


Communication

On How Crowdsourced Data and Landscape Organisation Metrics Can Facilitate the Mapping of Cultural Ecosystem Services: An Estonian Case Study

Oleksandr Karasov ^{1,*} , Stien Heremans ², Mart Külvik ¹, Artem Domnich ³ and Igor Chervanyov ⁴

¹ Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, 51006 Tartu, Estonia; mart.kylvik@emu.ee

² Research Institute for Nature and Forest (INBO), 1000 Brussels, Belgium; stien.heremans@inbo.be

³ Institute of Computer Science, University of Tartu, 51009 Tartu, Estonia; artem.domnich@tudeng.ut.ee

⁴ Physical Geography and Cartography Department, School of Geology, Geography, Recreation and Tourism, V. N. Karazin Kharkiv National University, 61022 Kharkiv, Ukraine; chervanyov@ukr.net

* Correspondence: oleksandr.karasov@student.emu.ee

Received: 9 April 2020; Accepted: 14 May 2020; Published: 19 May 2020



Abstract: Social media continues to grow, permanently capturing our digital footprint in the form of texts, photographs, and videos, thereby reflecting our daily lives. Therefore, recent studies are increasingly recognising passively crowdsourced geotagged photographs retrieved from location-based social media as suitable data for quantitative mapping and assessment of cultural ecosystem service (CES) flow. In this study, we attempt to improve CES mapping from geotagged photographs by combining natural language processing, i.e., topic modelling and automated machine learning classification. Our study focuses on three main groups of CESs that are abundant in outdoor social media data: landscape watching, active outdoor recreation, and wildlife watching. Moreover, by means of a comparative viewshed analysis, we compare the geographic information system- and remote sensing-based landscape organisation metrics related to landscape coherence and colour harmony. We observed the spatial distribution of CESs in Estonia and confirmed that colour harmony indices are more strongly associated with landscape watching and outdoor recreation, while landscape coherence is more associated with wildlife watching. Both CES use and values of landscape organisation indices are land cover-specific. The suggested methodology can significantly improve the state-of-the-art with regard to CES mapping from geotagged photographs, and it is therefore particularly relevant for monitoring landscape sustainability.

Keywords: cultural ecosystem services; automated image recognition; natural language processing; topic modelling; landscape coherence; colour harmony

1. Introduction

Almost 50 years ago, in the 1970s, Philippe Saint-Marc interpreted the outdoor environment as a social service supporting a good quality of life and public well-being [1]. Ever since then, this logic has been elaborated upon with the concept of cultural ecosystem services (CESs) [2,3] and a geographic perspective connecting the ecosystem (landscape) structure and functions with benefits and values [4]. Accordingly, the capacity of landscapes to provide CESs among other ecosystem services is now considered a prerequisite for landscape sustainability in connecting the Earth's patterns and processes to individual values and preferences [5–7].

However, CESs have proven difficult to quantify, and consequently they are difficult to manage. Therefore, many authors have discussed CESs in the context of metrics, including economic assessment

and quantitative mapping [8–12]. Currently, having a proper understanding, quantitative assessment, and an incorporation of CES into decision-making processes is considered crucial for achieving sustainable development goals and other policy targets [13–16]. The most advanced approach to the classification of the CES is being developed in the Common International Classification of Ecosystem Services (CICES) [17]. Our work examines, in the Estonian context, the following classes of CESs according to the CICES:

- a) characteristics of living systems that enable aesthetic experiences (experiencing landscape beauty, passive recreation);
- b) characteristics of living systems that enable activities promoting health, recuperation, or enjoyment through active or immersive interactions (active outdoor recreation); and
- c) characteristics of living systems that enable activities promoting health, recuperation, or enjoyment through passive or observational interactions (e.g., watching organisms: plants, animals and mushrooms).

The vast majority of CES assessments are based on surveys, interviews, participatory mapping, workshops, and other kinds of offline engagements with pre-selected individuals, such as local communities, key stakeholders, or experts [8,18–20]. However, the last two decades have seen a growing trend towards crowdsourcing applications in this field. In particular, the use of publicly available location-based social media (LBSM) data—mainly geotagged photographs—stored in online photo repositories (Flickr and Panoramio), applications (Instagram and Strava), and social networks (VK.com and Twitter) has proliferated [21]. Passively crowdsourced digital footprint has been used for (a) the assessment of touristic place visitation rates [22], (b) mapping landscape values across spatial scales [23,24], (c) mapping landscape aesthetic flow [25], (d) analysing spatial CES distributions [12,26], etc.

However, the amount of geotagged data in the online repositories of varying and often non-relevant content poses an issue for content selection and classification. The most common approaches of content analysis include manual selection [25–27] or photo-user-days mapping within the InVEST ecosystem service models [22,28,29]. Therefore, image recognition services and machine learning models have been gaining attention more recently. For instance, machine learning algorithms provided by Clarifai (Clarifai Inc., New York, NY, USA) and Google Cloud Vision were recently reported to be very promising for CES recognition and mapping [30,31], and natural language processing was applied to categorise social media users in relation to outdoor recreation [32].

In our study, the objectives are to (a) identify and map CES use in Estonia by using a combination of automated content image recognition and topic modelling on photos from selected social media platforms, and (b) quantify the association between two types of landscape attributes reflecting subjective landscape organisation, i.e., the landscape coherence and colour harmony of land cover, and CES flow. Landscape coherence is a landscape attribute, which, according to existing reports, rather positively influences landscape preferences by generalising order and organisation of recognisable elements of landscape pattern [33]. It can be mapped with a geographic information system (GIS)-based indicator in relation to photographing preferences [34]. Colour harmony is also discussed as an important aesthetic variable of visual landscape [35] and is recently mapped with satellite imagery and textural metrics [36], but it has received much less attention in literature compared to landscape coherence. Suggested objective indicators of landscape coherence and colour harmony of land cover remain understudied in the context of CES use and require testing across various environmental settings and scales.

The paper is developed around a simple framework of CES use classification and its linkage to landscape attributes, assessable with remote sensing- and GIS-based indicators. Section 2 justifies the study area choice and introduces the methods used to extract knowledge on CES use and landscape attributes from geolocated photographs and GIS data, respectively. Section 3 presents the results of CES use mapping in relation to the GIS- and remote sensing-based indicators, as well as land cover

types. Section 4 discusses the results in the wider context of added research value compared to existing research papers. Section 5 concludes with the main findings and directions of further work.

2. Materials and Methods

2.1. Study Area

According to DataReportal, 98% of Estonians are Internet users to some extent, and 57% are active users of social media [37]. This high level of Internet penetration, combined with a well-developed touristic policy and infrastructure, as well as the significant share of the Russian-speaking community in the total population (VK.com is based in Russia) render Estonia a good study area for social media and CES-related studies. Moreover, the diverse environmental conditions and numerous protected areas in Estonia enhance opportunities for analyses from geographic and nature conservation perspectives.

2.2. Mapping of Cultural Ecosystem Service (CES) Represented in Social Media in Estonia

To test the applicability of topic modelling for CES identification and classification, we used geotagged photo-series analysis [38], retrieving metadata by means of application programming interface (API) calls (including geographic coordinates, user and photo ID, date of taking, web-links to photographs) for publicly available images uploaded to Flickr.com and VK.com services from 2015 to 2018. Flickr and VK.com continue to provide access to their non-private geolocated content, while Panoramio discontinued its service and Instagram has not shared its data with third parties since 2015. We additionally used the GIS-data for buildings in Estonia [39] to remove the metadata for indoor photographs. In total, metadata for 21,242 geographically outdoor photographs were retrieved and combined into a single dataset. We then applied content image recognition to these photographs with automated Python API requests to Clarifai's service (Clarifai Inc., New York, NY, USA). We used the general model with a cut-off greater than 90% for the probability that the tag is correct.

We then tested topic modelling (Latent Dirichlet Allocation (LDA) algorithm) implemented in the Orange data mining software [40] to classify the tags into a number of topics and deleted the irrelevant ones (assuming that photographs sharing the same tags represent the same "topic"). As a result, the pre-processed dataset consisted of 9983 photographs. After some initial testing, we decided on three topics for the LDA analysis. The LDA algorithm was useful in two aspects: (a) identification of the non-relevant photographs (for example, we removed the photographs, sharing topics of tags related to driving and cars, indoor design, architecture, fashion and beauty, military service) and (b) identification of the relevant topics in the rest of the photographs' tags.

As the LDA algorithm calculates the probability score, indicating the likelihood of the set of tags for each photograph belonging to each topic, we assumed that the assigned photographs belong to the topic with the highest probability score. Owing to the potential overlap with this fuzzy distinction between the topics of each photograph, we decided to post-process the results manually by interpreting the context of each photograph in addition to its content. For instance, photographs of pets were transferred from the topic of wildlife watching to outdoor recreation, and photographs with a minor presence of people or their recreation-related equipment were moved from landscape watching to outdoor recreation. We devised an a priori hypothesis about the small number of relevant CESs, according to the CICES classes (3–5), and the very first test of LDA algorithm resulted in three relevant CES-related topics. In case we applied LDA with a higher number of intended topics, some minor subclasses of recreation appeared, but these minor classes of recreational CES are beyond the scope of this study. We identified the following topics corresponding to the groups of CESs:

- a. Landscape watching. This consists of the following tags: nature, outdoors, landscape, tree, nobody, wood, sky, travel, water, and summer (6154 photographs; 17 manually transferred from topic 3).

- b. Active outdoor recreation. This consists of the following tags: people, recreation, adult, fun, man, leisure, outdoors, one, sport, and action (2345 photographs; 770 manually transferred from topic 1, and 114 from topic 3).
- c. Wildlife watching. This consists of the following tags: nature, outdoors, no one, flora, leaf, wild, wildlife, season, animal, growth (1484 photographs; 124 manually transferred from topic 1, and 2 from topic 2).

We mapped CES use from the photo locations to examine whether people (subconsciously) consider some selected aesthetic landscape attributes that represent landscape organisation [34,36], and these attributes can be derived from remote sensing data [41].

2.3. Impact of Landscape Organisation on CES Use

The colour harmony of land cover is a landscape attribute often neglected in landscape studies [41] but is potentially responsible for visual landscape quality [35] and is assessable using remotely sensed data. We used Landsat 8 OLI cloudless summertime mosaics for the territory of Estonia with a 5 km buffer zone pre-processed with the Google Earth Engine. The red (B4), green (B3), and blue (B2) bands, corresponding to the natural colours band combination, were converted into the hue-saturation-value (HSV) colour space to quantify colour harmony. We assumed that the hue and chroma (saturation in HSV space) similarity, which is listed among the universal principles of colour harmony [42], can be quantified for the hue and saturation raster datasets. Such assessment can be done using the grey level co-occurrence matrix (GLCM) homogeneity index (GLCMH, Equation (1)) [43], which measures the similarity of image pixel pairs [44] (Figure 1a,b):

$$GLCMH = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1 + (i - j)^2} P(i, j) \quad (1)$$

where $P(i, j)$ is the probability of co-occurrence of pixels i and j , and N_g is the number of distinct grey levels in the quantised image (64 in this study).

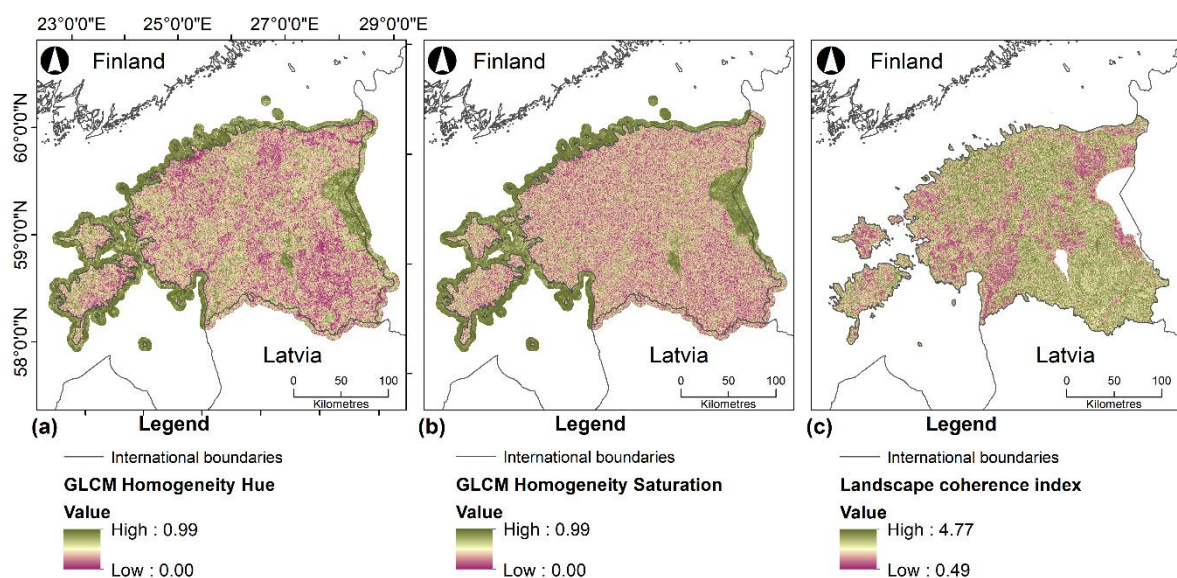


Figure 1. Spatial distribution of the values of the landscape organisation indices: (a) grey level co-occurrence matrix (GLCM) homogeneity for hue-saturation-value (HSV) hue component (colour harmony index); (b) GLCM homogeneity for the HSV saturation component (colour harmony index); (c) landscape coherence index. Higher values of colour harmony indices indicate water bodies (sea and lakes). Higher values of landscape coherence index indicate urban areas and, particularly, complex landscapes of Southern Eastern and Northern Estonia.

It should be mentioned that colour harmony depends on “how strongly an observer experiences the colours in the combination as going or belonging together, regardless of whether the observer likes the combination or not”, (p. 551, [45]). Therefore, it is rather a component of the formal landscape aesthetics and, in theory, does not necessarily reflect landscape preferences.

We interpret landscape coherence as a degree of order inherent to the landscape pattern that is composed of diverse and distinct landscape elements and features [46]. Landscape coherence is one of the classic subjective landscape attributes responsible for the emergence of landscape values [33]. Increasing the landscape coherence extent generally leads to a moderate increase in landscape preferences [34]. Therefore, we assess the vertical landscape coherence using the landscape coherence index (LCI, Figure 1c, Equation (2)) proposed by Karasov et al. [34], which is based on the concepts of the emergent theory of information, as presented by Lutsenko [47]. We calculate the LCI within a circular neighbourhood of seven pixels for the CORINE land cover model and the Topographic Position Index (TPI) landform classification, obtained with the respective SAGA GIS module [48].

$$LCI = \frac{I_{landscape}}{I_{land\ cover} + I_{land\ forms}} \quad (2)$$

where LCI is the landscape coherence index; $I_{land\ cover}$ and $I_{land\ forms}$ are the Hartley functions for the land cover/land use (LU/LC) model and the TPI-based landform classification based on the digital elevation model [43], respectively; and $I_{landscape}$ is the Hartley function for the parametric composite (digital landscape model) of the LU/LC model and TPI-based landforms.

The landscape coherence index benefits from the feature of additivity of the Hartley function (Equation (3)), which is a particular case of Shannon’s information entropy (Shannon diversity index):

$$I = n_i \log_2 m \quad (3)$$

where m is the total number of observations (landscape or land cover classes, types of landforms), and n is the number of observations in neighbourhood i .

The logic of landscape coherence calculation is based on the following assumption: for independent landforms and land cover, the algebraic sum of the amount of information, according to the Hartley function for landforms and land cover, will be equal for the amount of information for their parametric composite or digital landscape model. If the landforms and land cover models, which compose the digital landscape model, interact and are not independent, the summarised Hartley functions for these datasets will give a smaller value than the value of the Hartley function for the pixels of a digital landscape model. The ratio between Hartley functions for the digital landscape model and its components highlights the extent of systematic features of landscape and can be related to landscape preferences. Hypothetically, the increase in landscape coherence contributes to the visual landscape quality and therefore to CES use.

We then performed a viewshed analysis, identifying the visible surface from the set of observation points, namely the geolocations of the selected photographs from each group of CESs (see Section 2.2) and for the same number of randomly selected locations, which serve as pseudo-absence data (Figure 2). We used the PixScape software [49] on the European Digital Elevation Model (EU-DEM), version 1.1 [50] with the maximum visible distance and observer height set to 5 km and 1.6 m, respectively. The median LCI, hue homogeneity, and saturation homogeneity were calculated for each viewshed and compared between the actual (presence) and random (pseudo-absence) geolocations using Wilcoxon’s rank-sum test with continuity correction (see Appendix A for details), implemented in the Exploratory software (Exploratory Inc. (Delaware US) Sacramento, CA, USA).

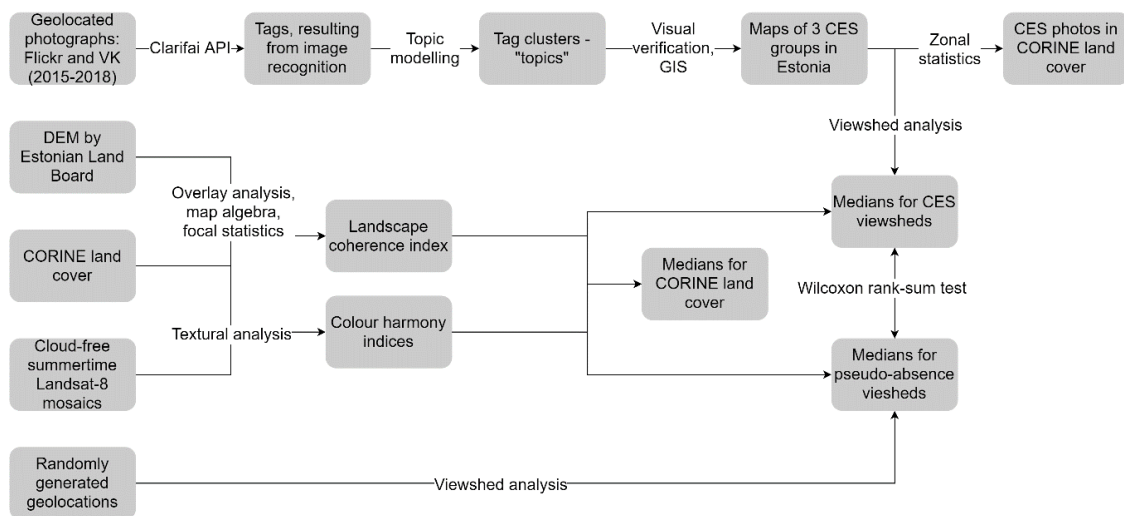


Figure 2. Research workflow of the cultural ecosystem services (CES) mapping in relation to the calculated landscape organisation indices.

3. Results

3.1. Mapping of CES Represented in Social Media in Estonia

Figure 3 presents the results of the CES mapping obtained with the application of topic modelling (geographical coordinates of the photographs from the combined Flickr and VK.com dataset, classified into three categories of CES groups). The clear linear patterns of the photographs highlight the main flows of people alongside the main roads and coastlines of Estonia. The exploratory buffer analysis for OpenStreetMap road data indicates that transport accessibility is extremely important for CES use. To be precise, 6148 out of 6153 landscape-watching photographs, 2311 out of 2345 outdoor recreation photographs, and 1483 out of 1484 wildlife-watching photographs have been taken no farther than 500 m from the roads and trails of all types. Although indoor photographs have been removed from the analysis (see Section 2.2), many photographs were taken in the main cities (Tallinn, Tartu, Narva, etc.), especially in their suburban zones. Additionally, the protected areas are conspicuous as approximately 59% of the total number of selected photographs were taken within these regions. A full list of the protected areas is presented in the Table S1 (Supplementary Materials).

An exploratory analysis of land cover (CORINE land cover 2018, Figure 4) shows that most of the photographs were taken in coniferous forests, agricultural areas, mixed forests, and transitional woodland-shrub areas. All the CES groups under consideration are well represented in these land cover classes. On the contrary, water bodies and courses, sea, peat bogs, inland marshes, and natural grasslands are frequented more for landscape watching than for the other groups of CES. Outdoor recreation is present in complex cultivation patterns and green urban areas. Wildlife watching frequently occurs in broad-leaved forests and pastures. In this way, more “natural” land cover classes are much better represented in the study datasets of passively crowdsourced photographs. However, land, which is principally occupied by agriculture, is among the leaders in enabling CES use.

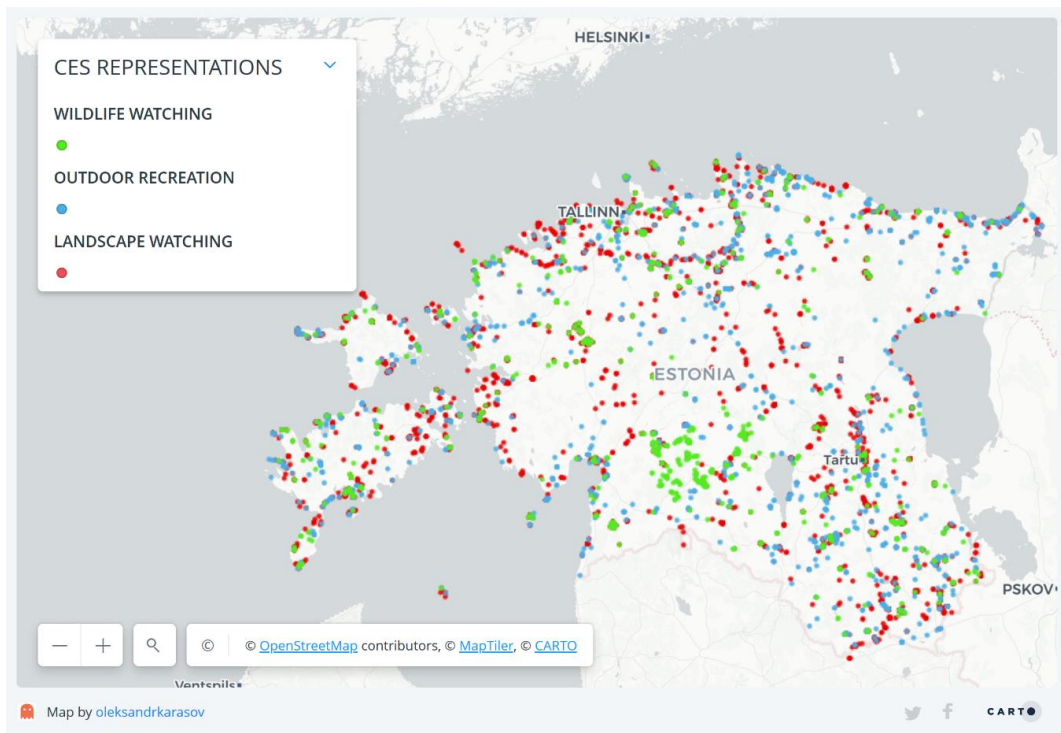


Figure 3. Geotagged photographs representing actual use of three groups of CES in Estonia (2016–2018): landscape watching (passive recreation), outdoor recreation activities, and wildlife watching. The web-map, designed in Carto, is available via the following link: <https://oleksandrkarasov.carto.com/builder/1e69e28a-9705-45a9-8276-471a330da2ff/embed>.

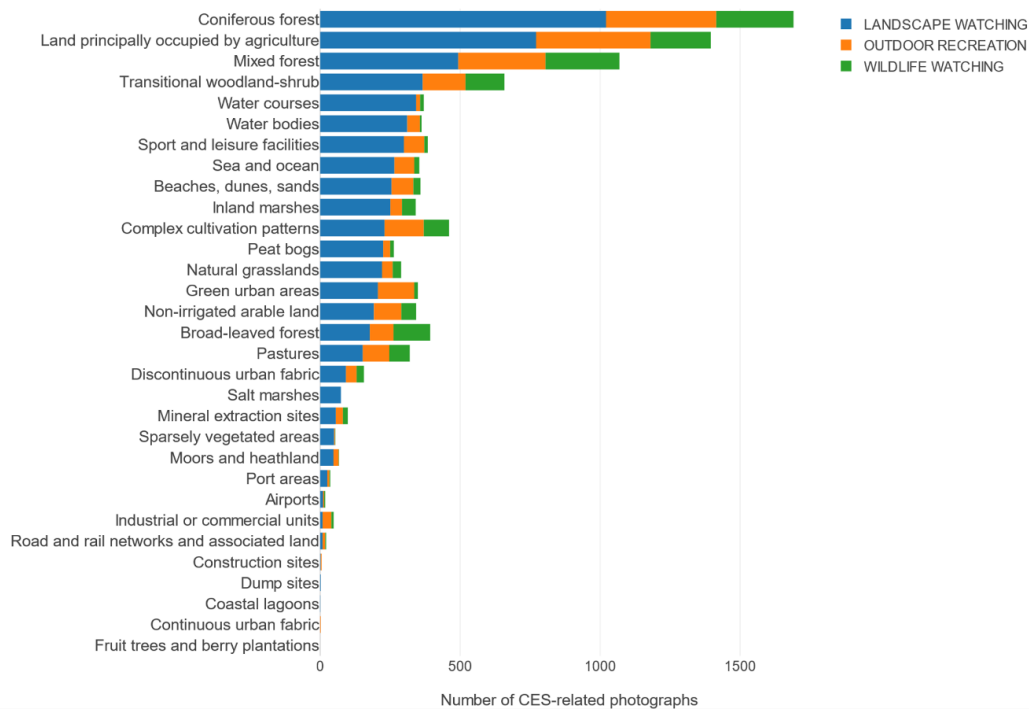


Figure 4. CES use in Estonia encompasses (with a few exceptions) predominantly natural and semi-natural land cover (CORINE land cover 2018). Land cover classes are ranked in order of decreasing number of landscape watching photographs.

3.2. Impact of Landscape Organisation on CES Use

As is clear from Figure 5, we see that the median hue and saturation similarity values are remarkably higher for the actual rather non-vegetated (median value of the normalized difference vegetation index (NDVI) lower than 0.1) viewsheds corresponding to landscape watching and outdoor recreation than for the pseudo-absence viewsheds. The indicators used exhibit similar behaviours for the landscape watching and outdoor recreation viewsheds, whilst colour harmony does not seem to influence wildlife watching.

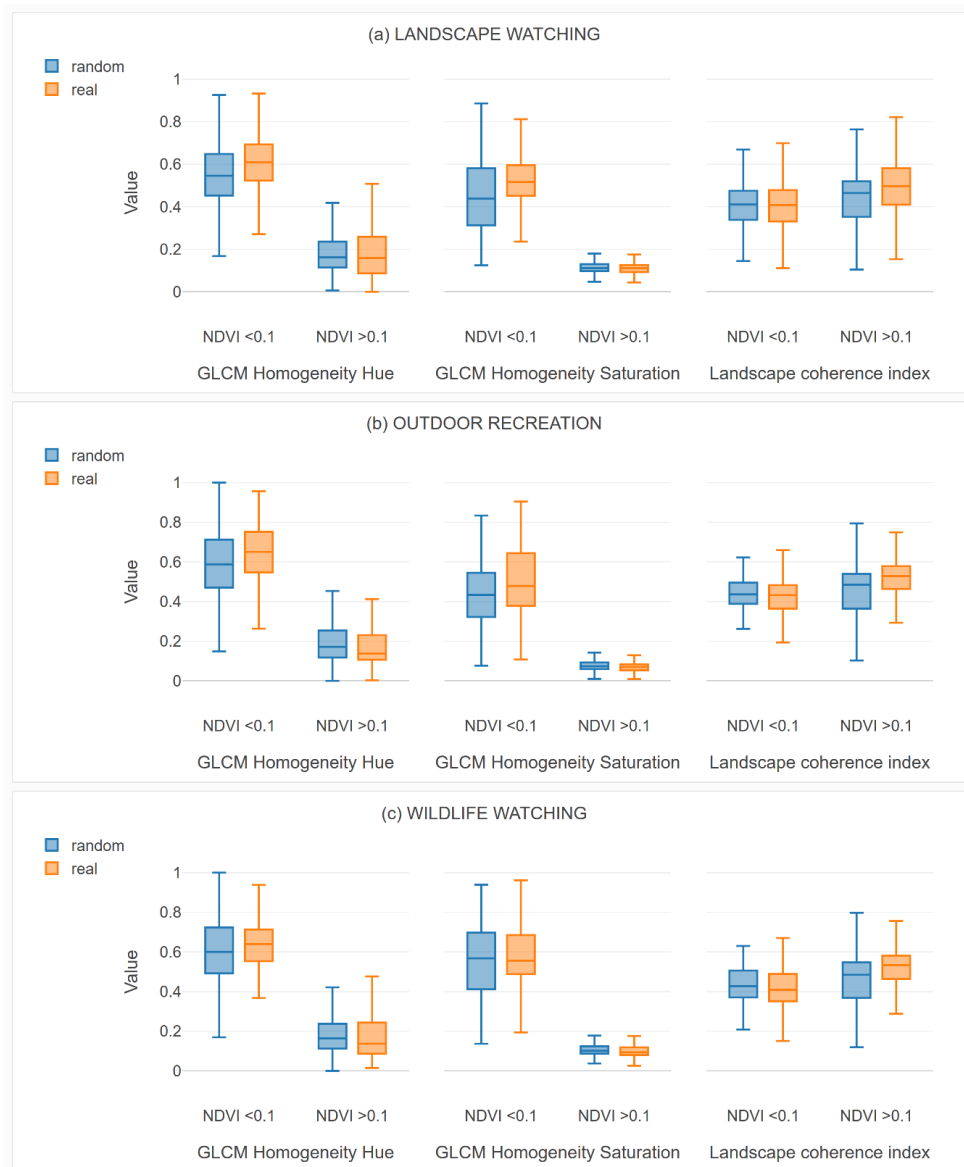


Figure 5. Comparison of medians of landscape coherence and harmony-based visual quality indices for each group of CESs within viewshed areas for actual geotagged photographs (“real”) and randomly simulated locations (“random”): (a) landscape watching; (b) outdoor recreation; (c) wildlife watching. Boxplots are designed separately for median normalized difference vegetation (NDVI) index values for each viewshed being higher 0.1 and lower 0.1 to present the index performance for rather vegetated and non-vegetated area (mainly water bodies and streams). Colour harmony indices are higher for actual CES viewsheds in the case of non-vegetated areas, while landscape coherence index is higher for photographs of vegetated areas. The GLCM homogeneity index for the saturation of pixel pairs does not indicate wildlife watching in any case.

According to the Wilcoxon rank sum test with continuity correction, all the distribution differences except for colour harmony indices for wildlife watching are statistically highly significant, suggesting that most CES-related photographs were taken with consideration for land cover of higher colour harmony and landscape coherence (Figure 6). It is highly likely that colour harmony values affect landscape watching and outdoor recreation, while landscape coherence seems to have a clear positive influence on wildlife watching and a weaker positive influence on landscape watching and outdoor recreation.

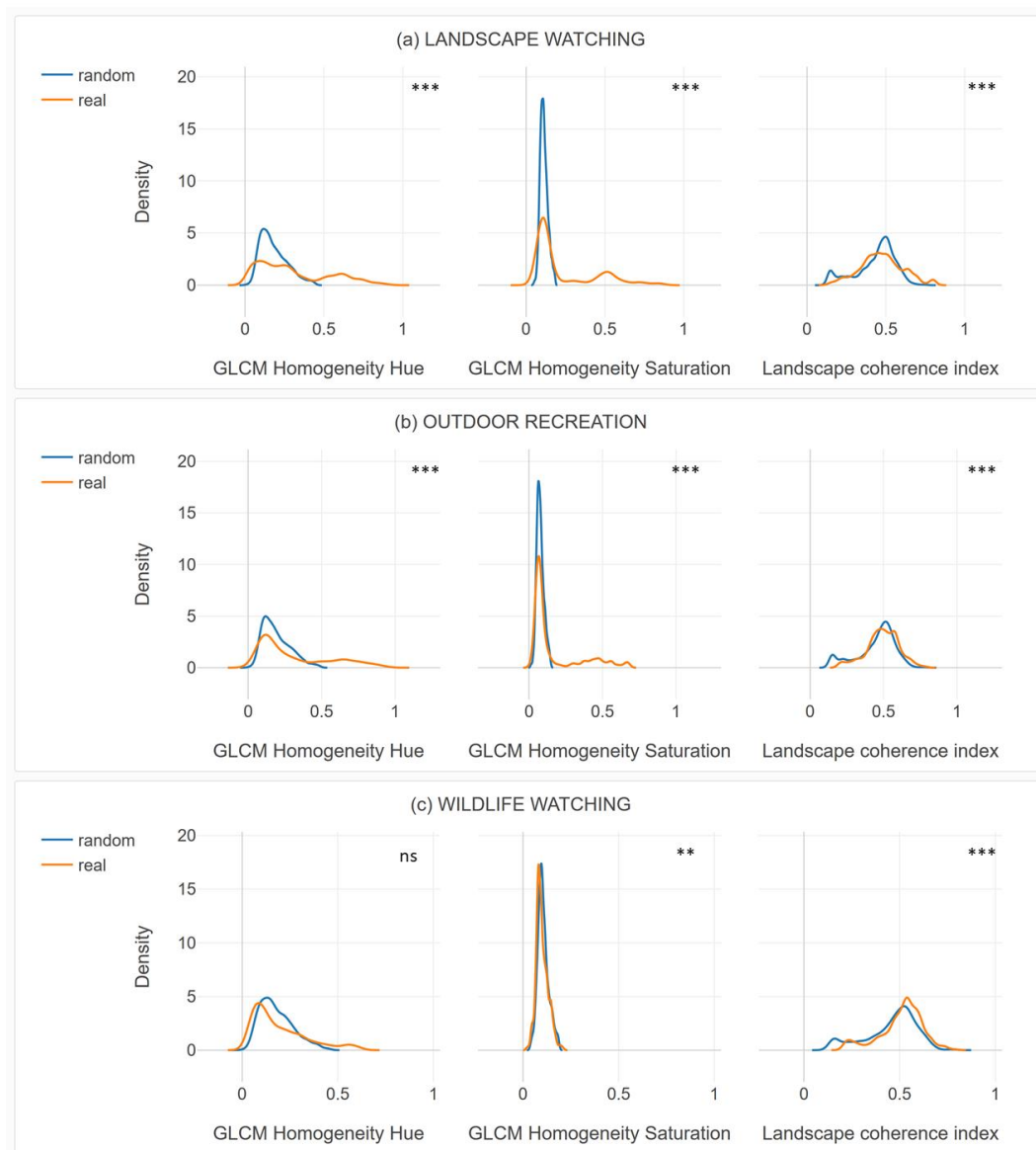


Figure 6. Density plots representing the results of the Wilcoxon rank sum test with continuity correction, applied to the medians of landscape coherence and harmony-based visual quality indices for each group of CESs within viewshed areas for actual geotagged photographs (“real”) and randomly simulated locations (“random”): (a) landscape watching; (b) outdoor recreation; (c) wildlife watching. Significance levels: *** p -value less than 0.001; ** p -value less than 0.01; ns—not significant. Alternative hypothesis: two-sided. Confidence level: 0.95.

A visual exploration of land cover with regard to LCI and colour harmony indices reveals a complementary character of the considered landscape organisation indices (Figure 7). Landscape coherence is the highest for culturally modified land covers—urban fabric, urban green areas, and

agricultural areas—and lower for natural areas, the minimum being observed for peat bogs and water bodies. Colour harmony, in contrast, is the highest for water forest and peat bogs. Therefore, colour harmony and landscape coherence extents are highly dependable on the land cover type: higher cultural modification of landscape results in the increasing orderliness and complexity, while colour harmony increases for homogeneous and predominantly natural (while often managed) land cover.

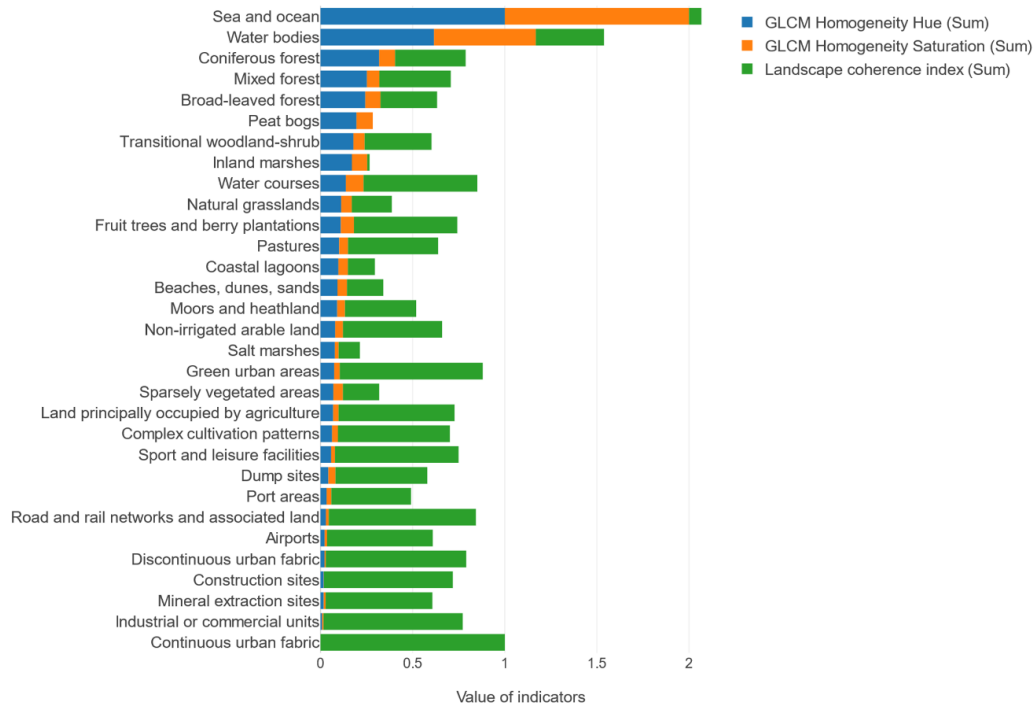


Figure 7. CORINE 2018 land cover classes ranged in order of decrease in GLCM homogeneity for the hue component of HSV colour space, indicating transition from water-related and natural land cover to urban-related areas. Landscape coherence generally increases in this direction.

4. Discussion

4.1. Mapping of CES Represented in Social Media in Estonia

Our results contribute to addressing the challenge of CES mapping by studying the relationships among the three categories of outdoor geotagged photographs and remote sensing-based landscape characteristics [24]. The distribution pattern of the CES-related photographs is in line with previous findings [23] as we confirm transport accessibility and naturalness to be the main factors influencing the probability of taking outdoor photographs [51]. Photographs from different CES groups often overlap spatially, indicating landscape multifunctionality. Landscape multifunctionality is important for the overall distribution of landscape values; hence, our approach can contribute to evidence-based trade-off analyses and the detection of hotspots of cultural landscape functions through CES patterns [52,53]. There is also a synergy between our nationwide CES mappings and the ESMERALDA project [54]. Our cross-disciplinary approach, integrating bio-physical and socio-cultural methods, allows for CES mapping and assessment across various spatial and temporal scales and is applicable to both urban and non-urban environments.

Much of the CES use seems concentrated within nature protection areas, revealing the efficiency and efficacy of the nature conservation policy in Estonia as well as the potential for further expansion of protected areas, which can contribute to an increase in nature-based tourism [55]. Thereby, we confirm the LBSM data as a valuable source of data for nature conservation as well as for CES mapping [21,56,57].

Our results continue the methodological approach that has been initiated in previous research [30–32]. The LDA topic modelling algorithm significantly facilitated the process of LBSM data assigned with the content-related tags as a result of automated image recognition with Clarifai. Therefore, we confirm that the LDA method of topic modelling is highly relevant and valuable for a rapid assessment of cultural ecosystem services use over large areas [31]. When all the photographs representing non-relevant topics have been removed, we applied the LDA algorithm to the relevant tags only, and some testing showed that three topics of tags sufficiently represent their diversity and meet our needs. The automated character of topic modelling often results in a meaningless classification, so the exact number of relevant topics (3) was found by trial and error (while a priori we assumed that there are just a few major CES categories). Further analysis would result in accounting for minor CES categories, such as picnicking, cycling, or playing tennis outdoors, but such detailed classification was beyond the scope of our study.

Obviously, the proposed methodological combination is not a complete substitute for traditional visual content analysis; instead, it should be used as an initial procedure for CES use assessment and followed with a quick visual verification. For example, we transferred to the outdoor recreation category those photographs that were automatically selected for landscape watching if they contained minor presence of people or their equipment; since presence of pets was automatically interpreted as wildlife (the general machine learning model provided by Clarifai does not account specifically for this distinction), we also manually moved these photographs to the category for outdoor recreation. Photographs with minor presence of wild animals classified as related to landscape watching were also manually transferred to the wildlife watching category.

4.2. Impact of Landscape Organisation on CES Use

In line with previous studies, landscape coherence was found to have a positive but rather weak association with CES use in our countrywide study [34]. Vertical landscape coherence increases for places of significant cultural modification (more legible urban and agricultural areas). Thereby, we confirm LCI performance as indicative of the orderliness of the landscape pattern, but unexpectedly it has a rather small impact on CES use. Wildlife watching occurs in places with higher LCI. This is potentially because people are more likely to take photographs of animals, plants, and mushrooms near their homes (such as green urban areas) in some understandable settings rather than in a more natural environment. Other authors have additionally explored the hotspots of wildlife watching near cities [29]. As some photographed areas have higher LCI, compared with the values for random locations, signs of anthropic modification (parks, suburban areas, agricultural fields, and other elements of cultural landscape) can be additionally important for CES use, complementing pure naturalness [58].

Colour harmony indicators (HSV hue and saturation similarities, indicated with GLCM homogeneity) showed a larger difference between the photographed viewsheds and random background viewsheds, suggesting that people tend to take photographs with a preference for land covers of greater colour harmony. However, as there is an association between land cover classes (CORINE land cover 2018) and colour harmony, the bias may be caused by the effect of the land cover itself; for instance, sea, water bodies, forests, and peat bogs additionally have powerful intrinsic and other values. Therefore, our results should be treated with caution, and colour harmony mappings should contribute to the general understanding of landscape rather than perform as the standalone indicators of landscape preferences.

4.3. Other Sources of Bias

It is most likely that elderly persons and children are the least represented age strata in LBSM. However, Flickr and VK.com were launched in 2004 and 2006, respectively, and have become very popular among diverse user groups, while general Internet penetration in Estonia is growing also [59]. We can expect that in the coming years, LBSM social media will become more orientated towards elderly people owing to the regular ageing of active Internet-users. Unfortunately, the LBSM data—unlike

surveys—provide little or no information on the individuals' sex and gender, age, education level, family status, ethnic origin, etc. Nevertheless, the LBSM data are free from some survey-specific issues, such as recollection and mind biases, which occur owing to intrusive surveying [52,60]. Therefore, in our opinion, social media data provide added value to CES studies.

5. Conclusions

Our results are based on photographs uploaded to the social media sites, Flickr and VK.com, which can be used to represent the actual use of some CESs (landscape watching, outdoor recreation, and wildlife watching), and are linked to spatial landscape indices in Estonia. Their spatial analysis enables a better understanding of the geographic organisation of the environment and its potential for providing CES and supporting nature appreciation in an urbanised society [61]. Evidence from our study suggests that social media users prefer taking photographs of landscapes and outdoor activities in areas with greater colour harmony, whilst landscape coherence is linked strongly only to wildlife watching and, to a lesser extent, other CESs.

Topic modelling significantly reduced the time needed for the content analysis of the photographs, and our CES mapping depends on the quality of this automated image content analysis. Therefore, future research could be targeted towards comparing different machine learning algorithms and including the temporal component. The suggested methodological combination of machine learning and natural language processing algorithms advances the existing common methods of CES assessment based on passively crowdsourced photographs, and it is sufficiently robust to be applied across the regional, continental, and global scales. In turn, the test of GIS-based landscape organisation metrics in relation to CES use shows that they can also facilitate the prospects of rapid and reliable landscape visual quality assessment up to the global scale, which does not depend on local subjective landscape evaluations and complements regional landscape character assessment. Drawbacks of the approach are related to the representativeness of the social media data as a source of knowledge about CES use and also to limitations of the GIS and remote sensing applicability for physio-gnomic landscape research. Notwithstanding, we have demonstrated that the combined usage of LBSM data, automated image recognition, natural language processing, satellite imagery, and GIS data is highly relevant for evidence-based ecosystem management and nature protection.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-445X/9/5/158/s1>, Table S1—Number of CES-related photographs per protected area.

Author Contributions: Conceptualization, O.K., M.K., I.C.; methodology, O.K.; software, O.K., A.D.; validation, O.K.; formal analysis, O.K.; investigation, O.K.; resources, M.K., S.H.; data curation, O.K.; writing—original draft preparation, O.K.; writing—review and editing, S.H., M.K.; visualization, O.K.; supervision, S.H., M.K.; project administration, S.H., M.K.; funding acquisition, M.K., O.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Social Fund's Dora Plus Programme, the ALTER-Net mobility fund, and the IMAGINE project (ERANET BIODIVERSA 3). The APC was partly funded by the IMAGINE project (ERANET BIODIVERSA 3).

Acknowledgments: This research was supported by the European Social Fund's Dora Plus Programme, ALTER-Net mobility fund, and the IMAGINE project (ERANET BIODIVERSA 3). We credit Olha Kaminska (University of Ghent, Belgium) for the methodological suggestion to use topic modelling, and Ms. Edith Chenault, who provided English proofreading for the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Results of the applied statistical analysis (Wilcoxon rank sum test with continuity correction) for median values of landscape organisation metrics within the viewsheds based on CES-related and

randomly generated geolocations. The landscape coherence index was rescaled (0; 1) to meet the scale of colour harmony estimations in Results.

Table A1. Summary statistics.

Indicator	U Statistic	p Value	Difference	Conf High	Conf Low
Landscape watching					
GLCM homogeneity hue	14,795,531.5	9.88×10^{-98}	-0.061	-0.055	-0.068
GLCM homogeneity saturation	14,594,742.5	2.91×10^{-107}	0.017	-0.015	-0.018
Landscape coherence index	16,273,215.5	2.01×10^{-41}	-0.033	-0.029	-0.039
Outdoor recreation					
GLCM homogeneity hue	2,356,245	2.21×10^{-17}	-0.032	-0.024	-0.040
GLCM homogeneity saturation	2,293,880	8.59×10^{-23}	-0.011	-0.008	-0.013
Landscape coherence index	2,280,950.5	5.18×10^{-24}	-0.035	-0.029	-0.042
Wildlife watching					
GLCM homogeneity hue	1,140,081	0.095	0.007	0.015	-0.001
GLCM homogeneity saturation	1,168,362	0.004	0.004	0.006	0.001
Landscape coherence index	898,071.5	3.34×10^{-18}	-0.037	-0.029	-0.046

Table A2. Detailed statistics.

Indicator	Type	Number of Rows	Mean	Confidence Low	Confidence High	Standard Error of Mean	Standard Deviation	Minimum	Maximum
Landscape watching									
GLCM homogeneity hue	random	6153	0.20	0.20	0.20	0.00	0.12	0.01	1.00
GLCM homogeneity hue	real	6153	0.31	0.30	0.31	0.00	0.23	0.00	0.93
GLCM homogeneity saturation	random	6153	0.13	0.13	0.13	0.00	0.08	0.01	1.00
GLCM homogeneity saturation	real	6153	0.24	0.24	0.24	0.00	0.21	0.00	0.93
Landscape coherence index	random	6153	0.42	0.41	0.42	0.00	0.16	0.00	0.87
Landscape coherence index	real	6153	0.47	0.47	0.48	0.00	0.14	0.00	1.00
Outdoor recreation									
GLCM homogeneity hue	random	2345	0.21	0.21	0.21	0.00	0.13	0.00	1.00
GLCM homogeneity hue	real	2345	0.31	0.30	0.32	0.01	0.25	0.00	0.96
GLCM homogeneity saturation	random	2345	0.09	0.09	0.10	0.00	0.09	0.00	1.00
GLCM homogeneity saturation	real	2345	0.20	0.19	0.20	0.00	0.22	0.01	0.90
Landscape coherence index	random	2345	0.44	0.43	0.44	0.00	0.15	0.00	0.87
Landscape coherence index	real	2345	0.49	0.49	0.49	0.00	0.12	0.00	1.00
Wildlife watching									
GLCM homogeneity hue	random	1484	0.20	0.20	0.21	0.00	0.14	0.00	1.00
GLCM homogeneity hue	real	1484	0.23	0.22	0.24	0.00	0.19	0.01	0.96
GLCM homogeneity saturation	random	1484	0.13	0.12	0.13	0.00	0.11	0.01	0.94
GLCM homogeneity saturation	real	1484	0.16	0.15	0.16	0.00	0.17	0.00	1.00
Landscape coherence index	random	1484	0.44	0.43	0.45	0.00	0.16	0.00	0.80
Landscape coherence index	real	1484	0.48	0.48	0.49	0.00	0.15	0.00	1.00

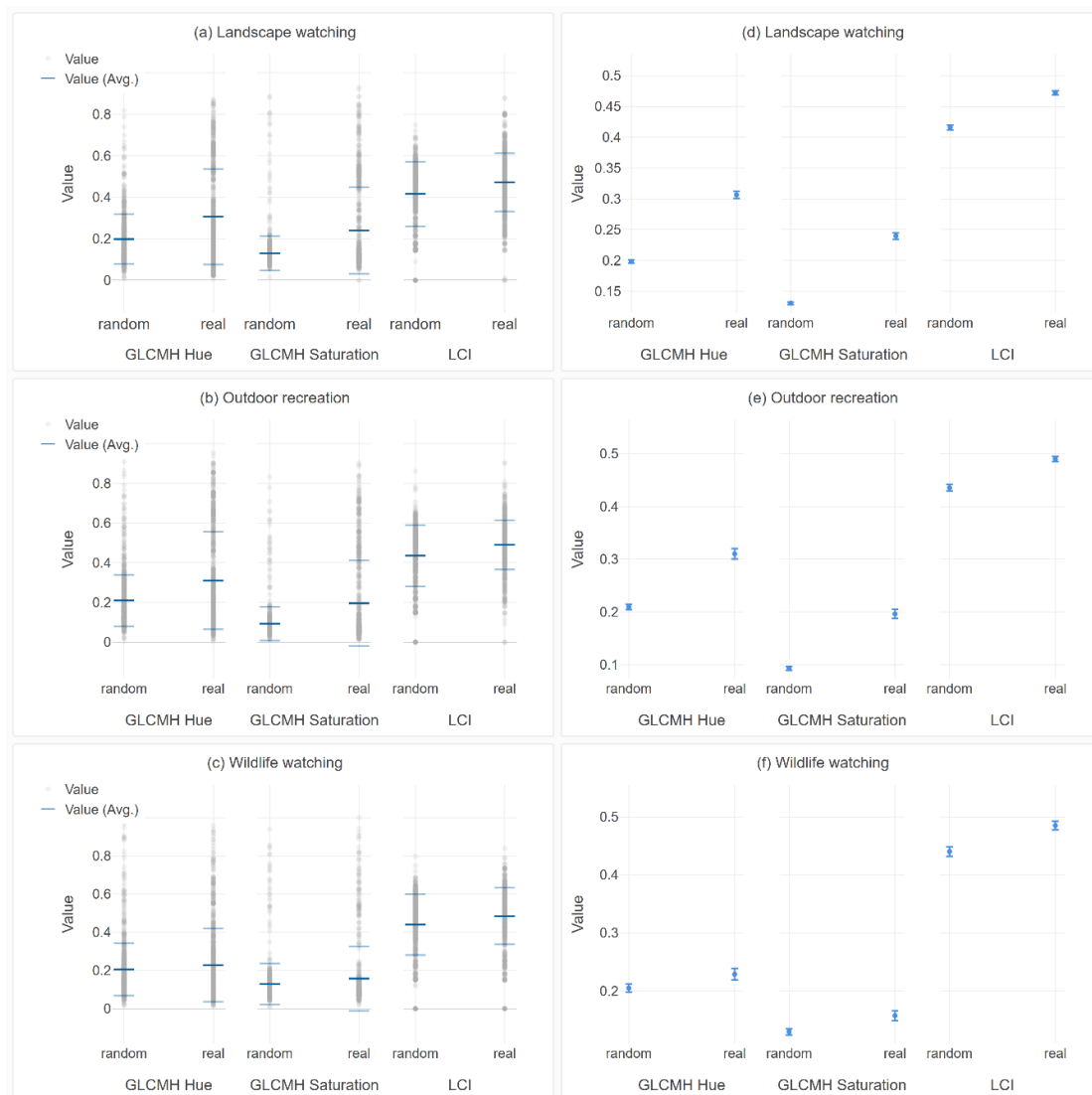


Figure A1. Complementary visualisation plots of the Wilcoxon test results. Scatters: (a) landscape watching; (b) outdoor recreation; (c) wildlife watching. Error plots: (d) landscape watching; (e) outdoor recreation; (f) wildlife watching.

References

1. Saint-Marc, P. *The Socialization of the Environment*; Stock: Paris, France, 1971.
2. Costanza, R.; D'Arge, R.; De Groot, R.; Farber, S.; Grasso, M.; Hannon, B.; Limburg, K.; Naeem, S.; O'Neill, R.V.; Paruelo, J.; et al. The value of the world's ecosystem services and natural capital. *Nature* **1997**, *387*, 253–260. [[CrossRef](#)]
3. Finlayson, M.; Cruz, R.D.; Davidson, N.; Alder, J.; Cork, S.; de Groot, R.S.; Lévêque, C.; Milton, G.R.; Peterson, G.; Pritchard, D.; et al. *Millennium Ecosystem Assessment: Ecosystems and Human Well-Being: Wetlands and Water Synthesis*; Island Press: Washington, DC, USA, 2005.
4. Potschin, M.B.; Haines-Young, R.H. Ecosystem services: Exploring a geographical perspective. *Prog. Phys. Geogr.* **2011**, *35*, 575–594. [[CrossRef](#)]
5. Wu, J. Landscape sustainability science: Ecosystem services and human well-being in changing landscapes. *Landsc. Ecol.* **2013**, *28*, 999–1023. [[CrossRef](#)]
6. Musacchio, L.R. Key concepts and research priorities for landscape sustainability. *Landsc. Ecol.* **2013**, *28*, 995–998. [[CrossRef](#)]

7. Plieninger, T.; Bieling, C.; Fagerholm, N.; Byg, A.; Hartel, T.; Hurley, P.; López-Santiago, C.A.; Nagabhatla, N.; Oteros-Rozas, E.; Raymond, C.M.; et al. *The Role of Cultural Ecosystem Services in Landscape Management and Planning*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 14, pp. 28–33.
8. Milcu, A.I.; Hanspach, J.; Abson, D.; Fischer, J. Cultural Ecosystem Services: A Literature Review and Prospects for Future Research. *Ecol. Soc.* **2013**, *18*, art44. [[CrossRef](#)]
9. Dickinson, D.C.; Hobbs, R.J. Cultural ecosystem services: Characteristics, challenges and lessons for urban green space research. *Ecosyst. Serv.* **2017**, *25*, 179–194. [[CrossRef](#)]
10. Tew, E.R.; Simmons, B.I.; Sutherland, W.J. Quantifying cultural ecosystem services: Disentangling the effects of management from landscape features. *People Nat.* **2019**, *1*, 70–86. [[CrossRef](#)]
11. Kopperoinen, L.; Luque, S.; Tenerelli, P.; Zulian, G.; Viinikka, A. 5.5. 3. Mapping cultural ecosystem services. *Mapp. Ecosyst. Serv.* **2017**, 197–209.
12. Figueroa-Alfaro, R.W.; Tang, Z. Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. *J. Environ. Plan. Manag.* **2017**, *60*, 266–281. [[CrossRef](#)]
13. Díaz, S.; Demissew, S.; Carabias, J.; Joly, C.; Lonsdale, M.; Ash, N.; Larigauderie, A.; Adhikari, J.R.; Arico, S.; Baldi, A.; et al. The IPBES Conceptual Framework—Connecting nature and people. *Curr. Opin. Environ. Sustain.* **2015**, *14*, 1–16. [[CrossRef](#)]
14. Pascual, U.; Balvanera, P.; Díaz, S.; Pataki, G.; Roth, E.; Stenseke, M.; Watson, R.T.; Başak Dessane, E.; Islar, M.; Kelemen, E.; et al. Valuing nature’s contributions to people: The IPBES approach. *Curr. Opin. Environ. Sustain.* **2017**, *26*, 7–16. [[CrossRef](#)]
15. Martín-López, B.; Barton, D.N.; Gomez-Baggethun, E.; Boeraeve, F.; McGrath, F.L.; Vierikko, K.; Geneletti, D.; Sevecke, K.J.J.; Pipart, N.; Primmer, E.; et al. A new valuation school: Integrating diverse values of nature in resource and land use decisions. *Ecosyst. Serv.* **2016**, *22*, 213–220.
16. Calcagni, F.; Amorim Maia, A.T.; Connolly, J.J.T.; Langemeyer, J. Digital co-construction of relational values: Understanding the role of social media for sustainability. *Sustain. Sci.* **2019**, *14*, 1309–1321. [[CrossRef](#)]
17. Haines-Young, R.; Potschin, M.B. *Common International Classification of Ecosystem Services (CICES) V5. 1 and Guidance on the Application of The revised Structure*; Fabis Consult. Ltd.: Nottingham, UK, 2018; Volume 53.
18. Dunford, R.; Harrison, P.; Smith, A.; Dick, J.; Barton, D.N.; Martin-Lopez, B.; Kelemen, E.; Jacobs, S.; Saarikoski, H.; Turkelboom, F.; et al. Integrating methods for ecosystem service assessment: Experiences from real world situations. *Ecosyst. Serv.* **2018**, *29*, 499–514. [[CrossRef](#)]
19. La Rosa, D.; Spyra, M.; Inostroza, L.; Rosa, D.L.; Spyra, M.; Inostroza, L. *Indicators of Cultural Ecosystem Services for Urban Planning: A Review*; Elsevier B.V.: Amsterdam, The Netherlands, 2016; Volume 61, pp. 74–89.
20. Bachi, L.; Ribeiro, S.C.; Hermes, J.; Saadi, A. Cultural Ecosystem Services (CES) in landscapes with a tourist vocation: Mapping and modeling the physical landscape components that bring benefits to people in a mountain tourist destination in southeastern Brazil. *Tour. Manag.* **2020**, *77*, 104017. [[CrossRef](#)]
21. Hausmann, A.; Toivonen, T.; Slotow, R.; Tenkanen, H.; Moilanen, A.; Heikinheimo, V.; Di Minin, E. Social Media Data Can Be Used to Understand Tourists’ Preferences for Nature-Based Experiences in Protected Areas. *Conserv. Lett.* **2018**, *11*, e12343. [[CrossRef](#)]
22. Wood, S.A.; Guerry, A.D.; Silver, J.M.; Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* **2013**, *3*, 2976. [[CrossRef](#)]
23. Van Zanten, B.T.; Van Berkel, D.B.; Meentemeyer, R.K.; Smith, J.W.; Tieskens, K.F.; Verburg, P.H. Continental-scale quantification of landscape values using social media data. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12974–12979. [[CrossRef](#)]
24. Oteros-Rozas, E.; Martín-López, B.; Fagerholm, N.; Bieling, C.; Plieninger, T. Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecol. Indic.* **2018**, *94*, 74–86. [[CrossRef](#)]
25. Langemeyer, J.; Calcagni, F.; Baró, F. Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. *Land Use Policy* **2018**, *77*, 542–552. [[CrossRef](#)]
26. Tenerelli, P.; Demšar, U.; Luque, S. Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecol. Indic.* **2016**, *64*, 237–248. [[CrossRef](#)]
27. Tieskens, K.F.; Van Zanten, B.T.; Schulp, C.J.E.; Verburg, P.H. Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landsc. Urban Plan.* **2018**, *177*, 128–137. [[CrossRef](#)]

28. Sharp, R.; Tallis, H.T.; Ricketts, T.; Guerry, A.D.; Wood, S.A.; Chaplin-Kramer, R.; Nelson, E.; Ennaanay, D.; Wolny, S.; Olwero, N.; et al. *InVEST 3.6.0 User's Guide*; Stanford University: Stanford, CA, USA, 2018.
29. Mancini, F.; Coghill, G.M.; Lusseau, D. Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. *PLoS ONE* **2018**, *13*, e0200565. [[CrossRef](#)] [[PubMed](#)]
30. Lee, H.; Seo, B.; Koellner, T.; Lautenbach, S. Mapping cultural ecosystem services 2.0—Potential and shortcomings from unlabeled crowd sourced images. *Ecol. Indic.* **2019**, *96*, 505–515. [[CrossRef](#)]
31. Richards, D.R.; Tunçer, B. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosyst. Serv.* **2018**, *31*, 318–325. [[CrossRef](#)]
32. Gosal, A.S.; Geijzendorffer, I.R.; Václavík, T.; Poulin, B.; Ziv, G. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosyst. Serv.* **2019**, *38*, 100958. [[CrossRef](#)]
33. Kaplan, R.; Kaplan, S. *The Experience of Nature: A Psychological Perspective*; Cambridge University Press: Cambridge, UK, 1989; ISBN 0521341396.
34. Karasov, O.; Vieira, A.A.B.; Külvik, M.; Chervanyov, I. Landscape coherence revisited: GIS-based mapping in relation to scenic values and preferences estimated with geolocated social media data. *Ecol. Indic.* **2020**, *111*, 105973. [[CrossRef](#)]
35. Sullivan, R.G.; Meyer, M.E. Environmental Reviews and Case Studies: The National Park Service Visual Resource Inventory: Capturing the Historic and Cultural Values of Scenic Views. *Environ. Pract.* **2016**, *18*, 166–179. [[CrossRef](#)]
36. Karasov, O.; Külvik, M.; Chervanyov, I.; Priadka, K. Mapping the extent of land cover colour harmony based on satellite Earth observation data. *GeoJournal* **2019**, *84*, 1057–1072. [[CrossRef](#)]
37. Kemp, S. Kepios Team Digital 2019: Estonia. Available online: <https://datareportal.com/reports/digital-2019-estonia?rq=estonia> (accessed on 29 January 2020).
38. Santos-Martin, F.; Viinikka, A.; Mononen, L.; Brander, L.M.; Vihervaara, P.; Liekens, I.; Potschin-Young, M. Creating an operational database for ecosystems services mapping and assessment methods. *One Ecosyst.* **2018**, *3*, e26719. [[CrossRef](#)]
39. OpenStreetMap Contributors Planet Dump. Available online: <https://planet.openstreetmap.org/> (accessed on 9 April 2020).
40. Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data mining toolbox in python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
41. Karasov, O.; Külvik, M.; Burdun, I. Deconstructing landscape pattern: Applications of remote sensing to physiognomic landscape mapping. *GeoJournal* **2019**, 1–27. [[CrossRef](#)]
42. Ou, L.-C.; Yuan, Y.; Sato, T.; Lee, W.-Y.; Szabó, F.; Sueeprasan, S.; Huertas, R. Universal models of colour emotion and colour harmony. *Color Res. Appl.* **2018**, *43*, 736–748. [[CrossRef](#)]
43. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, *6*, 610–621. [[CrossRef](#)]
44. Hall-Beyer, M. GLCM Texture: A Tutorial v. 3.0. Available online: <https://doi.org/10.13140/rg.2.2.12424.21767> (accessed on 17 May 2020).
45. Schloss, K.B.; Palmer, S.E. Aesthetic response to color combinations: Preference, harmony, and similarity. *Atten. Percept. Psychophys.* **2011**, *73*, 551–571. [[CrossRef](#)]
46. Antrop, M.; Van Eetvelde, V. Basic Concepts of a Complex Spatial System. In *Landscape Perspectives: The Holistic Nature of Landscape*; Springer: Dordrecht, The Netherlands, 2017; pp. 81–101.
47. Lutsenko, E.V. Conceptual principles of the system (emergent) information theory and its application for the cognitive modelling of the active objects (entities). In Proceedings of the IEEE International Conference on Artificial Intelligence Systems, ICAIS, Divnomorskoe, Russia, 5–10 September 2002; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2002; pp. 268–269.
48. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [[CrossRef](#)]
49. Sahraoui, Y.; Vuidel, G.; Joly, D.; Foltête, J.C. Integrated GIS software for computing landscape visibility metrics. *Trans. GIS* **2018**, *22*, 1310–1323. [[CrossRef](#)]

50. Copernicus Land Monitoring Service EU-DEM v1.1—Copernicus Land Monitoring Service. Available online: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata> (accessed on 13 September 2018).
51. Van Berkel, D.B.; Tabrizian, P.; Dorning, M.A.; Smart, L.; Newcomb, D.; Mehaffey, M.; Neale, A.; Meentemeyer, R.K. Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. *Ecosyst. Serv.* **2018**, *31*, 326–335. [[CrossRef](#)]
52. Ghermandi, A.; Sinclair, M. Passive crowdsourcing of social media in environmental research: A systematic map. *Glob. Environ. Chang.* **2019**, *55*, 36–47. [[CrossRef](#)]
53. Cao, Y.; Wu, Y.; Zhang, Y.; Tian, J. Landscape pattern and sustainability of a 1300-year-old agricultural landscape in subtropical mountain areas, Southwestern China. *Int. J. Sustain. Dev. World Ecol.* **2013**, *20*, 349–357. [[CrossRef](#)]
54. Burkhard, B.; Maes, J.; Potschin-Young, M.B.; Santos-Martín, F.; Geneletti, D.; Stoev, P.; Kopperoinen, L.; Adamescu, C.M.; Adem Esmail, B.; Arany, I.; et al. Mapping and assessing ecosystem services in the EU—Lessons learned from the ESMEALDA approach of integration. *One Ecosyst.* **2018**, *3*, e29153. [[CrossRef](#)]
55. Kim, Y.; Kim, C.K.; Lee, D.K.; Lee, H.W.; Andrada, R.I.T. Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tour. Manag.* **2019**, *72*, 249–256. [[CrossRef](#)]
56. Tenkanen, H.; Di Minin, E.; Heikinheimo, V.; Hausmann, A.; Herbst, M.; Kajala, L.; Toivonen, T. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Sci. Rep.* **2017**, *7*, 17615. [[CrossRef](#)] [[PubMed](#)]
57. Yoshimura, N.; Hiura, T. Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosyst. Serv.* **2017**, *24*, 68–78. [[CrossRef](#)]
58. Martínez Pastur, G.; Peri, P.L.; Lencinas, M.V.; García-Llorente, M.; Martín-López, B. Spatial patterns of cultural ecosystem services provision in Southern Patagonia. *Landsc. Ecol.* **2016**, *31*, 383–399. [[CrossRef](#)]
59. Statistics Estonia. The Majority of Enterprises use Information and Communication Technology (ICT) security measures—Statistics Estonia. Available online: <https://www.stat.ee/news-release-2019-111> (accessed on 7 February 2020).
60. Dunkel, A. Visualizing the perceived environment using crowdsourced photo geodata. *Landsc. Urban Plan.* **2015**, *142*, 173–186. [[CrossRef](#)]
61. Hermes, J.; Van Berkel, D.; Burkhard, B.; Plieninger, T.; Fagerholm, N.; von Haaren, C.; Albert, C. Assessment and valuation of recreational ecosystem services of landscapes. *Ecosyst. Serv.* **2018**, *31*, 289–295. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).