclair M, Fichtman E, Gish M (2020b) Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing. Glob Environ Chang 65:102189. https://doi.org/10.1016/j.gloenvcha.2020.102189

Gkotsis G, Oellrich A, Velupillai S, et al (2017) Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 7:1–11. https://doi.org/10.1038/srep45141

Gliozzo G, Pettorelli N, Muki Haklay 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 (2018) Geoheritage, geotourism and the cultural landscape: Enhancing the visitor experience and promoting geoconservation. Geosci 8:. https://doi.org/10.3390/geosciences8040136

Gordon JE, Barron HF (2013) The role of geodiversity in delivering ecosystem services and benefits in Scotland. Scottish J Geol 49:41–58. https://doi.org/10.1144/sjg2011-465

Gordon JE, Barron HF, Hansom JD, Thomas MF (2012) Engaging with geodiversity-why it matters. Proc Geol Assoc 123:1–6. https://doi.org/10.1016/j.pgeola.2011.08.002

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

Gosal AS, Geijzendorffer IR, Václavík T, et al (2019) Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosyst Serv 38:100958. https://doi.org/10.1016/j.ecoser.2019.100958

Goyal A, Gupta V, Kumar M (2018) Recent Named Entity Recognition and Classification techniques: A systematic review. Comput Sci Rev 29:21–43. https://doi.org/10.1016/j.cosrev.2018.06.001

Grab H, Branstetter MG, Amon N, et al (2019) Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. Science (80-) 363:282–284

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

Graham LJ, Spake R, Gillings S, et al (2019) Incorporating fine-scale environmental heterogeneity into broad-extent models. Methods Ecol Evol 10:767–778. https://doi.org/10.1111/2041-210X.13177

Gray M (2004) Geodiversity: valuing and conserving abiotic nature. John Wiley & Sons

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

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

Gray M (2012) Valuing Geodiversity in an "Ecosystem Services" Context. Scottish Geogr J 128:177–194. https://doi.org/10.1080/14702541.2012.725858

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

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

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 (2018) Geodiversity: The backbone of geoheritage and geoconservation. Elsevier Inc.

Gray M (2005) Geodiversity and Geoconservation: what, why, and how? George Wright Forum 22:4–12

Gray M, Gordon JE, Brown EJ (2013) Geodiversity and the ecosystem approach: The contribution of geoscience in delivering integrated environmental management. Proc Geol Assoc 124:659–673. https://doi.org/10.1016/j.pgeola.2013.01.003

Guerrero P, Møller MS, Olafsson AS, Snizek B (2016) Revealing cultural ecosystem services through instagram images: The potential of social media volunteered geographic information for urban green infrastructure planning and governance. Urban Plan 1:1–17. https://doi.org/10.17645/up.v1i2.609

Guimarães A, Balalau O, Terolli E, Weikum G (2019) Analyzing the traits and anomalies of political discussions on reddit. Proc 13th Int Conf Web Soc Media, ICWSM 2019 205–213

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

Gülçin D (2021) Predicting Visual Aesthetic Preferences of Landscapes Near Historical Sites By Fluency Theory Using Social Media Data and Gis. lnternational J Geogr Geogr Educ. https://doi.org/10.32003/igge.811658

Gülçin D (2020) Peyzajın Görsel Estetik Değeri ve Peyzaj Çeşitliliği Arasındaki İlişkinin Objektif Paradigma ile Değerlendirilmesi. Bartın Orman Fakültesi Derg. https://doi.org/10.24011/barofd.785561

Hadwen WL, Hill W, Pickering CM (2007) Icons under threat: Why monitoring visitors and their ecological impacts in protected areas matters. Ecol Manag Restor 8:177–181. https://doi.org/10.1111/j.1442-8903.2007.00364.x

Haines-Young R, Potschin MB (2010) The links between biodiversity, ecosystem services and human well-being. In: Raffaelli D, Frid C (eds) Ecosystem Ecology: a new synthesis. Cambridge, pp 110–139

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

Haines-Young R, Potschin MB (2012) Common international classification of ecosystem services (CICES, Version 4.1).

Hale RL, Cook EM, Beltrán BJ (2019) Cultural ecosystem services provided by rivers across diverse social-ecological landscapes: A social media analysis. Ecol Indic 107:105580. https://doi.org/10.1016/j.ecolind.2019.105580

Hallema DW, Moussa R, Sun G, McNulty SG (2016) Surface storm flow prediction on hillslopes based on topography and hydrologic connectivity. Ecol Process 5:1–13. https://doi.org/10.1186/s13717-016-0057-1

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

Harrigian K (2018) Geocoding Without Geotags: A Text-based Approach for reddit. arXiv. https://doi.org/10.18653/v1/w18-6103

Hart AG, Carpenter WS, Hlustik-Smith E, et al (2018) Testing the potential of Twitter mining methods for data acquisition: Evaluating novel opportunities for ecological research in multiple taxa. Methods Ecol Evol 9:2194–2205. https://doi.org/10.1111/2041-210X.13063

Hausmann A, Toivonen T, Fink C, et al (2019) Assessing global popularity and threats to Important Bird and Biodiversity Areas using social media data. Sci Total Environ 683:617–623. https://doi.org/10.1016/j.scitotenv.2019.05.268

Havinga I, Bogaart PW, Hein L, Tuia D (2020) Defining and spatially modelling cultural ecosystem services using crowdsourced data. Ecosyst Serv 43:101091. https://doi.org/10.1016/j.ecoser.2020.101091

Heikinheimo V, Minin E Di, Tenkanen H, et al (2017) User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. ISPRS Int J Geo-Information 6:. https://doi.org/10.3390/ijgi6030085

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

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

Hermes J, Van Berkel D, Burkhard B, et al (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

Hernández-Morcillo M, Plieninger T, Bieling C (2013) An empirical review of cultural ecosystem service indicators. Ecol Indic 29:434–444. https://doi.org/10.1016/j.ecolind.2013.01.013

Hirons M, Comberti C, Dunford R (2016) Valuing Cultural Ecosystem Services. Annu Rev Environ Resour 41:545–574. https://doi.org/10.1146/annurev-environ-110615-085831

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

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, Luoto M (2010) Geodiversity of high-latitude landscapes in northern Finland.

Geomorphology 115:109–116. https://doi.org/10.1016/j.geomorph.2009.09.039

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

Hooper DU, Chapin FS, Ewel JJ, et al (2005) Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecol Monogr 75:3–35. https://doi.org/10.1890/04-0922

Huang F, Zhang X, Zhao Z, et al (2019) Image–text sentiment analysis via deep multimodal attentive fusion. Knowledge-Based Syst 167:26–37. https://doi.org/10.1016/j.knosys.2019.01.019

Hueneke H, Baker R (2009) Tourist behaviour, local values, and interpretation at ulur{line below}u: "The sacred deed at Australia's mighty heart." GeoJournal 74:477–490. https://doi.org/10.1007/s10708-008-9249-2

Isaac NJB, Jarzyna MA, Keil P, et al (2020) Data Integration for Large-Scale Models of Species Distributions. Trends Ecol Evol 35:56–67. https://doi.org/10.1016/j.tree.2019.08.006

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

Jamnik MR, Lane DJ (2017) The use of Reddit as an inexpensive source for high-quality data. Pract Assessment, Res Eval 22:1–10

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

Jeawak SS, Jones CB, Schockaert S (2017) Using flickr for characterizing the environment: An exploratory analysis. Leibniz Int Proc Informatics, LIPIcs 86:. https://doi.org/10.4230/LIPIcs.COSIT.2017.21

Jeffers ES, Nogué S, Willis KJ (2015) The role of palaeoecological records in assessing ecosystem services. Quat Sci Rev 112:17–32. https://doi.org/10.1016/j.quascirev.2014.12.018

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

Johnson TF, Kent H, Hill BM, et al (2021) classecol: classifiers to understand public opinions of nature. Methods Ecol Evol 1:2041–210X.13596. https://doi.org/10.1111/2041-210X.13596

jReddit (2020) jReddit

Kahle D, Wickham H (2013) ggmap: Spatial visualization with ggplot2. R J 5:144–161. https://doi.org/10.32614/rj-2013-014

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

Kandziora M, Burkhard B, Müller F (2013) Mapping provisioning ecosystem services at the local scale using data of varying spatial and temporal resolution. Ecosyst Serv 4:47–59. https://doi.org/10.1016/j.ecoser.2013.04.001

Kang P, Chen W, Hou Y, Li Y (2018) Linking ecosystem services and ecosystem health to ecological risk assessment: A case study of the Beijing-Tianjin-Hebei urban agglomeration. Sci Total Environ 636:1442–1454. https://doi.org/10.1016/j.scitotenv.2018.04.427

Karasov O, Heremans S, Külvik M, et al (2020) On How Crowdsourced Data and Landscape Organisation Metrics Can Facilitate the Mapping of Cultural Ecosystem Services: An Estonian Case Study. Land 9:158. https://doi.org/10.3390/land9050158

Kearney M. (2020) . rreddit: Collecting reddit data. R package version 0.0.1.

Kearney SP, Fonte SJ, García E, et al (2019) Evaluating ecosystem service trade-offs and synergies from slash-and-mulch agroforestry systems in El Salvador. Ecol Indic 105:264–278. https://doi.org/10.1016/j.ecolind.2017.08.032

Kiernan K (2015) Landforms as Sacred Places: Implications for Geodiversity and Geoheritage. Geoheritage 7:177–193. https://doi.org/10.1007/s12371-014-0128-6

Kim H-Y (2017) Statistical notes for clinical researchers: Chi-squared test and Fisher’s exact test. Restor Dent Endod 42:152. https://doi.org/10.5395/rde.2017.42.2.152

Kim Y, Kim C ki, Lee DK, et al (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, et al (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

Kipp MEI, Beak J, Choi I (2017) Motivations and intentions of flickr users in enriching flick records for library of congress photos. J Assoc Inf Sci Technol 68:2364–2379. https://doi.org/10.1002/asi.23869

Kondo MC, Fluehr JM, McKeon T, Branas CC (2018) Urban green space and its impact on human health. Int J Environ Res Public Health 15:. https://doi.org/10.3390/ijerph15030445

Koto F, Adriani M (2015) A comparative study on twitter sentiment analysis: Which features are good? Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9103:453–457. https://doi.org/10.1007/978-3-319-19581-0_46

Kremen C (2005) Managing ecosystem services: What do we need to know about their ecology? Ecol Lett 8:468–479. https://doi.org/10.1111/j.1461-0248.2005.00751.x

Kubalíková L (2020) Cultural Ecosystem Services of Geodiversity : A Case. Land 9:15

Kubalíková L, Drápela E, Kirchner K, et al (2021) Urban geotourism development and geoconservation: Is it possible to find a balance? Environ Sci Policy 121:1–10. https://doi.org/10.1016/j.envsci.2021.03.016

Kubiszewski I, Costanza R, Anderson S, Sutton P (2017) The future value of ecosystem services: Global scenarios and national implications. Ecosyst Serv 26:289–301. https://doi.org/10.1016/j.ecoser.2017.05.004

Kumar P (2010) The economics of ecosystems and biodiversity (TEEB): ecological and economic foundations. Earthscan, London

La Notte A, D’Amato D, Mäkinen H, et al (2017) Ecosystem services classification: A systems ecology perspective of the cascade framework. Ecol Indic 74:392–402. https://doi.org/10.1016/j.ecolind.2016.11.030

Landuyt D, Broekx S, D’hondt R, et al (2013) A review of Bayesian belief networks in ecosystem service modelling. Environ Model Softw 46:1–11. https://doi.org/10.1016/j.envsoft.2013.03.011

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

Larondelle N, Haase D, Kabisch N (2014) Mapping the diversity of regulating ecosystem services in

European cities. Glob Environ Chang 26:119–129. https://doi.org/10.1016/j.gloenvcha.2014.04.008

Larwood JG, Badman T, McKeever PJ (2013) The progress and future of geoconservation at a global level. Proc Geol Assoc 124:720–730. https://doi.org/10.1016/j.pgeola.2013.04.001

Lautenbach S, Mupepele AC, Dormann CF, et al (2019) Blind spots in ecosystem services research and challenges for implementation. Reg Environ Chang 19:2151–2172. https://doi.org/10.1007/s10113-018-1457-9

Lavorel S, Bayer A, Bondeau A, et al (2017) Pathways to bridge the biophysical realism gap in ecosystem services mapping approaches. Ecol Indic 74:241–260. https://doi.org/10.1016/j.ecolind.2016.11.015

Lawler JJ, Ackerly DD, Albano CM, et al (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

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

Li L, Goodchild MF, Xu B (2013) Spatial, temporal, and socioeconomic patterns in the use of twitter and flickr. Cartogr Geogr Inf Sci 40:61–77. https://doi.org/10.1080/15230406.2013.777139

Lin S, Wu R, Yang F, et al (2018) Spatial trade-offs and synergies among ecosystem services within a global biodiversity hotspot. Ecol Indic 84:371–381. https://doi.org/10.1016/j.ecolind.2017.09.007

Liu X, He J, Yao Y, et al (2017) Classifying urban land use by integrating remote sensing and social media data. Int J Geogr Inf Sci 31:1675–1696. https://doi.org/10.1080/13658816.2017.1324976

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

Mace GM, Norris K, Fitter AH (2012) Biodiversity and ecosystem services: A multilayered relationship. Trends Ecol Evol 27:19–26. https://doi.org/10.1016/j.tree.2011.08.006

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

Maes J, Egoh B, Willemen L, et al (2012) Mapping ecosystem services for policy support and decision making in the European Union. Ecosyst Serv 1:31–39. https://doi.org/10.1016/j.ecoser.2012.06.004

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

Manosso FC, Zwoliński Z, Najwer A, et al (2021) Spatial pattern of geodiversity assessment in the Marrecas River drainage basin, Paraná, Brazil. Ecol Indic 126:107703. https://doi.org/10.1016/j.ecolind.2021.107703

Martí P, García-Mayor C, Nolasco-Cirugeda A, Serrano-Estrada L (2020) Green infrastructure planning: Unveiling meaningful spaces through Foursquare users' preferences. Land use policy 97:104641. https://doi.org/10.1016/j.landusepol.2020.104641

Martínez Pastur G, Peri PL, Lencinas M V., et al (2016) Spatial patterns of cultural ecosystem services provision in Southern Patagonia. Landsc Ecol 31:383–399. https://doi.org/10.1007/s10980-015-0254-9

Maseyk FJF, Mackay AD, Possingham HP, et al (2017) Managing Natural Capital Stocks for the Provision of Ecosystem Services. Conserv Lett 10:211–220. https://doi.org/10.1111/conl.12242

Matthews TJ (2014) Integrating Geoconservation and Biodiversity Conservation: Theoretical Foundations and Conservation Recommendations in a European Union Context. Geoheritage 6:57–70. https://doi.org/10.1007/s12371-013-0092-6

Mayer M, Woltering M (2018) Assessing and valuing the recreational ecosystem services of Germany's national parks using travel cost models. Ecosyst Serv 31:371–386. https://doi.org/10.1016/j.ecoser.2017.12.009

McDonald YJ, Schwind M, Goldberg DW, et al (2017) An analysis of the process and results of manual geocode correction. Geospat Health 12:. https://doi.org/10.4081/gh.2017.526

Mckay E (2021) Dis-origined Materials : The Role and Invisibilization of. SocArXiv

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

Mendoza M, Poblete B, Valderrama I (2019) Nowcasting earthquake damages with Twitter. EPJ Data Sci 8:. https://doi.org/10.1140/epjds/s13688-019-0181-0

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

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

Minkiewicz J, Bridson K, Evans J (2016) Co-production of service experiences: insights from the cultural sector. J Serv Mark 30:749–761. https://doi.org/10.1108/JSM-04-2015-0156

Miri S, Naghdi M, Rouissi T, et al (2019) Recent biotechnological advances in petroleum hydrocarbons degradation under cold climate conditions: A review. Crit Rev Environ Sci Technol 49:553–586. https://doi.org/10.1080/10643389.2018.1552070

Mitten D, Overholt JR, Haynes FI, et al (2018) Hiking: A Low-Cost, Accessible Intervention to Promote Health Benefits. Am J Lifestyle Med 12:302–310. https://doi.org/10.1177/1559827616658229

Mittermeier JC, Roll U, Matthews TJ, Grenyer R (2019) A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. PLoS Biol 17:e3000146. https://doi.org/10.1371/journal.pbio.3000146

Monkman GG, Kaiser MJ, Hyder K (2018) Text and data mining of social media to map wildlife recreation activity. Biol Conserv 228:89–99. https://doi.org/10.1016/j.biocon.2018.10.010

Moore R, Dattilo J, Devine MA (1996) A comparison of rail-trail preferences between adults with and without disabilities. Adapt Phys Act Q 13:27–37. https://doi.org/10.1123/apaq.13.1.27

Moreno-Llorca R, F. Méndez P, Ros-Candeira A, et al (2020) Evaluating tourist profiles and nature-based experiences in Biosphere Reserves using Flickr: Matches and mismatches between online social surveys and photo content analysis. Sci Total Environ 737:. https://doi.org/10.1016/j.scitotenv.2020.140067

Mullen LA, Bratt J (2018) USAboundaries: Historical and Contemporary Boundaries of the United States of America. J Open Source Softw 3:314. https://doi.org/10.21105/joss.00314

Muñoz L, Hausner VH, Runge C, et al (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

Naidoo R, Balmford A, Costanza R, et al (2008) Global mapping of ecosystem services and conservation priorities. Proc Natl Acad Sci U S A 105:9495–9500. https://doi.org/10.1073/pnas.0707823105

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

Noor T, Batool N, Mazhar R, Ilyas N (2015) Effects of Siltation, Temperature and Salinity on Mangrove Plants. Eur Acad Res 2:14172–14179

Norrman J, Ericsson LO, Nilsson KL, et al (2021) Mapping subsurface qualities for planning purposes : a pilot study. https://doi.org/10.1088/1755-1315/703/1/012011

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

Oakes MP, Farrow M (2007) Use of the chi-squared test to examine vocabulary differences in English language corpora representing seven different countries. Lit Linguist Comput 22:85–99. https://doi.org/10.1093/llc/fql044

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

Oteros-Rozas E, Martín-López B, Fagerholm N, et al (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

Pagnuco IA, Pastore JI, Abras G, et al (2016) Analysis of genetic association in Listeria and Diabetes using Hierarchical Clustering and Silhouette Index. J Phys Conf Ser 705:. https://doi.org/10.1088/1742-6596/705/1/012002

Palacios-Agundez I, Onaindia M, Barraqueta P, Madariaga I (2015) Provisioning ecosystem services supply and demand: The role of landscape management to reinforce supply and promote synergies with other ecosystem services. Land use policy 47:145–155. https://doi.org/10.1016/j.landusepol.2015.03.012

Palomo I, Felipe-Lucia MR, Bennett EM, et al (2016) Disentangling the Pathways and Effects of Ecosystem Service Co-Production, 1st edn. Elsevier Ltd.

Park A, Conway M, Chen AT (2018) Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. Comput Human Behav 78:98–112. https://doi.org/10.1016/j.chb.2017.09.001

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

Pastorella F, Giacovelli G, De Meo I, Paletto A (2017) People’s preferences for Alpine forest landscapes: Results of an internet-based survey. J For Res 22:36–43. https://doi.org/10.1080/13416979.2017.1279708

Penã-Aguilera P, Burguillo-Madrid L, Barve V, et al (2019) Niche segregation in Iberian Argiope species. J Arachnol 47:37–44. https://doi.org/10.1636/0161-8202-47.1.37

Pepper M, Doughty P, Keogh JS (2013) Geodiversity and endemism in the iconic Australian Pilbara region: A review of landscape evolution and biotic response in an ancient refugium. J Biogeogr 40:1225–1239. https://doi.org/10.1111/jbi.12080

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

Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. Ecol Lett 5:402–411. https://doi.org/10.1046/j.1461-0248.2002.00339.x

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

Plieninger T, Dijks S, Oteros-rozas E, Bieling C (2013) Land Use Policy Assessing , mapping , and quantifying cultural ecosystem services at community level. Land use policy 33:118–129. https://doi.org/10.1016/j.landusepol.2012.12.013

Plumwood V (2006) the Concept of a Cultural Landscape: Nature, Culture and Agency in the Land . Ethics Environ 11:115–150. https://doi.org/10.2979/ete.2006.11.2.115

Ponette-González AG, Brauman KA, Marín-Spiotta E, et al (2015) Managing water services in tropical regions: From land cover proxies to hydrologic fluxes. Ambio 44:367–375. https://doi.org/10.1007/s13280-014-0578-8

Potschin-Young M, Haines-Young R, Görg C, et al (2018) Understanding the role of conceptual frameworks: Reading the ecosystem service cascade. Ecosyst Serv 29:428–440. https://doi.org/10.1016/j.ecoser.2017.05.015

Potschin MB, Haines-Young RH (2011) Ecosystem services: Exploring a geographical perspective. Prog Phys Geogr 35:575–594. https://doi.org/10.1177/0309133311423172

Prosser CD (2013) Our rich and varied geoconservation portfolio: The foundation for the future. Proc Geol Assoc 124:568–580. https://doi.org/10.1016/j.pgeola.2012.06.001

Prosser CD (2008) The history of geoconservation in England: Legislative and policy milestones. Geol Soc Spec Publ 300:113–122. https://doi.org/10.1144/SP300.9

Prosser CD, Burek C V., Evans DH, et al (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

Quesada-Román A, Pérez-Umaña D (2020) State of the art of geodiversity, geoconservation, and geotourism in Costa Rica. Geosci 10:1–17. https://doi.org/10.3390/geosciences10060211

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

Raymond CM, Giusti M, Barthel S (2018) An embodied perspective on the co-production of cultural ecosystem services: toward embodied ecosystems. J Environ Plan Manag 61:778–799. https://doi.org/10.1080/09640568.2017.1312300

Reilly KH, Bennett EM, Adamowski JF, Hickey GM (2021) Reducing Nutrient Loading from Agriculture to Lake Ecosystems – Contributions of Resilience Principles. In: Water Resilience. Springer International Publishing, Cham, pp 91–111

Retka J, Jepson P, Ladle RJ, et al (2019) Assessing cultural ecosystem services of a large marine protected area through social media photographs. Ocean Coast Manag 176:40–48. https://doi.org/10.1016/j.ocecoaman.2019.04.018

Reverte FC, Garcia M da GM, Brilha J, Pellejero AU (2020) Assessment of impacts on ecosystem services provided by geodiversity in highly urbanised areas: A case study of the Taubaté Basin, Brazil. Environ Sci Policy 112:91–106. https://doi.org/10.1016/j.envsci.2020.05.015

Reynaud A, Lanzanova D (2017) A Global Meta-Analysis of the Value of Ecosystem Services Provided by Lakes. Ecol Econ 137:184–194. https://doi.org/10.1016/j.ecolecon.2017.03.001

Richards DR, Friess DA (2015) A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. Ecol Indic 53:187–195. https://doi.org/10.1016/j.ecolind.2015.01.034

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

Rinker TW (2015) entity: Named entity recognition version 0.1.0.

Rivera I (2019) RedditExtractoR: Reddit Data Extraction Toolkit. R package version 2.1.5.

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

Rosário IT, Rebelo R, Cardoso P, et al (2019) Can geocaching be an indicator of cultural ecosystem services? The case of the montado savannah-like landscape. Ecol Indic 99:375–386. https://doi.org/10.1016/j.ecolind.2018.12.003

Roucoux KH, Lawson IT, Baker TR, et al (2017) Threats to intact tropical peatlands and opportunities for their conservation. Conserv Biol 31:1283–1292. https://doi.org/10.1111/cobi.12925

Roux DJ, Smith MKS, Smit IPJ, et al (2020) Cultural ecosystem services as complex outcomes of people–nature interactions in protected areas. Ecosyst Serv 43:101111. https://doi.org/10.1016/j.ecoser.2020.101111

Ruban D (2014) Geodiversity. In: G T, T M, P C (eds) Encyclopedia of Mineral and Energy Policy. Springer, Berlin (Heidelberg), pp 1–2

Ruban DA (2017) Geodiversity as a precious national resource: A note on the role of geoparks. Resour Policy 53:103–108. https://doi.org/10.1016/j.resourpol.2017.06.007

Ruban DA (2010) Quantification of geodiversity and its loss. Proc Geol Assoc 121:326–333. https://doi.org/10.1016/j.pgeola.2010.07.002

Sankey TT (2017) Statistical Descriptions of Spatial Patterns. In: Encyclopedia of GIS. Springer International Publishing, Cham, pp 2196–2204

Schirpke U, Meisch C, Marsoner T, Tappeiner U (2018a) 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

Schirpke U, Scolozzi R, Da Re R, et al (2018b) Recreational ecosystem services in protected areas: A survey of visitors to Natura 2000 sites in Italy. J Outdoor Recreat Tour 21:39–50. https://doi.org/10.1016/j.jort.2018.01.003

Schirpke U, Timmermann F, Tappeiner U, Tasser E (2016) Cultural ecosystem services of mountain regions: Modelling the aesthetic value. Ecol Indic 69:78–90. https://doi.org/10.1016/j.ecolind.2016.04.001

Schmidt MWI, Torn MS, Abiven S, et al (2011) Persistence of soil organic matter as an ecosystem property. Nature 478:49–56. https://doi.org/10.1038/nature10386

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. Geogr Helv 62:140–147. https://doi.org/10.5194/gh-62-140-2007

Sharp R, Douglass J, Wolny S, et al (2020) InVEST 3.9.0.post32+ug.gbbc6b72 User's Guide.

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

Sinclair M, Ghermandi A, Moses SA, Joseph S (2019) Recreation and environmental quality of tropical wetlands: A social media based spatial analysis. Tour Manag 71:179–186. https://doi.org/10.1016/j.tourman.2018.10.018

Sinclair M, Mayer M, Woltering M, Ghermandi A (2020a) Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. J Environ Manage 263:110418. https://doi.org/10.1016/j.jenvman.2020.110418

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

Smale DA, Burrows MT, Moore P, et al (2013) Threats and knowledge gaps for ecosystem services provided by kelp forests: A northeast Atlantic perspective. Ecol Evol 3:4016–4038. https://doi.org/10.1002/ece3.774

Small N, Munday M, Durance I (2017) The challenge of valuing ecosystem services that have no material benefits. Glob Environ Chang 44:57–67. https://doi.org/10.1016/j.gloenvcha.2017.03.005

Sorichetta A, Ballabio C, Masetti M, et al (2013) A comparison of data-driven groundwater vulnerability assessment methods. Groundwater 51:866–879. https://doi.org/10.1111/gwat.12012

Spake R, Bellamy C, Graham LJ, et al (2019) An analytical framework for spatially targeted management of natural capital. Nat Sustain 2:90–97. https://doi.org/10.1038/s41893-019-0223-4

Spake R, Lasseur R, Crouzat E, et al (2017) Unpacking ecosystem service bundles: Towards predictive mapping of synergies and trade-offs between ecosystem services. Glob Environ Chang 47:37–50. https://doi.org/10.1016/j.gloenvcha.2017.08.004

Spangenberg JH, von Haaren C, Settele J (2014) The ecosystem service cascade: Further developing the metaphor. Integrating societal processes to accommodate social processes and planning, and the case of bioenergy. Ecol Econ 104:22–32. https://doi.org/10.1016/j.ecolecon.2014.04.025

Stephens PA, Pettorelli N, Barlow J, et al (2015) Management by proxy? The use of indices in applied ecology. J Appl Ecol 52:1–6. https://doi.org/10.1111/1365-2664.12383

Sun W, Shao Q, Liu J, Zhai J (2014) Assessing the effects of land use and topography on soil erosion on the Loess Plateau in China. Catena 121:151–163. https://doi.org/10.1016/j.catena.2014.05.009

Sun Y, Du Y, Wang Y, Zhuang L (2017) Examining associations of environmental characteristics with recreational cycling behaviour by street-level strava data. Int J Environ Res Public Health 14:. https://doi.org/10.3390/ijerph14060644

Swift MJ, Izac AMN, Van Noordwijk M (2004) Biodiversity and ecosystem services in agricultural landscapes - Are we asking the right questions? Agric Ecosyst Environ 104:113–134. https://doi.org/10.1016/j.agee.2004.01.013

Tallis H, Kareiva P, Marvier M, Chang A (2008) An ecosystem services framework to support both practical conservation and economic development. Proc Natl Acad Sci 105:9457–9464. https://doi.org/10.1073/pnas.0705797105

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

Tenkanen H, Di Minin E, Heikinheimo V, et al (2017) Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci Rep 7:1–11.

https://doi.org/10.1038/s41598-017-18007-4

Thiagarajah J, Wong SKM, Richards DR, Friess DA (2015) Historical and contemporary cultural ecosystem service values in the rapidly urbanizing city state of Singapore. Ambio 44:666–677. https://doi.org/10.1007/s13280-015-0647-7

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

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

Tieskens KF, Schulp CJE, Levers C, et al (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

Tieskens KF, Van Zanten BT, Schulp CJE, Verburg PH (2018) Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. Landsc Urban Plan 177:128–137. https://doi.org/10.1016/j.landurbplan.2018.05.002

UNEP-WCMC and IUCN (2021) Protected Planet: The World Database on Protected Areas (WDPA)

Van Berkel DB, Tabrizian P, Dorning MA, et al (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

Van Berkel DB, Verburg PH (2014) Spatial quantification and valuation of cultural ecosystem services in an agricultural landscape. Ecol Indic 37:163–174. https://doi.org/10.1016/j.ecolind.2012.06.025

van der Meulen ES, Braat LC, Brils JM (2016) Abiotic flows should be inherent part of ecosystem services classification. Ecosyst Serv 19:1–5. https://doi.org/10.1016/j.ecoser.2016.03.007

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 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 Zanten BT, Van Berkel DB, Meentemeyer RK, et al (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

Vaz AS, Gonçalves JF, Pereira P, et al (2019) Earth observation and social media: Evaluating the spatiotemporal contribution of non-native trees to cultural ecosystem services. Remote Sens Environ 230:111193. https://doi.org/10.1016/j.rse.2019.05.012

Verburg PH, Koomen E, Hilferink M, et al (2012) An assessment of the impact of climate adaptation measures to reduce flood risk on ecosystem services. Landsc Ecol 27:473–486. https://doi.org/10.1007/s10980-012-9715-6

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

Völske M, Potthast M, Syed S, Stein B (2018) TL;DR: Mining Reddit to Learn Automatic Summarization. 59–63. https://doi.org/10.18653/v1/w17-4508

Walden-Schreiner C, Rossi SD, Barros A, et al (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

Wang X, Xu Y (2019) An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. IOP Conf Ser Mater Sci Eng 569:. https://doi.org/10.1088/1757-899X/569/5/052024

Wang Z, Ye X, Tsou MH (2016) Spatial, temporal, and content analysis of Twitter for wildfire hazards. Nat Hazards 83:523–540. https://doi.org/10.1007/s11069-016-2329-6

Wetzel FT, Bingham HC, Groom Q, et al (2018) Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe. Biol Conserv 221:78–85. https://doi.org/10.1016/j.biocon.2017.12.024

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

Wild F (2015) lsa: Latent Semantic Analysis. R package version 0.73.1.

Willcock S, Hooftman DAP, Balbi S, et al (2019) A Continental-Scale Validation of Ecosystem Service Models. Ecosystems 22:1902–1917. https://doi.org/10.1007/s10021-019-00380-y

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

Wittich HC, Seeland M, Wäldchen J, et al (2018) Recommending plant taxa for supporting on-site species identification. BMC Bioinformatics 19:1–17. https://doi.org/10.1186/s12859-018-2201-7

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 87:106544. https://doi.org/10.1016/j.eiar.2020.106544

Xie G, Lundholm JT, MacIvor SJ (2018) Phylogenetic diversity and plant trait composition predict multiple ecosystem functions in green roofs. Sci Total Environ 628–629:1017–1026. https://doi.org/10.1016/j.scitotenv.2018.02.093

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

Yoshimura N, Hiura T (2017) Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. Ecosyst Serv 24:68–78. https://doi.org/10.1016/j.ecoser.2017.02.009

Zarnetske PL, Read QD, Record S, et al (2019) Towards connecting biodiversity and geodiversity across scales with satellite remote sensing. Glob Ecol Biogeogr 28:548–556. https://doi.org/10.1111/geb.12887

Zhang H, Huang R, Zhang Y, Buhalis D (2020) Cultural ecosystem services evaluation using geolocated social media data: a review. Tour Geogr 0:1–23. https://doi.org/10.1080/14616688.2020.1801828

Zhang S, Zhou W (2018) Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. Landsc Urban Plan 180:27–35. https://doi.org/10.1016/j.landurbplan.2018.08.004

Zhao C, Sander HA (2018) Assessing the sensitivity of urban ecosystem service maps to input spatial data resolution and method choice. Landsc Urban Plan 175:11–22. https://doi.org/10.1016/j.landurbplan.2018.03.007

Zwoliński Z (2009) The routine of landform geodiversity map design for the Polish Carpathian Mts. Landf Anal 11:77–85