Forest cultural ecosystem services assessment using social media data

by

Federico Lingua

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2023

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Forest cultural ecosystem services assessment using social media data

Submitted by Federico Lingua in partial fulfillment of the requirements for the degree

of Doctor of Philosophy

in Forestry

Examining Committee:

Nicholas C. Coops, Professor, Forestry, UBC

Supervisor

Verena C. Griess, Professor, Environmental Systems Science, Eidgenössische Technische Hochschule

Supervisory Committee Member

Valentine Lafond, Postdoctoral Researcher, Forestry, UBC

Supervisory Committee Member

Terry Sunderland, Professor, Forestry, UBC

University Examiner

Tamara Munzner, Professor, Computer Science, UBC

University Examiner

Additional Supervisory Committee Members:

Christopher Gaston, Associate Professor, Forestry, UBC

Supervisory Committee Member

Abstract

To achieve comprehensive sustainable forest management the entire suite of ecosystem services that forests provide must be considered. Centuries of research have equipped forest managers with numerous tools to assess the value of timber, the main focus of most management plans in many countries, while cultural ecosystem services (CES) are often neglected. This is partly due to the challenges that CES pose given the nature of non-marketable goods. The ubiquitous approach to study CES is through the use of *in-situ* surveys, which are expensive and time-consuming. The overarching objective of this thesis is to explore new ways to characterize CES and develop methodologies for forest managers to integrate CES into forest management plans. To this end, a novel data source is exploited throughout this dissertation: geotagged social media images, acquired by visitors to British Columbia's forested areas and provincial parks. These images and their metadata were crowdsourced from Flickr, analyzed with machine learning and deep learning techniques, and combined with various traditional economic methods such as travel cost and benefit transfer analysis. In doing so, it was possible to characterize, map, and estimate the value of the CES provided by forested areas across British Columbia. The number of recreational visits in BC's forests was estimated to be over 44 million per year, and the most frequently identified activities were hiking (30.2%), skiing (16.1%), and water related activities such as fishing and kayaking (13.1%). Findings suggest that crowdsourced social media images are a useful, versatile, and unique data source. If used appropriately, they have the potential to tackle the mapping and valuing issues that historically limited the inclusion of CES in forest management plans.

Lay Summary

Forests provide society with a wide range of services from timber to recreation. Forest management plans, however, often neglect non-marketable values of forests because these are difficult to map and value and typically require researchers to resort to costly and time-consuming surveys. Images shared on social media represent a low-cost and easily accessible data source to acquire insights into the recreational use of forests. The objective of my thesis is to explore the opportunities that arise from the analysis of images for the study of nature-based recreation, focusing on the forested areas of British Columbia. I found that it is possible to effectively use social media images to characterize, map and value forest recreation, by combining the latest development in the machine learning field and traditional economic theory techniques.

Preface

The research questions and objectives of this dissertation were originally conceived from a discussion between my supervisory committee and myself. Chapters of this dissertation are derived from co-authored studies in peer-reviewed journals. Co-authors Nicholas C. Coops, Verena C. Griess, Valentine Lafond, Christopher Gaston and Richard C. Hamelin were responsible for providing guidance and editing in the manuscript preparation phase. I performed the primary research, data analysis, interpretation, and preparation of final manuscripts. Main findings were published in 4 peer-reviewed research articles, which correspond to the following chapters:

Chapter 1: **Lingua, F.**, Coops, N.C., Lafond, V., Gaston, C., Hamelin, R. C., Griess, V.C. In prep. The evolution of the economics of forest biotic disturbances: outlining the future from the learnings of the past.

Chapter 3: **Lingua, F.**, Coops, N.C., Lafond, V., Gaston, C., Griess, V.C. 2022. Characterizing, mapping, and valuing the demand for forest recreation using crowdsourced social media data. Plos one, 17(8), e0272406

Chapter 4: **Lingua, F.**, Coops, N.C., Griess, V.C. 2022. Valuing Cultural Ecosystem Services Combining Deep Learning and Benefit Transfer Approach. Ecosystem Services 58 (2022): 101487.

Chapter 5: **Lingua, F.**, Coops, N.C., Griess, V.C. Under review. Assessing the potential and the value of the recreational ecosystem services provided by forests using deep learning and social media images.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API - Application Programming Interface

AUD - Activity User Day

BC - British Columbia

CAN$ - Canadian Dollars

CES - Cultural Ecosystem Services

CNN - Convolutional Neural Network

CS - Consumer Surplus

CV - Contingent Valuation

ES - Ecosystem Services

GADM - Database of Global Administrative Areas

GDP - Gross Domestic Product

GHII – Global Human Interference Index

LUD - Landscape User Day

MEA - Millennium Ecosystem Assessment

OLS - Ordinary Least Squares

OSM - Open Street Map

PUD - Photo User Day

TSA - Timber Supply Area

Acknowledgments

During my scientific journey, I was lucky enough to have the guidance of two supervisors, Verena Griess and Nicholas Coops. This research would not have been possible without their continuous support. I would like to thank Verena for nourishing my intellectual curiosity, allowing me plenty of freedom to perusing my interests during my first two years; and I would like to thank Nicholas for being so thoughtful and responsive and for being instrumental in helping me in stay afloat during the publication process of my research. A special thanks goes to Valentine Lafond and Chris Gaston who, as members of my committee, provided me with plenty of insights and food for thought. Lastly thank you to my family, especially my wife Sara and my daughters Cecilia and Matilde, for their endless love and understanding.

To my mother, to my father, to my sister and to my brother. To my wife, and to my daughters. I will love you, until a fragment of the universe will retain my shape

Chapter 1
Foundation for research

This chapter establishes the context for the greater body of work within this dissertation and outlines overarching objectives and research questions. Notable examples of the ecosystem services provided by forests are illustrated, as well as the traditional economic approaches to assess their value. Furthermore, the limitations and opportunities that arise from the social media data for the study of cultural ecosystem services are presented.

1.1 Ecosystem services

Forests and their associated goods and services have always been vital for humanity. Forest ecosystems provide food, filtration of fresh water, wood fuel, and building materials, as well as aesthetic, recreational, and spiritual experiences (Brockerhoff et al., 2017). The growing awareness of the importance of the services that forests and other ecosystems provide culminated in the UN-led Millennium Ecosystems Assessment (MEA, 2001), which ultimately identified and classified them into four categories: supporting, regulating, cultural, and provisioning ecosystem services (Figure 1). Another major outcome of the MEA was highlighting the importance of performing assessments of the economic value of the services provided by ecosystems worldwide. Since ecosystem services (ES) are often devoid of market value, they are often not adequately represented in the context of policy decisions. In a sense, the economic value of ecosystem services is indefinite since the worldwide economies would stop without them. However, for sustainable management of ecosystems, it is useful to estimate the marginal value of ecosystem services, i.e., the estimated change of value compared with a change in the quantity of the service provided (Costanza et al., 1997).

The economic valuation of ecosystem services is particularly important in the context of forested ecosystems that provide an array of services that must be thoroughly understood and recognized to create sustainable management plans (Ritter and Dauksta, 2013). In the next sections, notable examples of forest ES will be provided for each one of these four categories. These forest ES were selected based on their economic importance assessed by Krieger (2001).

1.1.1 Supporting forest ES

Supporting services are those that are needed for the provision of all the other services and they only indirectly impact humans. Supporting forest ES includes photosynthesis, nutrient cycling, biological diversity, and others. As for Krieger (2001) among the supporting forest ES the most economically valuable for human society is biological diversity. Biological diversity provides the genetic material that could potentially be used to improve the productivity and quality of commercial crops and livestock in the future. Furthermore, the biodiversity that forested ecosystems support and protect is an important genetic reservoir for the development of future products, such as wood derivates, drugs and technologies aimed at the improvement of the quality and the length of human life. Lastly, some animal species that depend on forests as their habitat are beneficial to the agricultural sector by acting as a biological control for pests (Daily, 1997) and as pollinators (Nabhan and Buchmann, 1997).

1.1.2 Regulating forest ES

Regulating services are the benefits that humans enjoy from the regulatory processes that occur in ecosystems. Examples of regulating forest ES include soil protection and stabilization, the improvement of air quality, climate regulation or carbon sequestration. Among the regulating services, the most valuable include climate regulation and carbon sequestration (Krieger, 2001). Forests influence the temperature of the surface of our planet by trapping moisture, hence lowering the earth's surface temperature (Costanza and Folke, 1997). Furthermore, trees significantly contribute to the capture of carbon dioxide, reducing greenhouse gas (GHG) concentration in the atmosphere (Albrecht and Kandji, 2003).

1.1.3 Provisioning forest ES

Provisioning ecosystem services are the products that humans extract from ecosystems. Numerous goods and products are sourced from forests worldwide, including timber, berries, mushrooms, or game meat. The provisioning ES that yields the highest economic benefit to society is timber (Krieger, 2001). The economic importance of timber can be fully appraised considering the share of gross national product and employment generated by the various timber-based activities, such as forest management, harvesting, manufacturing, and construction. As a whole, this sector contributes to over 2% of the world's gross domestic product (GDP) (FAO, 2020). As concerns Canada latest available statistics, the forest sector contributes 4.11% to its GDP in 2020, while in British Columbia it generates over 42,000 jobs and 2.2% of the provincial GDP (Statista, 2021).

1.1.4 Cultural forest ES

Lastly, cultural ecosystem services are the nonmaterial benefits that humans obtain from ecosystems. Examples of cultural ES include the aesthetic and spiritual experiences that people can enjoy in forested areas. In particular, recreational activities in forested ecosystems have shown the potential of mitigating several urban health concerns, such as lack of exercise and obesity (Rosenberger et al., 2005), as well as mental health problems (Bielinis et al., 2019). For this reason, recreation plays a role of growing importance in the array of ecosystem services that forests provide to humans

(Wilkes-Allemann et al., 2015). Therefore, providing access to forest recreational opportunities is of increasing importance on political agendas worldwide (Bell et al., 2009).

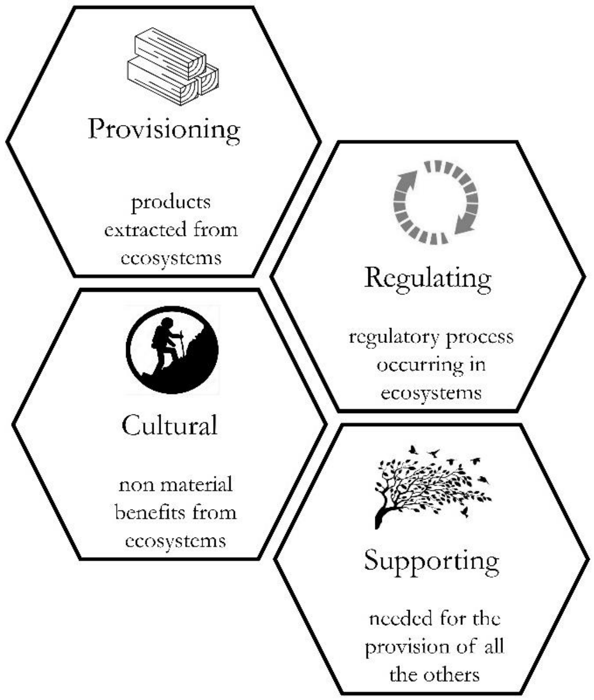

Figure 1. Categories of ecosystem services and their definitions.

1.1.5 Economic evaluation of forest ecosystem services

Economic evaluations of ES are designed to assess the change in human economic well-being, due to changes in the quantity or quality of the service considered. Some forest ES can be freely exchanged in markets (marketable ES), in which case the market price is a direct measure of the benefit that humans get from a unit of the service. However, most forest ES cannot be exchanged on markets and are therefore devoid of market prices (non-marketable ES). Cultural ES belong to the latter category.

Researchers have been developing and applying a set of economic techniques to estimate the monetary value of ES. Below, I will illustrate the most frequently used techniques to assess the value of the forest non-market ecosystem services.

1.2 Evaluation techniques for non-market ES

There have been excellent reviews on the economic techniques to assess the monetary value of nonmarketable ES, including (Rosenberger et al., 2012; Holmes et al., 2014; Champ et al., 2003). In particular, in Champ et al. (2003), the methods are illustrated with practical examples of their applications and it was used for the basis of the discussion below. To estimate the value of nonmarketable services, economists observe individuals' decisions both in hypothetical and actual markets. Techniques used to evaluate nonmarketable ES measuring behavioral data in hypothetical markets are called stated preference, whereas techniques used to evaluate nonmarketable ES in actual markets are called revealed preference. Forest cultural ES have been evaluated using both stated and revealed preference techniques. In the following sections, I will describe the two approaches and provide examples of studies that have adopted them.

1.2.1 Stated preference techniques

Stated preference techniques are used to infer people's preferences by asking them to choose between different hypothetical alternatives. The most frequently used technique to evaluate cultural ES is the contingent valuation method. Contingent valuation (CV) relies on surveys that explore the number of monetary units that people would be willing to pay for a modification in the quantity or the quality of environmental good, or ES. Since its introduction in a study on big game conducted by Davis (1963), CV is one of the most used techniques to value CES, because it is the only assessment technique available when markets for an ES do not exist, and the revealed preference methods are not relevant. In practice, CV is conducted asking the public questions such as: "*How much would you be willing to pay to avoid a forest pest outbreak in your region?*" (Chang et al., 2011); "*Suppose that managers treat a forest to make it healthy in 100 years, how much would you be willing to pay to fund this program?*" (Meldrum et al., 2013). The single bid, placed by the respondents, is a direct consequence of the hypothetical changes in the resource quality or quantity (Rosenberger et al., 2012).

1.2.2 Revealed preference techniques

In contrast to stated preference techniques, revealed preference techniques are used to infer people's preferences based on the observation of decisions made in actual markets for goods whose value is determined by nonmarketable ES. The main revealed preference methods that have been used to estimate the value of forest cultural ES are the removal and replacement cost method, the hedonic price method, the travel cost method, and the benefit transfer method.

The conceptual assumption behind the removal and replacement cost method is that the value of a non-marketable ES is equal to the costs necessary for its substitution (Dixon et al., 2013). For example, in the case of the aesthetic value of trees, the reference cost is the cost that the owner would incur to replace trees affected by a disturbance.

The hedonic price method relies on the assumption that the price of any good is determined by several internal and external characteristics of the good itself. When applying this method, economists assess the relative contribution of each one of these characteristics (Pearson et al., 2002). In the context of the forest's CES the hedonic price method was mainly used to estimate the value of the aesthetic ES provided by urban forests, using housing market data (Holmes et al., 2006) (Price et al., 2010). Specifically, the assumption is that given two identical houses, except for the presence of a forest view from one, the aesthetic value of the forest is the difference in value between the two houses. Therefore, using regression analyses on house prices, it is possible to determine the marginal contribution (implicit price) of the forest.

The travel cost method estimates the demand curve for a specific recreational site using variations in visitor travel costs to that location. Typically, the total cost of a trip includes the cost of traveling, access fees, equipment costs, and the time cost. While equipment costs and access fees are fixed, the traveling and time costs are a function of the distance traveled. Therefore, using this approach, the most influential variable in the assessment of the economic value of a recreational site is the distance traveled by visitors to get there. Besides being used to assess the value of a site, the travel cost method can also be applied to assess the impacts of disturbances (Loomis and Walsh, 1997). To do so it is assumed that the disturbance causes a downward shift of the demand curve for the considered recreational site. This, in turn, causes a contraction of the consumer surplus (see also partial equilibrium). With the travel cost method, the impacts of forest biotic disturbance are considered equal to the reduction in consumer surplus generated by the disturbance itself.

The basic premise behind the benefit transfer approach is that the economic benefit of a non-marketable service provided by an ecosystem can be estimated using values found in original studies conducted elsewhere (Loomis, 1992). For example, the aggregate benefit to users who come to forests for recreational purposes can be estimated by multiplying: (i) the consumer surplus per activity day that accrues to an individual engaged in a certain recreational activity; by (ii) the number of people engaging in that recreational activity in the considered area.

1.2.3 Limits of traditional techniques and alternatives

The economic techniques that have traditionally been applied to study forest CES have limitations that have been thoroughly analyzed by many authors (e.g. Champ et al., 2003; Holmes et al., 2014; Rosenberger et al., 2012).

Concerns about the contingent valuation method and its empirical nature have occurred since its inception (Scott, 1965). When applying this method, there are multiple potential sources of bias that require consideration. In particular, the CV method has been criticized for three main reasons (Cookson, 2003): (i) CV tends to be under-sensitive to the magnitude of the hypothetical benefit considered (Baron and Greene, 1996); (ii) CV tends to inflate the valuation of the specific intervention that respondents are asked about, relative to interventions that respondents are not asked about (Kemp and Maxwell, 1993); (iii) if the ES evaluated is unfamiliar to respondents, then responses are inaccurate (Svedsäter, 2003).

Among the revealed preference methods, the removal and replacement cost method provides no guarantee that the public would be willing to replace an ecosystem service. When there is no proof of the existence of replacement demand, this methodology is not an appropriate approach. Therefore, this method is likely only suitable for urban areas, where the active replacement of trees is a more feasible and realistic assumption.

The use of the hedonic price technique suffers from two main drawbacks. First, it requires a large database of sales transactions for the study area and these data are rarely available. Second, the technique is exclusively suitable for urban forests or wildland-urban interfaces where most of the transactions in the housing sector are made.

The travel cost method's main drawback is the introduction of biases and sampling errors during the surveys. Furthermore, those who place a high value on a recreational site could decide to live nearby. For those people, the travel cost method will systematically underestimate the value of the site. In addition, the process of collecting data *in situ* is time-consuming and costly, liming the number of applications of the method.

Lastly, the reliability of the benefit transfer method is high only when the new and the original study areas are ecologically and socio-economically similar. Furthermore, the accuracy of the values obtained with this approach can only be as good as the one of the initial studies.

Given the limitations of the traditional economic techniques for cultural ES, researchers and ecosystem managers have been exploring new approaches and data sources for the study of forest CES. One of these approaches is the use of crowdsourced social media data (Ciesielski and Stereńczak, 2021). Social media data can broadly be defined as any digital data (such as text, images, videos, or others) that users share on social media platforms. These data can be retrieved from social media application programming interfaces (API) and used to explore human interactions with ecosystems (Wood et al., 2013). In the next section I will illustrate in more detail existing examples of the use of social media images in the study of CES

1.3 Crowdsourced social media data for CES research

During the past decade, hundreds of studies have retrieved and analyzed data from social media to characterize the human-environment relationship (Lee et al., 2019). The most frequently used type of digital data from which researchers crowdsourced data for the study of CES are images shared on social media (Ghermandi and Sinclair, 2019), which can be processed in two broad ways: studies that analyze the metadata of geotagged images and those that analyze the content of the images themselves.

1.3.1 Previous analyses of image metadata

Metadata of images, in particular geotags and timestamps, have been used to different ends. For example, counts of geotagged pictures can be used to infer the number of visits to a natural landscape and hence to map the consumption of forest recreation at a large scale, with high temporal and spatial resolution. In fact, geotagged picture counts have shown a strong correlation, both spatially and temporally, with empirically determined numbers of visits across various natural ecosystems (Sessions et al., 2016; Tenkanen et al., 2017; Wood et al., 2013). Studies that infer the number of visits from counts of geotagged picture adopt regression models calibrated on empirically collected data e.g. (Keeler et al., 2015). More recently, Wood et al. (2020) used linear fixed effects models to estimate the number of visitors in unmonitored recreational sites, using weather conditions, time of year, and the number of social media posts as predictor variables. Crowdsourced social media data can also be successfully used to infer the place of origin of people recreating in the natural landscape, enabling the segmentation of visitors into domestic and international (Sinclair et al., 2020a). Lastly, recent studies conducted on wetlands (Sinclair et al., 2018) and National Parks (Sinclair et al., 2020b) have shown that crowdsourced social media data can be effective in informing traditional revealed preference valuation methods for non-market values (specifically the travel cost method).

1.3.2 Previous analyses of image content

Metadata alone cannot provide insights on why people are attracted to a site, or in which activity recreationists engage. To address this limitation, Richards and Friess (2015) proposed the use of content analysis of social media images. In their methodological framework, applied in four mangrove sites in Singapore, images were manually classified based on the type of CES depicted in them (e.g., landscape appreciation, social recreation, recreational fishing, etc.). The image categorization process enabled the authors to assess variations in CES using the different sites photographed. This research led to further studies that explored CES, relying on the manual classification of images (Heikinheimo et al., 2017; Maia et al., 2020; Pastur et al., 2016; Ros Candeira et al., 2020; Speak et al., 2021).

The manual classification of images however is extremely time-consuming and therefore not suited for the analysis of larger databases. For this reason, researchers turned to the use of artificial intelligence in an attempt to automate the classification process (Callau et al., 2019). Initially, studies used commercial pre-trained image classification models, such as Google Cloud Vision (https://cloud.google.com/vision) and Clarifai (https://www.clarifai.com/models/general-image-recognition) to classify the visual content of images with machine-generated classes (Ghermandi et al., 2022). Richards and Trunçer (2018) used Google Cloud Vision to automatically classify image content, applying approximately 20,000 potential classes. They then clustered the outputs into seven macro-categories: transport, plants, animals, food, people, sports, landscape, and miscellaneous. The automatic classification results were then compared to the results of a manual classification of a subset of images, resulting in an automatic classification accuracy of 85%. Lastly, the authors used outputs of the image classification and clustering process to model and map the probability of occurrence of images depicting animals and plants. This success led to the wider use of Google Cloud Vision in follow-up studies (e.g. Alemu et al., 2021; Ghermandi et al., 2020; Gosal and Ziv, 2020; Runge et al., 2020; Sottini et al., 2019) as well as other software solutions including Clarifai (Depietri and Orenstein, 2020; Karasov et al., 2020; Lee et al., 2019).

1.3.3 Previous studies in forested areas

The use of metadata of social media images in ES research has rapidly expanded, but the discipline is still in its infancy (Ghermandi and Sinclair, 2019). Only few studies have adopted this data source to explore the CES provided by forests. Bernetti et al. (2019) used the metadata of Flickr images to determine the relationship between forest and topographic variables, obtained *via* various remote sensing technologies, and picture density in Tuscany (Italy). Similarly, Ciesielski et al. (2021) used Flickr data and boosted regression tree models to determine which forest variables influenced picture acquisition. These variables were obtained by combining various sources including landcover (Hościło and Tomaszewska, 2014) and terrain information from the Shuttle Radar Topography Mission (Rodriguez et al., 2006). You et al., 2022 assessed the spatial-temporal dynamics of forest recreation values in the Zhejiang Province (China), integrating remote sensing data and text-mining of social media comments. In all the above studies, remote sensing technologies data were used to obtain topographic, socioeconomic, and forest biometrics data coincident with the location (and in some cases time) the social media images were acquired, hence identifying which of these variables had a positive effect on forest recreation. Lastly, Wartmann et al. (2021) explored the potential that social media data have in assessing visitor frequencies in urban and peri-urban forests in Switzerland.

1.3.4 Knowledge gaps

The previous applications of social media data for the studies of forest CES were carried out in Europe and Asia, while studies that have adopted this approach in North America are, based on my literature review, missing. Furthermore, in the existing literature there are no examples of estimates of the value of forest CES starting from social media data. Lastly, in the past social media data have only been used to explore quantitative aspects of forest CES (e.g. number of visitors, hotspots identification, visitation trends, etc.), while qualitative aspects (e.g. type of recreational activities are carried out in forests) have been ignored. As I have shown, social media data are widely used in CES research being unexpensive and quick to gather. However, the outcomes of the methodologies adopted in these studies are not usually implemented in forest planning. Therefore, currently there is a need for the development of new methodologies that could potentially contribute to transform crowdsourced social media data in a viable data source for the management of forest CES.

1.4 Objectives

The above review has confirmed the importance of including cultural ES in forest management decisions, and the existing knowledge gaps related to the use of crowdsourced social media data to examine cultural ES provided by forests. The objective of this thesis is to explore the opportunities that crowdsourced social media images offer for the characterization, mapping, and valuing of forests' cultural ES, and develop methodologies that could be, in the future, be used in forest management plans.

1.5 Research questions

The above objective will be addressed by answering the following research questions:

1. Can metadata of crowdsourced social media images be used to deduce trends of forest visitation and as a data source for the application of the travel cost method?
2. Is it possible to adopt deep-learning techniques to automatically recognize and classify the cultural ecosystem service depicted in social media images?
3. Can the combination of the outcomes of the image classification process and remote sensing technologies data be integrated to explore the recreational potential of forests?

1.6 Dissertation overview

Chapter 2 provides an overview of the study area and the social media data that I collected and analysed. **Chapter 3** is the first research chapter and explores how the metadata of the images can be used to analyze the temporal and seasonal patterns of recreation, estimate the number of annual visits to forests and apply the travel-cost method. In **Chapter 4** the content of the images is automatically analyzed with deep-learning techniques to detect recreational activities and the type of landscapes portrayed and apply an adaptation of the benefit transfer approach to crowdsourced data.

In **Chapter 5** the outcomes of the image classification process are coupled with remote sensing data to model and then map the recreational potential of forest sites. Lastly, **Chapter 6** illustrates the conclusion that can be drawn based on the scientific objectives pursued in this dissertation and highlights the innovative aspects and the limitations relative to this research.

Chapter 2
Site Overview and Data Curation

2.1 Study site

This study focuses on British Columbia's (BC), the westernmost province of Canada (Figure 2). BC is dominated by forested ecosystems, which cover 57% of the land area and are 95% publicly owned (Gilani and Innes, 2020). This province includes the traditional, ancestral and unceded territory of indigenous people. From an economic perspective, timber provisioning has traditionally been the main ecosystem service provided by BC forests. Timber supply areas (TSA) are located throughout the province based on the wood flow patterns from the management units to the main timber-using industries. BC has a system of 689 provincial parks that cover over 140,000 km2, protecting a variety of ecosystems from alpine to coastal biomes. These parks are of great socio-economic importance, and are visited over 23 million times annually. Almost 90% of BC residents have visited a provincial park at some point in time, and about 60% of them recreate in a provincial park each year (Parks, 2018). BC parks provide a variety of ES to humans, and CES are an essential component of this array as stated in BC park´s mission: "(…) *protect representative and special natural places within the province's Protected Areas System for world-class conservation, outdoor recreation, education, and scientific study.*". In the following three chapters, the study area is progressively narrowed. Chapter 3 deals with the CES provided by the entire forested area of BC. Chapter 4 focuses on the CES provided by BC's provincial parks. Lastly, Chapter 5 is concerned with the CES provided by BC's forested areas included in BC's provincial parks system.

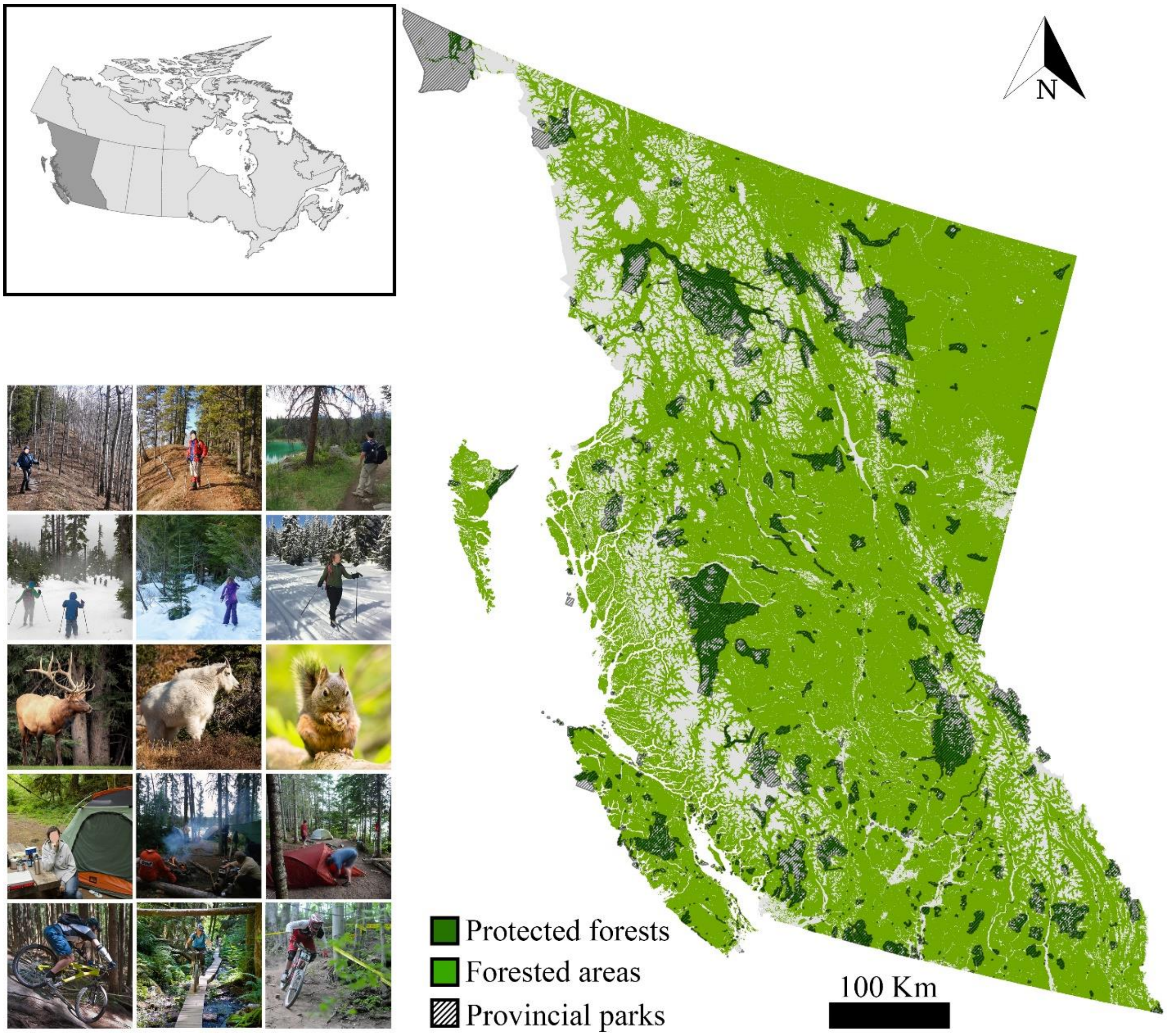

Figure 2. Map of BC forests and the BC Parks system and examples of the images used to study CES.

2.2 Data Overview

The data used throughout this study can be classified into three categories: (i) crowdsourced, (ii) geographic, and (iii) secondary, *i.e.,* data previously collected through primary sources and made available for research (Table 1). Additional data types that were used only in individual chapters are described in the relative materials and methods sections.

2.1.1 Crowdsourced Data

I developed a Python application and extracted the metadata of all the geotagged pictures acquired in BC between January 2005 and December 2020 from the Flickr Application Programming Interface (API). Flickr is the most widely used social media website in the study of nature-based recreation (Ghermandi and Sinclair, 2019) because of its easily accessible API. Moreover, Flickr was launched in 2004, currently offering the longest series of data among the photo-sharing social media platforms. Data were collected following the terms and conditions of Flickr's API. Furthermore, only the pictures uploaded by users that authorized their photos to be accessed by API-based searches done outside of the Flickr website were accessed and included in the analysis. Lastly, all the metadata were anonymized and analyzed in aggregate form.

Retrieved metadata included: (i) unique identifier code for the picture, (ii) unique identifier code for the author of the picture, (iii) the date on which the picture was taken, (iv) the coordinates of where the picture was taken.

2.1.2 Geographic Data

Geographic data are shapefiles defining: (i) the administrative boundaries of BC, (ii) the land cover map of BC, and (iii) the boundaries of BC provincial parks. The shapefiles are publicly accessible and were downloaded from BC's data catalogue (https://catalogue.data.gov.bc.ca/).

2.1.3 Secondary Data

Secondary data used in this study were: (i) visitation statistics for BC parks collected by the BC Ministry of Environment and Climate Change (Parks, 2018), and (ii) global human influence index (GHII) (Wildlife Conservation Society - WCS and Center for International Earth Science Information Network - CIESIN - Columbia University, 2005). Visitation statistics are included in the end of year reports that BC parks published annually from the fiscal year 2007/2008 until the fiscal year 2017/2018. These reports contain detailed park attendance and revenue tables and graphs, satisfaction survey information and financial tables. GHII (https://sedac.ciesin.columbia.edu) is a global dataset of 1-kilometer grid cells, created from different layers (accounting for human population pressure, land use, infrastructure, and human access) that provides an updated map of anthropogenic impacts useful for human-environment interactions research.

Table 1. Typologies and sources of the data used throughout the dissertation.

Description	**Typology**	**Source**
Geotagged pictures acquired in BC	Crowdsourced Data	Flickr API
BC provincial park boundaries	Geographic data	BC data catalogue
BC forested area	Geographic data	BC data catalogue
BC park boundaries	Geographic Data	BC data catalogue
Visitation data for BC parks	Secondary Data	BC Parks end of year reports
Global Human Influence Index	Secondary Data	SEDAC (NASA)

Chapter 3
Characterizing Mapping and Valuing the Demand for Forest Recreation Using Crowdsourced Social Media Data.

3.1 Background and Motivation

In this chapter I will explore the opportunities that the metadata of geotagged social media images offer in the study of forest cultural ecosystem services. Specifically, I will address the knowledge gaps relating to the use of this data source for the characterization, mapping and valuing of recreational values in managed forests. Despite the rapidly growing number of studies that use crowdsourced social media data in environmental research, this discipline is still in its infancy (Ghermandi and Sinclair, 2019), and managed forests have so far been the subject of a very limited number of such studies (Bernetti et al., 2019; Ciesielski and Stereńczak, 2021; Wartmann et al., 2021).

In this context, the overarching objective of our study is to evaluate the potential of crowdsourced social media data in characterizing forest recreation and overcoming the mapping and valuing issues that have historically limited the consideration of cultural ecosystem services in forest management planning. The analysis is structured on three levels. First, Flickr geotagged pictures will be used to characterize demand for recreation in BC forests, identifying temporal and geographic trends, as well as the place of origin of the recreationists. Second, forest recreation in BC will be mapped, assessing the average annual number of visits in BC forests from geotagged pictures. Third, the value of the recreational ecosystem service provided by BC forested provincial parks will be estimated, applying the travel cost method to crowdsourced data. Lastly, in the light of the obtained results, the opportunities and the limitations of this innovative data source in the study of forest recreation and its inclusion in forest management plans will be discussed.

3.2 Data and Methods

3.2.1 Data Overview

As shown in Table 2, along with the data adopted throughout the dissertation, in this chapter we gathered and used various specific data types and sources.

Crowdsourced data

Combining the metadata of the crowdsourced images, Photo-User-Days (PUDs) were derived. PUDs are unique combinations of author and day used to account for the fact that a single social media user can acquire and upload multiple pictures during the same recreational experience (Wood et al., 2013). For each PUD, a reference coordinate was then assigned as the geographic center (mean) of all grouped pictures.

Geographic data

Geographic data used in this study were shapefiles defining: (i) the BC road network downloaded from Open Street Map using the Python package OSMnx; (ii) a grid of 10 km^2 cells covering the entire study area created using the using the tessellation tool in ArcMap.

Secondary data

Secondary data used in this study were: (i) average costs of driving a private car in BC (CAA, 2021), (ii) the estimated hourly income rate from BC census data (British Columbia Government, 2020).

Table 2. Data used to estimate the seasonal trends, number of annual visits of British Columbia's (BC) forests and the value of the recreational ecosystem service. PUD = photo-user-days, API = Application Programming Interface.

Description	Typology	Source
PUDs in BC provincial parks for each Flickr user	Crowdsourced data	Flickr API
PUDs from geotagged pictures taken in BC	Crowdsourced Data	Flickr API
Open Street Map Road Network	Geographic data	https://planet.openstreetmap.org/
Grid with 10 km^2 cells	Geographic Data	ArcMap elaboration
Individual hourly income	Secondary data	BC data catalogue
Average cost of driving a private car (CAN$/km)	Secondary data	https://carcosts.caa.ca

3.2.2 Methods

The methodology that was designed to explore, map and value the recreational ecosystem service provided by BC forests from crowdsourced social media data can be subdivided in three steps. First, the crowdsourced geotagged pictures were used to infer a home location for the visitors of BC forests, as well as to identify temporal and geographic trends of recreational visits. Second, a model to assess the number of annual visits in BC forests based on the number of pictures extracted from Flickr was created. Third, recreational values in BC were estimated by calculating the average consumer surplus for various forested provincial parks combining pictures' metadata with secondary data.

Visitors home location and visitation trends

In many of the previous studies that have inferred social media users' home locations, locations were determined by extracting the information directly from each user's public profile. However, as previously found by Da Rugna et al. (2012), only 40-48% of Flickr users share a home location in their profile. In addition, shared home location may not be up to date. To overcome these issues, home locations were estimated using the max photo-user-days (max PUDs) method proposed by Bojic et al. (2015) and verified by Sinclair et al.(2020). The max PUDs method assumes that the user's home location can be estimated as the one in which they have spent the maximum number of PUDs. This methodology was applied to estimate the home location of each user with at least 10 geotagged pictures shared, while users with less pictures were excluded from further analyses. Retained visitors were divided into: (i) domestic (users having most PUDs within BC), and (ii) international (users having most PUDs outside BC borders). For domestic visitors, the max PUDs method was applied to infer the regional district of origin. Then the coordinates of the home location of each domestic visitor were estimated as the geographic center (mean) of the pictures taken within their regional district of origin. To test the precision of the estimated home locations, the results of the max PUDs method were compared with the available countries of origin declared by users in their profiles. To identify BC hotspots of forest recreation consumption, the optimized hotspot analysis tool in ArcGIS was applied, using the geotags of the retrieved pictures. This tool divides the study area in cells of equal area and then calculates for each cell the probability of it being a hotspot, using the Getis-Ord Gi* statistic (Getis and Ord, 2010). The cells area used by the optimized hotspot analysis tool was 5.4 km^2, calculated by multiplying the median nearest neighbor distance by two, and then using this value to construct a hexagon polygon grid. To analyze temporal trends, the pictures were grouped using their timestamps.

Annual visits in BC forests

To estimate the number of annual visits in BC forests from crowdsourced social media data, we applied a similar methodology to the one applied by Ghermandi (2016) and Sinclair et al. (2019), who estimated the number of visitors in natural systems worldwide and wetlands in Kerala (India), respectively. However, while in these studies the authors used univariate models that only considered PUDs to estimate the number of yearly visitors, in this study a second variable was introduced, the Global Human Influence Index (GHII). The decision to include GHII the model was made to account for the fact that the number of social media pictures acquired in the landscape is influenced by its accessibility (van Zanten et al., 2016). As shown in Figure 3, the statistical relationship between the dependent variable (number of annual visits) and independent variables: (i) number of annual PUDs and, (ii) global human influence index (GHII) was estimated in a calibration sample. Then, this relationship was applied to predict the number of forest visits outside the calibration sample. Specifically, a multivariate ordinary least squares (OLS) model was developed (Eq. 1), and it was calibrated using visitation statistics of BC provincial parks. The rationale of using an OLS regression was that a previous study (Ghermandi, 2016) that compared the performances of various models in estimating the yearly visitation from PUDs determined that forward OLS regression yielded the best results together with the standardized major axis regression. Furthermore, OLS regression was used in (Sinclair et al., 2019) where the authors estimated the number of yearly visitors from PUDs, after calibrating their model in the context of the study region.

Visitation statistics were empirically collected (from 2007 to 2018) by BC Parks and made publicly available in an annual statistics report that contained annual attendance data for each provincial park in BC. In this analysis, exclusively forest-dominated parks were used (forest cover ≥ 50%) to ensure consistency between the calibration sample and the areas in which we applied the OLS model. Therefore, the calibration sample included all the forest-dominated provincial parks in which BC Parks had published annual visitation data from 2007 to 2018 (N=100), and the OLS regression parameters were estimated using PUDs detected in the parks, within the same time window. GHII value for BC provincial parks were estimated averaging the GHII values of each cell that fell into the park boundaries. To this end the zonal statistic ArcGIS pro tool was used. Since the dependent variable cannot be negative, we performed a logarithmic transformation on both the dependent and independent variables. Following (Ghermandi, 2016) the PUDs count in each park is increased by one to avoid attempts to calculate the logarithm of zero in parks in which no geotagged pictures were detected.The OLS model had the following functional form:

Eq. 1 $$log(y_i) = \beta_0 + \beta_1 \, log(x_i) + \beta_2 \, log(x_{ii}) + \epsilon_i$$

where y_i is the average annual attendance for the i^{th} provincial park, and x_i is the average annual number of PUDs and x_{ii} is the average GHII value for the provincial park. Model coefficients were estimated using the Python package Statsmodels. The assumption of normality for the distribution of the residuals was tested with the Shapiro-Wilk test and the hypothesis of covariance between the two independent variables. To predict the number of visits from the log-transformed outputs, both the naïve forecast technique (exponential transformation of the logarithmic predictions) and the Duan's Smearing technique (Duan, 1983) were used. The visit estimates obtained with these two approaches were then compared using the actual parks visitation data to determine which one yielded the most accurate results, using the Root Mean Square Error (RMSE) metric. Using the same metric, the performance of the multivariate OLS regression model was compared with a traditional univariate model (calibrated on the same dataset). This was done to determine if the inclusion of GHII resulted in an improvement of the model accuracy. Lastly, the OLS model was applied to estimate the number of annual visits in each 10 km^2 cell of a grid covering the entire province. The GHII values in each cell were estimated using the zonal statistics tool of ArcGIS pro.

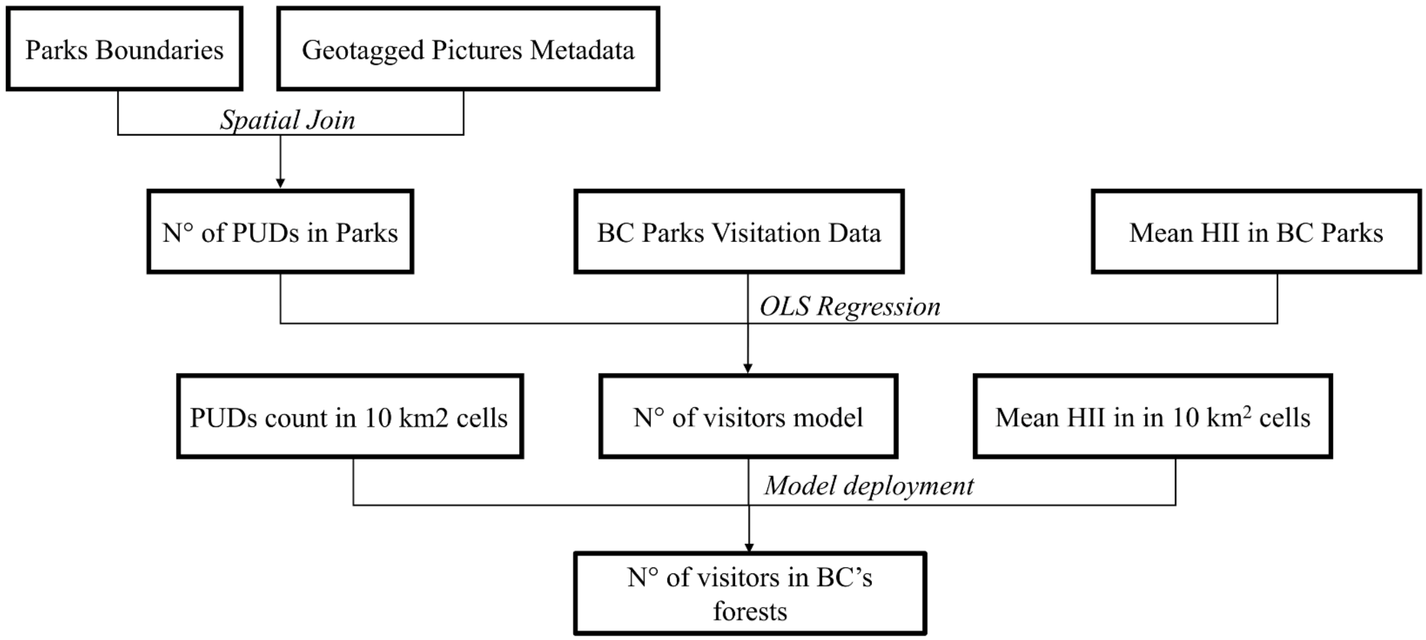

Figure 3. Outline of the method used for estimating the annual visitation rates in BC forests.

Value of forest recreational sites

To assess the monetary value associated with the recreational ecosystem service provided by BC forests, the single-site individual travel cost method was applied to crowdsourced social media data (Figure 4). The travel cost method (TCM) from crowdsourced social media data has been applied in the past by Ghermandi (2016) and Sinclair et al. (2018). The method was applied only in forest dominated (forest cover ≥ 50%) Provincial Parks. Demand functions (Eq. 2 and 3) were estimated using a negative binomial regression, with the number of visits made by each detected visitor as the dependent variable and the following independent variables: (i) total travel cost to the site, (ii) visitor's income, (iii) travel cost to the nearest substitute site. In an effort to reduce the incidence of multi-purpose and multi-day trips, which can bias the results of TCM analyses (Champ et al., 2003), four criteria that each PUD had to respect to be considered a visit were introduced: (i) the author had be a domestic visitor according to the max-PUD methods results, (ii) the distance traveled by the author for the round trip to the site had not exceed 400 km, (iii) the author must not have acquired pictures of the same provincial park for the following seven days, and, (iv) more than 50% of the pictures constituting the PUDs must have been acquired within provincial park boundaries. The total travel cost includes two components: (i) the costs of a round trip to the site, and (ii) the opportunity cost of the time spent traveling. To estimate the cost of traveling to the site, the distance traveled by each visitor was assessed using the Python package OSMnx. This package allows the user to download geospatial data from OpenStreetMap, and then model the drivable urban networks. The travel distance was estimated as the most efficient route between the estimated visitor home location and the coordinates of the PUD. To obtain the cost of traveling, the estimated travel distance was then multiplied by the average cost of driving a private car, retrieved from the Canadian Automobile Association (CAA, 2021). Lastly, the opportunity cost of the time spent traveling was calculated as 1/3 of the estimated hourly income rate derived from BC census data (Table 2), as is commonly done in travel cost analysis (Fezzi et al., 2014). To each park an alternative provincial park was assigned as the park closest to the visitor estimated home location. The travel cost for a round trip to the substitute site was then estimated applying the same travel cost estimation.

Once all the variables were obtained, all provincial parks with less than 40 detected visitors were excluded to ensure that our sample size respected a minimum threshold of 10 users per variable as suggested by Ghermandi (2016). The functional form of the estimated demand function was the following.

Eq. 2
$$log(y_i) = \beta_0 + \beta_{TC}x_{TCi} + \beta_I x_{Ii} + \beta_s x_{si} + \epsilon_i$$

Where y is the number of PUDs by visitor i^{th} for the recreational site considered, β0 is the intercept of the model, x_{TCi} is the cost for a round trip to the recreational site considered for visitor i^{th}, x_{Ii} is the individual income of the visitor i^{th}, and x_{si} is the travel cost for a round trip to the substitute site for visitor i^{th}. Lastly, the consumer surplus generated by a visit to the site was estimated following Creel and Loomis (1990):

Eq. 3

$$CS_{visit} = -\frac{1}{\beta_{TC}}$$

Where CS_{visit} is the consumer surplus generated by each visit to a site and β_{TC} is the coefficient of the travel cost variable in the model.

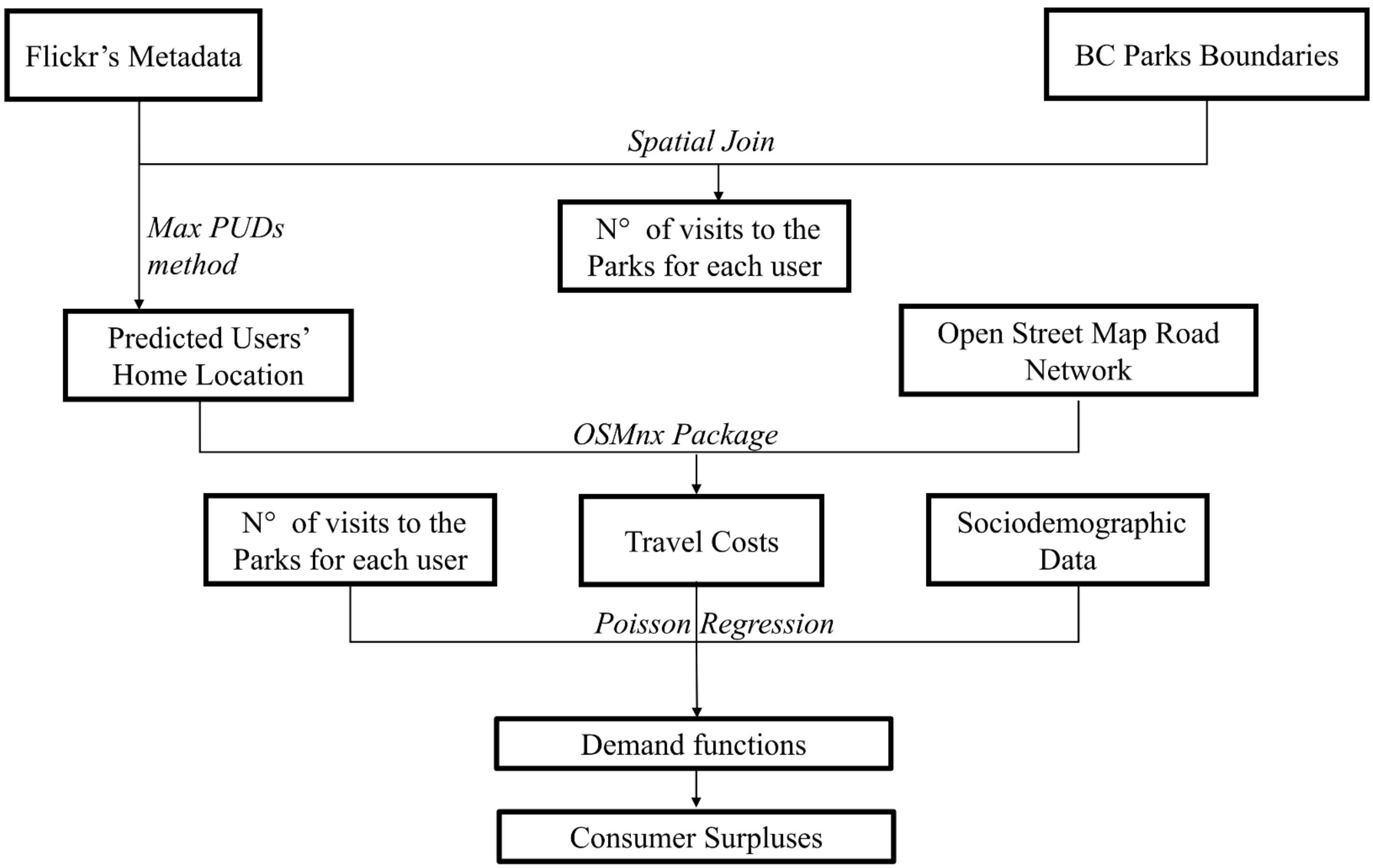

Figure 4. Outline of the methodology used for estimating the monetary value of the recreational ecosystem service provided by BC forested provincial parks.

3.3 Results

In total, the metadata of 1,719,130 geotagged pictures were collected *via* Flickr API. Only 365,477 of these pictures (21.26%) were acquired in forested areas and therefore retained for further analysis. These pictures were uploaded by 12,532 Flickr users and resulted in 44,102 PUDs.

3.3.1 Visitors home location and visitation trends

Among the 12,532 Flickr users that acquired at least one picture between 2005 and 2020, 10,399 satisfied the 10 pictures requirement needed to apply the max PUDs methodology and were included in the visitor's home location analysis. We compared the results of the max PUDs method with the country of origin indicated by the authors in their public profiles and found a 84% correspondence. The majority of the users are Canadian (55.2%). International visitors come from 92 countries, the 5 main countries being (in declining order): United States (48.5%), United Kingdom (13.3%), Australia (5.0%), Germany (3.9%) and Netherlands (2.5%) (Figure 5). Applying the max PUDs method on Canadian users, we found that most came from BC (81.6%) and Ontario (7.6%). Finally, among the visitors from BC, the most frequent districts of origin are: Greater Vancouver Area (22.4%), Whistler (4.1%) and Surrey (4.0%).

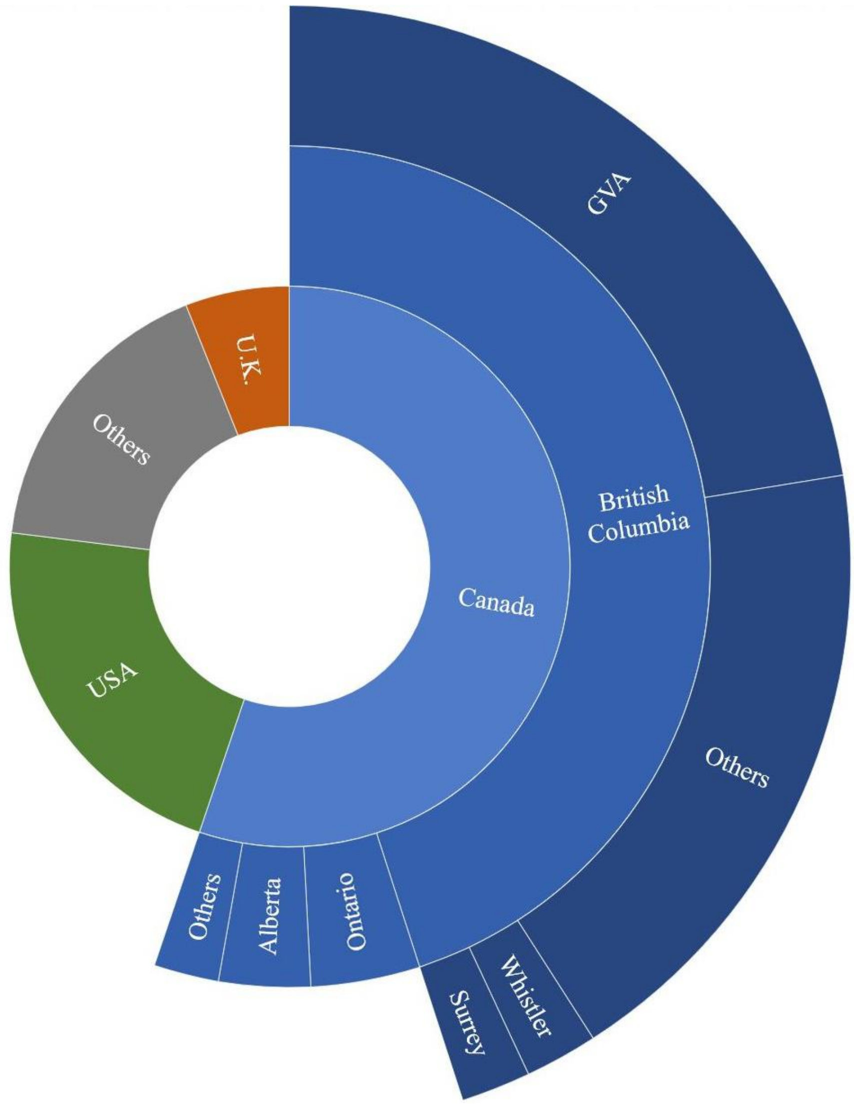

Figure 5. Place of origin of Flickr users that visited BC forests. Deduced from the maximum PUDs method. GVA = Greater Vancouver Area.

From the timestamps of all 365,477 gathered pictures acquired in forested areas we inferred the temporal pattern of forest recreation consumption in BC. As shown in Figure 5a, consumption of forest recreation peaked during the summer months and the end of December, while it was lowest in October and November. As expected, the consumption of forest recreation furthermore peaks during the weekend while remaining almost constant during the working week (Figures 6b and 6c).

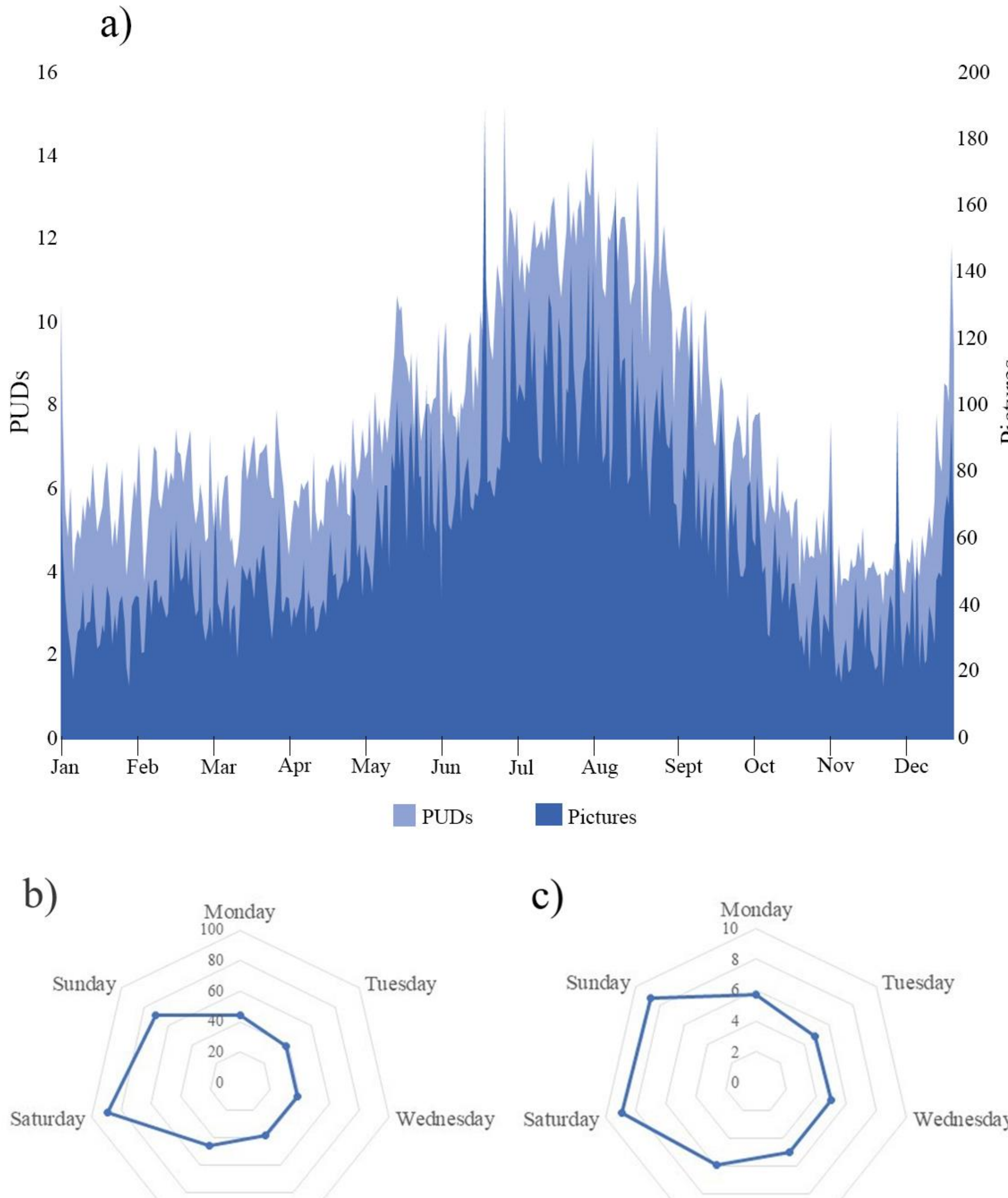

Figure 6. Temporal trends of forest recreation consumption in BC. Values are averages obtained grouping the timestamps of the entire dataset of relevant pictures: (a) monthly variations in numbers of pictures (dark blue) and PUDs (light blue), and average number of pictures (b) and PUDs (c) over weekdays.

We further used the timestamps to explore the hour of the day during which the pictures were acquired across the four seasons. As shown in Figure 7a (and in accordance with Figure 6a), most pictures were taken in summer months, followed by spring, winter, and fall. As expected, the majority of pictures were acquired during daylight between 10:00h and 15:59h, irrespective of the season. Figure 7b shows how the pictures acquired in each season are distributed during the day. Summer and fall months have similar picture distributions, while winter months have a distinctive peak between 11:00h and 12:59h and spring months between 14:00h and 15:59h.

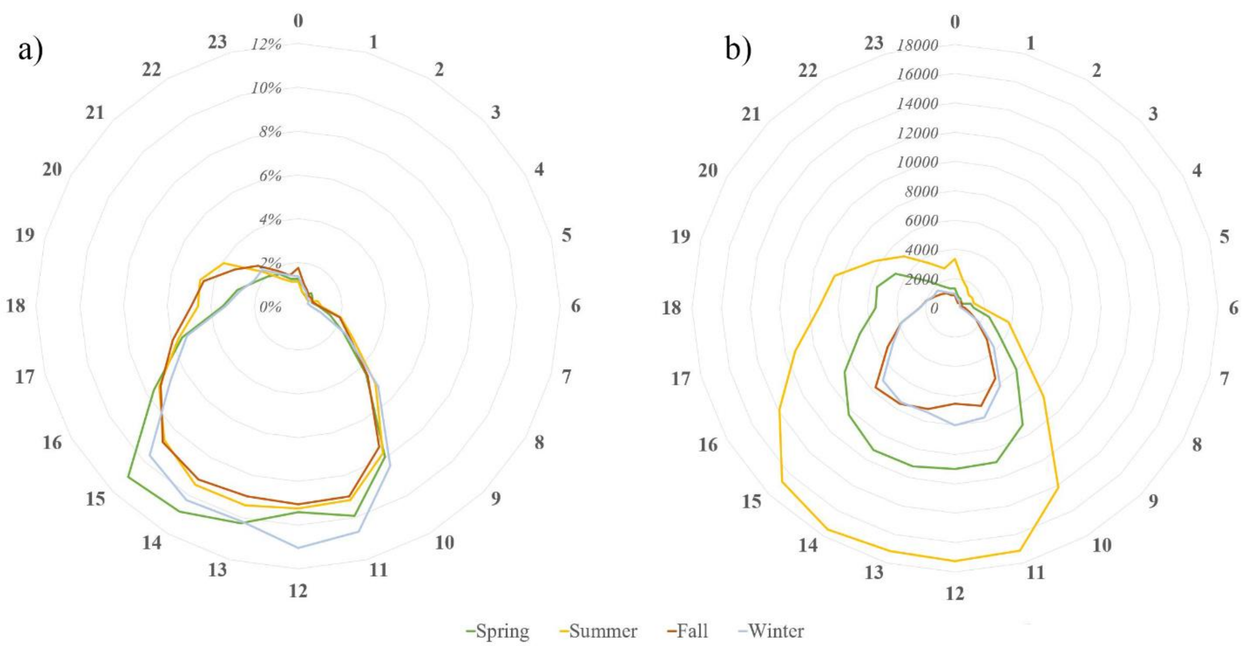

Figure 7. Temporal (hourly) trends of picture acquisition in BC forests, across the seasons: a) total number of pictures, b) percentage of pictures (over the total of the season). Time of acquisition 0 includes pictures acquired between midnight and 12:59 AM local time, 1 includes pictures acquired between 1:00 and 1:59 AM, etc.

Lastly, combining picture locations and dates, we identified hotspots of recreation consumption. Throughout the year, hotspots of forest recreation are located in the south-west of BC, specifically in TSAs 30, 31, 39 and 38 (Figure 8). Hotspots during spring and fall are more spatially constrained. However, spring hotspots expand further north into TSA 31. Summer and winter hotspots are distributed differently: Summer hotspots expand further into the inland of southern Vancouver Island (TSA 38) and winter hotspots expand further into the forests north of Whistler

(TSA 31). Recreational hotspots in the south of TSA 39 and surroundings are present throughout the year.

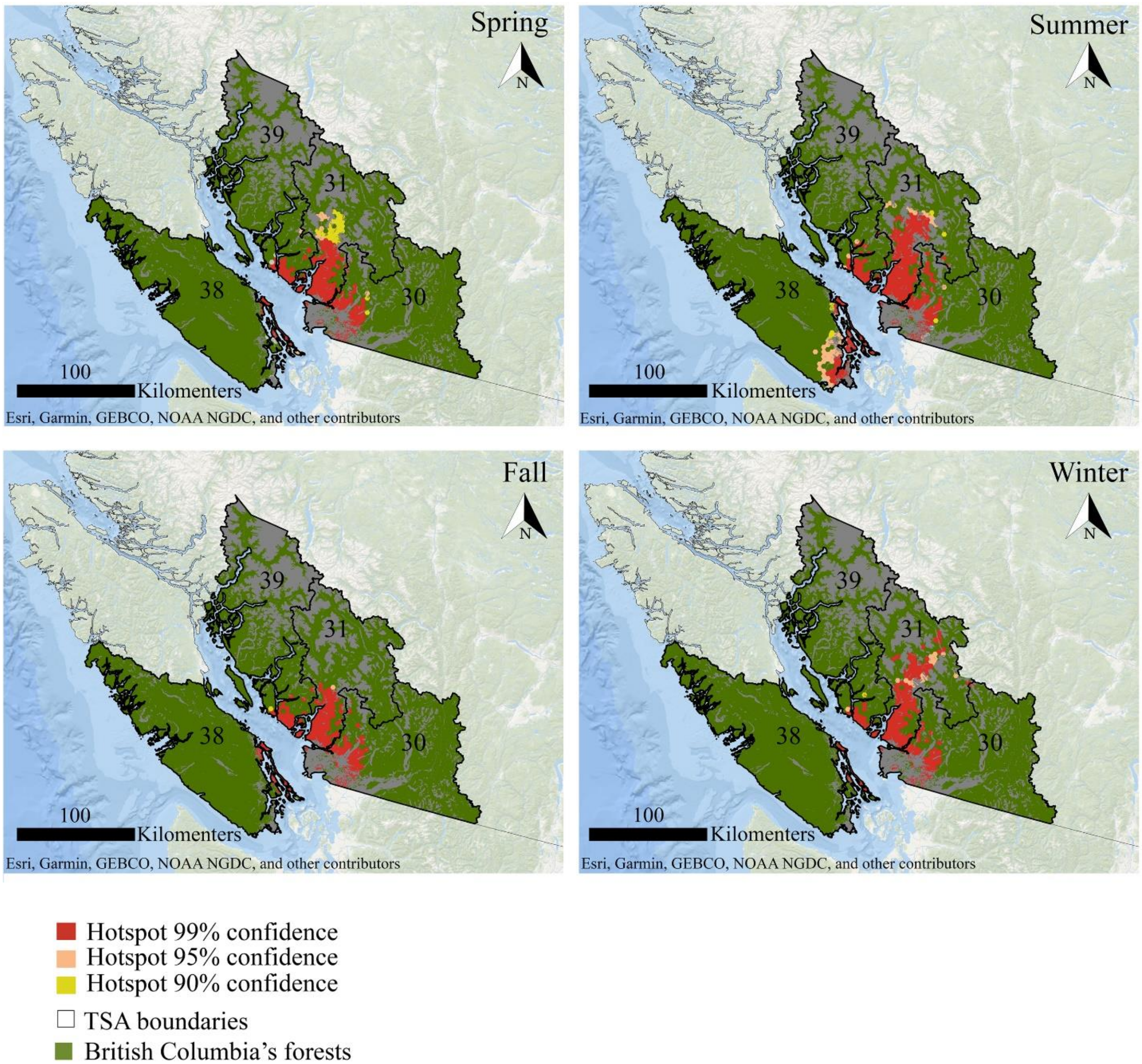

Figure 8. Seasonal hotspots of forest recreation in south-western BC. The confidence value indicates the probability of the cell being a hotspot according to the Getis-Ord Gi statistic. The size of the cells is 5.4 km2.*

3.3.2 Number of annual visits in British Columbia's forests

Of the 1,719,130 geotagged pictures gathered in BC, 19,176 were acquired within the boundaries of a forested provincial park by 1,641 unique visitors who, on average, visited the parks

2.38 times and took 4.91 pictures on each visit, resulting in a total of 3,908 PUDs. GHII ranges from 0 (low impacts) to 65 (high impacts) and the mean GHII value for BC provincial parks is 16.4 with Hemer park being the one with highest GHII (46.0), and Purcell wilderness conservancy park being the one with the lowest GHII (1.3). Table 3 presents the coefficients of the two OLS linear regression models calibrated on the provincial parks' visitation data. As shown, both models highlight a statistically significant relationship between the number of visitors reported by BC Parks statistics and the number of detected PUDs (p-value < 0.001). Furthermore, in the bivariate model the mean GHII value is also statistically significant (p-value < 0.001). Table 4 presents the performance of the two models and shows that model that include both the number of PUDs and the average GHII of the park (bivariate OLS) outperform the univariate OLS both in terms of variation of the dependent variable explained (R-squared) and root mean square error. Lastly, comparing the performances of the naïve transformation and Duan's smearing technique, results show that the smearing transformation is the best suited to transform the log predictions of the model to the actual number of annual visits.

Table 3. Results for the univariate and bivariate OLS log-log regression models, in parenthesis we report standard errors.

Coefficient	**Univariate OLS**	**Bivariate OLS**
Intercept ***	9.2086 (0.145)	7.8389 (0.357)
Slope PUDs***	1.1532 (0.091)	1.0628 (0.096)
Slope GHII***	-	0.5642 (0.138)
Akaike Information Criterion	327.6	274.3

Note: *, **, *** denote significance at the 10%, 5%, and 1% level, respectively, in parenthesis we report standard errors.

Table 4. Performances of the univariate and bivariate OLS log-log regression models.

Parameter	**Univariate OLS**	**Bivariate OLS**
R-squared:	0.537	0.654
Root Mean Square Error Naïve	127,194	106,607
Root Mean Square Error Smearing	118,731	101,922

The Shapiro-Wilk test did not reject the hypothesis of normal distribution of the residuals neither for the univariate nor for the bivariate model, also in the bivariate model the two independent variables do not present multicollinearity.

We then estimated the visitation rates of BC forests for the whole province (Figure 9), applying the OLS regression model to the count of PUDs and the mean GHII value inside each 10 km^2 cell covering the province and implementing the naïve transformation to the model predictions. For this analysis all the 44,102 PUDs detected in BC forests from 2005 to 2020 were considered. The total number of recreational experiences (by both national international visitors) from 2005 to 2020, was on average > 44.2 million per year. As expected, the geographic distribution of these recreational experiences is not homogeneous across the province. Over 54 % of all recreational visits to BC forest are concentrated in TSA 30 (surrounding Greater Vancouver) and TSA 38 (surrounding Greater Victoria).

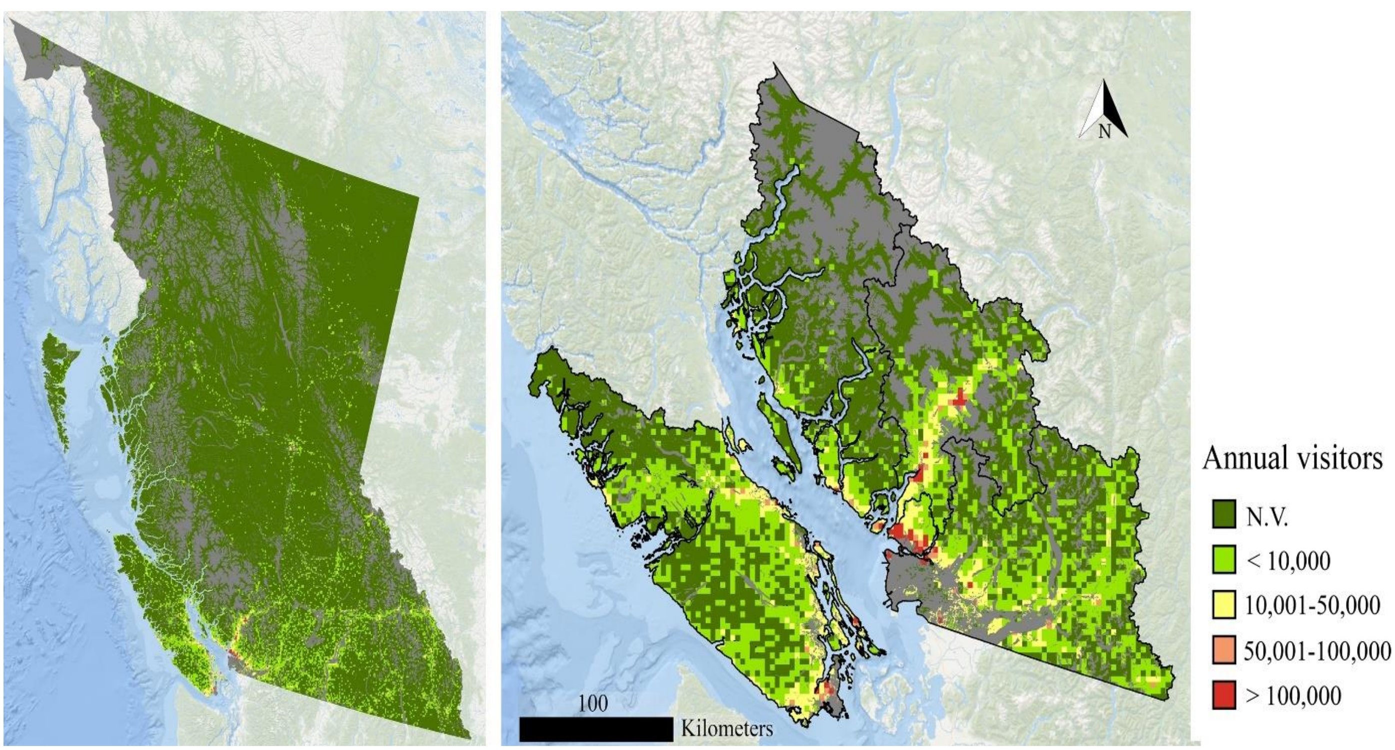

Figure 9. Estimated number of annual visits in BC forests (left) and south-western BC (right). Values are average annual number of visits between 2005 and 2020.

3.3.3 Value of forest recreational sites

The 3,908 PUDs detected in forest-dominated BC provincial parks resulted in 2,814 visits, made by 962 visitors. As shown in Table 5 the average income of these visitors is 71,237 CAN$/year, the average travel cost incurred by the visitors is CAN$ 48.5 while the average travel cost to the substitute site is CAN$ 22.6.

Table 5. Summary statistics for the data used in the negative binomial regressions.

Variable	**Average**	**Range**	**Std. Dev.**
Income (CAN$)	71,237	33,380 - 123,648	12,848.3
Travel Cost to provincial park (CAN$)	48	1 - 168	38.63
Travel Cost to substitute provincial park (CAN$)	23	1 - 322	21.98

In total, 8 forested provincial parks in BC had sufficient numbers of detected visitors (*i.e.*, at least 10 visitors per predictor variable in the model) to apply the crowdsourced travel cost method (Table 6). In each negative binomial regression, the coefficient of the travel cost variable (TC) has the expected negative sign, meaning that the number of visits made by a visitor to a site decreases when the cost of traveling to the site increases. However, the negative binomial regressions yielded statistically significant results only for 6 of 8 forested provincial parks initially included in the analysis (Table 6 and Figure 10). As noted in Table 6, all significant results can be found in parks where the number of detected total visitors exceeds 50. The other variables that we included in the model, income (Inc) and travel cost to the nearest substitute site (TC sub. Site see p. 44 for definition), have poorer performances, statistically significant in only three and two provincial parks respectively. However, the estimated coefficients had the expected positive signs indicating that the number of visits made to a site is expected to increase for visitors with higher income and when the cost of traveling to the closest substitute site increases. As expected, the performance of the negative binomial regression, measured with the Akaike Information Criterion (AIC), generally increases as the number of detected visitors increases. As can be seen in Figure 10, all the forested provincial parks in which we estimated statistically significant coefficients are located near the largest urban centers of BC, *i.e.,* Greater Vancouver Area and Greater Victoria. The only forested provincial park in which more than 50 visitors were detected that is not statistically significant is E.C. Manning Provincial Park, which is situated more than 150 km east of Vancouver. The average consumer surplus estimated with the Creel and Loomis approach ranges from CAN$ 29 estimated in Mount Seymour Provincial Park and Say Nuth Khaw Yum (Indian Arm) Provincial Park, to CAN$ 87 estimated in Goldstream Provincial Park.

Combining the estimates of the annual number of visits made by domestic visitors (obtained *via* the OLS linear regression) and the average consumer surplus per visit obtained with the negative binomial regression, it was possible to estimate the total annual consumer surplus of the seven parks, providing insights into the monetary value of the recreational ecosystems service provided by the park per year. The provincial park with the highest total annual consumer surplus is Cypress Provincial Park (~ CAN$ 34 million).

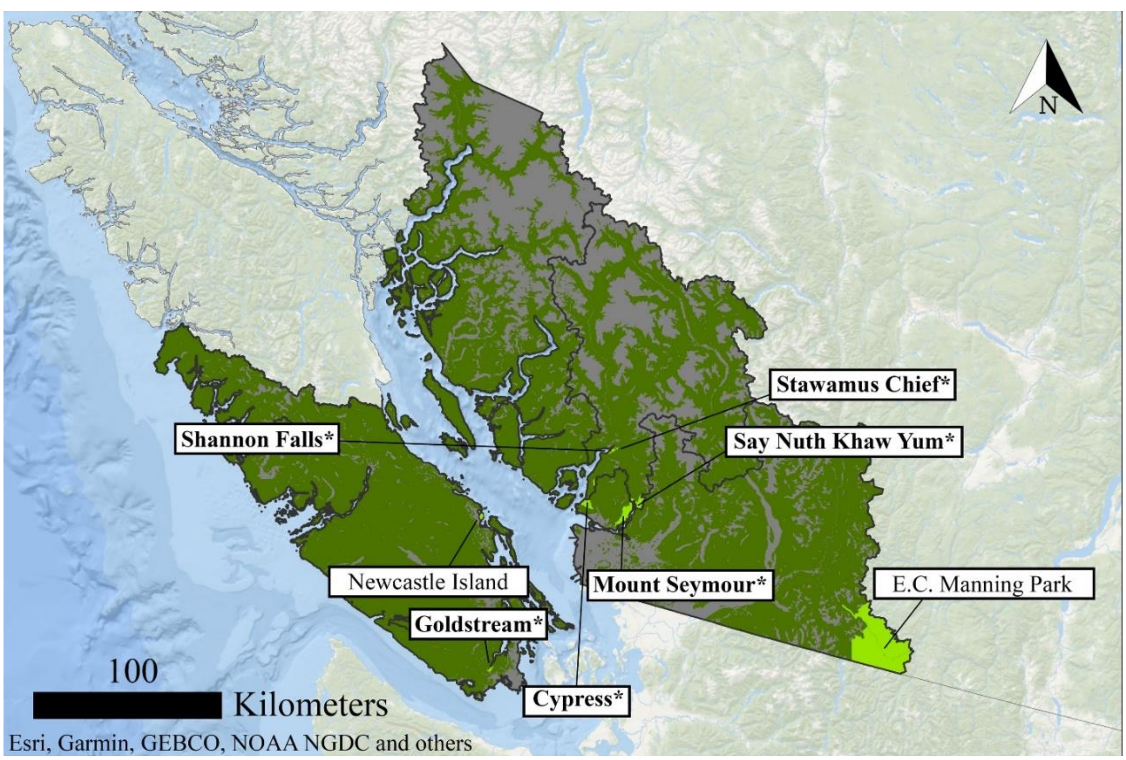

*Figure 10. Locations of the forested provincial parks in which we administered the negative binomial regression. * denotes the parks in which the regressions yielded statistically significant results.*

Table 6. Results of negative binomial regression for the detected visitors in BC forested provincial parks.

Park Name	dF	AIC	Int.	TC	Inc.	TC sub. site	CS (CAN$)	Tot. CS (M CAN$)
Stawamus Chief	197	986.4	1.1410 (0.873)	-0.0140*** (0.005)	0.002 (0.001)	-0.0087 (0.006)	71	21.286
Mount Seymour	129	896.4	1.1887*** (0.419)	-0.0328*** (0.005)	0.0301*** (0.007)	0.0084* (0.005)	30	19.379
Cypress	152	963.4	0.9765*** (0.308)	-0.028* (0.004)	0.0195** (0.007)	-0.0045 (0.005)	36	34.913
E.C. Manning	101	546.9	1.9166 (0.349)	-0.0011 (0.002)	-0.0188 (0.01)	-0.0075 (0.004)	N.S.	N.S.
Shannon Falls	84	301.7	1.2895 (0.877)	-0.013* (0.005)	-0.0033 (0.018)	-0.00114 (0.01)	77	27.150
Goldstream	53	230.15	-1.4021** (0.596)	-0.0115*** (0.003)	-0.0607*** (0.014)	0.0028* (0.002)	87	29.750
Say Nuth Khaw Yum	52	250.1	5.2387 (0.908)	-0.034* (0.015)	-0.1012 (0.019)	0.015 (0.02)	29	2.908
Newcastle Island	38	178.9	1.0695 (0.927)	-0.0058 (0.005)	-0.0096 (0.024)	0.0095 (0.011)	N.S.	N.S.

dF = degree of freedom, AIC = Akaike information criterion, Int. = Intercept, TC = travel cost, Inc. = income, TC sub. Site = Travel cost to substitute site, CS = Average consumer surplus, Tot. CS = total Consumer Surplus in millions of CAN$. The number of estimated visitors is derived from BC Parks visitation statistics.

*Note: *, **, *** denote significance at the 10%, 5%, and 1% level, respectively, in parenthesis we report standard errors.*

3.4 Discussion and conclusion

Crowdsourced social media data offers valuable insights into the cultural services provided by ecosystems. The high volume, low-cost, fast data collection, and large scale applicability of social media data (Ghermandi and Sinclair, 2019) provide new opportunities to study the demand for forest recreation, having the potential to drastically reduce the time and costs associated with on-site surveys in remote areas. Thus, their application in the study of nature-based recreation is a fast-growing research field. The overarching objective of our research was to assess this potential, especially in tackling the mapping and valuing challenges that have limited the inclusion of recreational ecosystem services in forest management plans. Our analysis was structured on three levels. First, we characterized the consumption of forest recreation, identifying inter- and intra-annual trends, recreation hotspots and the place of origin of visitors. Second, we quantified the consumption of forest recreation with an OLS regression model, calibrated with empirically collected visitation data in BC provincial parks, and applied it to the entire province. Third, we performed a travel-cost analysis to estimate recreational values in BC Parks, relying on crowdsourced and secondary data. For each of these three levels of analysis, we will discuss our findings and their limitations, as well as the opportunities for forest management associated with crowdsourced social media data. Lastly, we will discuss future directions for the use of crowdsourced social media data to analyze forest recreation ecosystem service.

3.4.1 Crowdsourced social media data to characterize forest recreation consumption

Using the timestamps of Flickr pictures we were able to identify the annual, weekly, and hourly trends in forest recreation consumption. Our findings show that the consumption of forest recreation peaks during the summer and during the weekends, in line with expectations and with the results of previous studies (M. Sinclair et al., 2020). The hourly distribution of the pictures acquired in BC forests is also similar to the one described by Ciesielski et al. (2021), with most of the pictures being acquired between 10:00h and 16:00h. Combining the geotags and the timestamps of the gathered pictures, we were able to identify and differentiate seasonal hotspots of forest recreation consumption. As expected, recreational hotspots are located in the southwest of the province, around the main urban centers: Vancouver and Victoria. The spatial distribution of the hotspots aligns with previous findings that the data volume obtained from crowdsourcing social media is higher in areas with large populations than in remote areas (Tenkanen et al., 2017). Information such as temporal and geographic trends of recreation consumption could help forest managers plan harvesting operations, so as to minimize impacts on the provision of recreational ecosystem service. However, the fact that remote areas tend to be underrepresented by crowdsourced social media data should be considered carefully, and an integration of crowdsourced social media data and traditional surveys methods could be a better option. Moreover, the pre-selection of pictures by social media users before posting content could introduce biases towards: (i) certain hours of the day (e.g., due to occurrence of sunrises and sunsets), (ii) days in which certain atmospheric phenomena occurs (e.g., rainbows, *aurora borealis*), and (iii) seasons characterized by certain phenological events (e.g., blooming and autumnal leaves color change). Lastly, the presence of wildlife, especially in the case of certain charismatic species such as grizzly bear (*Ursus arctos horribilis*) or wolf (*Canis lupus*), could also introduce biases, as already suggested by Tenkanen et al.(2017).

Our results confirm the reliability of the max PUDs methodology in predicting the home locations of social-media-users as found by Sinclair et al. (M. Sinclair et al., 2020). We obtained a slightly lower precision at the country level than Sinclair et al. (2020) (84% with a 10 PUDs threshold *versus* 90%). Few Flickr users declared their home location to a finer level of spatial detail than country, making it impossible to assess the reliability of the method for home cities or sub-national regions.

Comparing the predictions of the max PUDs methodology with the data on international visitors' arrivals from Statistics Canada (Corp., 2021), it appears that crowdsourced data from Flickr performed well in predicting the country of origin for Europeans, Americans and Australians, while systematically underestimating the number of visitors from Asia. In fact, while Asian visitors constitute on average over 16% of the international visitors of BC, they have acquired only 3% of the pictures gathered in this study (according to the results of the max PUDs methodology). This discrepancy could be explained with different accessibility of online photo-sharing services across different countries, in particular China (Ghermandi, 2016).

3.4.2 Crowdsourced social media data as a proxy for forest visitation rates

We found a strong correlation between the number of PUDs detected, the GHII value and the number of recreationists visiting forest dominated provincial parks. Various studies estimated univariate models in the past, using exclusively PUDs as the independent variable. Wood et al. (2013) observed correlations at 839 worldwide recreational sites (r^2=0.39), Keeler et al (2015) for North-American lakes (r^2=0.65 and 0.70) and Sinclair et al. (2019) for the Kerala's wetland ecosystems network (India) (r^2=0.63). In our study, the provincial parks that we used to calibrate our model range from highly frequented peri-urban areas such as Cypress Provincial Park (averaging over 1,200,000 annual visits observed between 2007 and 2018) to remote areas far from any major urban centers such as the Kinaskan Lake Provincial Park (averaging less than 15,000 annual visits observed between 2007 and 2018). Despite this wide range of empirically detected visits, the performances of our calibration model (r^2 = 0.65) lie in the upper limit of the above-mentioned models. The statistical significance of the GHII indicator in our model aligns with the notion that the number of geotagged pictures acquired in a landscape is influenced by its accessibility and closeness to urban areas (Ghermandi and Sinclair, 2019) (van Zanten et al., 2016). Comparing the performances of the bivariate model (with both PUDs and average GHII as independent variables) with the performances of a traditional univariate model (with only PUDs as independent variable) across our study area we found that the bivariate model outperforms the univariate model. This indicates that GHII is a potentially a useful indicator that could easily be integrated in future models attempting to estimate the number of visitors in the natural landscape from crowdsourced social media data.

We recognize that, when applying the OLS model outside its calibration sample (*i.e.*, provincial parks) to estimate the number of recreationists in forested areas in the whole province, the results are potentially subject to biases. Most notably, social media users may not be a representative sample of the entire population of recreationists, as they are usually younger (Chua et al., 2016), and different social media draw the interest of different sections of the general population (Ghermandi and Sinclair, 2019). However, empirical studies (Hausmann et al., 2018) indicate that Flickr is less biased towards younger people that Instagram and Twitter. Also, previous studies have shown that this limitation can be partially mitigated by crowdsourcing data from different social media (Tenkanen et al., 2017). Another limitation of crowdsourced social media data is the fact that not every recreational activity traditionally carried-out in forested areas is equally compatible with the acquisition of pictures to be shared on social media. In fact, among the main recreational activities in which forest visitors engage, as listed by Rosenberger et al. (2017), are activities such as hunting and motorized boating, which appear to be less likely to be portrayed in pictures than hiking or camping. Therefore, relying exclusively on crowdsourced social media data to estimate the number of visits in forested areas could result in failing to recognize important recreational areas devoted to specific forest recreational activities, or areas that are mostly used by a certain segment of the general population. Despite these limitations, crowdsourced social media data offer unparalleled opportunities to estimate forest visitation rates at large-scale with relatively high spatial resolution. Visitation rates maps - such as the ones that we were able to generate in our study - could provide a useful benchmark to forest managers.

For example, by overlaying a map of forest visitation rates with a map of timber values, it becomes possible to identify areas were the trade-off between timber production and forest recreation has to be carefully balanced. In conclusion, this innovative approach can produce valuable outputs that could help in tackling the issue of mapping nature-based recreation an effectively account for its values in forest management.

3.4.3 Crowdsourced social media data for travel-cost analysis in forest sites

In our final analysis, we used crowdsourced geotagged pictures as a data source to perform a monetary assessment of the recreational values of BC forested provincial parks. The use of the OSMnx package has allowed us to simultaneously estimate the travel distances of over 900 visitors across the entire BC provincial parks system. Previous studies that adopted the crowdsourced travel cost methodology have instead used the Google Maps API (Ghermandi, 2018; Sinclair et al., 2018) (Sinclair et al., 2018) (Ghermandi, 2018). However, Google Maps API implements a freemium strategy, offering for free a defined number of requests (1000), limiting the extent of the analyses. On the contrary OSMnx enables the user to make an unlimited number of requests making our approach a more scalable and innovative solution. Our results (ranging from CAN$ 29 to CAN$ 87 of consumer surplus per visit) are consistent with the values estimated with traditional surveys for forest recreation in forests in the North-West of the United States of America, where the average consumer surplus of forest recreation is 67 US$ (83 CAN$) (Rosenberger et al., 2017). In their application of the crowdsourced travel cost method to a set of wetlands, Ghermandi (2018) obtained significant results despite having fewer available data (10 visitors per variable included in the regression), while we obtained significant results only in forested provincial parks where we detected more than 12 visitors per variable. The poor performance of the crowdsourced travel cost method in our study area could be explained by the geographical distribution of BC population. In fact, over 57% of BC inhabitants reside in either Greater Vancouver or Greater Victoria. The coastal location of both Vancouver and Victoria is likely to introduce a bias in the determination of the exact coordinates of the home locations of the authors because of the ocean truncation effect. This, combined with the fact that the precision of the maximum PUDs method drops when predicting the place of origin of an author at the municipality level (M. Sinclair et al., 2020), hampers the reliability of distances traveled in our dataset. As a result, the use of the travel cost method from crowdsourced social media data appears to be especially appropriate for peri-urban forests that are accessible on a single-day trip from multiple municipalities. When estimating the monetary value of the recreation ecosystem service provided by forest located further from urban areas, traditional survey methods (*i.e., in-situ* and mail) might be more suitable. Another limitation of the crowdsourced travel cost method is the use of census data to estimate the income of recreationists. In Canada, census data are collected by Statistics Canada every five years. During these time windows people can move and change their occupation. In our study (as well as in previous applications of the crowdsourced travel cost method), the analysis spans over multiple years, therefore the ability of these data to approximate the recreationists incomes progressively decreases for visits made prior to and later than the census year. As mentioned above, another bias that can affect these analyses is that social media users are not necessarily representative of the general population. The development of techniques to deal with this issue is hampered by the limited socio-demographic data that can be gathered when performing a crowdsourced analysis in respect to traditional surveys. Furthermore, despite the implementation of weighing systems, such as the use of PUDs, the crowdsourced travel cost analysis may still be biased towards very active users. Despite these limitations the use of crowdsourced data, combined with traditional non-market evaluation approaches (such as travel cost) and openly accessible data (such as Open Street Map) is an innovative and promising solution to the issue of valuing the recreational services provided by an ecosystem. Lastly, it is important to note that whenever crowdsourced data are used, ethical concerns about the social media users' privacy arise.

3.4.4 Advantages and limitations of crowdsourced data in the study of forest recreation

The analyses that we carried-out allow us to identify both the advantages and limitations of the use of crowdsourced data in characterizing, mapping and valuing forest recreation. Our conclusions are briefly summarized below

Characterizing forest recreation

Crowdsourced social media data can provide valuable information to forest managers. However, pictures shared on social media may not be representative of all of the pictures taken by forest visitors. As a result, the screening process we used to select the most relevant pictures may have introduced some bias towards certain times of day, atmospheric or light phenomena, phenological events, and/or the presence of certain charismatic species. Lastly, when using crowdsourced social media data to estimate the visitors' Country of origin, the different accessibility and popularities of social media platforms globally could also introduce biases.

Mapping forest recreation

However, the results of these models should be interpreted as preliminary values that forest managers could use to include the recreational ecosystem service when planning forest operations at the strategic scale, recognizing that the estimates can be affected by biases. For example, pictures tend to be concentrated in forested areas near large population centers and in areas with well developed infrastructure, possibly causing an overestimation of the intensity of forest recreation consumption near cities and an underestimation in remote areas. Lastly, some forest recreational activities appear to be less easily depicted in pictures than others.

Valuing forest recreation

However, the applicability of the crowdsourced travel cost method may be limited to popular, peri-urban recreational sites for which it is possible to obtain statistically significant results. Conducting a crowdsourced travel analysis, the time horizon can be easily extended to multiple years. This, on one hand, has the advantage of allowing the collection of more data. On the other hand, socio-demographic data acquired from secondary sources, such as the census, progressively lose their ability to represent visitors adequately. Lastly, the performance of the method used to identify the home location of the visitors (max PUDs), may became poorer when determining specific administrative areas of origin (rather than Country) of a visitor.

3.4.5 Future perspectives

Analysing crowdsourced social media data represents a promising approach to include recreational values in forest management plans, nevertheless, this does not imply that these data could already completely replace traditional *in-situ* surveys. As discussed, the use of crowdsourced data may introduce a number of biases in the interpretation of the results. For example, the results of our research suggest that crowdsourced social media data can provide estimates of the visitation intensity in forested areas, where the use of empirical in-situ surveys is challenging and costly. However, future studies that test this hypothesis, by comparing the results of estimates based on crowdsourced social media data with empirically determined visitor numbers, are required. Moreover, new studies that explore the relationship between trends of forest recreation (inferred by crowdsourced social media data) and forest stand characteristics (such age, species, etc.) could provide insights on how the consumption of forest recreation is impacted by stand dynamics and forest management, and evolves over time and space. These studies could ultimately facilitate the integration of recreational ecosystem services into forest models and decision support systems. Additional studies that promote the integration of crowdsourced social media data and the evaluation of non-market ecosystem services are needed. Other traditional non-market evaluation techniques, such as the benefit transfer approach, could be successfully integrated with crowdsourced social media data. To this end, the content of crowdsourced social media images could provide valuable information on the recreational activities popular at specific locations. Having access to this information could potentially allow forest economists to conduct fine-grained application of the benefit-transfer approach, since the various recreational activities that can take place in forests generate different economic surpluses. We believe that, in the future, modern deep learning algorithms used for image classification tasks, such as convolutional neural networks, could play an important role in such studies.

Chapter 4
Valuing Cultural Ecosystem Services Combining Deep Learning and Benefit Transfer Approach

4.1 Background and motivation

Most of the current literature uses commercial pre-trained image classification models to explore the content of crowdsourced social media data in the context of CES research (Ghermandi et al., 2022). These models present clear advantages compared to the manual classification (most notably time-saving), but it also presents two major limitations: (i) these services are not for free (e.g., Clarifai and Google Vision), but they operate under a freemium strategy, offering free classification only for a limited number of images; (ii) the categories (classes) that these services can identify are fixed and cannot be modified to meet the researchers' needs. Freely available pre-trained convolutional neural networks (CNN) offer an alternative to commercial solutions. A CNN is a deep learning neural network that is capable of processing structured data arrays, such as digital images and other images. CNNs are widely used in computer vision and have become the state of the art for image classification (Howard and Gugger, 2020). They are made of two main components: an architecture, and a set of weights. The architecture is the fixed structure of the CNN, composed of multiple interconnected layers. Weights are numerical parameters, that are modified during CNN training, whose values describe the strength of the connections between the nodes composing the different layers. A pre-trained CNN was for example used by Seresinhe et al. (2018). In their study, the authors used the ResNet50 architecture (He et al., 2016) and the parameters resulting from training on the Places365 database (Zhou et al., 2017) on a dataset of images taken all around Great Britain. In particular, the CNN was successfully used as a relevance filter, dividing the images into two classes: images acquired indoors and pictures acquired outdoors. More recently, a similar approach was used by Payntar et al. (2021) and Havinga et al. (2021) that used again the ResNet50 architecture, pre-trained on the ImageNet database (Deng et al., 2009) to cluster images based on their content.

However, in these studies, the classes used in the image classification process were not adapted to the specific user requirements. To be able to freely define all classes, as would be required for automatic image classification for the analysis of CES, there are two available approaches: training a CNN from scratch, and transfer-learning. Training of a CNN requires large datasets of manually classified images, a skilled analyst, and substantial computational power. These requirements make also the training from scratch of ad hoc CNNs an impractical solution.

Alternatively, transfer learning, a more recent development in the field of machine learning, could allow researchers to easily customize freely available pre-trained CNN models. The basic idea of transfer learning is to adapt (fine-tune) pre-trained models, to a new classification task (Torrey and Shavlik, 2010). In practice, the last layers (head) of the pre-trained CNN are discarded and substituted by a new head designed for the intended classification task. This approach allows researchers to freely modify the output classes. The newly obtained CNN is then re-trained using an ad hoc populated training set. The transfer learning approach allows the development of CNNs capable of providing good performances even with access to limited training data (Howard and Gugger, 2020). Furthermore, transfer learning represents also an answer to the growing concerns about the environmental impacts of training convolutional neural networks (Strubell et al., 2019). This approach has been recently adopted by Cardoso et al. (2022) to identify and classify elements relevant to the study of CES in social media images taken in two areas of the Iberian Peninsula (Peneda-Gerês in Northern Portugal and the Sierra Nevada in Southern Spain). Specifically, the authors used transfer learning to train models to classify social media images based on the type of CES depicted on them: species, landscape, nature, human activities, human structure, or posing. The authors then compared the performances of CNNs trained with a transfer learning approach and CNNs trained from scratch, finding that the firsts outperformed the seconds. This study shows how transfer learning can successfully be used to detect and classify CES elements from social media images, suggesting that social media images could potentially help researchers to understand CES distributions, benefits, and values.

Despite the results obtained in the studies illustrated above, the use of social media image content to explore CES provisioning and consumption is still in its infancy. In particular, it appears that while previous studies have shown how information can be extracted from the analysis of the images, little is known about the potential applications of this information for ecosystem management. In this context, the overall objective of this study will be to bridge this knowledge gap, exploring the potential of social media image content analysis in characterizing and valuing CES provision. To do so, images taken in BC's provincial parks system will be retrieved from social media and analyzed by combining freely available pre-trained CNNs and purposely trained CNNs. First, CNNs will be used to determine, for each analyzed image: its relevance in the context of CES (images taken outdoors in a natural context vs images taken indoors in an urban context); the type of CES depicted (aesthetic vs recreational); and the landscape or recreational activity depicted. Then, the outcomes of the image classification process and the image metadata will be combined to assess the value of the recreational ES provided by BC parks, adopting an innovative crowdsourced benefit transfer approach. Lastly, the outcomes obtained, as well as limitations and opportunities that arise from the use of social media data and CNN to extract insights on CES management will be discussed.

4.2 Methods

4.2.1 Methods overview

The method applied in this study follows three main steps. First, the data required for the analyses were gathered via the social media of choice (Flickr) Application Programming Interface (API). Flickr is the most widely used social media website in CES research because of its easily and freely accessible API (Ghermandi and Sinclair, 2019). Flickr API was queried to gather pictures used both for CES characterization, and CNNs training. To do so a purposely developed Python application was used.

Second, Convolutional Neural Networks (CNNs) were trained and deployed to classify social media pictures acquired in BC. This made it possible to automatically detect and characterize the cultural ecosystem services provided by BC provincial parks system.

Lastly, the outcomes of the image classification process were combined with the image metadata to assess the monetary value of the recreational ES provided by the BC provincial parks system to BC residents. To do so, an adaptation of the benefit transfer method (Champ et al., 2003) to crowdsourced data was applied. In the following section, the individual methodological steps will be detailed.

4.2.2 Data gathering

As shown in Table 7, along with the data adopted throughout the dissertation, in this chapter we gathered and used various specific data types and sources. These data can be subdivided into three categories: (i) crowdsourced; (ii) geographic; (iii) secondary.

Crowdsourced data: In this study, in addition to the crowdsourced images were used in the previous chapter, other Flickr images were used to populate the dataset used for the training of the

CNNs. To collect these images the *tags* argument of the "*flickr.photos.search*" function of Flickr's API was used. This argument enables the selection of images that were tagged with a certain word when uploaded by the user. This approach was used to facilitate the collection of relevant training images for each category included in the analysis (Table A2). Furthermore, the query for training images included a bounding box enclosing all Canada Provinces, BC excluded. The bounding box was specified using the *bbox* argument of the "*flickr.photos.search*". The bounding box allowed for consistency between training and analyzed images, while avoiding duplicates between the two sets. The following coordinates were used as the bottom right corner and upper corner respectively: (49.0 N, -114.0 E) and (60.0 N, -70.0 E).

Lastly, combining the image metadata and the outcomes of the image classification process two additional data types were obtained: Landscape-User-Days (LUDs) and Activity-User-Days (AUDs). A LUD is a unique combination of user, date, and landscape typology portrayed (e.g., forest, mountain, etc.). In the case of multiple landscape typologies portrayed during the same day, by the same author, the one portrayed with the highest frequency was considered the primary landscape. An AUD is a unique combination of user, date, and recreational activity portrayed (e.g., hiking, skiing, etc.). In the case of multiple recreational activities portrayed during the same day by the same author, the one portrayed with the highest frequency was considered the primary activity. Both these data types, introduced in this study, are built on the concept of Photo-User-Day introduced by Wood et al. (2013).

Geographic data: Shapefiles defining the boundaries of BC's Provincial Parks were downloaded from the BC data catalog (https://catalogue.data.gov.bc.ca/). To avoid the exclusion of relevant images taken just outside park boundaries, a 200m buffer was applied to each polygon. The buffered park boundaries were then used to select the relevant images among the ones downloaded from the Flickr API.

Secondary data: Lastly, different types of secondary data were used: (i) two versions of the ResNet-152 CNN architecture (He et al., 2016), one pre-trained on the Places365 database (Zhou et al., 2017), and one pre-trained on the ImageNet database (Deng et al., 2009); (ii) Provincial Parks visitation data from BC's Parks statistics; and (iii) consumer surpluses reported by Rosenberger et al. (2017) for various recreational activities. Consumer surplus is the excess of social valuation of product (in this case the recreational experience) over the price actually paid. Table 1 summarizes the data utilized in the analysis. In the following sections, further information about the used CNN architectures and parameters will be provided.

Table 7. Data used in the exploration of the cultural ecosystem services provided by BC's Parks.

Description	Typology	Source
Geotagged images taken in BC	Crowdsourced Data	Flickr API
Relevant images for CNNs training	Crowdsourced Data	Flickr API
Landscape User Days (LUDs)	Crowdsourced Data	Flickr API
Activity User Days (AUDs)	Crowdsourced Data	Flickr API
BC Provincial Parks boundaries	Geographic Data	BC's data catalog
Visitation data for BC Parks	Secondary Data	BC Parks end of year reports
ResNet-152 trained on Places365	Secondary Data	http://places2.csail.mit.edu/
ResNet-152 trained on ImageNet	Secondary Data	https://www.image-net.org/

4.2.3 Image classification

The objective of the image classification process was to automatically detect and characterize the CES depicted in the gathered images. Four image classifiers were used: relevance CNN, Cultural Ecosystem Services (CES) CNN, aesthetic CNN, and recreation CNN. All these classifiers were based on ResNet-152 (He et al., 2016), a residual neural network architecture, 152 layers deep, that employs residual learning units easing the training of deep neural networks. The common architecture was adapted to answer different questions on the gathered images. The relevance CNN was used to answer the question: "*is this image relevant in the context of CES research?*"; the CES CNN was used to answer the question: "*is this image depicting an aesthetic or a recreational experience?*". Lastly, in the case of an aesthetic image, the aesthetic CNN was used to answer the question: "*what type of landscape is depicted in this image?*"; and in case of a recreational image, the recreational CNN was used to answer the question: "*which recreational activity is depicted in this image?*"

To answer these questions the ResNet-152 architecture was coupled with a different set of weights. The relevance CNN and the aesthetic CNN both relied on the weights previously obtained by Zhou et al. (2017) when training ResNet-152 on the Places365 image database. Places365 is an image database used for scene recognition. Places365 contains over 1.8 million images, classified into 365 scenes subdivided into three macro-categories: indoor (e.g., living room and butcher shop), urban (e.g., parking lot and residential neighborhood), and natural (e.g., rain forest and waterfall). Instead, the CES CNN and the recreation CNN were both created *ad hoc* for this study using a transfer learning approach, starting from the weights previously obtained by Deng et al. (2009) when training ResNet-152 on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) image classification and localization dataset. ImageNet contains over 14 million images divided into 1,000 classes. Further details on how these two sets of weights were obtained with the transfer learning approach are provided in section 2.4.1.

The image classification process, illustrated in Figure 11, can be described as follows. First, each image was fed into the relevance CNN to be classified as either relevant or not relevant. Images classified as belonging to classes of the natural macro-category were deemed relevant, while images classified as belonging to indoor and urban macro-categories were deemed not relevant. Second, relevant images were fed into the CES CNN that separated them into two classes based on the CES depicted: aesthetic ES and recreational ES. The distinction between aesthetic and recreation ES adopted here broadly follows that of Richards and Friess (2015), who distinguish between "Landscape" and "Social Recreation". The rationale for focusing on aesthetics and recreation was that these two ES appear to be the most likely to be depicted in social media images. In the aesthetic category, were included: (i) pictures depicting natural landscapes and skyscapes (ii); pictures depicting vegetation; (iii) pictures depicting natural features (e.g., mountains, lakes, streams, waterfalls, etc.). In the recreation category, were included: (i) pictures depicting people engaging in a recreational activity; (ii) pictures depicting equipment related to recreational activities, but not people; (iii) pictures depicting wildlife. In the last step of the analysis, both aesthetic and recreational ES images were further classified to extract information on the type of landscape and the recreational activities that were depicted on them. Aesthetic ES images were fed into aesthetic CNN that grouped them into six landscapes typology, based on the outputs of the ResNet-152 CNN architecture pre-trained on the Places365 database. The landscape typologies were the following: (i) anthropic; (ii) water; (iii) forest; (iv) mountain; (v) ice and snow; (vi) sky. The pairings between the CNN outputs and the landscapes categories are reported in the supplementary material (Table A1). Images classified as recreation were fed into the recreation CNN to be further distinguished based on the recreational activities depicted in them. Seven recreational activities were considered relevant for the study area: (i) hiking; (ii) climbing; (iii) skiing; (iv) camping; (v) biking; (vi) wildlife viewing; (vii) water-related (e.g., non-motorized boating, fishing, etc.). This list was obtained starting from the activities included in the Recreation Use Values Database by Rosenberger et al. (2017). The adaptation process was carried out by the authors visually analyzing a subset of 3,000 of the gathered images and harmonizing the most frequently depicted activities with the ones included in Rosenberger et al. (2017). The recreational CNN uses ResNet-152 architecture, coupled with another set of parameters, obtained with a transfer learning approach.

The criteria used in the image classification process are summarized in Table 8. Further details regarding the training and testing of the CES and recreation CNNs will be provided below.

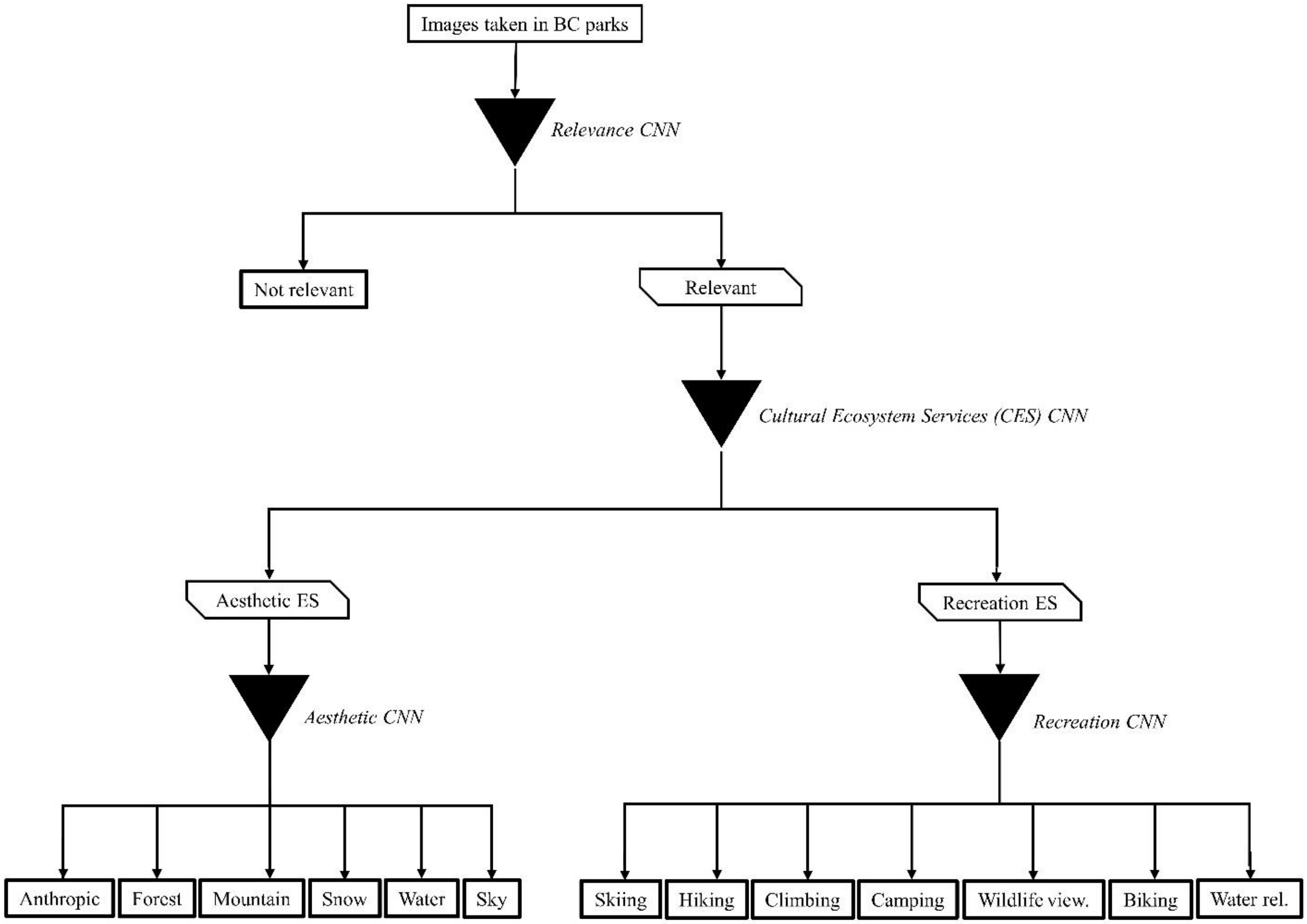

Figure 11. Workflow of the image classification process. First, each image is passed through the relevance CNN (ResNet-152 trained on Places365 database). Second, relevant images are fed into the CES CNN (ResNet-152 fine-tuned on aesthetic vs recreation database) that determine if they depict recreational or aesthetic experiences. Third, recreation images were classified with the recreation CNN (ResNet-152 fine-tuned on recreational activities database) while the aesthetic images were classified by the aesthetic CNN ((ResNet-152 trained on Places365 database).

Table 8. Classes used in the classification process.

Classes	Criteria of inclusion
Not relevant	Images acquired indoor in an urban context
Relevant	Images acquired outdoor in a natural context
Aesthetic	Images depicting vegetation, natural features or natural landscapes
Anthropic	Images depicting anthropic elements in the natural landscape (e.g. bridge)
Forest	Images depicting forest covered landscapes
Mountain	Images depicting mountainous landscapes
Snow	Images depicting snow covered landscapes
Water	Images depicting water covered landscapes
Sky	Images depicting skyscapes
Recreational	Images depicting people engaging in recreational activities in the natural landscape
Skiing	Images depicting people skiing or skiing equipment
Hiking	Images depicting people hiking or posing
Climbing	Images depicting people climbing or climbing equipment
Camping	Images depicting people camping or camping equipment
Wildlife viewing	Images of wildlife in the natural context
Biking	Images depicting people biking or biking equipment
Water related	Images depicting other recreational activities related to water (fishing, kayaking, etc.)

4.2.4 Training and testing of the CNN models

The training of both the CES CNN and the recreation CNN was performed with Colab (Google Research) using the Pytorch framework. Google Colab is a freely available tool, operated by Google, which allows the user to run Python code in a web browser, while providing free access to graphic processing units (GPUs). In the training process, a transfer learning approach was used, starting from the ResNet-152 architecture pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) image classification and localization dataset. To fine-tune the CES CNN, a training dataset populated with 7,000 images (3,500 aesthetic and 3,500 recreation) was used. As mentioned in section 2.3, the training images were obtained using the tags argument of the "*flickr.photos.search*" function of the Flickr's API. Training images of the aesthetic class were obtained using tags that could describe the ecosystems and landscapes present in BC such as: "*rainforest*", "*beach*", "*lake*", "*waterfall*", etc. Training images of the recreation class were also obtained querying Flickr API with appropriate tags. Specifically, 500 images were collected for each one of the seven recreational activities considered. For example, to collect training images for the class "skiing" the following tags were used: "*skiing*", "*skis*", "*snowboarding*", "*snowboard*", "*downhill*", "*skislope*", "*chalet*", "*snow*", "*backcountry*", "*resort*". The 3,500 recreation images were later used as the training set for the recreation CNN. The complete list of the tags used in the query of Flickr API is presented in the supplementary material (Table A2). Before initializing the training of the CNNs, the images obtained querying the Flickr API were visually analyzed by the authors to ensure that the content of the images reflected their tags.

To train the CES CNN, the first step was to create a validation set randomly selecting 20% of the images of both classes using the *RandomSplitter* function of the *fast.ai* library (Howard and Gugger, 2020). Therefore, the final training set for the CES CNN contained 2,800 aesthetic and 2,800 recreation images, while the validation set contained 700 images of both classes.

Similarly, for the training of the recreation CNN, 20% of the images of each class were randomly selected and used as a validation set, and the remaining 80% was used as a training set. Therefore, the final training set for the recreation CNN contained 400 images for each recreational activity and the validation set contained 100 images for each recreational activity. For the training of both CNNs we used the ADAM algorithm (Kingma and Ba, 2014) as an optimizer and a batch size of 10. To identify the optimal learning rate, the "*lr_find*" method of the fast.ai library was used. This method calculates the optimal learning rate using the learning rate finder approach introduced by Smith (2017). For the CES CNN the optimal learning rate was 2 e-4, while for the recreation CNN the optimal learning rate was 7 e-4. Both CNNs were trained for 20 epochs. Lastly, the performances of the image classification process were evaluated on a separate test database, composed of all the geotagged images (2,099) taken in BC provincial parks in 2020. The images of the test set were manually classified by the authors by applying the classes presented in Table 2 and then compared with the results of the CNNs. The performances of the four CNNs were evaluated both individually and jointly. The following metrics were estimated: (i) accuracy; (ii) precision; (iii) recall; (iv) F1-score. Accuracy is defined as the ratio between correctly predicted labels by the model and the total images. Precision is defined as the ratio between correctly predicted labels by the model and the images of each category. Recall is defined as the ratio between correctly predicted labels by the model to all images in the actual positive labels. Lastly, F1-score is the harmonic mean of the precision and recall. The choice of the CNNs evaluation metrics was made taking into account previous studies that used CNNs in the context of CES research, such as Cardoso et al. (2022).

4.2.5 Value assessment of recreational CES

Combining the outcomes of the image classification process described above and the secondary data, an assessment of the value of the recreational CES provided by the BC provincial parks system was performed. The method followed in this analysis was an adaptation of the traditional benefit transfer approach (as described by Champ et al. 2003) to crowdsourced social media data. Since the analysis presented in this study focuses exclusively on the benefit provided by the parks to BC residents, before applying the crowdsourced benefit transfer method Flickr users' home locations were estimated. In the following sections, the approaches used to estimate users' home location and the value of the recreational CES value will be explained.

Estimating users' home locations

In many previous studies that used social media users' home locations, those were extracted directly from information provided on public user profiles. However, as previously found by Da Rugna et al. (2012) only 40-48% of Flickr users share a home location in their profile. In addition, provided information may not be up to date. To overcome these issues, home locations were estimated using the max photo-user-days (max PUDs) methodology proposed by Bojic et al. (2015) and verified by Sinclair et al. (2020). The max PUDs method assumes that the user's home location can be estimated as the one in which they have spent the maximum number of PUDs. This methodology was applied to estimate the country and province of origin for each user who shared at least 10 geotagged images, while users with fewer images were excluded from further analyses. Furthermore, we excluded those users that took all their images within a single BC park to avoid overestimating the number of BC residents. To define countries and provinces, the shapefile provided by the GADM project (https://gadm.org/) was used. Retained visitors were divided into: (i) BC residents, (users with most PUDs in BC); and (ii) non-BC residents (users with most PUDs outside BC borders). To test the precision of the estimated home locations, the results of the max PUDs method were compared with the available countries and provinces of origin declared by the users in their profiles, using the accuracy metric. Finally, the home location estimated for each Flickr user was used to calculate the ratio of BC's residents vs non-BC's residents among the visitors in each BC provincial park.

Estimating CES value

To estimate the monetary value of the recreation ES provided by the provincial parks an adaptation of the traditional benefit transfer approach to crowdsourced social media data was applied. The basic premise behind the benefit transfer approach is that the economic benefit of a non-marketable service provided by an ecosystem can be estimated using values found in original studies conducted elsewhere (Loomis, 1992). Specifically, the aggregate benefit to users who come to forests for recreational purposes can be estimated by multiplying: (i) the consumer surplus per activity day that accrues to an individual engaged in a certain recreational activity; by (ii) the number of people engaging in that recreational activity in the considered area.

In the study at hand, to estimate the consumer surpluses generated by the various recreational activities, the values reported by Rosenberger et al. (2017) were adopted. Rosenberger et al. performed a meta-analysis of 342 recreation economic studies carried out in North America (the United States and Canada) between 1958 and 2015. This meta-analysis considers different recreational activities that can take place in natural ecosystems, reporting average consumer surpluses for each one of them. The number of annual visits in each park was estimated based on the publicly available BC Parks annual reports that include visitation statistics (https://bcparks.ca/research/). Specifically, the annual number of visits in each park was estimated by averaging the estimate of yearly visits for each provincial park from 2007 to 2018.

The approach used to adapt the benefit transfer method to crowdsourced social media data involves the following steps. First, images classified as not relevant and aesthetic were excluded. Second, images acquired by the same user, in the same park, during the same day were grouped to be transformed into AUDs. The activity assigned to each AUD (primary activity) was the most frequently depicted one among the images composing the same AUD. For example, if an AUD was composed of 3 images classified as hiking and 1 image classified as wildlife viewing the primary activity for that AUD was assumed to be hiking. In the case of two activities being depicted in the same number of images, the primary activity for the AUD was assumed to be the one that was depicted in the first image that the user acquired during the day. Third, in each park and for each recreational activity, was estimated the ratio with the following formula:

$$ratio\ activity_i = \frac{n°\ AUDs\ activity_i}{n°\ total\ AUDs}$$

Fourth, total consumer surpluses (CS) generated by the recreation ES in each provincial park for BC residents were estimated by applying the following formula:

$$CS = \sum_{i=0}^{n} (value\ act._i \times\ (\ avg.n°\ annual\ visitors\ \ \times ratio\ activity_i \times ratio\ BC\ visitors\ ))$$

Where $value\ act._i$ is the CS associated with the recreational activity i indicated by Rosenberger et al. (2017); $avg.n°\ annual\ visitors$ is the average number of visits estimated from

BC Parks statistics reports; $ratio\ activity_i$ is the ratio between the AUD for activity i and the total number of detected AUDs in the park, and $ratio\ BC\ visitors$ is estimated based on the outcomes of the max PUDs method. The choice of multiplying by the ratio of BC visitors was made to obtain estimates of the benefit that the BC provincial park system provides to BC residents.

4.2 Results

A total of 1,398,737 images, taken in BC between 2005/01/01 and 2020/12/31, were gathered accessing Flickr API. Among these images, 91,819 (6.6%) were taken within the boundaries of a provincial park and were retained for further analysis. These images were taken by 4,923 authors that on average took ~ 19 images each.

4.2.1 Image Classification

As shown in Figure 12, according to the results of the Relevance CNN 6,268 images (6.8%) were not relevant in the context of CES and therefore excluded from further analyses, while 85,551 images (93.2%) were relevant. The CES CNN classified the relevant images as follows: 55,132 images (64.4%) as images illustrating aesthetic ES and 30,419 images (35.6%) as images illustrating recreational ES.

The clustering of the aesthetic CES images yielded the following results: water 25.1%; anthropic 23.7%; forest 20.0%; snow and ice 16.5%, mountain and rocks 12.6%, and sky 2.1%. Grouping images taken by the same user on a single day portraying the same landscape typology we obtained 13,874 LUDs.

The recreation CNN classified images as follows: 30.2% as hiking, 16.1% as skiing, 13.1% as water-related, 12.6% as wildlife viewing, 10.9% as camping, 9.3% as biking, and lastly 7.8% as climbing. Grouping images taken by the same user on a single day portraying the same recreational activity 8,067 AUDs were obtained.

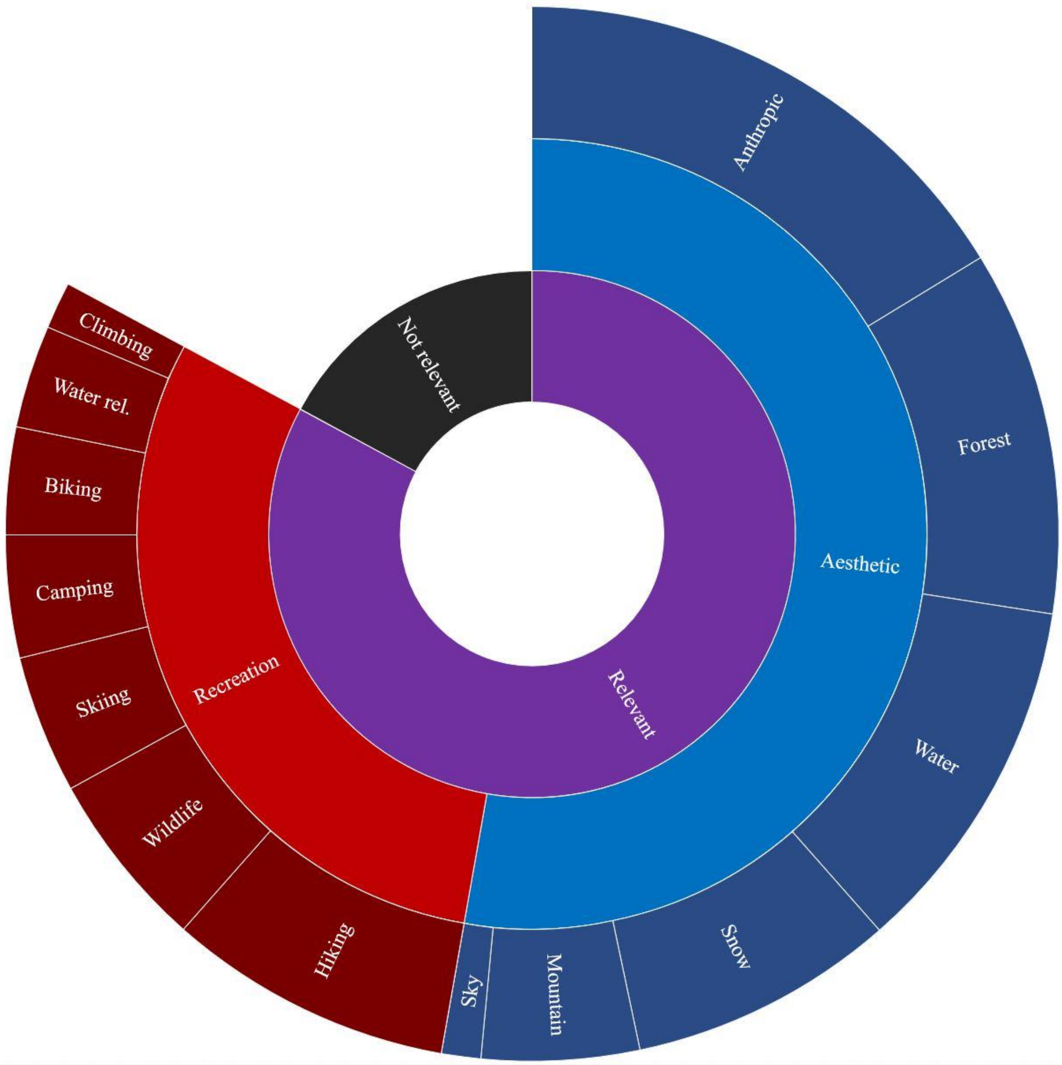

Figure 12. The inner layer represents relevance filter CNN, the mid layer represents CES CNN, external layer represents aesthetic and recreation CNNs. Outcomes of the image classification process undertaken on Flickr images acquired in the BC parks system from 2005 to 2020.

Table 9 shows the performance of the individual CNNs used in the image classification process on the test set. The relevance CNN had the worst performance with the not relevant images being classified with a precision of 0.69 and a recall of 0.67. The CES 1 CNN, used to distinguish between images depicting aesthetic and recreational ES performed well, correctly identifying aesthetic and recreational images with a precision and recall >85%. The aesthetic CNN had strong overall performances, with a weighted average of precision and recall of its classes >90%, but it is underrepresented classes of the test set (i.e., "anthropic" and "sky") had relatively poorer performance. Lastly, with respect to the recreation CNN, the overall performance was strong with the weighted averages of precision and recall higher than 85%. However, while the CNN was able to identify images depicting skiing and wildlife viewing with high precision and recall, other classes (i.e., camping, climbing and water related) were more frequently misclassified.

Table 9. Performances of the adopted CNNs.

	Class	Precision	Recall	F1-score
Relevance CNN	Not Relevant	0.69	0.67	0.68
	Relevant	0.99	1	0.99
	accuracy			0.99
	macro avg	0.84	0.83	0.84
	weighted avg	0.99	0.99	0.99
CES CNN	Aesthetic	0.94	1	0.97
	Recreation	0.99	0.87	0.93
	accuracy			0.96
	macro avg	0.97	0.93	0.95
	weighted avg	0.96	0.96	0.95
Aesthetic CNN	Forest	0.94	0.91	0.92
	Anthropic	0.69	0.94	0.8
	Mountain	0.94	0.83	0.88
	Sky	0.95	0.67	0.78
	Snow	0.92	0.99	0.95
	Water	0.93	0.91	0.92
	accuracy			0.91
	macro avg	0.89	0.88	0.88
	weighted avg	0.92	0.91	0.91
Recreation CNN	Biking	0.7	1	0.82
	Camping	0.78	0.84	0.81
	Climbing	0.62	0.7	0.66
	Hiking	0.91	0.85	0.88
	Water rel.	0.68	0.85	0.76
	Skiing	0.95	0.9	0.92
	Wildlife	0.92	0.93	0.93
	accuracy			0.86
	macro avg	0.79	0.87	0.83
	weighted avg	0.87	0.86	0.87

Regarding the performances of the image classification process as a whole, an overall accuracy of 0.85 was obtained (Table 10). In the same table are reported the statistics for each class, as well as the number of images on which these statistics were calculated. In Figure 13 is presented the confusion matrix of the various classes. As can be observed in the confusion matrix, there is little misclassification happening between aesthetic and recreation images (see quadrant b and quadrant c in Figure 13), meaning that the majority of the images that were manually classified as either aesthetic or recreation were correctly attributed to that class by the CES CNN. Instead, more misclassification occurred within the aesthetic (quadrant a) and recreation (quadrant d) class. The classes among which the most frequent class attribution errors occur are hiking and climbing and hiking and water related.

Table 10. Overall performances of the image classification process calculated on the test set.

Class	Precision	Recall	F1-score	N° images
Not relevant	0.69	0.67	0.68	33
Forest	0.92	0.91	0.91	469
Anthropic	0.60	0.94	0.73	99
Mountain	0.89	0.83	0.86	232
Sky	0.90	0.67	0.77	27
Snow	0.84	0.99	0.91	243
Water	0.85	0.91	0.88	303
Biking	0.67	0.93	0.78	15
Camping	0.72	0.68	0.70	31
Climbing	0.62	0.67	0.65	46
Hiking	0.91	0.75	0.82	309
Water rel.	0.68	0.62	0.65	55
Skiing	0.94	0.74	0.83	155
Wildlife	0.92	0.83	0.87	82
accuracy			0.85	2099
macro avg	0.80	0.80	0.79	2099
weighted avg	0.86	0.85	0.85	2099

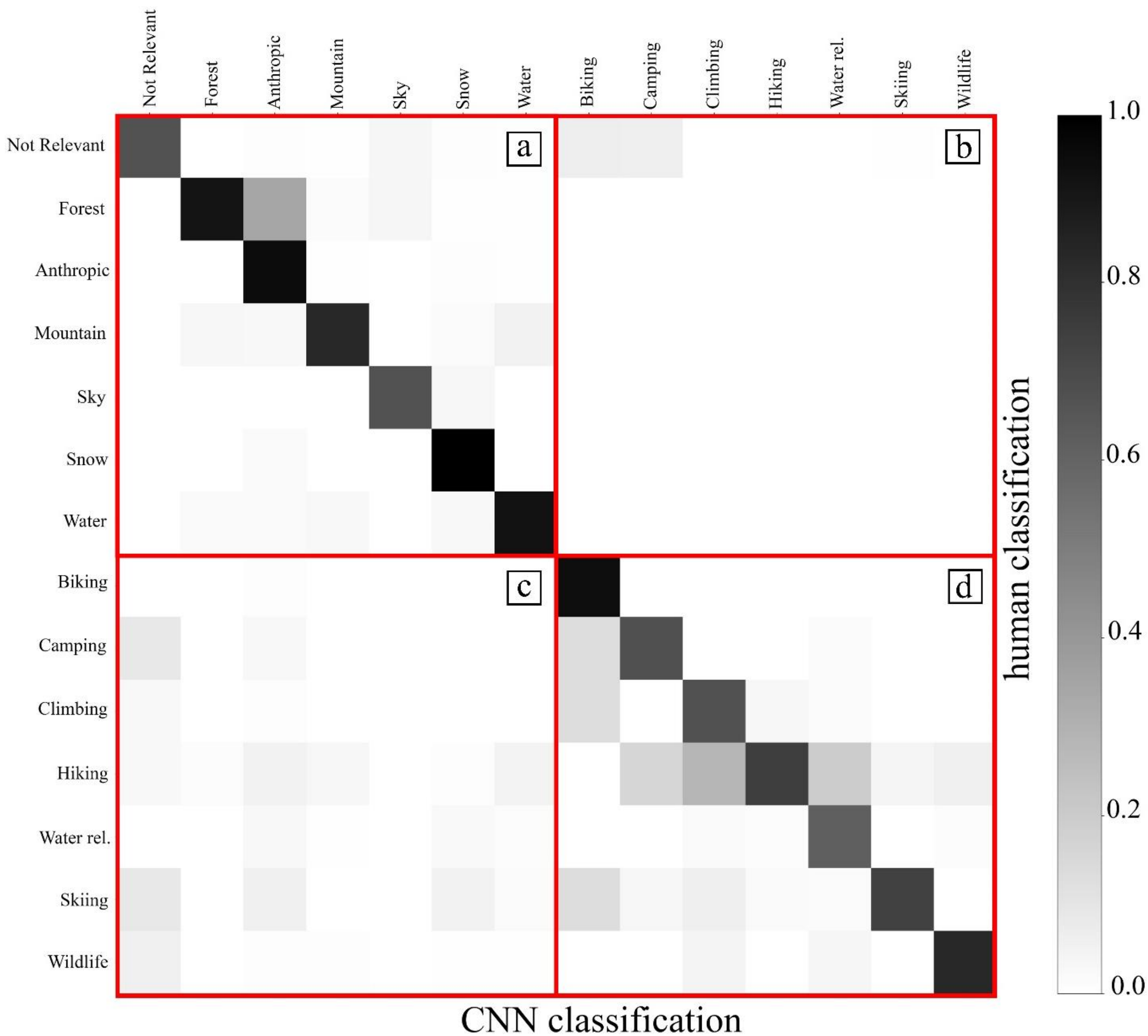

Figure 13. Confusion matrix created comparing the distribution of all the predicted responses and showing how they compare to their true classes. Values in the cells are the ratio between the nº images belonging to that class and the nº images assigned to that class by the CNN. These values were estimated analyzing the 2,099 images of the test set. Quadrant (a) shows the performances of the image classification process within the classes of the relevance CNN and the aesthetic CNN. Quadrant (b) and quadrant (c) show the performances of the image classification process between the classes of the aesthetic CNN and recreation CNN. Quadrant (d) shows the performances of the image classification process within the classes of the recreation CNN.

4.2.2 Visitors home location

Among the 4,923 Flickr users that uploaded at least one image taken within the boundaries of a BC provincial park, 3,992 (81.09%) had more than 10 PUDs in their Flickr libraries and therefore were included in the application of the max-PUD method. Overall, according to the results of the max PUDs method, the users came from 63 countries and five continents. As shown in Figure 14, the majority (58.2%) were Canadian, mostly from BC (80.4%), Alberta (7.3%), and Ontario (7.0%). The second most frequent home location were the United States (23.6% of the users), followed by United Kingdom (4.8%), Australia (1.8%), and Germany (1.6%).

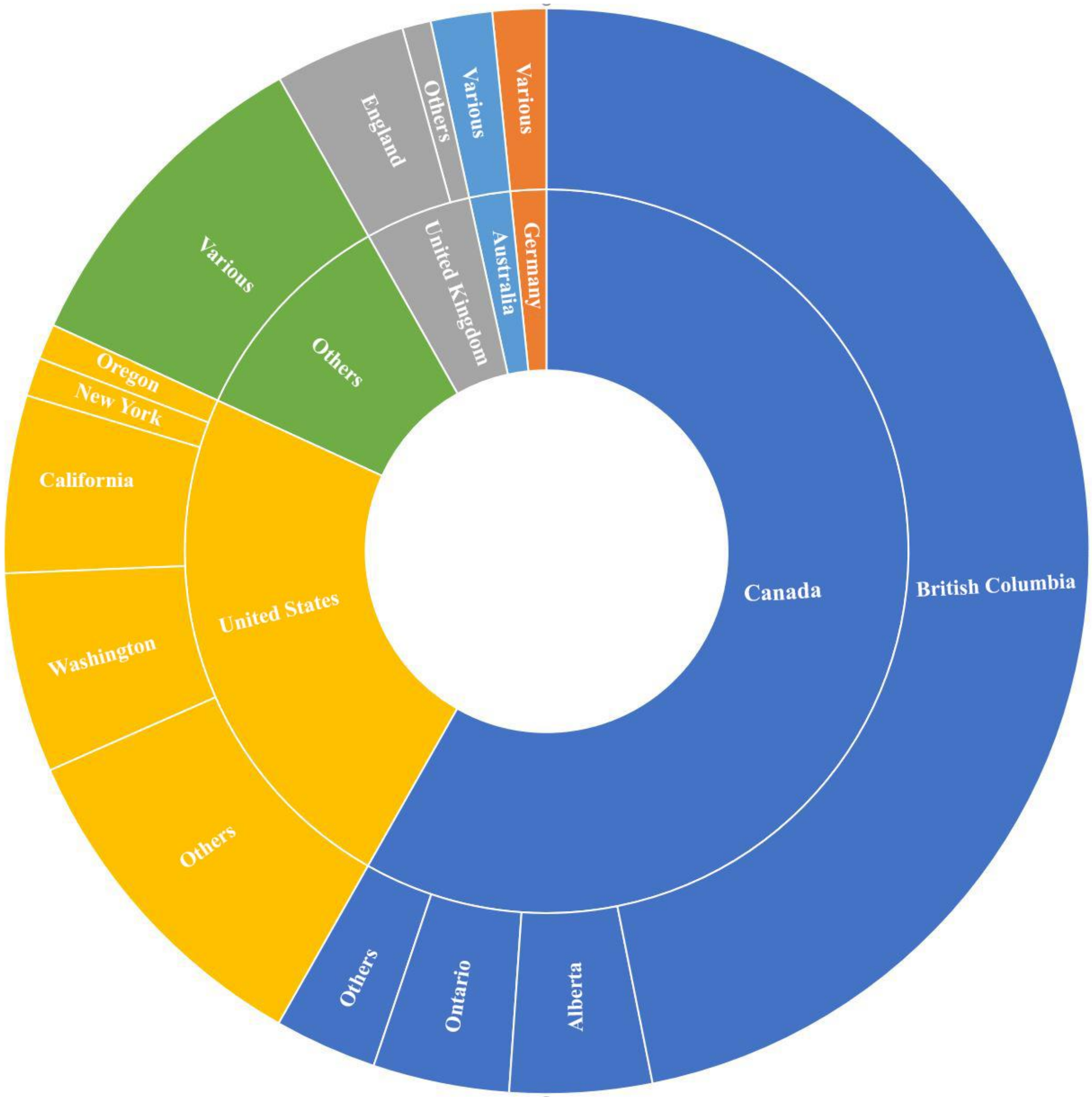

Figure 14. Countries (GADM 1) and provinces/states (GADM 2) of origin of the Flickr users included in the analysis, as predicted by the max PUDs method.

Among the users that were included in the application of the max-PUDs method, 2,406 (57.38%) declared their home location in their Flickr accounts. 93 authors did not provide existing countries and provinces in their profiles, bringing the total number of authors used in the evaluation of the performances of the max-PUDs method to 2,313. Comparing the results obtained with the max-PUDs method with the home locations reported by the authors in their accounts, an accuracy of 87% at the country level (GADM 1), and 70% at the provincial level (GADM 2) was determined.

4.2.3 Characterization of BC provincial parks system

In each provincial park were calculated: the proportions of the landscape categories depicted, the proportions of the recreational activities depicted, and the proportion between BC and non-BC visitors. Figure 15 shows the results obtained in the 10 parks which had the highest number of detected relevant images, while Table 11 presents results for all parks in which the sum of LUDs and AUDs was greater than 200. Parks where fewer LUDs and AUDs were detected were excluded from this analysis, to avoid estimates based on nonrepresentative data. As shown, the ratios between the various categories differ greatly across the various parks. For example, the images acquired in Garibaldi and Blackcomb glacier parks mostly portrayed snow-covered landscapes and people skiing, while images acquired in Shannon falls, Wells Grey, Little Qualicum falls, and Brandywine falls, mostly portrayed water-covered landscapes.

Lastly, applying the crowdsource benefit transfer method the consumer surpluses generated by the recreational ES provided by BC parks were estimated. The parks in which the highest values for recreational ES were estimated with the crowdsourced benefit transfer approach are the following: Cypress (105.2 million CAN$/year), E.C. Manning (69.3 million CAN$/year), and Mount Seymour (56.7 million CAN$/year). In the appendix are reported the consumer surpluses from Rosenberger et al. (2017) that were used in this analysis (Table A3) and the average number of yearly visits for the evaluated parks (Table A4).

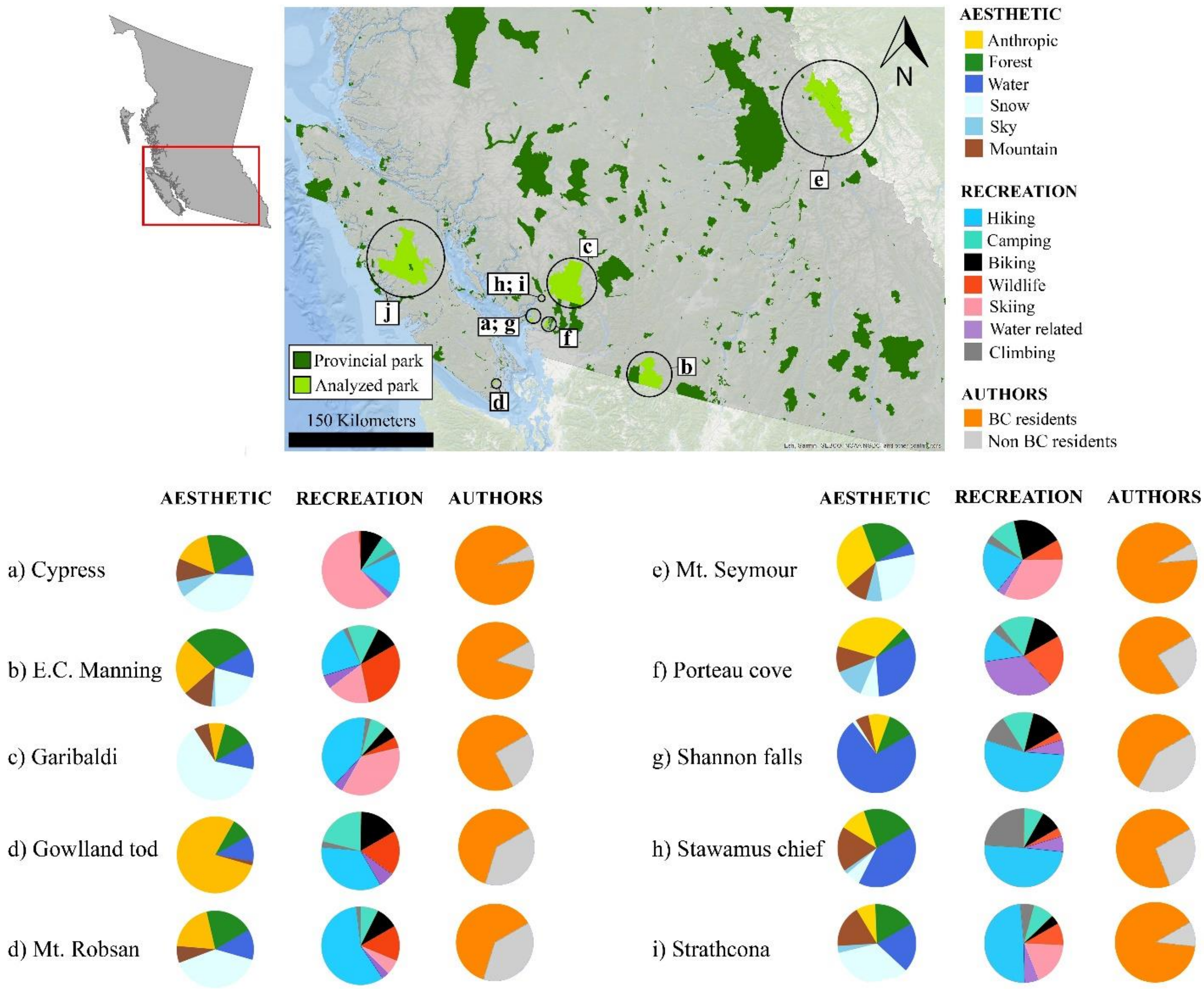

Figure 15. Characterization of the CES provided by the 10 parks in which the highest number of relevant pictures was detected. The pie charts illustrate the share of LUDs and AUDs in the parks.

Table 11. Characterization of the CES provided by BC provincial parks with more than 200 LUDs + AUDs. All the reported values are percentages. Percentages for the aesthetic CES are calculated on the total number of LUDs in each park, while the recreation CES percentages are calculated on the total number of AUDs. Percentages > 50% are in bold. The "BC" and "Non-BC" columns represent the percentages of BC and Non-BC visitors according to the max PUDs methodology. Lastly, in the "value" column are reported the average yearly consumer surplus generated by the parks in million CAN$/year assessed with the benefit transfer approach, from 2007 to 2018.

			Aesthetic CES						**Recreation CES**								
Provincial Park	PUD	AUD	Forest	Anthropic	Mountain	Sky	Snow	Water	Biking	Camping	Climbing	Hiking	Water rel.	Skiing	Wildlife	BC	N
Cypress	754	644	20	15.4	9.7	6.8	38.9	9.3	9.2	6.8	2.2	17.2	2.6	**61.1**	0.9	93.8	6.
Stawamus Chief	683	465	22	10.8	18.6	1.6	6.1	40.8	8.4	8	24.1	49	6.2	0.6	3.7	72.6	27
Garibaldi	637	501	12.4	6.9	6.3	0.6	**62.3**	11.5	5.4	7.2	2.2	39.9	3.8	36.9	4.6	74.4	25
Gowlland Tod	725	281	8.4	**78.2**	1.7	0.4	0.3	11	16.4	21.4	2.5	34.9	6.3	0	18.5	61.9	38
Mt. Seymour	510	380	22.4	30.8	9.4	6.7	25.7	5.1	20.3	10.8	3.4	21.3	3.2	32.8	8.2	92.9	7.
Mt. Robson	427	205	20.1	19.9	7	0.7	39.6	12.6	9.3	7.3	2	**57.6**	2.9	6.3	14.6	59.1	40
Shannon Falls	386	179	11.1	8.8	5.4	0.3	1.6	**72.8**	12.8	12.8	11.3	**53.6**	6.1	0	3.4	58.9	41
E.C. Manning	296	269	29.4	23.6	12.2	1.7	20.6	12.5	9.3	13	1.9	22.3	5.6	17.8	30.1	87.5	12
Strathcona	277	209	17.3	7.9	17.3	2.9	34.3	20.2	3.8	8.6	5.7	48.8	5.8	18.2	9.1	90	10
Porteau Cove	330	108	4.5	32.7	10.6	12.4	7.6	32.1	12	14.8	3.7	13	34.3	0.9	21.3	75.9	24
Golden Ears	252	134	38.9	8.3	10.3	0	11.5	31	2.3	23.1	8.2	36.6	21.6	1.5	6.7	86.7	13
Wells Gray	264	116	9.8	8	4.9	0.8	2.7	**73.9**	4.3	10.3	7.8	43.1	15.5	3.4	15.6	42.4	57
Juan De Fuca	244	121	14.3	13.9	38.5	1.6	2	29.5	4.1	9.9	9.9	30.6	27.3	0	18.2	85.6	14
Goldstream	235	106	37.4	23.8	3.4	0.9	0.9	33.6	9.4	11.4	2.8	28.3	10.4	0	37.7	83.9	16
Lac Du Bois	177	138	40.7	14.1	27.7	1.7	10.2	5.6	10.9	18.8	13	33.3	15.9	3.7	4.4	94.6	5.
Brandywine Falls	210	64	21.9	11.4	3.3	1	3.3	**59**	9.4	15.6	9.4	**57.8**	0	3.1	4.7	50.5	49
Macmillan	195	**77**	**88.2**	8.2	0.5	0	0	3.1	18.2	10.4	7.8	**62.3**	1.3	0	0	61.5	38
Joffre Lakes	159	93	9.4	2.5	3.1	1.3	49.7	34	2.2	5.4	4.2	**66.7**	3.2	9.7	8.6	61.3	38
Mount Assiniboine	151	101	26.5	11.9	6.6	0.7	46.4	7.9	1	5	1	**51.5**	2	31.6	7.9	73.7	26
Rathtrevor Beach	128	88	21.9	24.2	8.6	125	7	25.8	4.5	11.4	1.2	10.2	40.9	0	31.8	81.6	18
Blackcomb Glacier	95	116	0	4.2	3.2	0	**89.5**	3.2	2.6	0	0.9	22.4	2.6	**70.7**	0.8	67.1	32
Qualicum Falls	168	42	21.4	9.5	5.4	1.2	1.2	**61.3**	7.1	19	0	54.8	4.8	0	14.3	60.2	39

4.2 Discussion and conclusion

Images shared on social media can offer valuable insights on CES to ecosystems and park managers. In particular, the visual content of the images can be used to obtain a qualitative understanding of the relationship between humans and the natural landscape (Richards and Tunçer, 2018). However, performing such analyses manually severely limits the number of images that can be considered, hence recently researchers have been using machine learning tools (CNNs) to automatize the image classification process. This study aimed to develop a method to characterize CES provisioning via CNNs deployed on geotagged images taken in the study area. Geotagged Flickr images were used both as a source of training data and a source of crowdsourced data for CES analysis; and two versions of the ResNet-152 architecture, one trained on the Places365 dataset and one trained on Imagenet, were used as a starting point of the transfer learning approach. Ultimately, the outputs of the image classification process and image metadata were used to characterize and value the CES provided by BC provincial parks system.

4.2.1 Performances and outcomes

The deployed CNNs had an overall accuracy of 85%, a result that aligns with the findings of Terry et al. (2020), Gosal and Ziv (2020), and Cardoso et al. (2022) that showed how CNNs could be successfully adopted to classify images based on the CES depicted on them. This suggests that freely available resources such as Google Colab can be effectively used to train state-of-the-art CNNs, thanks to the transfer learning approach. The four classification levels (Relevance CNN, CES CNN, aesthetic CNN, and recreation CNN) had different performances, with the CES and the aesthetic CNNs being the best performing, followed by the recreation CNN, and by the relevance CNN. The fact that the relevance CNN had the poorest performances could indicate that using purposely designed CNN allows researchers to obtain more accurate results. To the best of our knowledge, Cardoso et al. (2022) is the only study that applied the transfer learning approach to train CNNs for CES classification. Ultimately, the performances of the CNNs trained by Cardoso et al. align with the performances of the CNNs trained in our study. Also, image metadata appears to be a reliable data source for assessing the home locations of natural parks visitors (~ 87% accuracy at GADM1 and ~ 70% at GADM2 when compared with the home locations shared by the users in their Flickr profile) as previously demonstrated by Sinclair (2020). While the margin of error is still significant, the alternative approach (i.e. using only the images of authors that declared their Country of origin in their Flickr account) would have caused the discard of more than 50% of the total images.

Furthermore, a traditional survey administrated to BC residents by Kux and Haider (2014) supports the trends found in this study. For example, in this survey hiking was identified as the recreational activity in which BC residents most frequently engaged, and it is also the most frequently depicted among the images classified. However, according to the results of the abovementioned survey, the popularity of water-related activities seems to be underrepresented in Flickr images. This phenomenon could be explained by: (i) the inherent difficulties in acquiring images while engaging in activities that involve water; (ii) the fishing bans enforced in some BC provincial parks. Furthermore, it appears that images can represent the landscape features that attract visitors to a natural area. For example, provincial parks that have remarkable waterfalls within their borders (e.g., Shannon falls, Brandwyne falls, and Little Qualicum falls provincial parks) have water as the most photographed landscape feature. MacMillan park, which protects a famous old-growth forest on Vancouver Island, has forest as the most photographed landscape feature. These findings support the hypothesis that by combining the metadata of images with image classification tools, it is possible to explore several aspects of the CES provided by a study area. To the best of our knowledge, there are no economic value assessments of the recreation CES provided by BC provincial parks; hence, it is not possible to compare the estimates obtained using the crowdsourced benefit transfer approach with estimates obtained with other more traditional techniques.

4.4.2 Innovativeness, limitations and future perspectives

The use of crowdsourced images in the study of CES is a growing area and in the last decade, hundreds of studies have used social media images to explore the provision and the consumption of various ES as shown by Ghermandi and Sinclair (2019). However, the analysis of the content of the images via CNN presented in this paper is an innovative approach that has still unexplored potential. For example, the automated analysis of the images acquired in an ecosystem could enable the mapping of CES consumption without having to conduct expensive and time-consuming *in-situ* surveys. Also, by knowing where people recreate and which activities are more popular among recreationists, ecosystem managers could better tailor the creation of trails and infrastructure to CES demand. Furthermore, the crowdsourced benefit transfer approach introduced and adopted in this study could allow ecosystem managers worldwide to obtain cheap and fast exploratory estimates of the monetary value of recreation. Finally, these estimates could allow for a better understanding of the trade-offs between CES and the other ecosystem services.

However, this approach has limitations. These limitations can be divided into two categories: those associated with using Flickr as a data source, and those associated with the applied methodological approach. The limitations of using Flickr as a data source in the study of CES have been thoroughly discussed in the past by various authors (e.g. (Ghermandi and Sinclair, 2019)). Generally, the biggest concerns arise from user demographics and the underrepresentation of areas located further away from cities and population centers. Furthermore, the popularity of social media platforms is subject to fluctuations and Flickr's popularity is currently declining (Gao and Mou, 2021). Nevertheless, at the moment Flickr appears to still be the best option for the study of CES, providing one of the longest time series of geotagged images that are easily and freely accessible. Also, while transfer learning has several advantages (such as flexibility of class definition, and the possibility of training CNNs from a small dataset of images), it has also limitations. First of all, transfer learning is a suitable approach only if the two classification tasks (the one for which the initial model was developed and the one for which the transfer learning approach is applied) are similar enough (Howard and Gugger, 2020). However, currently, no standardized measures of similarity between classification tasks exist. Furthermore, when adopting a transfer learning approach, the developer cannot modify the inner layers of the CNN to better adapt it to the classification task.

As specifically concerns the methodology applied in this study, the estimated ratios of visitors' participation in the recreational activities in each park, obtained via image classification, could introduce biases towards certain activities. For example, as discussed above, water-related activities (e.g., kayaking) could be underrepresented; and similar concerns arise for extractive recreational activities (e.g., hunting) and activities popular among demographics that are less likely to use Flickr, or social media in general. Furthermore, certain recreational activities such as picnicking, despite being popular, are difficult to recognize by analyzing images alone. Likewise, niche activities such as mushroom gathering would show up very infrequently in images crowdsourced from social media. These limitations imply that the assessment of CES conducted exclusively relying on crowdsourced social media data could result in an underestimation of the diversity of recreational activities carried out in an area and their total benefit. Lastly, the authors recognize that some of the classification choices made in this study are subjective (e.g., separating aesthetic and recreational ES based on the presence of human subjects in the image) and future research could apply different class definitions of recreational activities. At the same time, transfer learning warrants such flexibility as to potentially allow future researchers to easily adapt the classification structure of this study to their needs.

New studies are needed to tackle these limitations. In particular, these studies should explore the ability of social media pictures to faithfully describe the ratio between the recreational activities in which visitors of natural areas engage. To this end, the outcomes of the image classification process could be compared to *in-situ* surveys. Another area of research that should be further investigated is the differences in types of CES depicted in images shared on various social media platforms (e.g., Flickr vs Instagram). This would help researchers to get a better understanding of the potential benefits associated with integrating multiple data sources that have been suggested by various authors (Tenkanen et al., 2017). Lastly, new studies that compare the values obtained with traditional approaches to nonmarket values assessment, and the ones obtained in this study, would help in validating the crowdsourced benefit transfer approach.

At present, the most important limiting factor for the use of CNNs in the study of CES is the lack of freely accessible classified images that can be used in training. Our study demonstrated how Flickr images can be effectively used to obtain training data or CES image classification tasks. However, to further facilitate the creation of new CNNs aimed to classify CES, a first step could be the sharing of programming codes and image datasets among research groups as already suggested by Cardoso et al. (2022).

Once the potential biases introduced by the use of the crowdsourced benefit transfer method will be better explored and understood, this method has the potential of tackling the mapping and valuing issues that have historically limited the inclusion of CES in ecosystem management plans. This method could prove itself useful in obtaining estimates of the recreational values of remote study areas where the use of traditional valuation methods based on surveys is expensive and time-consuming. However, even applying the crowdsourced benefit transfer method, little-visited recreational areas will remain difficult to adequately value due to the lack of available social media images. Lastly, in this study, the crowdsourced benefit transfer approach was applied by multiplying the ratio of recreational activities by the number of annual visits as reported in BC park's statistic report. However, in the past, numerous authors (e.g. Lingua et al., 2022; Sinclair et al., 2019; Wood et al., 2013) have obtained estimates of annual visits by using regression models and the number of geotagged images as the independent variable. In future studies, the crowdsourced benefit transfer approach could be applied also in areas for which the number of annual visits is unknown, estimating it from the number of geotagged images.

In conclusion, CNNs and transfer learning offer new and promising ways to use social media images to explore CES. Relying on these approaches and this data source, researchers and ecosystem managers could obtain, quickly and cost-effectively, information on CES provision and consumption to an unprecedented level of detail, almost in real time. Ultimately, this would allow for a better harmonization of CES with the other services provided by ecosystems. However, before the outcomes of social media image classifications can be used extensively and reliably to this end, future studies must investigate the biases that they introduce and their shortcomings.

Chapter 5
Assessing forest recreational potential from remote sensing and social media data with deep learning

5.1 Background and motivation

Analyses of the metadata of geotagged social media images has been undertaken across various ecosystems around the world for differing objectives including to: (i) identify temporal trends and hotspots of recreational activities (e.g. Schirpke et al., 2018), (ii) estimate the number of visits into natural areas e.g. (Lingua 2022a; Tenkanen et al., 2017), and (iii) assess the monetary value of recreational sites (e.g. (Ghermandi, 2018); (Sinclair et al., 2018)). These studies have demonstrated how the metadata associated with social media images can provide highly useful quantitative information for CES management (e.g. number of yearly visits and visitation trends). However, metadata alone cannot shed light on the reasons why people choose to visit an area, or the activities that recreationists carry out. To answer these qualitative questions, researchers have adopted two strategies. The first is the content inspection of crowdsourced social media images. The second is the integration of said metadata and remotely sensed data.

Relevant examples of studies that have adopted the first approach have been thoroughly discussed in Chapter 4. As concerns the combined use of crowdsourced social media data and remote sensing technologies data, the following examples can be made. Bernetti et al. (2019) used the metadata of Flickr images to determine the relationship between forest and topographic variables, obtained via various remote sensing technologies, and picture density in Tuscany (Italy). Similarly, Ciesielski et al. (2021) used Flickr data and boosted regression tree models to determine which forest variables influenced picture acquisition by forest visitors. These variables were obtained combining various sources including land cover (Hościło and Tomaszewska, 2014) and the Shuttle Radar Topography Mission (Rodriguez et al., 2006). Lastly, You et al. (2022) assessed the spatial-temporal dynamics of forest recreation values in the Zhejiang Province (China) integrating remote sensing data and text-mining of social media comments. In all the above studies, remote sensing data were used to obtain topographic, socioeconomic, and forest biometric data. Coincident with the location (and in some cases time) social media images were acquired, hence allowing the identification of those variables that had a positive effect on forest recreation.

Remote sensing data and the output of automated image content analyses have not often been combined, therefore the potential of their integration is not fully explored. The overarching objective of this study therefore is to address these knowledge gaps, developing a methodology to effectively integrate CES in forest management planning, relying on remote sensing and crowdsourced social media data. To do so, the outcomes of the image classification process are paired with their corresponding topographic, socioeconomic and forest biometrics variables obtained using remote sensing technologies. Results will provide insights into which recreational activities are most popular in the observed forests and which natural and social attributes may potentially drive their popularity. Maps of the recreational potential of forests as well as of forest recreation economic value will be produced and the limitations and future perspectives of our approach discussed.

5.1 Data and methods

The methodology applied follows three steps. First, to each relevant image, classified in Chapter 4, the following characteristics were attributed based on its geographic coordinates: topography (slope and elevation), forest biometrics (canopy height, forest cover, gross stem volume, total biomass), and anthropogenic impacts (global human influence index) variables. Second, a random forest classifier model was trained to identify the most likely forest recreational activity to occur in an area, based on the above-mentioned variables. Finally, two types of maps were produced: (i) maps showing recreational potential; and (ii) maps showing recreational value. The recreational potential maps were created using the random forest classifier model to predict the most suitable recreational activity for points regularly distributed in a grid. The recreational value maps were created by applying a crowdsourced benefit transfer approach (Champ et al., 2003). In the following sections, further information on each of these steps will be provided.

5.1.1 Variables attribution and remote sensing technologies data

To assign to each image its corresponding topographic, anthropogenic impacts and forest biometrics variables, the Python package "rasterstats" (https://pythonhosted.org/rasterstats/) was used. The temporal variable was directly extracted from the images metadata.

Topographic data have previously been found to influence the provision forest recreation (Abildtrup et al., 2013; Roovers et al., 2002). In particular two topographic variables were found relevant in these studies and therefore used here: elevation and slope, both derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation model (GDEM V2, 30 m) (Tachikawa et al., 2011).

Several authors have shown the impacts that forest structure variables (e.g. stand density and basal area) have on forest recreation (Agimass et al., 2018; Carvalho-Ribeiro and Lovett, 2011). To examine the influence of forest structure on the provision of ecosystem services, we used four wall-to-wall, 30-m forest structure metrics, namely Lorey's height (the average height of all trees in a stand weighted by tree basal area), basal area, volume and above ground biomass, derived from annual composites of Landsat satellite imagery using the imputation method described in Matasci et al. (2018a; 2018b). This method used airborne laser scanning (ALS) and field plot data to estimate forest structure from topographic- and Landsat spectral predictors, using a k-Nearest Neighbor approach. Lastly, previous studies conducted in the same study area have shown how forest recreation patterns are influenced by the passing of the seasons (Lingua et al., 2022a; Lingua et al., 2022b). For this reason, the day of the year in which the images were acquired was also extracted from the metadata.

5.1.2 Random forest classifier model

A random forest classifier was used to identify if and how the independent variables affect the use of forest areas for recreation purposes, and predict the recreational activity most likely to occur in a set of regularly distributed points. The random forest model was created using the "RandomForestClassifier" module of the scikit-learn library.

To train the random forest classifier, the entire dataset of images depicting recreational activities was used. The dependent variables of the model were the five recreational activities identified by the recreation CNN (hiking, skiing, camping, biking and wildlife viewing) while the independent variables were the topographic, forest biometrics, anthropogenic impact, and seasonal data previously illustrated. The dataset was divided randomly selecting 30% of the samples to be used as training data, while the remaining 70% was used as validation dataset. The model included 30 fully-grown decision trees.

To identify which variables influenced the outcome of the model the "feature_importances_" attribute of the "RandomForestClassifier" module was used. The feature importance is the estimate of how much of the predicting power of the model is given by a variable and it is calculated as the decrease in node impurity weighted by the likelihood of reaching that node. The Gini impurity of a node can be defined as the likelihood of a random datum being misclassified if it were attributed to a random class (according to the class distribution in the dataset). Gini impurity can be calculated using the following formula

$$Gini\ impurity = \sum_{i=1}^{C} f_i(1 - f_i)$$

Where f_i is the frequency of label *i* at a node and *C* is the number of unique labels.

To explore the relationship between the dependent variable and independent variables partial dependency plots were created. Partial dependency plots show the marginal effect that a variable has on the outcome of the model. These plots were created for each recreational activity and variable pair.

Mapping the recreational potential

To demonstrate the potential of the method, the random forest classifier model was used to map the recreational potential of Cypress and Golden Ears provincial park across all four seasons. Cypress and Golden Ears provincial parks are among the most visited provincial parks in BC, with an average of ~1,200,000 and ~800,000 annual visits respectively (BC Parks statistics). To do so, a net regular of 200,000 distributed points was created with each point assigned the independent variables. The model was then applied to predict the most likely forest recreational activity to be carried out in each season at any given point.

Mapping the value of recreation

To estimate the monetary value of the recreational service provided by the forests included in BC`s provincial parks system, two data sources were used: (i) the consumer surpluses for forest recreational activities reported by Rosenberger et al. (2017); (ii) and the yearly visitation statistics by BC Parks. Rosenberger et al. (2017) reported the consumer surpluses that people enjoy from participating in forest recreational activities in North America, estimated performing a meta-analysis of 342 studies. BC Parks estimated the number of daily visitors in all of BC Parks from 2012 to 2018. These estimates are reported as the number of annual visitors available in the end of the year reports (https://bcparks.ca/research/).

The approach used for the monetary valuation was the crowdsourced benefit transfer applied in the previous chapter. When applying benefit transfer, the consumer surplus generated in a year by a recreational site is calculated using the following formula

$$CS = \sum_{i=0}^{n} (value\ act._i \times (\ avg.n°\ annual\ visits\ \times ratio\ activity_i\ ))$$

Where $value\ act._i$ is the CS associated with the recreational activity *i;* $avg.n°\ annual\ visitors$ is the average number of annual visits; $ratio\ activity_i$ is the ratio between the number of people engaging in activity *i* and the total number of people visiting the site.

In this study, this formula was applied to each cell of the same grid used for mapping the recreational potential of Cypress and Golden Ears provincial parks. The first step of the valuation process was to estimate the average number of visits in each cell, in every season. To do so, an assumption of linear relationship between the number of visitors and the number of images acquired was made. Therefore, the number of visitors in each cell, for every season, was estimated using the following formula:

$$avg.n°\ annual\ visits = \frac{n°\ img\ in\ cell\ during\ season}{tot\ n°\ img\ in\ park} \times avg.n°\ vistors\ in\ park$$

The ratios of the various recreational activities were estimated by grouping together the images acquired by the same user, during the same day, in activity user days (AUD). The activity assigned to each AUD (primary activity) was the most frequently depicted one among the images composing the same AUD. Then in each cell of the grid the ratio for each recreational activity was estimated using the following formula:

$$activity\ ratio_i = \frac{n°\ AUDs\ activity_i}{n°\ total\ AUDs}$$

Lastly, eq. 2 was applied in each cell of the grid, obtaining the annual consumer surplus generated.

5.3 Results

Combining the outcomes of the image classification process with the variables assigned to each image, it was possible to estimate the averages and standard deviations of the topographic variables, anthropogenic impacts, temporal variable, and biometrics variables within each recreational activity. Table 12 shows the averages and standard deviations, while Table 13 shows the differences between the averages of the activity and their statistical significance assessed using an ANOVA test.

Biking and skiing are the activities characterized by the highest GHII (Global Human Influence Index), while hiking, camping and wildlife viewing had the lowest GHII. Furthermore, images depicting hiking and skiing were acquired on average at the highest elevations and on steepest slopes.

Images depicting skiing and wildlife viewing were acquired in less dense forests (lower basal area, total biomass and gross stem volume) while there are no statistically significant differences in the canopy height variable.

Table 12. Characteristics describing the forest in which pictures of various recreational activities were taken, including topographic information, anthropogenic impacts (GHII) and forest biometrics data. Standard deviations are reported in brackets.

Activity	GHII	Slope (°)	Altitude (m)	Canopy height (m)	Basal Area (m^2/ha)	Gross Stem (m^3/ha)	Total Biomass (tonnes/ha)
Hiking	19.4 (16.0)	6.3 (13.8)	1034.1 (809.2)	25.5 (8.1)	45.0 (22.7)	621.8 (454.1)	241.0 (153.4)
Biking	23.6 (14.1)	9.1 (9.7)	731.2 (605.2)	24.9 (7.7)	43.4 (22.3)	588.4 (429.7)	240.0 (154.3)
Skiing	24.8 (13.7)	13.3 (12.7)	1325.7 (523.1)	23.6 (7.4)	36.7 (21.0)	468.0 (378.1)	194.0 (132.8)
Camping	20.0 (13.8)	9.1 (11.1)	611.4 (614.1)	23.7 (8.1)	40.8 (22.0)	535.5 (416.9)	221.2 (150.5)
Wildlife	19.7 (12.3)	12.9 (10.0)	633.6 (617.6)	22.4 (8.0)	37.3 (22.3)	471.4 (408.0)	199.3 (150.3)

Table 13. Differences in absolute value between the averages of the considered variables between the various recreational activities and statistical significance according to ANOVA test. Differences reported in bold are statistically significant.

Activ. 1	Activ. 2	GHII	Slope (°)	Altitude (m)	Canopy Height (m)	Basal Area (m^2/ha)	Gross Stem (m^3/ha)	Total Biomass (tonnes/ha)
Skiing	Wildlife	**5.0**	**3.8**	**692.1**	1.8	**4.2**	**86.3**	1.0
Skiing	Camping	**4.7**	**3.8**	**714.3**	3.0	**7.7**	**153.8**	18.8
Skiing	Hiking	**5.4**	**3.4**	**291.6**	1.9	1.6	**150.3**	**40.7**
Skiing	Biking	1.2	0.5	**594.4**	0.5	**8.3**	33.4	**46.0**
Wildlife	Camping	0.3	0.0	22.2	1.2	**3.5**	**67.5**	**19.8**
Wildlife	Hiking	0.3	**7.2**	**400.5**	0.1	2.6	**64.0**	**41.8**
Wildlife	Biking	**3.8**	**4.3**	**97.7**	1.3	**4.1**	52.9	**47.0**
Camping	Hiking	0.6	**7.2**	**422.7**	1.1	**6.1**	3.5	**21.9**
Camping	Biking	**3.5**	**4.3**	**119.9**	2.5	0.6	**120.4**	**27.2**
Hiking	Biking	**4.2**	**2.9**	**302.8**	1.4	**6.7**	**116.9**	5.3

The random forest classifier, trained on almost 18,000 images classified as: hiking, skiing, camping, biking or wildlife viewing by the level 2 CNN had an accuracy of 74%. As shown in Figure 16, the most important variable in improving the accuracy of the predictions is seasonality (*i.e.,* the day of the year in which the picture was acquired) contributing to ~25% of the decrease in Gini impurity. Topographic variables are the second most important, contributing to a decrease of ~ 18% (elevation) and ~13% (slope). Anthropogenic impacts (estimated using GHII) contributed to ~12% of the decrease in Gini impurity, while forest biometrics contributed the least (<10%).

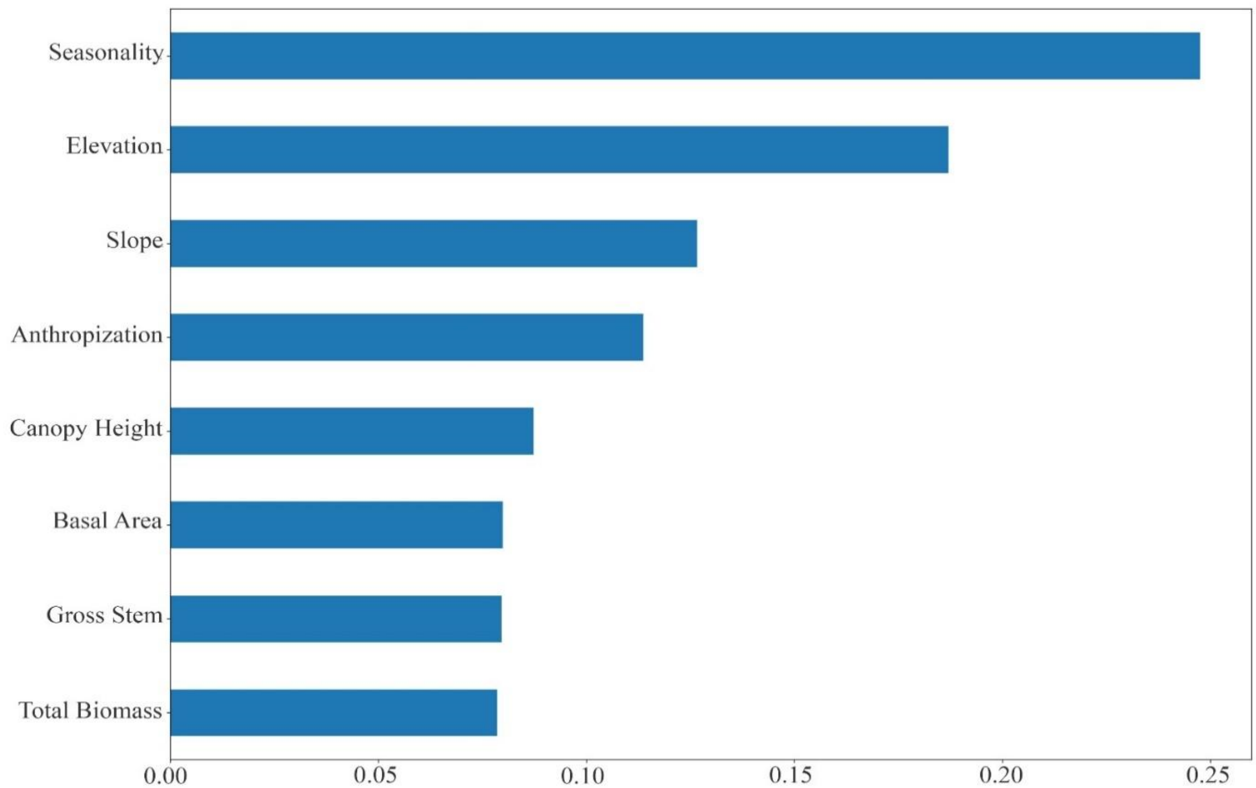

Figure 16. Variables importance. x-axis represents the percentage decrease in node impurity. The variables that contribute the most to the reduction of node impurity are the most important in determining the model accuracy.

Figure 17 shows the partial dependency plots generated from the random forest classifier model, for the four most influential variables. Only the variables that contributed the most to model performance were included (Gini impurity reduction > 10 %). The graphs illustrate how the variables influence the likelihood of an image to be classified as depicting one of the recreational activities.

As shown in Figure 17, the images acquired during summer months (from day 172 to day 265) are more likely to depict hiking and camping, while the likelihood of an image to depict wildlife viewing activities increases during spring (days 80-171) and autumn (days 266-355). As expected, the likelihood of an image to depict skiing decreases rapidly during summer months and peaks during winter (days 355-365 and 1-79).

Most of the forest recreational activities are more likely to occur in forests at lower elevations, except for skiing, where the likelihood peaks at around 1,000 m. Furthermore, forests characterized by low slopes are more often associated with camping and wildlife viewing, intermediate slopes favor biking, and high slopes favor hiking.

Lastly, anthropogenic impacts show a less clear trend: biking activity is more likely to occur where the anthropogenic impact on the forest is highest, while camping activities have the opposite behavior.

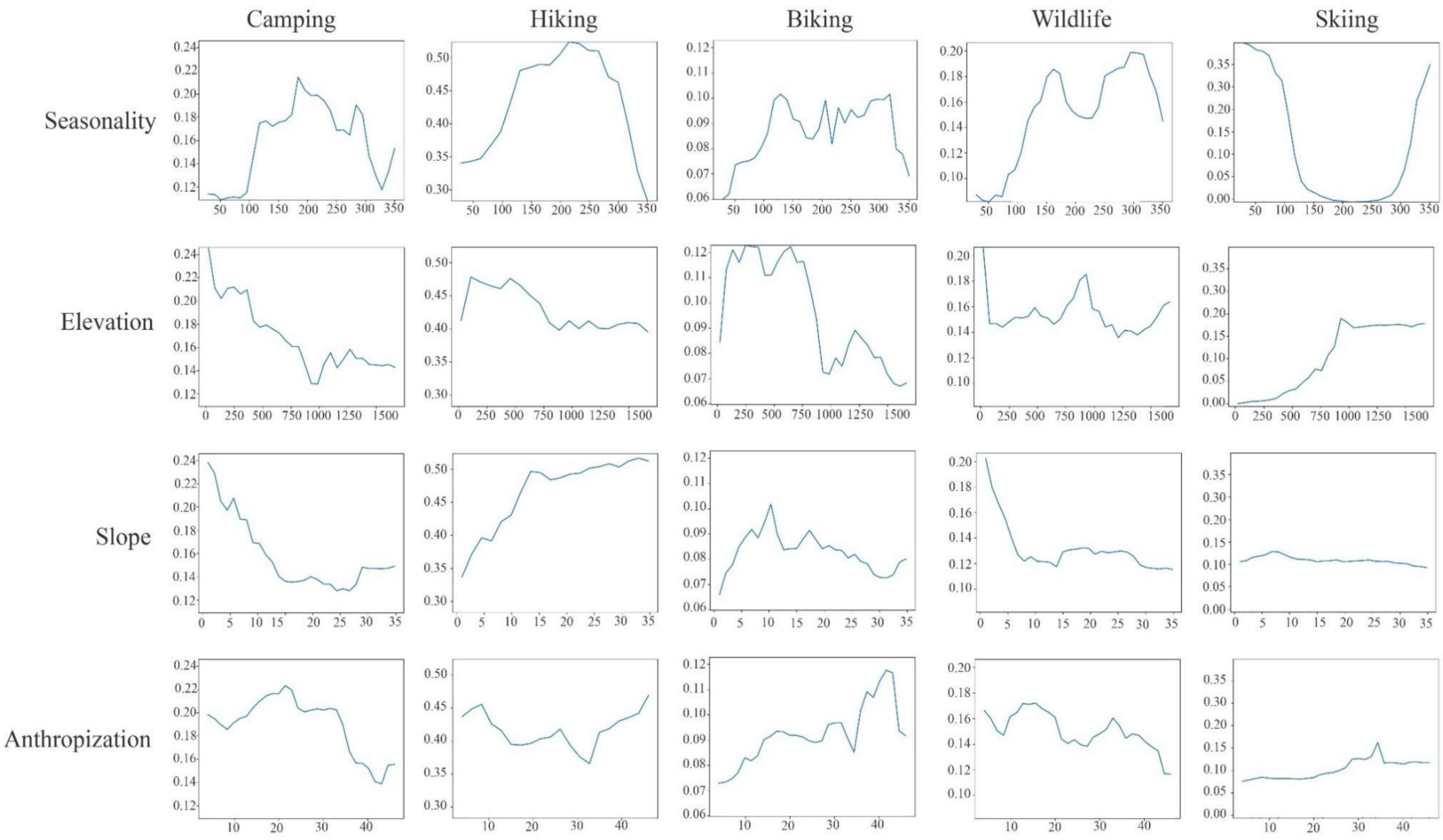

Figure 17. Partial dependency plots. On the x-axes are placed the different variables (day of the year for seasonality, m above sea level for elevation, degrees for slope, and GHII for anthropogenic impacts) on the y-axes is instead the predicted probability of the image to be classified as belonging to the class.

Figure 18 shows the recreational activities with the highest potential to occur across the four seasons in Cypress provincial park and Golden Ears provincial park. In both parks, hiking is by far the activity with the highest potential in spring, summer and fall, while skiing has the highest potential during winter. In particular, skiing has the highest potential in the south-east portion of Cypress park, especially during spring and fall and winter, while during summer it is absent. Other recreational activities, such as wildlife viewing, camping and biking have low potential in this provincial park.

In the last step of the analysis, an assessment of the monetary value of the recreational ecosystem service provided by Cypress park and Golden Ears park was undertaken. Using a crowdsourced benefit transfer approach, it was possible to estimate in every season, for each cell, the value of recreation expressed in CAN$/ha/day. As shown in Figure 18, for Cypress park it is possible to identify two hotspots of recreational value, one in the southeast of the park and one in the central part of the park; as concerns Golden Ears park the southern portion is the one in which the value of recreation is highest. Both these areas are the ones where most of the infrastructures (parking lots, hiking trails, Nordic ski and sledging areas, etc.).

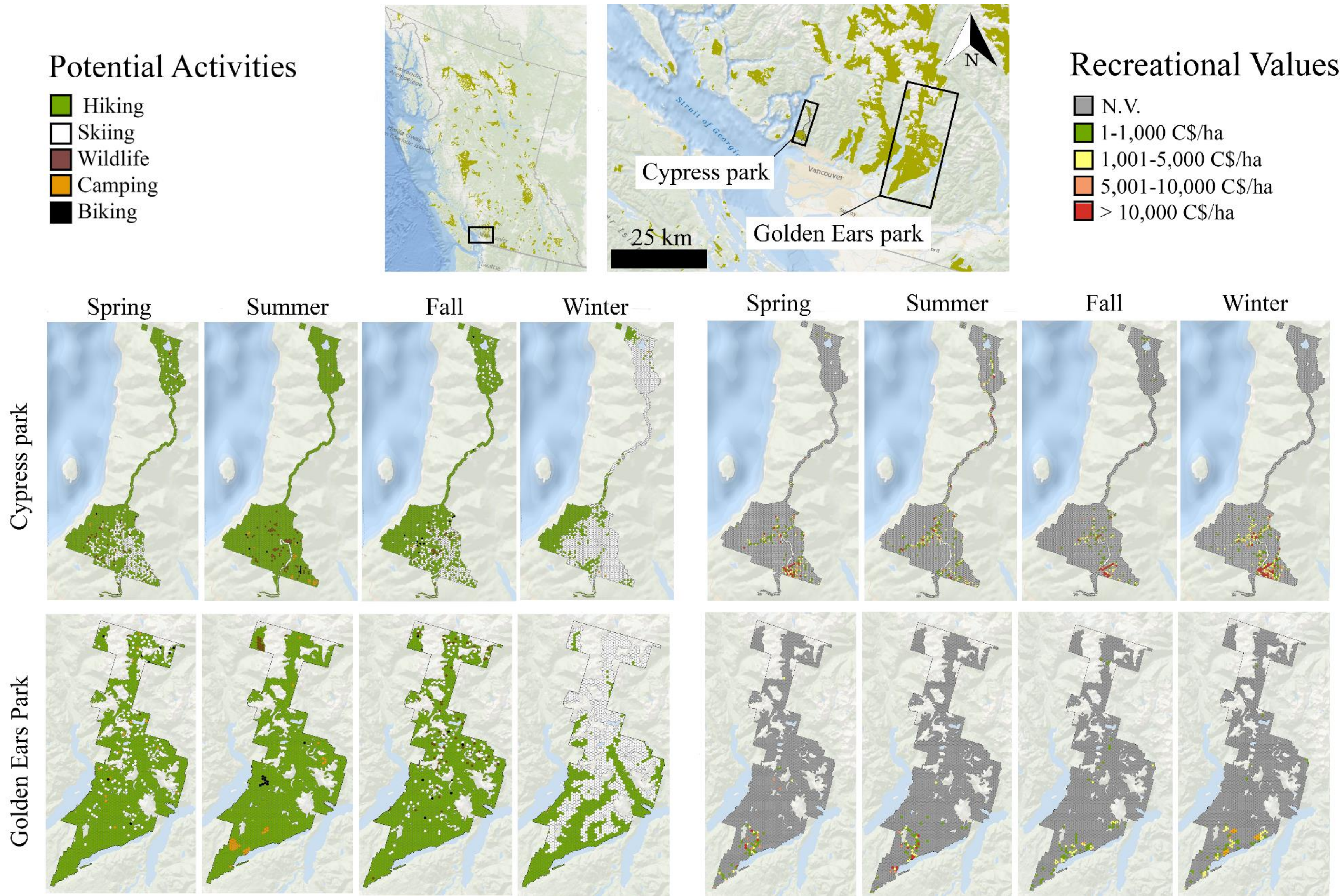

Figure 18. Potential activities map (left) and recreational value maps (right) of the forested areas of Cypress park and Golden Ears park.

5.4 Discussion

Crowdsourced social media images are a valuable data source to explore CES provision and consumption, however, to date, many of the existing studies consider exclusively the images metadata, not taking full advantage of the possibilities that crowdsourced social media data offer. In this study of forested BC provincial parks, Flickr images were automatically classified with purposely developed CNN and coupled with topographic, socioeconomic and forest biometric variables obtained via remote sensing technologies. This approach allowed us to examine if and how these variables influence the popularity of various forest recreational activities and map both the recreational potential and the value of the recreation CES.

5.4.1 Outcomes and performance

The performances of the CNN adopted to automatically classify the images used in this study align with the ones obtained by the previous applications of transfer learning to the study of CES (Cardoso et al., 2022) (Gosal and Ziv, 2020). The performances however were not homogenous for all classes. In particular, the classification process performed poorly when correctly identifying the not relevant images and the images depicting people biking. Based on the automated image classification process, the most popular recreational activity in BC forests is hiking, followed (in descending order of popularity) by skiing, wildlife viewing, camping and biking. These results compare well with those of conventional survey administrated to BC residents by Kux and Haider (2014), where BC residents were asked to indicate in which recreational activities they participated during 2012, and found that the most popular recreational activities among BC residents is hiking, followed by skiing, fishing and biking. Although the options given to the respondents of the survey differ from the activities considered in this analysis, the two most popular activities (hiking and skiing) match.

Previous studies suggested that recreational attractiveness of forested areas is influenced by: (i) topographic variables such as slope and elevation (Abildtrup et al., 2013); (ii) stand characteristics such as stand stocking and crown closure (Filyushkina et al., 2017; Weller and Elsasser, 2018); and anthropogenic impacts in the area (Roovers et al., 2002). While these variables seem to affect the number of visitors in forests, in this study not all of them had a significant effect on the type of recreational activities that people engage in. Specifically, it appears that the variables that have the most influence in determining the type of forest recreation are temporal and topographic variables, while forest biometrics play a lesser role.

5.4.2 Innovativeness, limitations and future perspectives

The literature around the use of social media data for CES exploration has been focusing on their use to describe CES demand and provision. Less attention has been given to how these data could be used at an operational level in ecosystem management. In this study, combining social media data with remote sensing data, and applying machine learning techniques, it was possible to obtain fine-grained information on how forest and landscape variables influence the type of recreation in which forest visitors engage. Previously, to obtain such data the most common approach was to resort to *in-situ* surveys that are costly and time-consuming (Richards and Tunçer, 2018). Instead, the methodology used in this study has shown the potential to be a cheap and fast alternative of such surveys. Furthermore, the produced maps of recreational potential and recreational values provide useful insights to forest managers. Recreational potential maps could be used to plan and locate the implementation of forest recreational infrastructures, such as hiking or biking trails, ensuring that the chosen forested area is suitable for that activity. Recreational values maps could instead be useful in providing detailed insights on the potential costs of partial park closure or the impacts of forests disturbances.

Despite the promise of the approach, however, future in-field applications do have some limitations. Ciesielski and Stereńczak (2021) have argued that social media user demographics could substantially differ from the ones of forest recreationists, causing the selection of a nonrepresentative sample. Flickr data appear to be less prone to this bias compared to other social media data (Hausmann et al., 2018). Furthermore, the various forest recreational activities could be characterized by different frequency of images acquisition. For example, hikers could be more inclined to take pictures than bikers. This could limit the availability of images portraying certain forest recreational activities, and affect the ratio of activities used in the assessment of recreational values. In addition, the possibility of freely accessing social media data may not be granted, since changes in social media platform policies can happen abruptly. Lastly, despite the fact that crowdsourced social media data are often referred to as "big data", the share of relevant images among the ones gathered is only 4.6%. This indicates that social media data should not be seen as a panacea for the study of forest CES, but rather as a useful source of insights on CES provision and demand, especially in peri-urban forests and forests with high recreational value.

So far, the research around the use of social media data in the study of forest CES has been mostly focusing on the development of methods and approaches to extract spatial and quantitative information on recreational fluxes. This study suggests that social media data, and in particular images, often overlooked opportunities for exploring the demand for forest CES from a qualitative point of view. To unlock the full potential of social media data for forest management, additional studies are needed. In particular, we believe that future research should focus on two objectives. The first objective is to obtain a better understanding of the relationship between *in-situ* and crowdsourced social media data. Is the ratio of activities estimated *via* social media data coherent with the one obtained *via in-situ* surveys? If biases are introduced by analyzing social media images, are they consistent in different study areas? Answering these questions would allow forest managers to confidently use the automated analyses of social media images to gather detailed, and almost real-time information, on the value of the recreational ecosystem services that the forests provide applying the crowdsourced benefit transfer approach adopted in this study. The second objective is the creation and the sharing of databases of images depicting forest recreational activities among researchers. These databases would allow research groups worldwide to train new and more accurate CNN for the classification of images based on the depicted recreational activities. Ultimately, this would allow for the application of the methodology developed in this study to new contexts in which alternative approaches are difficult to apply, such as developing countries and remote areas.

Chapter 6
Conclusions

6.1 Dissertation Objectives

The objective of this dissertation was to tackle the knowledge gaps around the use of crowdsourced social media data in the management of forest CES. Despite a significant number of studies having used social media data to study CES in the past, the understanding around how to use this type of data to meet the needs of ecosystem managers is still relatively rudimentary. This is especially true for forested ecosystems, where social media data have seldom been used to estimate quantitative information such as the number of yearly visitors or the consumer surplus per hectare generated by the forest's recreational function. Therefore, progressively narrowing the study area from all of BCs forested lands to specific provincial parks, allowed me to focus on developing new methodologies to exploit social media data to extract both qualitative and quantitative information on forest CES.

In the three research chapters, I adopted various approaches to characterize, map and value CES in forested ecosystems, using crowdsourced social media data from Flickr. While performing these tasks, I answered to the following research questions.

1) **Can the metadata of crowdsourced social media images be used to deduce trends of forest visitation and as a data source for the application of the travel cost method?**

The results obtained in Chapter 3 suggest that images time-stamps can be used to explore hourly, daily, and seasonal trends of forest recreation, while picture coordinates can be used to identify the hotspots of recreation and the countries of origin of visitors. These techniques have also been applied in the subsequent chapters.

By exploring the relationship between PUDs (Photo-User-Days), GHII (Global Human Index Interference) and the empirically determined number of visitors to forest recreational sites, it is possible to calibrate regression models to estimate the number of visitors in forested areas based on crowdsourced data. Using this approach, forest managers could map and quantify the use of forest recreational ecosystem services. Furthermore, such an approach can be applied across large areas, at fine spatial resolution and in near real time.

Lastly, the crowdsourced travel cost approach offers the possibility to obtain rapid and cheap estimates of the monetary value of forest recreational opportunities. The results obtained are consistent with the economic theory (down-sloping demand curves) and similar to the one obtained with traditional surveys.

Therefore, social media images metadata appear to be a valuable data source in the study of forest CES provision and consumption, especially at large-scales, where conducting *in-situ* surveys it is not a feasible way to proceed.

2) Is it possible to adopt deep-learning techniques to automatically recognize and classify the cultural ecosystem service depicted in social media images?

In Chapter 4, four CNNs were used to automatically classify the gathered social media images, based on the CES depicted in them. These CNNs were a combination of pre-trained freely available models and purposely trained models obtained with transfer learning approach. The overall performances of this process, assessed on an independent test-set, show that CNNs can be successfully used to characterize recreational activities and landscapes that forest visitors photograph.

Despite the limited training images collected and used, the CNN created to distinguish images depicting aesthetic experiences and recreational experiences had an accuracy of over 90%. Therefore, the transfer learning approach is a promising development in the study of CES trough social media data, since it could potentially enable ecosystem managers worldwide to obtain insights on the activities that recreationists carried out in a site, as well as the species and the landscapes that are photographed. This information could help ecosystem managers in tailoring the management of the forest to the needs of the visitors, while having a clear idea of which are the most disturbed plant and animal species.

Lastly, the output of the image classification process can be used as a starting point for the application of the benefit transfer method, providing an estimate of the value of the recreational ES provided by natural recreational areas. These value estimates that can be obtained quickly and cheaply appear to be especially promising for remote recreational sites in which *in-situ* surveys are particularly challenging and costly to carry-out.

3) Can the combination of the outcomes of the image classification process and remote sensing technologies data be integrated to explore the recreational potential of forests?

The analysis carried-out in Chapter 5 supports the hypothesis that the combination of crowdsourced social media data and remote sensing data can be a valuable approach for the management of forest CES. Combining the outcomes of the image classification process with topographic, biometrics, and socioeconomic variables, it was possible to train a random forest model to predict the recreational activity most likely to be carried-out in a given area at a certain time of the year. This model has a good accuracy and the potential of providing useful insights to parks managers. Estimating which activities are most likely to be carried out across a specific forest recreational site during the various seasons could allow the microplanning of forest recreational areas. In fact, having this information, forests and parks managers could make informed decision on the opening of a hiking trail or the creation of a new birdwatching tower.

In addition, by applying the crowdsourced benefit transfer method, it was possible to estimate the yearly value of recreational ES per hectare. Having a monetary value of the CES provided by each hectare of a forest allows for the consideration of CES at the stand level, recognizing their monetary value and therefore taking it into account when planning harvesting or partial park closures.

6.2 Innovations

The research undertaken in this dissertation presents several innovations relevant to the study of forest CES.

- So far, crowdsourced social media data have been rarely used in the study of forest CES, and the existing literature has focused on the use of image metadata to identify recreational trends, hotspots and the variables that causes them. While this type of study is certainly useful to map the provision and the consumption of forest CES at the landscape level, it cannot provide any information on the monetary value of CES. To tackle this challenge, in **Chapter 3**, I applied for the first time the travel cost method from crowdsourced social media data to a forested area. This approach can in the future be applied for the valuation of recreational sites, avoiding costly and time-consuming *is-situ* surveys. Ultimately, this could facilitate the recognition of CES in ecosystem management at the large scale. In many previous studies, researchers highlighted the issue of the difficulties that arise from the estimation of the number of visits in natural ecosystems from social media images counts when dealing with areas characterized by different levels of anthropic impacts. When creating the number of annual visits model, I used for the first time the GHII (Global Human Interference Index) to account for the unevenness of image distribution. The inclusion of GHII variable in the regression model improved its performances.

- In the last decade, deep learning models have seen application in countless fields, including in the automated classification of social media images acquired in natural landscapes. However, until very recently, only pre-trained commercial CNNs (Convolutional Neural Networks) have been used to this end. This has limited the number of applications for two reasons, first because these services are not for free, and second because the labels used in the classification process cannot be freely modified. To tackle these limitations, in **Chapter 4** I used the transfer learning approach to train the CNNs to classify the content of crowdsourced social media images. This approach, a recent development in the deep learning field, has allowed the training and deployment of purposely created CNNs to identify the type of CES depicted in an image, as well as types of recreational activities. The strong performances of the classification process allowed for the design and the application of an innovative approach: the crowdsourced benefit transfer method. This method is an adaptation of the traditional economic technique benefit transfer that could enable parks managers to obtain a detailed knowledge of the recreational activities in which parks' visitors engage and rapidly estimate the consumer surplus that they generate.
- Previous studies have explored several variables that influence the number of recreational visits in a forest. However, all the existing literature has used to this end all the geotagged images that were acquired in an area, regardless to their relevance or their content. In **Chapter 5** I for the first time explored how topographic, socioeconomic and biometrics variables influence the type of recreational activity in which forest visitors engage into. The innovative recreational potential model, created in this chapter, could allow forest managers to get a preliminary understanding of which areas are indicated for which recreational activities, helping in managing recreation. Furthermore, the attempts that have been made in the past to value the CES provided by natural ecosystems starting from crowdsourced social media data provided a unique value for the entire

study area. Chapter 5 builds on the crowdsourced benefit transfer method previously designed and perform the first spatial explicit assessment of the value of forest CES.

- Despite the current increase in studies that adopted crowdsourced social media data in the study of CES, most of the literature have been focusing in quantitative characterization of CES (e.g. seasonal trends of visitation and estimates of the number of annual visits). These types of applications are certainly useful in capturing and mapping CES provision, however their usefulness for practical ecosystem management is limited. In this dissertation, I instead developed tools and methodologies that could contribute in extracting, from crowdsourced social media data, outcomes that could prove to be useful in forest management, such as CES values assessments and maps. Overall, this thesis deepens our understanding on how crowdsourced social media data can be used to map and value CES in forest ecosystems, challenging tasks that have historically limited the inclusion of CES in forest management plans. Lastly, this dissertation shows that the most recent deep learning techniques can be used to fully exploit the potential of crowdsourced social media data.

6.3 Limitations

6.3.1 Big but uneven data

Crowdsourcing all geotagged images acquired in BC and uploaded on Flickr between 2005 and 2020 yielded over 1.3 million images. However over 60% of these images were acquired in the Greater Vancouver Area alone, and only around 22% were acquired in forested areas. The original objective of this dissertation was to map the CES provided by all of BC's forested lands. However, due to the unevenness of image distribution, enough data was available to perform a statistically sound analysis only for the southern portion of BC. This fact is illustrative of one of the main limitations of crowdsourced social media images for the study of forest CES: intensely visited forests and recreational sites tend to be over-represented in social media, while less frequented areas are generally underrepresented. Therefore, the use of crowdsourced social media data is particularly promising for peri-urban forests and popular recreational destinations, while it appears to be less appropriate for the study of remote forested areas. This limitation is further amplified by the methodological choice of excluding images that were acquired during multi-days trips by international visitors. This exclusion could be overcome by a further refinement of the crowdsourced travel cost method that could be based on traditional surveys. These surveys would be instrumental in correctly assign the consumer surpluses generated by multi-days trips.

6.3.2 Crowdsourced social media biases and limits

Crowdsourced social media data have inherent limitations. The first limitation is sampling bias, as the users of social media are not necessarily a representative sample of the general population. Hence, certain demographic tends to be overrepresented (typically the young), while other are overlooked. The second limitation is data availability, as social media are operated by private companies and changes to platforms and policies can happen overnight. Therefore, there is no guarantee that geotagged Flickr images will continue to be available to researchers. Also, the acquisition of images is not equally compatible with each type of recreational activities and each type of cultural ecosystem services. For example, extractive recreational activities such as hunting are underrepresented in social media images; similarly, recreational activities related to water (kayaking, fishing, boating, etc.) are less likely to be photographed due to the difficulties of carrying cameras in those circumstances. Another limitation of crowdsourced social media data is the lack of quality assurance, since the images are freely acquired and shared by social media users, without any consideration of best-practice standard of data acquisition. In this regard, the lack of the accuracy of the geotags is a troublesome aspect, since the spatial accuracy tends to be lower in remote areas.

6.3.3 Ethical concerns

Images and metadata included in the analyses presented were obtained and used within the standard Flickr-user agreements. However, the authors of the images are unaware of their indirect participation in the study, raising the issue of data privacy rights. To address these concerns, in this study, images including personal information (such as faces or structures) were treated as confidential and not used in public presentations. Furthermore, user data was stored in aggregate form and de-identified whenever possible. Despite the abovementioned precautions, several criticalities arise from crowdsourced images, especially when used in the training of convolutional neural network for image classification tasks. For example, it is impractical to obtain the consent by the authors of the images, and a privacy issue occur due to the use of images of people for the training of CNNs. This area is currently almost unregulated, and future laws and regulations could make the replication of the approaches used in this dissertation unfeasible.

6.3.4 Comparison with traditional techniques

Due to cost and time constraints, the values assessed in this dissertation were not compared to values obtained with traditional methods. Therefore, I was not able to perform a validation of the methodologies applied throughout the dissertation. For example, the regression model used to estimate the number of visits in forested areas was not tested against empirically obtained data. Similarly, the monetary values assessments obtained with both the crowdsourced travel cost method and the crowdsourced benefit transfer method were not compared with values obtained with traditional economic techniques. In absence of a systematic validation that explores the differences between the results obtained with traditional and crowdsourced data, it is impossible to know if the biases introduced by the use of the latter significantly alter the outcomes. Therefore, I believe that the methodologies proposed in this dissertation are currently immature for practical applications, and future studies are needed to tackle this limitation.

6.4 Future Directions

6.4.1 Data mashups

One of the most promising approaches to tackle the limitations posed by the unevenness of social media data availability is the integration of data from multiple origins. Combining data from different social media, it could be possible to partially overcome the lack of data in remote areas. To this regard, social media dedicated to specific activities and hobbies appear to be particularly promising. Social media such as eBird (https://ebird.org) could provide valuable information on wildlife viewing activities in remote areas. Similarly, mountain biking and hiking blogs could provide fine-grained information on these recreational activities. Finally, crowdsourcing data from other popular social media such as Instagram, Twitter, and Facebook could help in addressing the biases due to non-representativeness of Flickr users.

6.4.2 Data sharing

A promising horizon for the future of this research field is the creation of a shared database of images depicting CES consumption. Such a database could aid in the creation of more precise tools to automatically classify images and could help overcome issues arising from data becoming suddenly unavailable. Bringing together image databases obtained in other ecosystems and other continents could allow the creation of CNN applicable worldwide, helping in the arduous task of global CES assessment. Lastly, large amount of available data would result in more precise image classification models further increasing the attractiveness of their use in practical applications.

6.4.3 Citizen science projects

To overcome some of the ethical concerns of using images in the study of CES, one possible approach could be the creation of citizen science projects aimed to the collection of images acquired in specific recreational areas. Park visitors could volunteer the images acquired on their phone and cameras during their recreational experiences with researchers. These images would ease ethical concerns and could be freely used to better explore the dynamics of interactions between parks visitors and the landscape using the methodologies described in this dissertation. The process of collection of images from volunteers could be facilitated by the creation of experimental social media platforms created and managed directly by universities and research centres.

6.4.4 Comparison of crowdsourced and traditional approaches

An important unanswered question remains in how the results obtained with crowdsourced social media data compare with the results of traditional approaches. For example, future studies could validate the performances of the model used for estimating the number of visits in forested areas with empirically obtained data. Similar systematic comparisons are also needed for the techniques I used to estimate the monetary value of CES and the performances of the automated classification process. These analyses would enable a better understanding of the biases that crowdsourced social media data introduce and potentially confirm that crowdsourced social media data can indeed be used for the management of the cultural ecosystem services provided by forests.

6.5 Closing Statement

The work presented in this dissertation addressed the mapping and valuing issues that hampers the recognition of CES in forest management plans. I demonstrated how social media data, especially when analyzed with deep learning techniques, represent a valuable data source that can be used to improve our understanding of CES demand. It is my hope that the methodology that I designed and applied in this dissertation could, in the future, help forest managers to cheaply and quickly assess the values of forests' CES, contributing to the recognition and proper representation of their importance.

Bibliography

Abildtrup, J., Garcia, S., Olsen, S. B., and Stenger, A. (2013). Spatial preference heterogeneity in forest recreation. *Ecological economics*, *92*, 67-77.

Agimass, F., Lundhede, T., Panduro, T. E., and Jacobsen, J. B. (2018). The choice of forest site for recreation: A revealed preference analysis using spatial data. *Ecosystem Services*, *31*, 445-454.

Albrecht, A., and Kandji, S. T. (2003). Carbon sequestration in tropical agroforestry systems. *Agriculture, ecosystems and environment*, *99*(1-3), 15-27.

Alemu, J. B., Richards, D. R., Gaw, L. Y.-F., Masoudi, M., Nathan, Y., and Friess, D. A. (2021). Identifying spatial patterns and interactions among multiple ecosystem services in an urban mangrove landscape. *Ecological Indicators*, *121*, 107042.

Baron, J., and Greene, J. (1996). Determinants of insensitivity to quantity in valuation of public goods: Contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied*, *2*(2), 107.

Bell, S., Simpson, M., Tyrväinen, L., Sievänen, T., and Pröbstl, U. (2009). *European forest recreation and tourism: a handbook*. Taylor and Francis.

Bernetti, I., Chirici, G., and Sacchelli, S. (2019). Big data and evaluation of cultural ecosystem services: an analysis based on geotagged photographs from social media in Tuscan forest (Italy). *Iforest-Biogeosciences and Forestry*, *12*, 98-105. https://doi.org/10.3832/ifor2821-011

Bielinis, E., Bielinis, L., Krupińska-Szeluga, S., Łukowski, A., and Takayama, N. (2019). The effects of a short forest recreation program on physiological and psychological relaxation in young polish adults. *Forests*, *10*(1), 34.

Bojic, I., Massaro, E., Belyi, A., Sobolevsky, S., and Ratti, C. (2015). Choosing the right home location definition method for the given dataset. International Conference on Social Informatics, British Columbia Government. (2020). *Household Income Summaries for BC Census Subdivisions - 2016 Census* Retrieved 19-01-2021 from https://catalogue.data.gov.bc.ca/dataset/household-income-summaries-for-bc-census-subdivisions-2016-census

Brockerhoff, E. G., Barbaro, L., Castagneyrol, B., Forrester, D. I., Gardiner, B., Gonzalez-Olabarria, J. R., Jactel, H. (2017). Forest biodiversity, ecosystem functioning and the provision of ecosystem services. *Biodiversity and Conservation*, *26*(13), 3005-3035. https://doi.org/10.1007/s10531-017-1453-2

CAA. (2021). Canadian Automobile Association. Retrieved 2021-2-12 from https://carcosts.caa.ca/

Callau, A. À., Albert, M. Y. P., Rota, J. J., and Giné, D. S. (2019). Landscape characterization using photographs from crowdsourced platforms: content analysis of social media photographs. *Open Geosciences*, *11*(1), 558-571.

Cardoso, A. S., Renna, F., Moreno-Llorca, R., Alcaraz-Segura, D., Tabik, S., Ladle, R. J., and Vaz, A. S. (2022). Classifying the content of social media images to support cultural ecosystem service assessments using deep learning models. *Ecosystem Services*, *54*, 101410.

Carvalho-Ribeiro, S. M., and Lovett, A. (2011). Is an attractive forest also considered well managed? Public preferences for forest cover and stand structure across a rural/urban gradient in northern Portugal. *Forest Policy and Economics*, *13*(1), 46-54.

Champ, P. A., Boyle, K. J., Brown, T. C., and Peterson, L. G. (2003). *A primer on nonmarket valuation* (Vol. 3). Springer.

Chang, W. Y., Lantz, V. A., and MacLean, D. A. (2011). Social Benefits of Controlling Forest Insect Outbreaks: A Contingent Valuation Analysis in Two Canadian Provinces. *Canadian Journal of Agricultural Economics-Revue Canadienne D Agroeconomie*, *59*(3), 383-404. https://doi.org/10.1111/j.1744-7976.2010.01208.x

Chua, A., Servillo, L., Marcheggiani, E., and Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, *57*, 295-310. https://doi.org/10.1016/j.tourman.2016.06.013

Ciesielski, M., and Stereńczak, K. (2021). Using Flickr data and selected environmental characteristics to analyse the temporal and spatial distribution of activities in forest areas. *Forest Policy and Economics*, *129*, 102509.

Cookson, R. (2003). Willingness to pay methods in health care: a sceptical view. *Health economics*, *12*(11), 891-894.

Corp., D. B. (2021). *Industry performance*. Retrieved 2021-3-13 from https://www.destinationbc.ca/research-insights/type/industry-performance/

Costanza, R., dArge, R., deGroot, R., Farber, S., Grasso, M., Hannon, B., . . . vandenBelt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*(6630), 253-260. https://doi.org/10.1038/387253a0

Costanza, R., and Folke, C. (1997). Valuing ecosystem services with efficiency, fairness and sustainability as goals. *Nature's services: Societal dependence on natural ecosystems*, 49-70.

Creel, M. D., and Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American journal of agricultural economics*, *72*(2), 434-441.

Da Rugna, J., Chareyron, G., and Branchet, B. (2012). Tourist behavior analysis through geotagged photographies: a method to identify the country of origin. *13th Ieee International Symposium on Computational Intelligence and Informatics (Cinti 2012)*, 347-351.

Daily, G. C. (1997). Introduction: what are ecosystem services. *Nature's services: Societal dependence on natural ecosystems*, *1*(1).

Davis, R. K. (1963). Recreation planning as an economic problem. *Nat. Resources J.*, *3*, 239.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition,

Depietri, Y., and Orenstein, D. E. (2020). Managing fire risk at the wildland-urban interface requires reconciliation of tradeoffs between regulating and cultural ecosystem services. *Ecosystem Services*, *44*, 101108.

Dixon, J. A., Carpenter, R. A., Fallon, L. A., Sherman, P. B., and Manipomoke, S. (2013). *Economic analysis of the environmental impacts of development projects*. Routledge.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, *78*(383), 605-610.

Fezzi, C., Bateman, I. J., and Ferrini, S. (2014). Using revealed preferences to estimate the value of travel time to recreation sites. *Journal of Environmental Economics and Management*, *67*(1), 58-70.

Filyushkina, A., Agimass, F., Lundhede, T., Strange, N., and Jacobsen, J. B. (2017). Preferences for variation in forest characteristics: Does diversity between stands matter? *Ecological Economics*, *140*, 22-29.

Gao, H., and Mou, N. (2021). Changes in the spatial and temporal characteristics of inbound tourism flows in Tibet based on geotagged photographs. E3S Web of Conferences,

Getis, A., and Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on spatial data analysis* (pp. 127-145). Springer.

Ghermandi, A. (2016). Analysis of intensity and spatial patterns of public use in natural treatment systems using geotagged photos from social media. *Water Research*, *105*, 297-304. https://doi.org/10.1016/j.watres.2016.09.009

Ghermandi, A. (2018). Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands. *Ecosystem Services*, *31*, 351-357. https://doi.org/10.1016/j.ecoser.2017.12.012

Ghermandi, A., Camacho-Valdez, V., and Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism Management*, *77*, 104002.

Ghermandi, A., Depietri, Y., and Sinclair, M. (2022). In the AI of the beholder: A comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green spaces. *Landscape and Urban Planning*, *217*, 104261.

Ghermandi, A., Sheela, A. M., and Justus, J. (2016). Integrating similarity analysis and ecosystem service value transfer: Results from a tropical coastal wetland in India. *Ecosystem Services*, *22*, 73-82. https://doi.org/10.1016/j.ecoser.2016.09.014

Ghermandi, A., and Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. *Global environmental change*, *55*, 36-47.

Gilani, H. R., and Innes, J. L. (2020). The State of British Columbia's Forests: A Global Comparison. *Forests*, *11*(3), 316.

Gosal, A., and Ziv, G. (2020). Landscape aesthetics: Spatial modelling and mapping using social media images and machine learning. *Ecological Indicators*, *117*, 106638.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., and Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, *11*(1), e12343.

Havinga, I., Marcos, D., Bogaart, P. W., Hein, L., and Tuia, D. (2021). Social media and deep learning capture the aesthetic quality of the landscape. *Scientific reports*, *11*(1), 1-11.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., and Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, *6*(3), 85.

Holmes, T. P., Aukema, J., Englin, J., Haight, R. G., Kovacs, K., and Leung, B. (2014). Economic analysis of biological invasions in forests. *In: Handbook of Forest Resource Economics, Earthscan (Routledge): New York. P. 369-386*, 369-386.

Holmes, T. P., Murphy, E. A., and Bell, K. P. (2006). Exotic forest insects and residential property values. *Agricultural and Resource Economics Review*, *35*(1), 155-166.

Howard, J., and Gugger, S. (2020). *Deep Learning for Coders with fastai and PyTorch*. O'Reilly Media.

Howarth, C. (2014). *Where to Go?: Using Social Media to Assess the Spatial Distribution of Recreation on the Great Barrier Reef* Citeseer].

Hościło, A., and Tomaszewska, M. (2014). CORINE Land Cover 2012-4th CLC inventory completed in Poland. *Geoinformation Issues*, *6*(1), 49-58.

Karasov, O., Heremans, S., Külvik, M., Domnich, A., and Chervanyov, I. (2020). On how crowdsourced data and landscape organisation metrics can facilitate the mapping of cultural ecosystem services: An Estonian case study. *Land*, *9*(5), 158.

Keeler, B. L., Wood, S. A., Polasky, S., Kling, C., Filstrup, C. T., and Downing, J. A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, *13*(2), 76-81.

Kemp, M. A., and Maxwell, C. (1993). Exploring a budget context for contingent valuation estimates. In *Contingent valuation: A critical assessment* (pp. 217-269). Emerald Group Publishing Limited.

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kux, S., and Haider, W. (2014). Non-Motorized Outdoor Recreation in British Columbia in 2012: Participation and Economic Contributions. mountainclubs.org.

Lee, H., Seo, B., Koellner, T., and Lautenbach, S. (2019). Mapping cultural ecosystem services 2.0– Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators*, *96*, 505-515.

Lingua, F., Coops, N. C., Lafond, V., Gaston, C., and Griess, V. C. (2022a). Characterizing, mapping and valuing the demand for forest recreation using crowdsourced social media data. *Plos one*, *17*(8), e0272406.

Lingua, F., Coops, N.C., and Griess, V.C. (2022b). Valuing Cultural Ecosystem Services Combining Deep Learning and Benefit Transfer Approach. *Ecosystem Services*.

Loomis, J. B. (1992). The evolution of a more rigorous approach to benefit transfer: benefit function transfer. *Water Resources Research*, *28*(3), 701-705.

Loomis, J. B., and Walsh, R. G. (1997). *Recreation economic decisions; comparing benefits and costs*. Venture Publishing Inc.

Maia, A. T. A., Calcagni, F., Connolly, J. J. T., Anguelovski, I., and Langemeyer, J. (2020). Hidden drivers of social injustice: uncovering unequal cultural ecosystem services behind green gentrification. *Environmental Science and Policy*, *112*, 254-263.

Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., . . . Bater, C. W. (2018). Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots. *Remote Sensing of Environment*, *216*, 697-714.

Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., and Zald, H. S. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote sensing of environment*, *209*, 90-106.

MEA, M. E. A. (2001). *Millennium ecosystem assessment*. Millennium Ecosystem Assessment.

Meldrum, J. R., Champ, P. A., and Bond, C. A. (2013). Heterogeneous nonmarket benefits of managing white pine bluster rust in high-elevation pine forests. *Journal of forest economics*, *19*(1), 61-77.

Nabhan, G. P., and Buchmann, S. L. (1997). Services provided by pollinators. *Nature's Services: societal dependence on natural ecosystems*, 133-150.

Parks BC. (2018). *BC Parks End of Year Reports*. Retrieved 2020-12-17 from https://bcparks.ca/research/

Pastur, G. M., Peri, P. L., Lencinas, M. V., García-Llorente, M., and Martín-López, B. (2016). Spatial patterns of cultural ecosystem services provision in Southern Patagonia. *Landscape ecology*, *31*(2), 383-399.

Payntar, N. D., Hsiao, W.-L., Covey, R. A., and Grauman, K. (2021). Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. *Tourism management*, *82*, 104165.

Pearson, L., Tisdell, C., and Lisle, A. (2002). The impact of Noosa National Park on surrounding property values: An application of the hedonic price method. *Economic Analysis and Policy*, *32*(2), 155-171.

Price, J. I., McCollum, D. W., and Berrens, R. P. (2010). Insect infestation and residential property values: A hedonic analysis of the mountain pine beetle epidemic. *Forest Policy and Economics*, *12*(6), 415-422.

Produit, T., Tuia, D., de Morsier, F., and Golay, F. (2014). Do geographic features impact pictures location shared on the Web? Modeling photographic suitability in the Swiss Alps. EMR@ ICMR,

Richards, D. R., and Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*, *53*, 187-195.

Richards, D. R., and Tunçer, B. (2018). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services*, *31*, 318-325.

Ritter, E., and Dauksta, D. (2013). Human–forest relationships: ancient values in modern perspectives. *Environment, development and sustainability*, *15*(3), 645-662.

Rodriguez, E., Morris, C. S., and Belz, J. E. (2006). A global assessment of the SRTM performance. *Photogrammetric Engineering and Remote Sensing*, *72*(3), 249-260.

Roovers, P., Hermy, M., and Gulinck, H. (2002). Visitor profile, perceptions and expectations in forests from a gradient of increasing urbanisation in central Belgium. *Landscape and Urban Planning*, *59*(3), 129-145.

Ros Candeira, A., Moreno Llorca, R., Alcaraz Segura, D., Bonet García, F. J., and Vaz, A. S. (2020). Social media photo content for Sierra Nevada: a dataset to support the assessment of cultural ecosystem services in protected areas.

Rosenberger, R. S., Bell, L. A., Champ, P. A., and Smith, E. L. (2012). Nonmarket economic values of forest insect pests: An updated literature review. *Gen. Tech. Rep. RMRS-GTR-275WWW. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 46 p.*, *275*.

Rosenberger, R. S., Sneh, Y., Phipps, T. T., and Gurvitch, R. (2005). A spatial analysis of linkages between health care expenditures, physical inactivity, obesity and recreation supply. *Journal of Leisure Research*, *37*(2), 216-235.

Rosenberger, R. S., White, E. M., Kline, J. D., and Cvitanovich, C. (2017). Recreation economic values for estimating outdoor recreation economic benefits from the National Forest System. *Gen. Tech. Rep. PNW-GTR-957. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 33 p.*, *957*.

Runge, C. A., Hausner, V. H., Daigle, R. M., and Monz, C. A. (2020). Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis. *Environmental Research Communications*, *2*(7), 075001.

Schirpke, U., Meisch, C., Marsoner, T., and Tappeiner, U. (2018). Revealing spatial and temporal patterns of outdoor recreation in the European Alps and their surroundings. *Ecosystem services*, *31*, 336-350.

Scott, A. (1965). The valuation of game resources: some theoretical aspects. *Canadian Fisheries Report*, *4*, 27-47.

Seresinhe, C. I., Moat, H. S., and Preis, T. (2018). Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*, *45*(3), 567-582.

Sessions, C., Wood, S. A., Rabotyagov, S., and Fisher, D. M. (2016). Measuring recreational visitation at US National Parks with crowd sourced photographs. *Journal of Environmental Management*, *183*, 703-711. https://doi.org/10.1016/j.jenvman.2016.09.018

Sinclair, M., Ghermandi, A., Moses, S. A., and Joseph, S. (2019). Recreation and environmental quality of tropical wetlands: A social media based spatial analysis. *Tourism Management*, *71*, 179-186.

Sinclair, M., Ghermandi, A., and Sheela, A. M. (2018). A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. *Science of the total environment*, *642*, 356-365.

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

Sinclair, M., Mayer, M., Woltering, M., and Ghermandi, A. (2020). Valuing nature-based recreation using a crowdsourced travel cost method: A comparison to onsite survey data and value transfer. *Ecosystem Services*, *45*, 101165.

Smith, L. N. (2017). Cyclical learning rates for training neural networks. 2017 IEEE winter conference on applications of computer vision (WACV),

Sottini, V. A., Barbierato, E., Bernetti, I., Capecchi, I., Fabbrizzi, S., and Menghini, S. (2019). Winescape perception and big data analysis: An assessment through social media photographs in the Chianti Classico region. *Wine Economics and Policy*, *8*(2), 127-140.

Speak, A., Usher, M., Solly, H., and Zerbe, S. (2021). # urbanforest: cultural ecosystem services of urban trees through the lens of Instagram. *Journal of Place Management and Development*.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.

Svedsäter, H. (2003). Economic valuation of the environment: how citizens make sense of contingent valuation questions. *Land Economics*, *79*(1), 122-135.

Tachikawa, T., Hato, M., Kaku, M., and Iwasaki, A. (2011). Characteristics of ASTER GDEM version 2. 2011 IEEE international geoscience and remote sensing symposium, Tenerelli, P., Puffel, C., and Luque, S. (2017). Spatial assessment of aesthetic services in a complex mountain region: combining visual landscape properties with crowdsourced geographic information. *Landscape Ecology*, *32*(5), 1097-1115. https://doi.org/10.1007/s10980-017-0498-7

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), 1-11.

Terry, J. C. D., Roy, H. E., and August, T. A. (2020). Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods in Ecology and Evolution*, *11*(2), 303-315.

Torrey, L., and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242-264). IGI global.

van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., and Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(46), 12974-12979. https://doi.org/10.1073/pnas.1614158113

Wartmann, F. M., Baer, M. F., Hegetschweiler, K. T., Fischer, C., Hunziker, M., and Purves, R. S. (2021). Assessing the potential of social media for estimating recreational use of urban and peri-urban forests. *Urban Forestry and Urban Greening*, *64*, Article 127261. https://doi.org/10.1016/j.ufug.2021.127261

Weller, P., and Elsasser, P. (2018). Preferences for forest structural attributes in Germany–Evidence from a choice experiment. *Forest policy and economics*, *93*, 1-9.

Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. (2005). *Last of the Wild Project, Version 2, 2005 (LWP-*

2): Global Human Influence Index (HII) Dataset (Geographic) NASA Socioeconomic Data and Applications Center (SEDAC).

Wilkes-Allemann, J., Pütz, M., and Hirschi, C. (2015). Governance of Forest Recreation in Urban Areas: Analysing the role of stakeholders and institutions using the institutional analysis and development framework. *Environmental Policy and Governance*, *25*(2), 139-156.

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), 1-7.

Wood, S. A., Winder, S. G., Lia, E. H., White, E. M., Crowley, C. S. L., and Milnor, A. A. (2020). Next-generation visitation models using social media to estimate recreation on public lands. *Scientific Reports*, *10*(1), Article 15419. https://doi.org/10.1038/s41598-020-70829-x

You, S., Zheng, Q., Chen, B., Xu, Z., Lin, Y., Gan, M., . . . Wang, K. (2022). Identifying the spatiotemporal dynamics of forest ecotourism values with remotely sensed images and social media data: A perspective of public preferences. *Journal of Cleaner Production*, *341*, 130715.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, *40*(6), 1452-1464.

Appendix

Table A1. Pairings between the classes of the Places365 convolutional neural networ and the landscape typologies adopted in this study.

places_365_class	landscape_typology
airfield	anthropic
airplane_cabin	not_relevant
airport_terminal	not_relevant
alcove	not_relevant
alley	not_relevant
amphitheater	anthropic
amusement_arcade	not_relevant
amusement_park	anthropic
apartment_building/outdoor	not_relevant
aquarium	not_relevant
aqueduct	anthropic
arcade	not_relevant
arch	anthropic
archaelogical_excavation	anthropic
archive	not_relevant
arena/hockey	not_relevant
arena/performance	not_relevant
arena/rodeo	not_relevant
army_base	anthropic
art_gallery	not_relevant
art_school	not_relevant
art_studio	not_relevant
artists_loft	not_relevant
assembly_line	not_relevant
athletic_field/outdoor	anthropic
atrium/public	not_relevant
attic	not_relevant
auditorium	not_relevant
auto_factory	not_relevant
auto_showroom	not_relevant
badlands	mountain
bakery/shop	not_relevant
balcony/exterior	not_relevant
balcony/interior	not_relevant
ball_pit	not_relevant
ballroom	not_relevant
bamboo_forest	forest
bank_vault	not_relevant
banquet_hall	not_relevant
bar	not_relevant
barn	anthropic
barndoor	anthropic
baseball_field	not_relevant
basement	not_relevant

basketball_court/indoor	not_relevant
bathroom	not_relevant
bazaar/indoor	not_relevant
bazaar/outdoor	not_relevant
beach	water
beach_house	anthropic
beauty_salon	not_relevant
bedchamber	not_relevant
bedroom	not_relevant
beer_garden	anthropic
beer_hall	not_relevant
berth	not_relevant
biology_laboratory	not_relevant
boardwalk	anthropic
boat_deck	anthropic
boathouse	anthropic
bookstore	not_relevant
booth/indoor	not_relevant
botanical_garden	forest
bow_window/indoor	not_relevant
bowling_alley	not_relevant
boxing_ring	not_relevant
bridge	anthropic
building_facade	not_relevant
bullring	anthropic
burial_chamber	not_relevant
bus_interior	not_relevant
bus_station/indoor	not_relevant
butchers_shop	not_relevant
butte	anthropic
cabin/outdoor	anthropic
cafeteria	not_relevant
campsite	anthropic
campus	anthropic
canal/natural	water
canal/urban	anthropic
candy_store	not_relevant
canyon	mountain
car_interior	not_relevant
carrousel	anthropic
castle	not_relevant
catacomb	not_relevant
cemetery	not_relevant
chalet	anthropic
chemistry_lab	not_relevant
childs_room	not_relevant
church/indoor	not_relevant
church/outdoor	not_relevant
classroom	not_relevant
clean_room	not_relevant
cliff	mountain
closet	not_relevant

clothing_store	not_relevant
coast	mountain
cockpit	not_relevant
coffee_shop	not_relevant
computer_room	not_relevant
conference_center	not_relevant
conference_room	not_relevant
construction_site	anthropic
corn_field	anthropic
corral	anthropic
corridor	not_relevant
cottage	anthropic
courthouse	not_relevant
courtyard	not_relevant
creek	water
crevasse	water
crosswalk	not_relevant
dam	anthropic
delicatessen	not_relevant
department_store	not_relevant
desert/sand	mountain
desert/vegetation	mountain
desert_road	mountain
diner/outdoor	not_relevant
dining_hall	not_relevant
dining_room	not_relevant
discotheque	not_relevant
doorway/outdoor	not_relevant
dorm_room	not_relevant
downtown	not_relevant
dressing_room	not_relevant
driveway	anthropic
drugstore	not_relevant
elevator/door	not_relevant
elevator_lobby	not_relevant
elevator_shaft	not_relevant
embassy	not_relevant
engine_room	not_relevant
entrance_hall	not_relevant
escalator/indoor	not_relevant
excavation	anthropic
fabric_store	not_relevant
farm	anthropic
fastfood_restaurant	not_relevant
field/cultivated	anthropic
field/wild	forest
field_road	anthropic
fire_escape	not_relevant
fire_station	not_relevant
fishpond	water
flea_market/indoor	not_relevant
florist_shop/indoor	not_relevant

food_court	not_relevant
football_field	not_relevant
forest/broadleaf	forest
forest_path	forest
forest_road	forest
formal_garden	anthropic
fountain	anthropic
galley	not_relevant
garage/indoor	not_relevant
garage/outdoor	anthropic
gas_station	not_relevant
gazebo/exterior	anthropic
general_store/indoor	not_relevant
general_store/outdoor	not_relevant
gift_shop	not_relevant
glacier	snow
golf_course	anthropic
greenhouse/indoor	not_relevant
greenhouse/outdoor	anthropic
grotto	mountain
gymnasium/indoor	not_relevant
hangar/indoor	not_relevant
hangar/outdoor	anthropic
harbor	anthropic
hardware_store	not_relevant
hayfield	anthropic
heliport	anthropic
highway	anthropic
home_office	not_relevant
home_theater	not_relevant
hospital	not_relevant
hospital_room	not_relevant
hot_spring	water
hotel/outdoor	not_relevant
hotel_room	not_relevant
house	anthropic
hunting_lodge/outdoor	anthropic
ice_cream_parlor	not_relevant
ice_floe	snow
ice_shelf	snow
ice_skating_rink/indoor	not_relevant
ice_skating_rink/outdoor	anthropic
iceberg	snow
igloo	snow
industrial_area	not_relevant
inn/outdoor	not_relevant
islet	water
jacuzzi/indoor	not_relevant
jail_cell	not_relevant
japanese_garden	anthropic
jewelry_shop	not_relevant
junkyard	anthropic

kasbah	anthropic
kennel/outdoor	anthropic
kindergarden_classroom	not_relevant
kitchen	not_relevant
lagoon	water
lake/natural	water
landfill	anthropic
landing_deck	anthropic
laundromat	not_relevant
lawn	anthropic
lecture_room	not_relevant
legislative_chamber	not_relevant
library/indoor	not_relevant
library/outdoor	anthropic
lighthouse	anthropic
living_room	not_relevant
loading_dock	anthropic
lobby	not_relevant
lock_chamber	anthropic
locker_room	not_relevant
mansion	anthropic
manufactured_home	anthropic
market/indoor	not_relevant
market/outdoor	not_relevant
marsh	water
martial_arts_gym	not_relevant
mausoleum	anthropic
medina	not_relevant
mezzanine	not_relevant
moat/water	water
mosque/outdoor	anthropic
motel	not_relevant
mountain	mountain
mountain_path	mountain
mountain_snowy	snow
movie_theater/indoor	not_relevant
museum/indoor	not_relevant
museum/outdoor	anthropic
music_studio	not_relevant
natural_history_museum	not_relevant
nursery	not_relevant
nursing_home	not_relevant
oast_house	not_relevant
ocean	water
office	not_relevant
office_building	not_relevant
office_cubicles	not_relevant
oilrig	anthropic
operating_room	not_relevant
orchard	anthropic
orchestra_pit	not_relevant
pagoda	anthropic

palace	anthropic
pantry	not_relevant
park	anthropic
parking_garage/indoor	not_relevant
parking_garage/outdoor	not_relevant
parking_lot	not_relevant
pasture	anthropic
patio	anthropic
pavilion	anthropic
pet_shop	not_relevant
pharmacy	not_relevant
phone_booth	not_relevant
physics_laboratory	not_relevant
picnic_area	anthropic
pier	anthropic
pizzeria	not_relevant
playground	anthropic
playroom	not_relevant
plaza	not_relevant
pond	water
porch	not_relevant
promenade	anthropic
pub/indoor	not_relevant
racecourse	anthropic
raceway	anthropic
raft	anthropic
railroad_track	anthropic
rainforest	forest
reception	not_relevant
recreation_room	not_relevant
repair_shop	not_relevant
residential_neighborhood	not_relevant
restaurant	not_relevant
restaurant_kitchen	not_relevant
restaurant_patio	not_relevant
rice_paddy	anthropic
river	water
rock_arch	mountain
roof_garden	anthropic
rope_bridge	anthropic
ruin	anthropic
runway	anthropic
sandbox	anthropic
sauna	not_relevant
schoolhouse	not_relevant
science_museum	not_relevant
server_room	not_relevant
shed	anthropic
shoe_shop	not_relevant
shopfront	not_relevant
shopping_mall/indoor	not_relevant
shower	not_relevant

ski_resort	anthropic
ski_slope	snow
sky	sky
skyscraper	not_relevant
slum	not_relevant
snowfield	snow
soccer_field	not_relevant
stable	not_relevant
stadium/baseball	not_relevant
stadium/football	not_relevant
stadium/soccer	not_relevant
stage/indoor	not_relevant
stage/outdoor	not_relevant
staircase	not_relevant
storage_room	not_relevant
street	not_relevant
subway_station/platform	not_relevant
supermarket	not_relevant
sushi_bar	not_relevant
swamp	water
swimming_hole	water
swimming_pool/indoor	not_relevant
swimming_pool/outdoor	anthropic
synagogue/outdoor	anthropic
television_room	not_relevant
television_studio	not_relevant
temple/asia	anthropic
throne_room	not_relevant
ticket_booth	not_relevant
topiary_garden	anthropic
tower	anthropic
toyshop	not_relevant
train_interior	not_relevant
train_station/platform	not_relevant
tree_farm	forest
tree_house	anthropic
trench	anthropic
tundra	forest
underwater/ocean_deep	water
utility_room	not_relevant
valley	forest
vegetable_garden	anthropic
veterinarians_office	not_relevant
viaduct	anthropic
village	not_relevant
vineyard	anthropic
volcano	mountain
volleyball_court/outdoor	not_relevant
waiting_room	not_relevant
water_park	not_relevant
water_tower	not_relevant
waterfall	water

watering_hole	water
wave	water
wet_bar	not_relevant
wheat_field	anthropic
wind_farm	not_relevant
windmill	not_relevant
yard	anthropic
youth_hostel	not_relevant
zen_garden	anthropic

Table A2. List of the tags used as argument in the queries for training images

CLASS	TAG_1	TAG_2	TAG_3	TAG_4	TAG_5	TAG_6	TAG_7	TAG_8	TAG_9	TAG_10
Aestethic	forest	lake	waterfall	river	mountain	snow	stream	praire	ocean	coast
Skiing	skiing	skis	snowboarding	snowboard	downhill	skislope	chalet	snow	backcountry	resort
Hiking	hiking	walk	backpack	hike	countryside	mountain	mountais	excursion	walk	promenade
Climbing	climbing	rock-climbing	ascension	freeclimbing	bouldering	mountaineering	summit	rock	rope	scaling
Camping	camping	campsite	tent	retreat	campground	campfire	camper	backcountry	wildrness	backpacking
Wildlife viewing	wildlife	fauna	bear	bird	animals	deer	environment	wildrness	wolf	elk
Biking	biking	bike	cycling	mountain biking	mountain bike	trail	enduro	mud	downhill	bmx
Water related	fishing	kayaking	beach	swimming	canoe	rafting	rowing	kayak	canoe	beaching

Table A3 consumer surpluses used in the assessment of the consumer surpluses generated by the recreational ecosystem serivice provided by BC provincial parks

Recreational Activity	Consumer Surplus (CAN$)
Biking	98.9
Camping	40.6
Climbing	81.6
Hiking	96.3
Water related	102.9
Skiing	79.3
Wildlife	68.5

Table A4. Estimated annual visits

Provincial Park	Annual number of visitors
Cypress Park	1519323
E.C. Manning Park	1011314
Mt. Seymour Park	924896
Golden Ears Park	771346
Rathtrevor Beach Park	631541
Goldstream Park	607264
Shannon Falls Park	553545
Porteau Cove Park	544283
Stawamus Chief Park	499985
Macmillan Park	486394
Juan De Fuca Park	373883
Little Qualicum Falls Park	322511
Wells Gray Park	288317
Blackcomb Glacier Park	231845
Brandywine Falls Park	203356
Mt. Robson Park	175551
Joffre Lakes Park	115575
Garibaldi Park	102257
Strathcona Park	92068
Gowlland Tod Park	42698
Mount Assiniboine Park	28423
Lac Du Bois Park	4282