UNIVERSITY OF BIRMINGHAM

University of Birmingham Research at Birmingham

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

Fox, Nathan; Graham, Laura; Eigenbrod, Felix; Bullock, James M.; Parks, Katherine E.

Link to publication on Research at Birmingham portal

1   **Enriching social media data allows a more robust representation of cultural ecosystem services**

2   **Abstract**

3   Images and textual metadata from social media sites such as Flickr have been used to understand

4   the drivers and distributions of cultural ecosystem services (CES). However, using all available data

5   from social media sites may not provide an accurate representation of individual services. For

6   example, an image of nature might be described negatively in the image's description. Here, we

7   present a novel approach to refining social media data to represent CES better, including filtering by

8   keywords, photograph content and enriching the data by including a measure of the sentiment

9   expressed in the textual metadata. We demonstrate that the distribution of an enriched dataset of

10   Flickr images representing hiking in the USA can contribute to different results and conclusions than

11   the full dataset. Furthermore, we classified the contents of these hiking images and, using latent

12   semantic analysis, clustered the images into ten groups based on the similarity of their content. The

13   groups provide rich information, such as the importance of geodiversity and biodiversity in

14   supporting a positive hiking experience. The application of this method can help to enrich social

15   media data for CES studies, allowing researchers to further untangle the complex socio-ecological

16   interactions that drive CES distributions, benefits and values.

17   **Keywords**: Cultural ecosystem services, geosystem services, social media, Flickr, hiking

18   **1 Introduction**

19   Cultural ecosystem services (CES) are non-material benefits obtained from human-nature

20   interactions including through recreation, cognitive development, aesthetic views and spiritual

21   enrichment (Millennium Ecosystem Assessment 2005; Milcu et al. 2013). Though there are many

22   different definitions and classifications used for CES, most literature focuses on the links between

23   the biophysical environment and human wellbeing, while recognising that CES are intangible (Fish et

24   al. 2016; Dickinson and Hobbs 2017). Here, we classify CES following Milcu et al., (2013), who divided

25   CES into 11 subcategories; recreation and tourism, aesthetic values, spiritual and religious values,

26   educational values, cultural heritage values, bequest, intrinsic and existence, inspiration, sense of

27   place, knowledge systems, social relations, and cultural diversity. Furthermore, we note that CES

28   benefits can be delivered through multiple pathways. King et al. (2017) identified six pathways to

29   CES benefits that reoccur across CES literature: cognitive (benefits from the development of

30   knowledge), creative (benefits from influences on aesthetic appreciation and artistic expression),

31   intuitive (benefits from the influence on instincts and senses), retrospective (benefits from reflecting

32   on past experiences), regenerative (benefits from opportunities for recreation, leisure and tourism)

33   and communicative (benefits from social relations, cultural identity, and sense of place). However,

34   for these pathways to be actualised, there is a need for people to first recognise the potential

35   benefits of biophysical features and then utilise these potential benefits (Spangenberg et al. 2014).

36   Therefore, CES, as with other ecosystem services (ES), are not provided by ecosystems

37   independently of humans but are co-produced through our interactions with them (Fischer and

38   Eastwood 2016).

39   Though the value of ES is generally provided by an economic metric, ES values can also be measured

40   by its societal and cultural values and therefore the value of CES is often non-monetary (Haines-

41   Young and Potschin, 2018; Reynaud and Lanzanova, 2017; Small, Munday and Durance, 2017). There

42   are multiple methods of assessing CES value both, monetary (e.g. travel cost or willingness to pay)

43   and non-monetary (photograph analysis or ranking methods) (Hirons et al. 2016). However,

44   quantifying the benefits and values of CES are more difficult due to the perceptions of CES benefits

45 being unique to individuals based on their social and cultural norms (Daniel et al. 2012; Havinga et
46 al. 2020). Therefore, CES have been comparably under-researched compared to other ES and we,
47 therefore, need to develop our understanding of the human-nature interactions that provide these
48 services (Milcu et al. 2013; Dickinson and Hobbs 2017).

49 Previous literature assessing CES relationships, as well as wider ES relationships, tend to focus on the
50 role of nature on CES production (Fischer and Eastwood 2016). Because people, culture and nature
51 are so inherently interlinked, it can be difficult to disentangle what constitutes nature in the context
52 of ES (Plumwood 2006; Hirons et al. 2016). Here, we view nature as the biophysical features of an
53 ecosystem (Haines-Young and Potschin 2018), comprised of the interactions of biodiversity and
54 geodiversity (Gray 2012; Gordon and Barron 2013; Fox et al. 2020a). There has been a particular
55 focus on the role of biodiversity in CES production, with the role of geodiversity often omitted from
56 studies (Fox et al. 2020a). Geodiversity can be viewed as the abiotic equivalent to biodiversity,
57 representing the diversity of geological structures and processes, including rocks and minerals;
58 geomorphology, including landforms and topography; sediments and soils, including formation
59 processes; and hydrology, including marine, surface and subsurface waters (Gray 2004; Hjort et al.
60 2015; Fox et al. 2020a). Geodiversity can also provide CES in the absence of biodiversity. These
61 "geosystem services" include CES such as recreational activities (e.g. water-based sports, rock
62 climbing and caving), spiritual sites (e.g. Uluru, Australia and the Torres del Paine, Chile), as well as
63 providing opportunities for advancing scientific knowledge (e.g. the record of past climates and
64 ecosystems contained in sediment, rock and ice cores) (Gray 2012; Kiernan 2014; Fox et al. 2020a).
65 Often geodiversity is only assessed through landscapes types which are prescribed a general CES
66 value, however, CES are not distributed randomly within a landscape, but are concentrated in
67 hotspots that have specific features of biodiversity (e.g. forests and hedgerows) and geodiversity
68 (waterbodies and geological formations) (Plieninger et al. 2013; van Berkel and Verburg 2014). It is
69 therefore important that the relationship of the individual features of biodiversity and geodiversity
70 to CES be assessed.

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

84 To understand CES there is a need to for suitable datasets that can assess the complex relationship
85 between biodiversity, geodiversity and society. However, globally, datasets on ES are sparse,
86 meaning services are often mapped through proxies (Stephens et al. 2015), with the primary sources
87 of data for CES mostly coming from either monetary assessments, social surveys such as stated
88 preferences, or onsite surveys (Tenerelli et al. 2016; Figueroa-Alfaro and Tang 2017; Mayer &
89 Woltering, 2018). Due to labour-intensive methods and high financial costs, implementing these
90 over large spatial and time scales is not always feasible (Wood et al. 2013; Kim et al. 2019).

91 Furthermore, management decisions need to be better informed through methods that reliably
92 understand, identify, quantify and map CES (Tenerelli et al. 2016; Byczek et al. 2018). As CES are
93 dynamic, methods also need to be able to reliably investigate changes over time (Figieroa-Alfaro and
94 Tang 2017).

95 The potential of social media sites such as Flickr, Twitter and Facebook as a source of data for CES
96 questions is starting to be realised (Kim et al. 2019). In contrast to social surveys, social media data is
97 inexpensive, quick to gather and provides a means of mapping the distribution of CES and assessing
98 changes over space and time (Fox et al. 2020b). Social media data has been used for CES studies,
99 such as wildlife watching (Mancini et al. 2019), recreational services (Graham and Eigenbrod 2019;
100 Sinclair et al. 2020a), aesthetic views (Van Berkel et al. 2018) and visitation rates in protected areas
101 (Tenkanen et al. 2017, Kim et al. 2019; Sinclair et al. 2020b), providing key information for both
102 tourism and conservation. However, social media data is often messy (Ghermandi and Sinclair 2019;
103 Chen et al. 2020). Issues such as unknown or inaccurate spatial references (Figieroa-Alfaro and Tang
104 2017), unreliable image contents due to mistagged images or a mismatch between the content of a
105 photograph and the location it was taken (Oteros-Rozas et al. 2018), and biases introduced by user
106 groups (Langemeyer et al. 2018; Chen et al. 2020) therefore need to be accounted for.

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

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

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

137 used the Google Cloud Vision API to label the contents of Flickr images and subsequently used this
138 information to map the distribution of plants and animals, whilst Gosal et al. (2019) labelled
139 photograph contents using the Google Cloud Vision API to find groups of recreational beneficiaries.

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

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

168 Though the position of recreation as a final service has been questioned (Haines-Young and
169 Potschin, 2018), recreational activities provide restorative benefits (e.g. increased physical
170 wellbeing) and are therefore generally considered a CES (Millennium Ecosystem Assessment 2005;
171 Milcu et al. 2012; Plieninger et al. 2013; King et al. 2017; Balzan and Debono 2018). In this article we
172 will focus on hiking, a recreational activity that involves walking over an extended period, typically
173 through natural or rural areas (Mitten et al. 2016). We consider hiking a CES here as it can directly
174 provide restorative benefits, however, as the benefits from recreation may also be indirect (Balzan
175 and Debono 2018), we recognize hiking could also be considered as a pathway to other CES (King et
176 al. 2017) through co-production between human-nature interactions (Fischer and Eastwood 2016).
177 For example, hiking may be undertaken for spiritual and religious motivations such as a pilgrimage,
178 to experience aesthetic qualities of nature, or as a social activity with a sense of belonging (Collins-
179 Kreiner and Kliot 2017; Wilcer et al. 2019). From this viewpoint, hiking can also provide multiple
180 pathways to other CES benefits, for example, providing a means of access to aesthetic views
181 (creative pathway) or providing a sense of place (communicative pathway) (King et al. 2017). Hiking
182 is one of the most popular recreational activities, both in the USA and worldwide, with participants

183 from all age categories (Wilcer et al. 2019). Despite its popularity, research on the drivers of hiking
184 remains limited (Wilcer et al. 2019). Hiking, as with other ES, is driven by co-production between
185 ecosystems and people, such as the influences of cultural landmarks and landscapes or through the
186 provision of infrastructure such as signposted trails (Plieninger et al. 2013; Fischer and Eastwood
187 2016).

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

194 **2 Methods**

195 *2.1 Data collection*

196 *A reproducible R file for the data collection methods has been included in the supplementary material*
197 *(S1). To comply with API terms and privacy policies all data sets were anonymised, stored with*
198 *multiple layers of security and any unnecessary metadata was deleted.*

199 We queried the Flickr API for photographs containing the text "hiking" in a photograph's title,
200 description or tag metadata, in the contiguous 48 states of the USA, between 2015-01-01 and 2020-
201 01-01. To ensure consistency and reproducibility in the study we used the photosearcher package
202 (Fox et al. 2020b) within the R environment (R Core Team 2020). The photosearcher package allows
203 searches of Flickr to be constrained by a shapefile. Here the search was limited to photographs taken
204 and geotagged in the contiguous 48 states using a modified shapefile from the USAboundaries R
205 package (Mullen and Bratt 2018). As the number of posts meeting these criteria can change over
206 time (e.g. new photographs were taken during the study period but uploaded at a later date), we
207 limited photographs to those uploaded before 2020-06-01 to increase the reproducibility of the
208 search.

209 *2.2 Content analysis*

210 To ensure an image captures human-nature interactions, we categorized the contents of images to
211 help filter out images that do not contain any biophysical features in their contents (e.g. indoor
212 images). Here, the features within each image were automatically labelled using the Google Cloud
213 Vision API (Google Cloud Vision 2020), accessed through the imgrec R package (Schwemmer 2019).
214 The Google Cloud Vision API is a pre-trained machine learning model that can detect image contents
215 including objects, faces and text. For each image, we requested the API to identify and label the 10
216 most dominant features of the image. The generated labels had an associated confidence score
217 between 0 and 1. Here we only kept labels with a confidence score of >0.6 (Gosal et al. 2019). We
218 assessed each unique label identified by the Google Cloud Vision API and categorized them as either
219 biophysical nature or not. Here, we identified non-biophysical nature words as synthetic objects (e.g.
220 buildings and cars), relating to people (e.g. a person), and descriptive terms (e.g. black, text), as well
221 as non-biophysical aspects of nature such as weather phenomenon (e.g. sky and sun). Biophysical
222 nature labels were any label that was a feature of biodiversity (e.g. tree or bird of prey), geodiversity
223 (e.g. lake or geology), or an ecosystem (e.g. rainforest or grassland). Furthermore, we included
224 generic descriptions of landscapes (e.g. wilderness and natural landscape) as biophysical nature.
225 Supporting Information (Tables S2.1, S2.2) provides a full list of how each word was categorized. For
226 each image, we calculated the percentage of labels categorized as a biophysical feature of nature.

227 *2.3 Sentiment analysis*

228 To ensure an image captures a positive human-nature interaction, and therefore a CES, we calculate
229 the sentiment value expressed in the textual metadata. Here, we used the AFINN dictionary (Nielsen
230 2011) to summarise the sentiment of the textual metadata. The dictionary ranks words on a scale of
231 -5 (the most negative words) to +5, (the most positive words), and can be an effective method for
232 assessing social media datasets (Koto and Adriani 2015). For each image, the associated textual
233 metadata was analysed using the AFINN dictionary and the overall sentiment value was calculated as
234 the sum of all the positive and negative sentiment scores for that image. Each image was then
235 categorized into two groups, (1) positive images - those with an overall positive sentiment score, (2)
236 non-positive images - those with an overall negative sentiment score, an overall neutral sentiment
237 score or where no sentiment was expressed in the textual metadata.

238 *2.4 Impacts of refining the data*

239 To understand the ramifications of refining data by contents and sentiment, we carried out two
240 filtering processes: first selecting all images where the percentage of labels classified as biophysical
241 nature exceeded a given threshold (e.g. 25% of the Google Vision Cloud API labels were of
242 biophysical nature features), and second filtering only images with positive sentiment to ensure
243 images were associated with a positive experience (Fig. 1). The final part of the analysis applied both
244 filters, resulting in a final dataset representing images of human-natural interactions *AND* reflecting
245 a positive experience. We extracted a random sample of 100 images from each category
246 (*confirmedNature, nonNature, confirmedPositive, nonpositive*) to manually validate the automated
247 process.
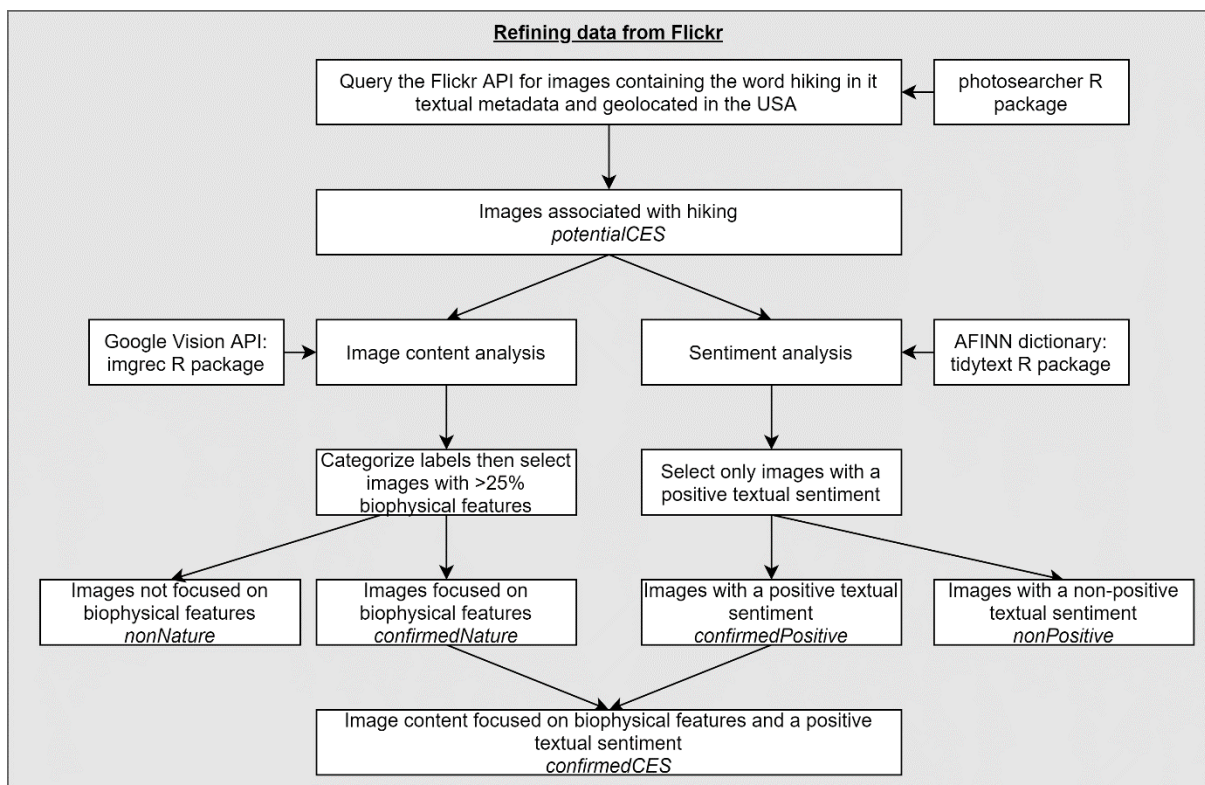
248

249 Figure 1. Processes applied to refining the Flickr dataset.

250 First, the full Flickr dataset was refined to images that were *confirmedNature* (thus removing any
251 images that, whilst including the relevant activity in the associated textual data, were not taken

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

270   As landscape characteristics can drive recreational activities at a range of scales, and Flickr is a good
271   proxy for recreation at a range of scales up to 50km (Graham and Eigenbrod 2019), here we map the
272   difference in distribution when the number of uploads is aggregated to 25km^2. We also map the
273   differences in distribution using a 3 x 3-pixel window (where a pixel is the size of one pixel of the
274   underlying raster as determined by the spatial resolution), a standard size for aggregating fine-scale
275   data (Graham et al., 2019). Furthermore, we map the differences based on refining by a threshold of
276   25% biophysical labels. This threshold was chosen to ensure that the refined set of images captured
277   human-nature interactions, without completely excluding images containing human features as
278   these could help to provide insight into the co-production of CES. However, as any changes in
279   distribution may be a facet of the filtering method, or the method of mapping and calculating the
280   deviance, we conducted a sensitivity analysis using all possible combinations of three thresholds for
281   considering an image to be of nature (images with 25%, 50% and 75% labels classed as biophysical
282   nature), three different spatial resolutions (5km^2, 10km^2 and 25km^2), and three sizes of moving
283   window (3 x 3 pixels, 5 x 5 pixels and 7 x 7 pixels). This resulted in 27 datasets.
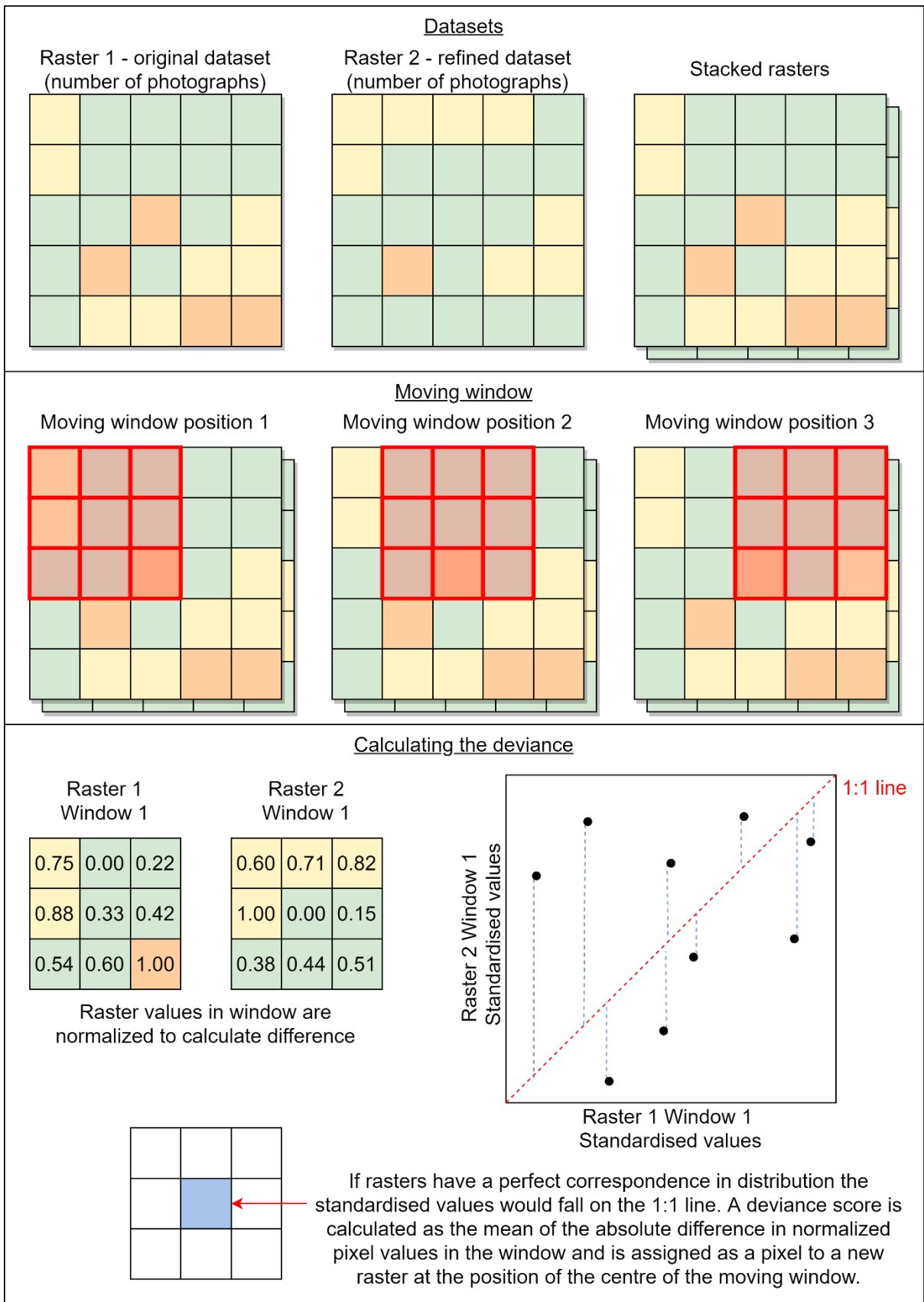
Figure 2. Methods for calculating the difference in distributions between the unrefined and refined datasets.

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

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

302 Finally, we refined the original dataset to images containing biophysical features and positive textual
303 sentiment (thus capturing CES as images of positive experiences occurring within a natural, or semi-
304 natural setting). Here, we mapped the distribution of these *confirmedCES* images versus the full
305 number of images and calculated the local deviance values of the two datasets, again using all the 27
306 possible combinations of three thresholds for considering an image to be of nature (images with
307 25%, 50% and 75% labels classed as nature), three different spatial resolutions ($5km^2$, $10km^2$ and
308 $25km^2$), and three sizes of moving window (3 x 3 pixels, 5 x 5 pixels and 7 x 7 pixels). Here, if the
309 original full dataset was a good proxy for CES we would expect deviance values to be closer to 0
310 between the *confirmedCES* images versus all the images. We also tested the spatial uniformity of the
311 local deviance values using Moran's I.

312 *2.5 Mapping distributions and sentiment*

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

320 *2.6 Assessing human-nature interaction in images*

321 To assess the relationship between biodiversity and geodiversity we carried out latent semantic
322 analysis (LSA) on the content labels (Fig. 3). LSA was carried out in the R environment, primarily
323 using the lsa R package (Wild 2015). LSA is a natural language processing technique that is used to
324 assess the relationship between a collection of documents (in this case a user's photographs from
325 one day) and the terms used in them (in this case Google Vision Cloud API labels) as a term-
326 document matrix (TDM) and can be used to examine how closely terms are related in use (Gefen et
327 al. 2017). LSA on a TDM has been previously used to help describe recreational activities (Monkman
328 et al. 2018; Gosal et al. 2019). Here, where the LSA shows terms are more closely related, this
329 indicates that those Google Vision Cloud API labels are more commonly photographed together, for
330 example, one might expect the labels "forest" and "tree" to be frequently photographed together.
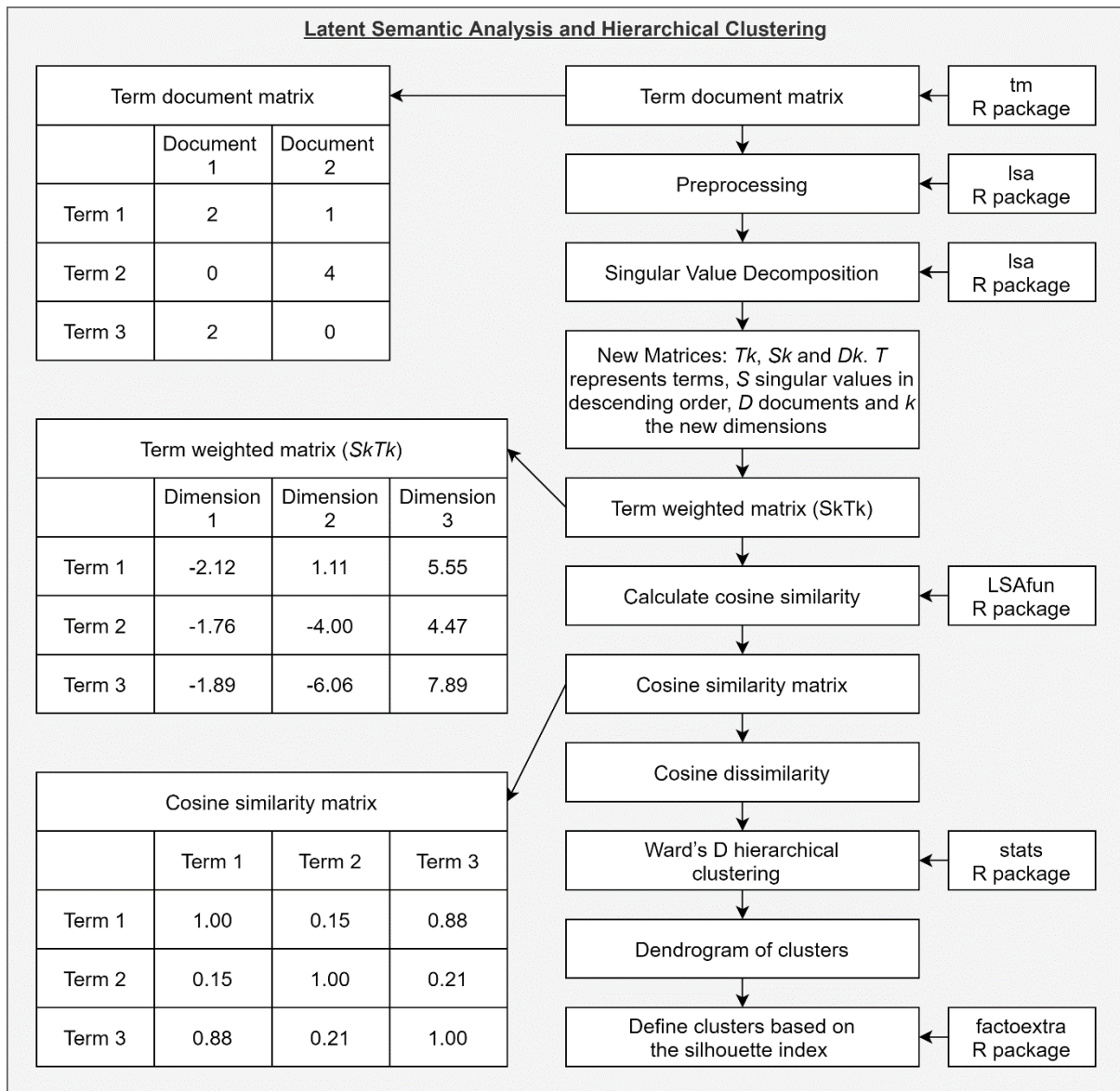
Figure 3. Example workflow for the methods of clustering photographs.

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

348 appear far more frequently than others (Evangelopoulos 2013; Christian et al. 2016; Gefen et al.
349 2017).

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

366 By carrying out the SVD, the terms can now be projected in multidimensional space and the
367 similarity between Google Vision Cloud API labels can be calculated using cosine similarity
368 (Evangelopoulos 2013). Cosine similarity measures the angle between vectors in multi-dimensional
369 space, with the resulting values ranging from 0 to 1, with 1 representing total similarity. The diagonal
370 in the matrix will always be 1 as a word is always the same as itself. Here we calculated the cosine
371 similarity between all terms. Cosine similarity of 1 means that the labels always appear together and
372 0 means they never appear together.

373 To understand what features (including biophysical and non-biophysical nature as well as human
374 features) are most often taken in the same images, we used hierarchical clustering to group the
375 Google Vision Cloud API labels based on their cosine similarity. As hierarchical clustering uses a
376 distance measure, we calculated cosine *dissimilarity* (1 - cosine similarity). We then carried out
377 hierarchical clustering using Ward's D method, which has previously been shown to create
378 unambiguous clusters for labels generated by the Google Cloud Vision API (Gosal et al. 2019). The
379 Ward's D method builds a dendrogram through a bottom-up approach to clustering. Each element in
380 the tree starts as an individual cluster, two clusters are then merged so that variance within clusters
381 is minimized. This process is repeated until all elements are clustered on the tree.

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

390 We summarised the types of images photographed when hiking based on the clusters. For each
391 user's daily images, we calculated the number of Google Vision Cloud API labels belonging to each
392 cluster then categorised the users daily images as which cluster was the most dominant (where

393 labels of two or more clusters appear equally dominant in an image, that image was classified as a
394 combination of those clusters). The mean sentiment of the images belonging to each cluster or
395 combination of clusters was calculated.

396 **3 Results**

397 *3.1 Full and enriched datasets*

398 There were 179,700 geotagged photographs containing the text "hiking" in a photograph's title,
399 description or tag metadata for the years 2015-2020 in the USA. The distribution of hiking images
400 was largely concentrated in the west, the northeast, and some southern states (Fig. 4a). Here, the
401 final refined dataset chosen to represent *confirmedCES* were images containing more than 25% of
402 the image content labels being classified as nature and a positive sentiment expressed in textual
403 metadata. There were 43,427 images that met these selection criteria, 24.17% of the full hiking
404 dataset. The distribution of these *confirmedCES* images follows some similar patterns to the full
405 dataset with the largest concentration of uploads along the west coast of the US in states such as
406 Washington, southwest states, such as Arizona and Utah, northeastern states around the Great
407 Lakes and along the Appalachians (Fig. 4b). Though there are similarities in the distribution of the
408 *confirmedCES* images and the full dataset images, there are also areas where the distributions have
409 different patterns for example, particularly with many regions having fewer uploads, such as
410 southern California, Arkansas and the west coast of Florida.

411
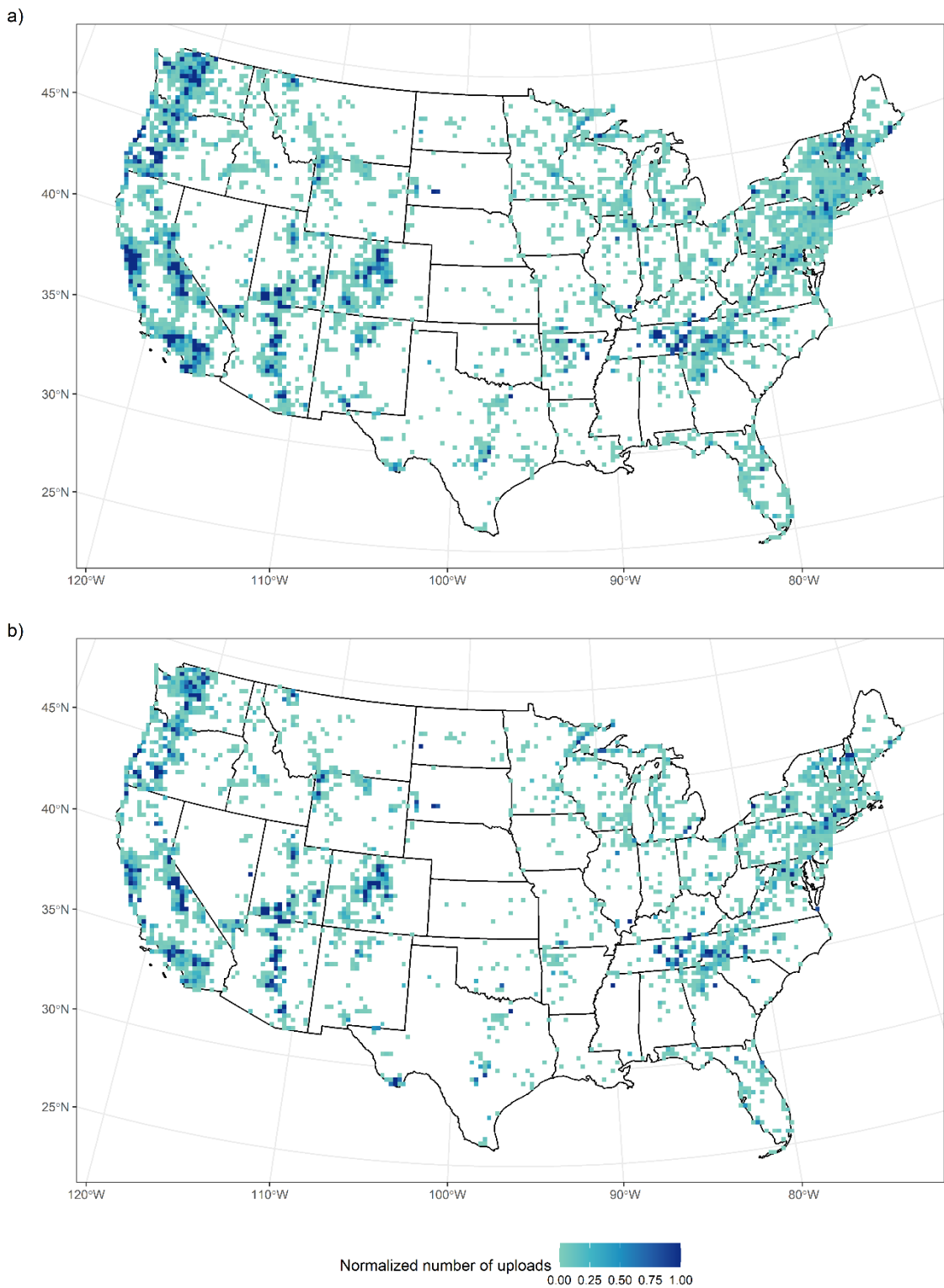
Figure 4. Number of images from Flickr in the USA between 2015-2020 with the term "hiking" in the images title, tag or description, results were normalized (spatial resolution 25km^2). a) the full Flickr dataset, b) images categorized as *confirmedCES* (images where the percentage of content labels classified 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. 5). 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.
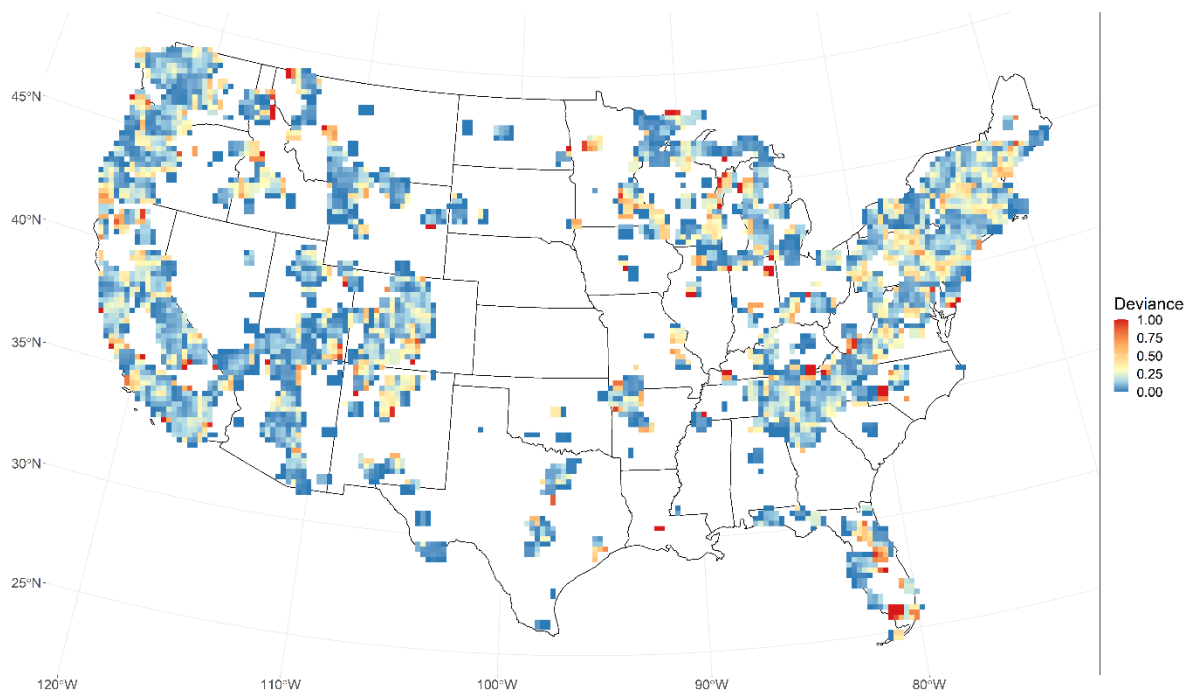
Figure 5. The local deviance in the scaled number of images from the full Flickr and *confirmedCES* images dataset (images where the percentage of content labels classified as nature were >25% and a positive sentiment was expressed in textural metadata) from the 1:1 line. Pixel size 25km^2; window size 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 (see supplementary material SI.3). 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.

445

3.2 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. 6). 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 (see supplementary material SI.4).

458

Figure 6. 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.

463

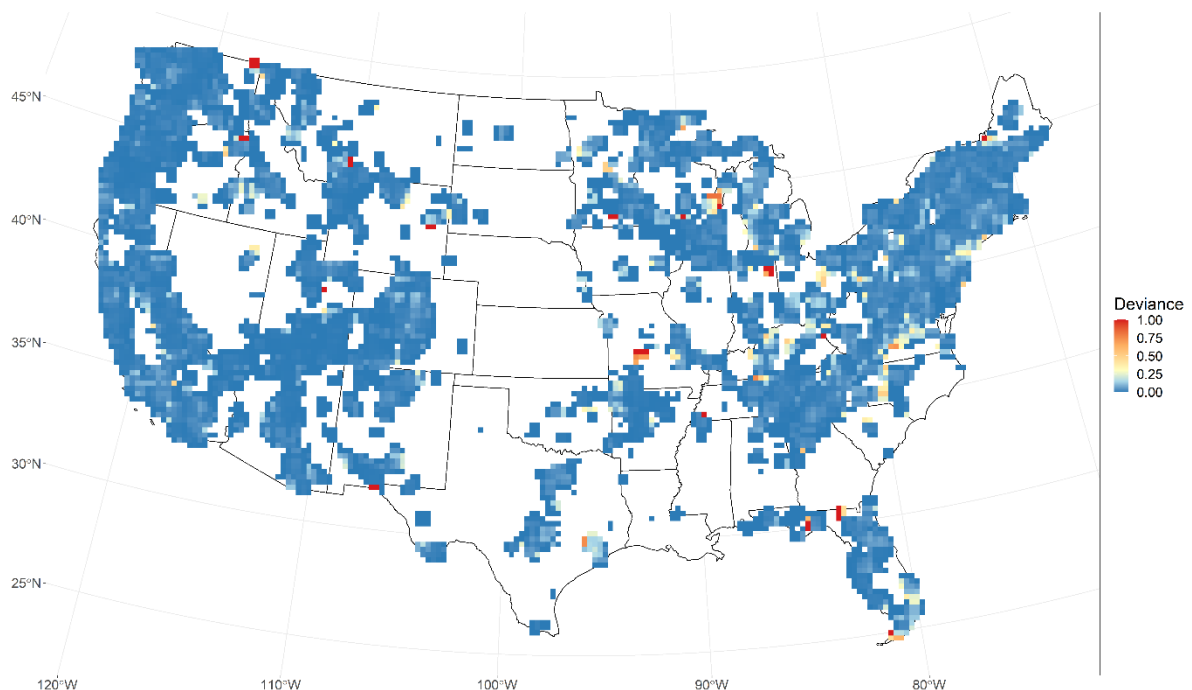*3.3 Individual influence of filtering by sentiment only*

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.

471   Here, 19.08% of the pixels have a deviance value of greater than 0.3, indicating that in these
472   locations, photographs of hiking on Flickr may not be a good proxy for positive images of hiking.
473   Furthermore, the deviation of uploads from the *confirmedPositive* and full dataset is not uniform
474   (Fig. 7), so it may not be possible to differentiate between areas where Flickr photographs are a
475   good proxy for a positive experience and those where they are not. The random structure to the
476   distribution (Moran's I = 0.27) suggests that there is no spatial structure to the distribution of the
477   locations which deviate between the full and filtered dataset. As with the previous examples, the
478   patterns of local deviance are similar across spatial resolution and window size and therefore the
479   choices in mapping the distribution have not impacted the results (see supplementary material
480   SI.5). As there was little to no difference in spatial distribution when refining by nature threshold
481   alone, but a larger difference when refining by sentiment alone, this indicates that differences
482   observed in the analysis of the *confirmedCES* versus full dataset are primarily driven by the inclusion
483   of refining by sentiment.

484

485   Figure 7. The local deviance in the scaled number of images from the full Flickr and
486   *confirmedPositive* images dataset (images where a positive sentiment was expressed in textural
487   metadata) from the 1:1 line. Pixel size 25km^2; window size 3 x 3; a break point of 0.3 was applied to
488   indicate areas of high deviance.

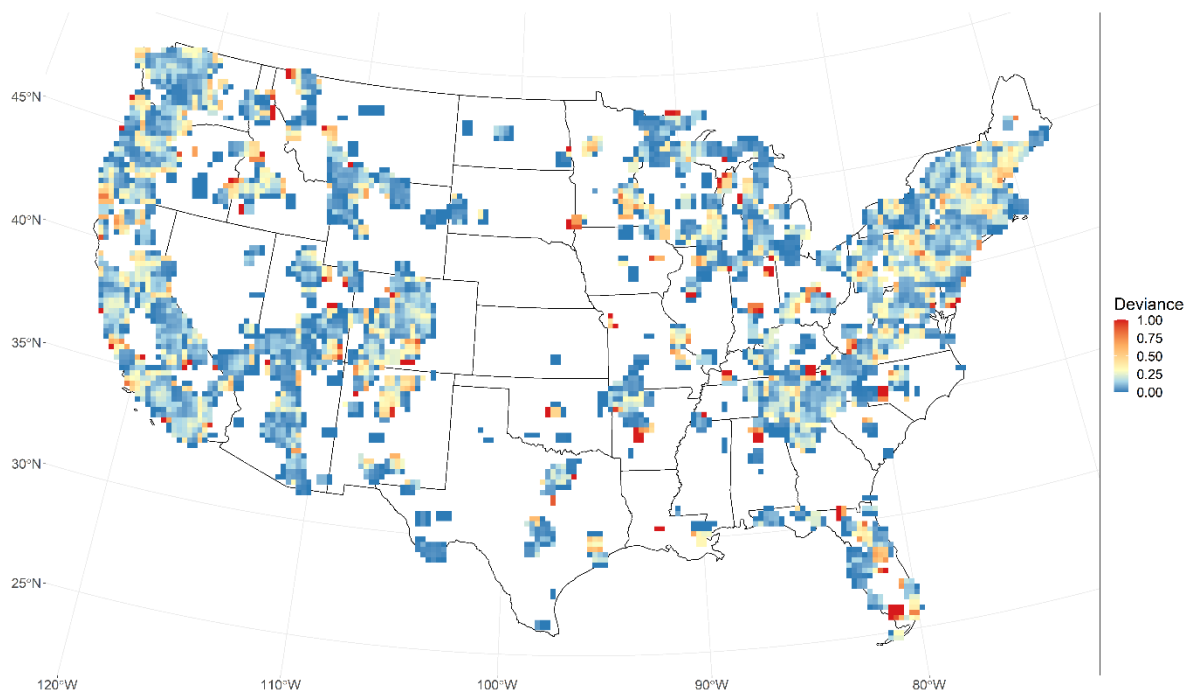489   *3.4 Validation of methods*

490   The manual validation of the image content and textual sentiment analysis indicates that the
491   automated methods have high accuracy. When assessing images classified as presenting human-
492   nature interactions we agreed with 98% of the images categorised this way at the 25% biophysical
493   nature label thresholds (*confirmedNature*): and 100% at 50% and 75% threshold (Table 1). However,
494   when assessing images deemed not to be focused on human-nature interactions (*nonNature)* the
495   method incorrectly included some images containing human-nature interactions, particularly when
496   using a threshold of 50% and 75% biophysical labels. When using a 25% threshold, the images that
497   were incorrectly labelled as *confirmedNature* were of artwork or indoors artificial water features.
498   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 1: 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.

3.5 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. 8). 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.

533
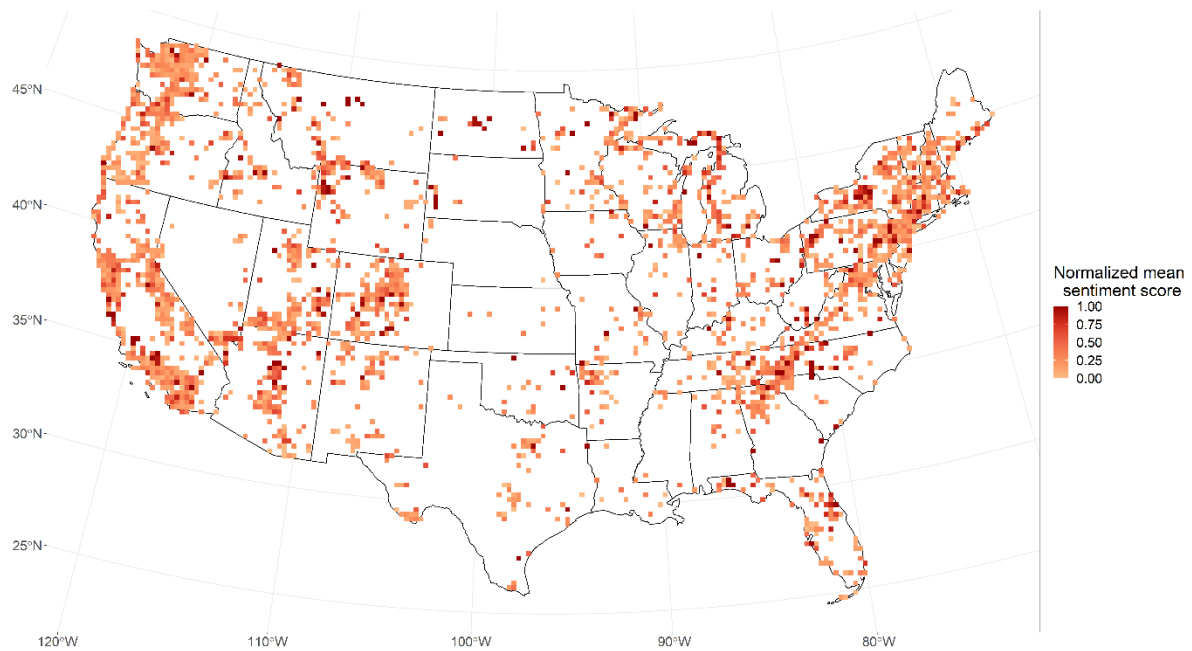
Figure 8. Mean sentiment value of *confirmedCES* images (images where the percentage of content labels classified as nature were >25% and a positive sentiment was expressed in textural metadata), results were normalized (pixel size 25km^2)

3.6 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 2). 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 2. 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

547 When categorising images based on which cluster's labels were most dominant in the photograph,
548 there were 165 different combinations of the most dominant cluster (Figure 8). The clusters that
549 were most frequently photographed were "mountains and hills", "forests", "geological or arid
550 landscapes" and "hydrological features". Many of the cluster combinations have large standard
551 deviations, though some combination of clusters (e.g. "views of lake/human-geology interactions")
552 have relatively low standard deviation values.

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

Number of images

Mountains and hills: 2750
Forests: 2705
Geological or arid landscapes: 1292
Hydrological features: 1247
Vegetated trails and parks: 618
Views of lakes: 559
Forests/Vegetated trails and parks: 330
Mountains and hills/Views of lakes: 262
Mountains and hills/Forests: 242
Mountains and hills/Geological or arid landscapes: 236
Forests/Views of lakes: 190
Flowers: 166
Views of lakes/Hydrological features: 162
Mountains and hills/Snow covered mountains: 144
Vegetated trails and parks/Flowers: 142
Snow covered mountains: 119
Mountains and hills/Forests/Views of lakes: 99
Mountains and hills/Vegetated trails and parks: 94
Forests/Hydrological features: 92
Mountains and hills/Hydrological features: 84
Mountains and hills/Forests/Vegetated trails and parks: 72
Mountains and hills/Views of lakes/Hydrological features: 63
Forests/Vegetated trails and parks/Geological or arid landscapes: 60
Forests/Vegetated trails and parks/Flowers: 57
Views of lakes/Human-geology interactions: 56

Mean sentiment of images

Forests/Vegetated trails and parks/Geological or arid landscapes: 6.35 (SD 9.39)
Mountains and hills/Views of lakes: 6.07 (SD 6.44)
Mountains and hills/Hydrological features: 6.05 (SD 5.62)
Mountains and hills/Forests/Vegetated trails and parks: 5.74 (SD 9.03)
Mountains and hills/Forests/Views of lakes: 5.73 (SD 5.37)
Hydrological features: 5.71 (SD 5.69)
Vegetated trails and parks/Flowers: 5.64 (SD 5.4)
Views of lakes/Hydrological features: 5.27 (SD 4.25)
Mountains and hills/Views of lakes/Hydrological features: 5.19 (SD 3.69)
Forests/Views of lakes: 5.14 (SD 4.48)
Views of lakes: 5.07 (SD 5.04)
Geological or arid landscapes: 5.06 (SD 4.72)
Forests/Vegetated trails and parks/Flowers: 5.05 (SD 4.63)
Mountains and hills: 4.94 (SD 5.26)
Snow covered mountains: 4.59 (SD 4.31)
Mountains and hills/Geological or arid landscapes: 4.58 (SD 3.47)
Vegetated trails and parks: 4.47 (SD 5.05)
Mountains and hills/Snow covered mountains: 4.46 (SD 5.21)
Forests: 4.38 (SD 4.35)
Forests/Hydrological features: 4.33 (SD 3.85)
Mountains and hills/Forests: 4.29 (SD 5.28)
Forests/Vegetated trails and parks: 4.29 (SD 3.88)
Flowers: 4.07 (SD 4.04)
Mountains and hills/Vegetated trails and parks: 3.56 (SD 2.79)
Views of lakes/Human-geology interactions: 3.32 (SD 1.57)

571
572 Figure 8. Count of images, their mean sentiment value and standard deviation of sentiment value for
573 the 25 most frequently photographed clusters, where a user's images from a single day were
574 categorised based on which cluster most of their assigned Google Vision Cloud API labels belonged
575 to.

4 Discussion

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

592 For the content analysis, we aimed to refine our dataset to represent images of human-nature
593 interactions, we therefore included images containing any aspects of biophysical nature. Studies
594 with different aims should filter their images accordingly, for example, Oteros-Rozas et al. (2018)
595 chose to include only images of landscapes in their study and removed images that featured animals.
596 Some CES studies have stated a preference for using images from Panoramio as it prevents uploads
597 of images of people and synthetic objects and therefore did not need to be refined (Casalegno et al.
598 2013; Pastur et al. 2016). However, as of 2016 Panoramio is no longer available. Here, we found little
599 difference in the distribution of the *confirmedNature* and full datasets, with low deviance between
600 the two groups for most of the US. This suggests that Flickr may therefore be a good proxy for
601 images of biophysical features across a range of spatial scales. However, studies that use Flickr, or
602 other social media sources, need to acknowledge that the returned dataset may contain images that
603 are unsuitable for their specific CES assessment and should ensure these images are refined out
604 accordingly. Overall, we found that a threshold of 25% of biophysical nature labels has provided a
605 suitable set of images representing human-nature interactions, while generally excluding images
606 that are not of CES. By not omitting all human features (e.g. Panoramio), a threshold of 25%
607 biophysical nature labels can provide results that allow for more robust recommendations for
608 improving the sustainable management of these complex human-nature interactions (Bennet et al.
609 2015; Palomo et al. 2016).

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

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

639 could be used in conjunction with the temporal metadata to assess changes in visitor opinions over
640 time (Becken et al. 2017).

641 Some of the clusters of image contents identified as being frequently photographed or having a high
642 associated sentiment have already been widely explored in CES literature. For example, Van Zanten
643 et al., (2016) found that some of the best predictors of recreational value at a landscape scale were
644 geomorphological features such as hills and mountains and Oteros-Rozas et al. (2018) found that
645 mountains are particularly associated with hiking. Furthermore, the impact of being close to water
646 and vegetation on hiking has also been widely explored (e.g. Pastorella et al. 2017; Schirpke et al.
647 2018; Aiba et al. 2019). Photographs that contain a combination of clusters can provide information
648 on the interactions between different biophysical features which give rise to CES. For instance, some
649 hikers may prefer natural mountainous areas with forest cover and others may prefer natural
650 mountainous areas closer to water (Pastorella et al. 2017; Schirpke et al. 2018).

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

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

679 Here, our analysis has shown that the grouped images (based on a single user's photographs on a
680 single day) can contain multiple different content label clusters. We note that here the individual
681 clusters of labels focus on biodiversity and geodiversity features separately, though this may be a
682 facet of the clustering procedure or the method in which the machine learning classifier labels
683 images. As with other studies which clustered photograph content (e.g. Lee et al. 2019) photographs
684 were frequently assigned multiple equally dominant clusters, indicating an interaction between

685   geodiversity, biodiversity and human features. It may not always be possible to include features of
686   the multiple clusters in a single image. For example, a hiker may be interested in photographing
687   large scale vegetation such as a forest from the top of a mountain (Aiba et al. 2019), but these
688   images may omit the mountain itself. We suggest that these relationships may not be properly
689   captured by assessing single photographs and therefore, recommend that future studies assess each
690   Flickr user's images from a single day, such as used here, or by grouping all photographs from a
691   smaller study site as a collective.

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

702   This study further demonstrates that social media data can easily provide large spatial-temporal
703   dataset compared to traditional methods (Fox et al. 2020b), however, future work could implement
704   this over small spatial scales to assess regional or local trends. One limitation of social media data is
705   that not all demographics are well represented (Oteros-Rozas et al. 2018) and management and
706   policy decisions should acknowledge this. Studies could therefore combine results of social media
707   studies with other datasets such as survey data (Graham and Eigenbrod 2019; Moreno-Llorca et al.
708   2020; Sinclair et al. 2020). A further potential limitation of this multi-faceted filtering process is that
709   although we can be more confident that the images we are assessing represent CES, there are
710   substantially fewer suitable images in the refined dataset compared to the original dataset
711   downloaded from Flickr. However, we still had a large sample size of 43,427 images for one activity
712   over 5 years. Other studies have found success when using smaller subsamples of the full dataset
713   (e.g. Langemeyer et al. 2018; Chen et al. 2020). We recommend that social media data can best
714   capture CES when researchers refine their dataset by the content of the images and enrich the data
715   with a measure of sentiment, which can provide a reliable representation of CES.

716   **5 Conclusion**

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

729   **6 References**

730 Aiba, M., Shibata, R., Oguro, M. and Nakashizuka, T., 2019. The seasonal and scale-dependent
731 associations between vegetation quality and hiking activities as a recreation service. *Sustainability*
732 *Science*, *14*(1), pp.119-129. https://doi.org/10.1007/s11625-018-0609-7

733 Becken, S., Stantic, B., Chen, J., Alaei, A.R. and Connolly, R.M., 2017. Monitoring the environment
734 and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. *Journal*
735 *of environmental management*, *203*, pp.87-97. https://doi.org/10.1016/j.jenvman.2017.07.007

736 Bian, L. and Butler, R., 1999. Comparing effects of aggregation methods on statistical and spatial
737 properties of simulated spatial data. *Photogrammetric engineering and remote sensing*, *65*, pp.73-
738 84.

739 Balzan, M.V. and Debono, I., 2018. Assessing urban recreation ecosystem services through the use of
740 geocache visitation and preference data: a case-study from an urbanised island environment. One
741 Ecosystem, 3, p.e24490.

742 Bennett, E.M., Cramer, W., Begossi, A., Cundill, G., Díaz, S., Egoh, B.N., Geijzendorffer, I.R., Krug, C.B.,
743 Lavorel, S., Lazos, E. and Lebel, L., 2015. Linking biodiversity, ecosystem services, and human well-
744 being: three challenges for designing research for sustainability. Current opinion in environmental
745 sustainability, 14, pp.76-85.

746 Brindley, P., Cameron, R.W., Ersoy, E., Jorgensen, A. and Maheswaran, R., 2019. Is more always
747 better? Exploring field survey and social media indicators of quality of urban greenspace, in relation
748 to health. *Urban Forestry & Urban Greening*, *39*, pp.45-54.
749 https://doi.org/10.1016/j.ufug.2019.01.015

750 Byczek, C., Longaretti, P.Y., Renaud, J. and Lavorel, S., 2018. Benefits of crowd-sourced GPS
751 information for modelling the recreation ecosystem service. *PloS one*, *13*(10), p.e0202645.
752 https://doi.org/10.1371/journal.pone.0202645

753 Casalegno, S., Inger, R., DeSilvey, C. and Gaston, K.J., 2013. Spatial covariance between aesthetic
754 value & other ecosystem services. *PloS one*, *8*(6), p.e68437.
755 https://doi.org/10.1371/journal.pone.0068437

756 Chan, K.M., Goldstein, J., Satterfield, T., Hannahs, N., Kikiloi, K., Naidoo, R., Vadeboncoeur, N. and
757 Woodside, U., 2011. Cultural services and non-use values. Natural capital: Theory and practice of
758 mapping ecosystem services, pp.206-228.

759 Chen, Y., Caesemaecker, C., Rahman, H.T. and Sherren, K., 2020. Comparing cultural ecosystem
760 service delivery in dykelands and marshes using Instagram: A case of the Cornwallis (Jijuktu'kwejk)
761 River, Nova Scotia, Canada. *Ocean & Coastal Management*, *193*, p.105254.
762 https://doi.org/10.1016/j.ocecoaman.2020.105254

763 Chhetri, P., 2015. A GIS methodology for modelling hiking experiences in the Grampians National
764 Park, Australia. *Tourism Geographies*, *17*(5), pp.795-814.
765 https://doi.org/10.1080/14616688.2015.1083609

766 Christian, H., Agus, M.P. and Suhartono, D., 2016. Single document automatic text summarization
767 using term frequency-inverse document frequency (TF-IDF). *ComTech: Computer, Mathematics and*
768 *Engineering Applications*, *7*(4), pp.285-294.

769 Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J.O. and Martins, M.J., 2019. Combining
770 social media photographs and species distribution models to map cultural ecosystem services: The

771   case of a Natural Park in Portugal. *Ecological Indicators*, *96*, pp.59-68.

772   https://doi.org/10.1016/j.ecolind.2018.08.043

773   Collins-Kreiner, N. and Kliot, N., 2017. Why Do People Hike? Hiking the Israel National

774   Trail. *Tijdschrift voor economische en sociale geografie*, *108*(5), pp.669-687.

775   https://doi.org/10.1111/tesg.12245

776   Daniel, T.C., Muhar, A., Arnberger, A., Aznar, O., Boyd, J.W., Chan, K.M., Costanza, R., Elmqvist, T.,

777   Flint, C.G., Gobster, P.H. and Grêt-Regamey, A., 2012. Contributions of cultural services to the

778   ecosystem services agenda. Proceedings of the National Academy of Sciences, 109(23), pp.8812-

779   8819. https://doi.org/10.1073/pnas.1114773109

780   Do, Y., 2019. Valuating aesthetic benefits of cultural ecosystem services using conservation

781   culturomics. *Ecosystem Services*, *36*, p.100894. https://doi.org/10.1016/j.ecoser.2019.100894

782   Evangelopoulos, N.E., 2013. Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive*

783   *Science*, *4*(6), pp.683-692. https://doi.org/10.1002/wcs.1254

784   Figueroa-Alfaro, R.W. and Tang, Z., 2017. Evaluating the aesthetic value of cultural ecosystem

785   services by mapping geo-tagged photographs from social media data on Panoramio and

786   Flickr. *Journal of Environmental Planning and Management*, *60*(2), pp.266-281.

787   https://doi.org/10.1080/09640568.2016.1151772

788   Fischer, A. & Eastwood, A. (2016). Coproduction of ecosystem services as human-nature

789   interactions—An analytical framework. Land Use Policy 52, 41-50.

790   doi:10.1016/j.landusepol.2015.12.004

791   Fish, R., Church, A. and Winter, M., 2016. Conceptualising cultural ecosystem services: A novel

792   framework for research and critical engagement. Ecosystem Services, 21, pp.208-217.

793   https://doi.org/10.1016/j.ecoser.2016.09.002

794   Fox, N., Graham, L.J., Eigenbrod, F., Bullock, J.M. and Parks, K.E., 2020a. Incorporating geodiversity in

795   ecosystem service decisions. *Ecosystems and People*, *16*(1), pp.151-159.

796   https://doi.org/10.1080/26395916.2020.1758214

797   Fox, N., Mancini, F., Graham, L.J., Sutter, L., August, T., 2020. photosearcher: Photo Searcher. R

798   package version 1.0. https://docs.ropensci.org/photosearcher

799   Gefen, D., Endicott, J.E., Fresneda, J.E., Miller, J. and Larsen, K.R., 2017. A guide to text analysis with

800   latent semantic analysis in R with annotated code: Studying online reviews and the stack exchange

801   community. *Communications of the Association for Information Systems*, *41*(1), p.21.

802   https://doi.org/10.17705/1CAIS.04121

803   Ghermandi, A., Camacho-Valdez, V. and Trejo-Espinosa, H., 2020. Social media-based analysis of

804   cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism*

805   *Management*, *77*, p.104002. https://doi.org/10.1016/j.tourman.2019.104002

806   Ghermandi, A. and Sinclair, M., 2019. Passive crowdsourcing of social media in environmental

807   research: A systematic map. *Global environmental change*, *55*, pp.36-47.

808   https://doi.org/10.1016/j.gloenvcha.2019.02.003

809   Google Cloud Vision, 2020. Documentation for the Google Cloud Vision API [WWW Document]. URL

810   www.cloud.google.com/vision/.

811   Gordon, J.E. and Barron, H.F., 2013. The role of geodiversity in delivering ecosystem services and
812   benefits in Scotland. *Scottish Journal of Geology*, *49*(1), pp.41-58. https://doi.org/10.1144/sjg2011-
813   465

814   Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B. and Ziv, G., 2019. Using social media, machine
815   learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem*
816   *Services*, *38*, p.100958. https://doi.org/10.1016/j.ecoser.2019.100958

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

819   Gray, M., 2008. Geodiversity: developing the paradigm. *Proceedings of the Geologists'*
820   *Association*, *119*(3-4), pp.287-298. https://doi.org/10.1016/S0016-7878(08)80307-0

821   Gray, M., 2004. *Geodiversity: valuing and conserving abiotic nature*. John Wiley & Sons.

822   Gray, M., 2012. Valuing geodiversity in an 'ecosystem services' context. *Scottish Geographical*
823   *Journal*, *128*(3-4), pp.177-194. https://doi.org/10.1080/14702541.2012.725858

824   Havinga, I., Bogaart, P.W., Hein, L. and Tuia, D., 2020. Defining and spatially modelling cultural
825   ecosystem services using crowdsourced data. *Ecosystem Services*, *43*, p.101091.
826   https://doi.org/10.1016/j.ecoser.2020.101091

827   Hirons, M., Comberti, C. and Dunford, R., 2016. Valuing cultural ecosystem services. Annual Review
828   of Environment and Resources, 41, pp.545-574.

829   Hjort, J., Gordon, J.E., Gray, M. and Hunter Jr, M.L., 2015. Why geodiversity matters in valuing
830   nature's stage. *Conservation Biology*, *29*(3), pp.630-639. https://doi.org/10.1111/cobi.12510

831   Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Tenkanen, H., Butchart, S.H., Brooks, T.M. and
832   Di Minin, E., 2019. Assessing global popularity and threats to Important Bird and Biodiversity Areas
833   using social media data. *Science of the Total Environment*, *683*, pp.617-623.
834   https://doi.org/10.1016/j.scitotenv.2019.05.268

835   Huang, F., Zhang, X., Zhao, Z., Xu, J. and Li, Z., 2019. Image–text sentiment analysis via deep
836   multimodal attentive fusion. *Knowledge-Based Systems*, *167*, pp.26-37.
837   https://doi.org/10.1016/j.knosys.2019.01.019

838   Kiernan, K., 2015. Landforms as sacred places: implications for geodiversity and
839   geoheritage. *Geoheritage*, *7*(2), pp.177-193. https://doi.org/10.1007/s12371-014-0128-6

840   Kim, Y., Kim, C.K., Lee, D.K., Lee, H.W. and Andrada, R.I.T., 2019. Quantifying nature-based tourism in
841   protected areas in developing countries by using social big data. *Tourism Management*, *72*, pp.249-
842   256. https://doi.org/10.1016/j.tourman.2018.12.005

843   King, H.P., Morris, J., Graves, A., Bradbury, R.B., McGinlay, J. and Bullock, J.M., 2017. Biodiversity and
844   cultural ecosystem benefits in lowland landscapes in southern England. Journal of environmental
845   psychology, 53, pp.185-197. https://doi.org/10.1016/j.jenvp.2017.08.002

846   Kondo, M.C., Fluehr, J.M., McKeon, T. and Branas, C.C., 2018. Urban green space and its impact on
847   human health. *International journal of environmental research and public health*, *15*(3), p.445.
848   https://doi.org/10.3390/ijerph15030445

849   Koto, F. and Adriani, M., 2015, June. A comparative study on twitter sentiment analysis: Which
850   features are good?. In *International Conference on Applications of Natural Language to Information*
851   *Systems* (pp. 453-457). Springer, Cham. https://doi.org/10.1007/978-3-319-19581-0_46

852   Langemeyer, J., Calcagni, F. and Baro, F., 2018. Mapping the intangible: Using geolocated social
853   media data to examine landscape aesthetics. *Land use policy*, *77*, pp.542-552.
854   https://doi.org/10.1016/j.landusepol.2018.05.049

855   Lee, H., Seo, B., Koellner, T. and Lautenbach, S., 2019. Mapping cultural ecosystem services 2.0–
856   Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators*, *96*, pp.505-
857   515. https://doi.org/10.1016/j.ecolind.2018.08.035

858   Mayer, M. and Woltering, M., 2018. Assessing and valuing the recreational ecosystem services of
859   Germany's national parks using travel cost models. Ecosystem services, 31, pp.371-386.

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

863   Milcu, A.I., Hanspach, J., Abson, D. and Fischer, J., 2013. Cultural ecosystem services: a literature
864   review and prospects for future research. Ecology and society, 18(3).
865   https://www.jstor.org/stable/26269377

866   Millennium Ecosystem Assessment., 2005. Millennium Ecosystem Assessment. Washington, DC: New
867   Island Press, 13, p.520.

868   Minkiewicz, J., Bridson, K. and Evans, J. (2016), "Co-production of service experiences: insights from
869   the cultural sector", Journal of Services Marketing, Vol. 30 No. 7, pp. 749-761.
870   https://doi.org/10.1108/JSM-04-2015-0156

871   Mitten, D., Overholt, J.R., Haynes, F.I., D'Amore, C.C. and Ady, J.C., 2018. Hiking: A low-cost,
872   accessible intervention to promote health benefits. *American Journal of Lifestyle Medicine*, *12*(4),
873   pp.302-310. https://doi.org/10.1177/1559827616658229

874   Monkman, G.G., Kaiser, M.J. and Hyder, K., 2018. Text and data mining of social media to map
875   wildlife recreation activity. *Biological Conservation*, *228*, pp.89-99.
876   https://doi.org/10.1016/j.biocon.2018.10.010

877   Moreno-Llorca, R., Méndez, P.F., Ros-Candeira, A., Alcaraz-Segura, D., Santamaría, L., Ramos-Ridao,
878   Á.F., Revilla, E., García, F.J.B. and Vaz, A.S., 2020. Evaluating tourist profiles and nature-based
879   experiences in Biosphere Reserves using Flickr: Matches and mismatches between online social
880   surveys and photo content analysis. *Science of The Total Environment*, p.140067.
881   https://doi.org/10.1016/j.scitotenv.2020.140067

882   Mullen, L.A., Bratt, J., 2018. "USAboundaries: Historical and Contemporary Boundaries of the United
883   States of America," Journal of Open Source Software 3(23), p.314.
884   https://doi.org/10.21105/joss.00314

885   Nielsen, F.Å., 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv
886   preprint arXiv:1103.2903.

887   Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C. and Plieninger, T., 2018. Using social
888   media photos to explore the relation between cultural ecosystem services and landscape features

889 across five European sites. *Ecological Indicators*, *94*, pp.74-86.
890 https://doi.org/10.1016/j.ecolind.2017.02.009

891 Pagnuco, I.A., Pastore, J.I., Abras, G., Brun, M. and Ballarin, V.L., 2016, April. Analysis of genetic
892 association in Listeria and Diabetes using Hierarchical Clustering and Silhouette Index. In *Journal of*
893 *Physics: Conference Series* (Vol. 705, p. 012002). https://doi.org/10.1088/1742-6596/705/1/012002

894 Palomo, I., Felipe-Lucia, M.R., Bennett, E.M., Martín-López, B. and Pascual, U., 2016. Disentangling
895 the pathways and effects of ecosystem service co-production. Advances in ecological research, 54,
896 pp.245-283.

897 Pastorella, F., Giacovelli, G., De Meo, I. and Paletto, A., 2017. People's preferences for Alpine forest
898 landscapes: results of an internet-based survey. *Journal of Forest Research*, *22*(1), pp.36-43.
899 https://doi.org/10.1080/13416979.2017.1279708

900 Pastur, G.M., Peri, P.L., Lencinas, M.V., García-Llorente, M. and Martín-López, B., 2016. Spatial
901 patterns of cultural ecosystem services provision in Southern Patagonia. *Landscape Ecology*, *31*(2),
902 pp.383-399. https://doi.org/10.1007/s10980-015-0254-9

903 Plieninger, T., Dijks, S., Oteros-Rozas, E. and Bieling, C., 2013. Assessing, mapping, and quantifying
904 cultural ecosystem services at community level. Land use policy, 33, pp.118-129.
905 https://doi.org/10.1016/j.landusepol.2012.12.013

906 R Core Team (2019). R: A language and environment for statistical computing. R Foundation for
907 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

908 Raymond, C.M., Giusti, M. and Barthel, S., 2018. An embodied perspective on the co-production of
909 cultural ecosystem services: toward embodied ecosystems. Journal of Environmental Planning and
910 Management, 61(5-6), pp.778-799.

911 Richards, D.R. and Tunçer, B., 2018. Using image recognition to automate assessment of cultural
912 ecosystem services from social media photographs. *Ecosystem services*, *31*, pp.318-325.
913 https://doi.org/10.1016/j.ecoser.2017.09.004

914 Sankey, T., 2017. *Statistical Descriptions of Spatial Patterns*. In: Shekhar S., Xiong H., Zhou X. (eds)
915 *Encyclopedia of GIS. Springer, Cham*. https://doi.org/10.1007/978-3-319-17885-1_1351

916 Schirpke, U., Meisch, C., Marsoner, T. and Tappeiner, U., 2018. Revealing spatial and temporal
917 patterns of outdoor recreation in the European Alps and their surroundings. *Ecosystem services*, *31*,
918 pp.336-350. https://doi.org/10.1016/j.ecoser.2017.11.017

919 Schwemmer, C., 2019. imgrec: An Interface for Image Recognition. R package version 0.1.1.
920 https://github.com/cschwem2er/imgrec

921 Sharp, R., Douglass, J., Wolny, S., Arkema, K., Bernhardt, J., Bierbower, W., Chaumont, N., Denu, D.,
922 Fisher, D., Glowinski, K., Griffin, R., Guannel, G., Guerry, A., Johnson, J., Hamel, P., Kennedy, C., Kim,
923 C.K., Lacayo, M., Lonsdorf, E., Mandle, L., Rogers, L., Silver, J., Toft, J., Verutes, G., Vogl, A. L., Wood,
924 S, and Wyatt, K. 2020, InVEST 3.8.7.post9+ug.ga50c7f5 User's Guide. The Natural Capital Project,
925 Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund.
926 https://storage.googleapis.com/releases.naturalcapitalproject.org/invest-
927 userguide/latest/index.html

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

Sinclair, M., Mayer, M., Woltering, M. and Ghermandi, A., 2020b. Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. Journal of Environmental Management, 263, p.110418. https://doi.org/10.1016/j.jenvman.2020.110418

Spangenberg, J., 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. Ecological Economics 104, 22-32. doi:10.1016/j.ecolecon.2014.04.025

Stephens, P.A., Pettorelli, N., Barlow, J., Whittingham, M.J. and Cadotte, M.W., 2015. Management by proxy? The use of indices in applied ecology. *Journal of applied ecology.*, *52*(1), pp.1-6. https://doi.org/10.1111/1365-2664.12383

Tenerelli, P., Demšar, U. and Luque, S., 2016. Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, *64*, pp.237-248. https://doi.org/10.1016/j.ecolind.2015.12.042

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L. and Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific reports*, *7*(1), pp.1-11. https://doi.org/10.1038/s41598-017-18007-4

Van Berkel, D.B., Tabrizian, P., Dorning, M.A., Smart, L., Newcomb, D., Mehaffey, M., Neale, A. and Meentemeyer, R.K., 2018. Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. Ecosystem services, 31, pp.326-335. https://doi.org/10.1016/j.ecoser.2018.03.022

Wang, X. and Xu, Y., 2019, July. An improved index for clustering validation based on silhouette index and Calinski-Harabasz index. In *IOP Conference Series: Materials Science and Engineering* (Vol. 569, No. 5, p. 052024). IOP Publishing.

Wilcer, S.R., Larson, L.R., Hallo, J.C. and Baldwin, E., 2019. Exploring the Diverse Motivations of Day Hikers: Implications for Hike Marketing and Management. *Journal of Park & Recreation Administration*, *37*(3). https://doi.org/10.18666/JPRA-2019-9176

Wild, F. 2015. lsa: Latent Semantic Analysis. R package version 0.73.1. https://CRAN.R-project.org/package=lsa

Willcock, S., Hooftman, D.A., Balbi, S., Blanchard, R., Dawson, T.P., O'Farrell, P.J., Hickler, T., Hudson, M.D., Lindeskog, M., Martinez-Lopez, J. and Mulligan, M., 2019. A continental-scale validation of ecosystem service models. *Ecosystems*, *22*(8), pp.1902-1917. https://doi.org/10.1007/s10021-019-00380-y

Wilson, T., Lovelace, R. and Evans, A.J., 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. *Applied spatial analysis and policy*, *12*(1), pp.71-99. https://doi.org/10.1007/s12061-016-9201-7

Wood, S.A., Guerry, A.D., Silver, J.M. and Lacayo, M., 2013. Using social media to quantify nature-based tourism and recreation. *Scientific reports*, *3*(1), pp.1-7. https://doi.org/10.1038/srep02976

969   Yan, Y., Schultz, M. and Zipf, A., 2019. An exploratory analysis of usability of Flickr tags for land

970   use/land cover attribution. *Geo-spatial Information Science*, *22*(1), pp.12-22.

971   https://doi.org/10.1080/10095020.2018.1560044

972

973