PLACE-BASED CONSERVATION FOR PROTECTED AREA MANAGEMENT:

GEOTAGGED USER-GENERATED CONTENT ANALYSIS

by

INHYE KONG

(Under the Direction of FAUSTO O. SARMIENTO & LAN MU)

ABSTRACT

Place-based conservation is a recent paradigm in protected area management that incorporates human dimensional values into conservation. Defined as landscape values or cultural ecosystem services (CES), these values include aesthetic appreciation, recreational opportunities, social cohesion, spiritual experiences, and place identity. While a lack of relevant data has been a significant challenge to identify humanistic values and incorporate them in landscape conservation, the advent of geotagged user-generated content and big data analytics has revolutionized the data-driven analysis of personal experiences associated with landscapes.

The overarching research objective is to investigate the mechanism of place-based conservation in protected areas using geotagged user-generated content. Throughout four case studies, I aim to disentangle the role of humans and national parks in addressing place experiences and cultural ecosystem services in the United States and Ecuador. The analytic approaches include geospatial and text content analysis using Flickr, TripAdvisor, and Google Maps. In addition, I present a crowdsourced phrasal lexicon as a reference dictionary to define eight CES values from text content.

The findings demonstrate the great potential of geotagged user-generated content to elaborate the interactive role of humans and nature in protected areas. Tangible examination of intangible landscape values and cultural ecosystem services can enable data-driven communication across the general public, managers, and stakeholders in achieving place-based conservation. In addition, national park management can take advantage of knowing site-specific values as tourist niches in order to build competitive tourism destinations. Overall, this work showcases a netnographic approach for digital humanities that may suggest a creative conservation strategy for the future generation.

INDEX WORDS: social media; text analysis; GIScience; tourism geography; cultural ecosystem services; national park management; conservation policy

PLACE-BASED CONSERVATION FOR PROTECTED AREA MANAGEMENT: GEOTAGGED USER-GENERATED CONTENT ANALYSIS

by

INHYE KONG

B.L.A., Seoul National University, 2012

M.L.A., Seoul National University, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

PLACE-BASED CONSERVATION FOR PROTECTED AREA MANAGEMENT:

GEOTAGGED USER-GENERATED CONTENT ANALYSIS

by

INHYE KONG

Major Professors:	Fausto O. Sarmiento
	Lan Mu
Committee:	Sergio Bernardes
	Xiaobai Angela Yao
	Kyle M. Woosnam

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

ACKNOWLEDGEMENTS

This dissertation is indebted to many people. First of all, I am deeply thankful for my co-advisors, Dr. Fausto Sarmiento and Dr. Lan Mu, who have shown me wholehearted support and guidance throughout this journey. Words cannot do justice to describe how grateful I am. My gratitude also goes to my amazing committee members, Dr. Sergio Bernardes and Dr. Xiaobai Angela Yao from Geography and Dr. Kyle Woosnam from Warnell School of Forestry. I feel grateful for their meticulous feedback, warm encouragement, and positive perspectives.

I would also like to thank all of the Geography professors and staff members who have always shown genuine kindness and professionalism. I also thank the generosity of the Graduate School at the University of Georgia, which has allowed me to complete my doctorate work with success from the honorable Presidential Fellowship and Summer Research Grant. I also have to express my deepest gratitude to all of my friends. It is hard to imagine going through all of these years without the time we spent together, which was filled with laughter, joy, and support.

Last but not least, everything I have achieved thus far would not have been possible without the support of my parents, who have always encouraged me to pursue my dream and believed in my own decision.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER 1. INTRODUCTION 1

PARADIGM SHIFTS IN PROTECTED AREA MANAGEMENT 1

PLACE-BASED CONSERVATION 4

GEOTAGGED USER-GENERATED CONTENT 11

OVERARCHING GOALS AND RESEARCH OBJECTIVES 13

REFERENCES 16

CHAPTER 2. USING GEOTAGGED SOCIAL MEDIA TO ANALYZE VISITOR PREFERENCES ASSOCIATED WITH MULTIVARIATE TRIPOGRAPHIC SEGMENTS 23

ABSTRACT 24

INTRODUCTION 24

LITERATURE REVIEW: 26

STUDY SITE AND METHODS 28

RESULTS 32

DISCUSSION 38

CONCLUSION 43

REFERENCES 44

CHAPTER 3. ONLINE REVIEWS AND CROWDSOURCED PHRASAL LEXICON TO IDENTIFY CULTURAL ECOSYSTEM SERVICES: A CASE STUDY OF EL CAJAS NATIONAL PARK, ECUADOR 50
ABSTRACT 51
INTRODUCTION 51
METHODS 55
RESULTS 62
DISCUSSION 76
CONCLUSION 80
ACKNOWLEDGEMENT 81
REFERENCES 81
CHAPTER 4. CROWDSOURCED ANALYSIS OF CULTURAL ECOSYSTEM SERVICES ACROSS THE U.S. NATIONAL PARKS 87
ABSTRACT 88
INTRODUCTION 88
METHODS 90
RESULTS 95
DISCUSSION 101
CONCLUSION 103
REFERENCES 103
CHAPTER 5. ANALYZING DESTINATION COMPETITIVENESS CONSIDERING TOBLER'S FIRST LAW OF GEOGRAPHY: A CROSS-COMPARISONS OF U.S. NATIONAL PARKS 106
ABSTRACT 107
INTRODUCTION 107
LITERATURE REVIEW 111

DATA AND METHODS 114
RESULTS 120
DISCUSSION 127
CONCLUSION 131
REFERENCES 132
CHAPTER 6. CONCLUSION 138
SUMMARY OF THE FINDINGS 138
IMPLICATIONS, UNCERTAINTY, AND FUTURE DIRECTIONS 140
REFERENCES 145
APPENDICES. 148
Appendix A. The part-of-speech (POS) library to retrieve crude noun phrases and verb phrases (Chapter 3 and 4) 148
Appendix B. The summary of collected data (Chapter 4 and 5) 149
Appendix C. The operation of crowdsourced survey to develop a crowdsurced phrasal lexicon (Chapter 3 and 4) 151
Appendix D. The frequency of eight CES values across 48 U.S. National Parks (Chapter 4) 153
Appendix E. Supplementary materials for Chapter 5 155

LIST OF TABLES

Page

Table 1-1 Definition of cultural ecosystem services in MEA (2005) 6

Table 1-2. Typologies of landscape values and cultural ecosystem services 8

Table 2-1. Details of Flickr data based on tripographic subset variables. 33

Table 2-2. Details of the 25 key attractions in the GSMNP 34

Table 3-1. The details of collected online reviews by languages and platforms 63

Table 3-2. The number of viable phrases by CES categories, languages, and phrasal types 64

Table 4-1. The definitions and exemplary phrases for eight CES 93

Table 4-2. The frequency of CES values across 48 U.S. National Parks 96

Table 4-3. Correlation matrix of eight CES values 97

LIST OF FIGURES

Page

Figure 1-1. The relationship among national parks, visitors, and user-generated content 12

Figure 1-2. Flow of the study within the mechanism of place-based conservation..... 14

Figure 2-1. The locations of 26 top attractions in the GSMNP. 34

Figure 3-1. El Cajas National Park in Ecuador as a study site. 56

Figure 3-2. CES frequencies by CES and the phrasal types.. 66

Figure 3-3. CES frequencies by CES and the platforms..... 67

Figure 3-4. Networks of the 30 most frequent phrases in describing CES between English and Spanish reviews 76

Figure 4-1. Four map units from SOM clustering 98

Figure 4-2. Word clouds of frequent phrasal expressions from national parks that represent four SOM map units 99

Figure 4-3. Spatial distribution of national parks associated with four SOM map units and probabilistic ellipses..... 100

Figure 5-1. The study sites of 48 U.S. National Parks..... 115

Figure 5-2. Flow of data processing 117

Figure 5-3. Four patterns to describe the gravity of destinations in association with travel distances..... 121

Figure 5-4. Four patterns to describe the quality of experiences in association with travel distances..... 123

Figure 5-5. Scatter plots of destination gravity. 124

Figure 5-6. Scatter plots of experience quality. 125

Figure 5-7. A matrix map showing the spatial distribution of 48 national parks based on four clusters of destination gravity and four clusters of experience quality 126

CHAPTER 1

INTRODUCTION

PARADIGM SHIFTS IN PROTECTED AREA MANAGEMENT

Protected areas (PAs) are a key component of global landscape conservation strategy (Dudley, 2008; Worboys et al., 2015). The global paradigm of PA management has evolved for the last century, expanding key concerns and managerial scopes (Phillips, 2003; Shafer, 2015). The original concept of PAs emerged in the U.S., to set aside a large wilderness area for the long-term protection of scenic, archaeological, geological, and biological distinctions, as shown in the establishment of Yellowstone National Park in 1872 and Yosemite National Park in 1890. With the increasing threats for nature conservation posed by radical population growth and subsequent urbanization in the early to mid-twentieth century, there has been a growing number of PAs at global scales (Maxwell et al., 2020). Increasing concerns for conservation further heightened the urgency of developing an institutional framework for PAs. Accordingly, the International Union for Conservation of Nature (IUCN) declared the need for "the protection and maintenance of biological diversity, and of the natural and associated cultural resources, managed through legal or other effective means" and addressed four categories of PAs: Strict Nature Reserve and Wilderness Area (Category Ia and Ib), National Park (Category II), Natural Monument (Category III), and Habitat/species Management Area (Category IV) (IUCN, 1994). Key managerial practices in such PAs were prone to regulate, control, and minimize human activities from the standpoint of viewing humans as a negative detriment for the protection of natural integrity (Phillips, 2003; Lawton and Weaver, 2017).

Meanwhile, the maturation of PA management systems has instigated a demand for conservation policy addressing the human dimension of nature conservation (Phillips, 2003; Taylor and Lennon, 2011). In the 1970s, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) launched the Man and Biosphere Programme to address the role of people and communities in the implementation of successful and comprehensive conservation strategies. In addition, UNESCO declared to encompass cultural heritage in the World Heritage Convention of 1972 to ensure the conservation of cultural landscapes as the heritage of humanity (Taylor and Lennon, 2011). However, it was not until the 1990s that global consent was aggregated to incorporate the human component such as the livelihood of the local community, cultural landscape, local knowledge, and intangible heritage into the framework of PAs (Phillips, 2003). After years of discussion, the IUCN framework decided to add two new categories of PAs, Protected Landscape and Seascape (Category V) and Managed Resource Protected Area (Category VI), that acknowledge the dynamic interplay between humans and nature in the creation of cultural landscapes (Phillips, 2003; Shafer, 2015). As a consequence of global paradigm shift in PA management, institutional efforts at country-level also expanded to incorporate natural, cultural, and historic context in conservation planning (Jones, Shipley, and Ul-Hasan, 2017).

With the growing concerns for cultural and humanistic consciousness in conservation policy, a novel conservation paradigm called the framework of ecosystem service was proposed to incorporate nature's benefit into conservation planning (Costanza et al., 1997; Daily, 1997). In their work published in Nature, Costanza et al. (1997) elaborated 17 ecosystem services and calculated their monetary values throughout the world. The framework of ecosystem services was then succeeded by the Millennium Ecosystem Assessment (MEA) of the United Nations Environmental Programme (UNEP) in 2005, detailing four categories of ecosystem services:

provisioning, regulating, supporting, and cultural ecosystem services (MEA, 2005). MEA (2005) also highlighted the significance of the framework of ecosystem services to make conservation relevant to people by acknowledging nature's benefits on human well-being. Overall, the paradigm has brought substantial innovations in conservation policy to facilitate multidisciplinary approaches from a group of biologists, geographers, economists, psychologists, and social scienitsts (Bennett et al., 2017; Cord et al., 2017) In addition, the paradigm sheds light on human-centric benefits of nature into institutional conservation policy (Chan et al., 2006; Egoh et al., 2007).

Since the MEA (2005), several frameworks have been followed to elaborate the mechanism of ecosystem services for practical policy implementation. A framework from the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) is acclaimed to supplement the MEA (2005) by supplementing the typologies of ecosystem services with socio-cultural and relational values that may not have direct use-values but are critical to human well-being (IPBES, 2016). The Common International Classification of Ecosystem Services (CICES) is a framework to elaborate the ecological functions and flows based on a cascade model (Haines-Young and Potschin, 2010). Meanwhile, The Economics of Ecosystems and Biodiversity (TEEB) and Kareiva et al. (2011) proposed a concept of natural capital and a novel conservation program called Payment for Ecosystem Services that involves an monetary reimbursement of ecosystem services to mitigate catastrophic environmental deterioration.

Of significant accomplishments, the framework of ecosystem services has been praised for institutionalizing the conservation of intangible landscape values – cultural ecosystem services (Chan et al., 2012; Daniel et al., 2012; Fish et al., 2016). Although previous PA management and conservation policy has evolved to include a variety of target features such as biological, cultural,

geological, archaeological distinctions, the scopes of conservation have been focused on tangible remnants and landscapes due to the legal and managerial liability. The insufficient data and methods in measuring intangible values have been another major obstacle. As a result, CES values such as cultural heritage and practices, traditional knowledge, spirituality, and place identity have been largely neglected in conventional conservation schemes despite their crucial roles in supporting human well-being and maintaining landscape integrity (Chan et al., 2012; Daniel et al., 2012; Sarmiento and Hitchner, 2017). The importance of the human dimension for the achievement of integrative conservation has been heightened with the advancement of PA management in the last decades (Bennett et al., 2017).

PLACE-BASED CONSERVATION

Definition of place-based conservation

Geographers have long been interested in the study of space and place. Early works of geography had centered on exploring physical environmental settings, such as biogeography, archaeology, and geology, yet the disciplines of geography expanded to explore the role of humans in modifying and interpreting nature in the early 20th century. Carl O. Sauer was a pioneer in human geography and the Berkeley School to investigate cultural entities and cultural landscape, with a focus on the analysis of the humanistic effect on the landscape. With the advance of discourses in human geography, the scope has expanded in the mid-twentieth century to reveal intangible components of a landscape, solidifying the notion of a ‘place’. Yi-Fu Tuan, a prominent scholar in place studies, describes a ‘space’ as a basic physical container, but a ‘place’ is a notion that encompasses human experiences and meanings associated with a space (Tuan, 1977). Edward Relph is another prominent researcher who distinguishes place from space in his writings regarding

a sense of place and place identity, which entail attachment and meaning derived from our mind's interaction and interpretation of physical surroundings (Relph, 1976). As a result, the definition of a place in geography indicates a spatial location that entails collective meanings, memories, and values from the cognitive process of human-nature interactions. In the late 1990s, the views for PA management also made significant progress in recognizing sociocultural traits such as religious, artistic, spiritual, and cultural domains as essential priorities in the framework of PAs (Taylor and Lennon, 2011).

The subsequent emergence of a place-based approach in environmental governance in the late twentieth century, therefore, succeeded the definition of a place in geography to incorporate integrative and holistic management of the environment considering spatial, social, cultural, and biophysical components (Brown and Weber, 2013; Reed et al., 2017). Place-based conservation, therefore, succeeds the place-based approach in conservation by considering both biophysical and sociocultural contexts (William et al., 2013). However, a place-based strategy was not systematically addressed in institutional conservation planning until the 1990s, with the introduction of the ecosystem service framework. The ecosystem services paradigm was praised for demonstrating the direct and indirect advantages that people perceive from nature, making conservation efforts relevant to humans. In addition, the framework proposed to consider intangible benefits from nature, such as aesthetic appreciation, physical contact, familiarity, and attachment, which have been largely ignored in conventional conservation planning but were highlighted by human geographers in the 1970s to constitute a place. Defined as cultural ecosystem services (CES), the intangible benefits from nature range from aesthetic appreciation and recreational opportunities to social cohesion, spiritual experiences, and place identity (MEA, 2005) (Table 1-1).

Table 1-1 Definition of cultural ecosystem services in MEA (2005)

Categories of CES	Definitions
Aesthetic value	Ecosystems provide a rich source of inspiration for art, folklore, national symbols, architecture, and advertising.
Inspirational value	Ecosystems provide a rich source of inspiration for art, folklore, national symbols, architecture, and advertising.
Recreation and ecotourism value	People often choose where to spend their leisure time based in part on the characteristics of the natural or cultivated landscapes in a particular area.
Social relational value	Ecosystems influence the types of social relations that are established in particular cultures. Fishing societies, for example, differ in many respects in their social relations from nomadic herding or agricultural societies.
Spiritual and religious value	Many religions attach spiritual and religious values to ecosystems or their components.
Sense of place	Many people value the "sense of place" that is associated with recognized features of their environment, including aspects of the ecosystem.
Cultural heritage value	Many societies place high value on the maintenance of either historically important landscapes ("cultural landscapes") or culturally significant species.
Cultural diversity	The diversity of ecosystems is one factor influencing the diversity of cultures.
Knowledge system	Ecosystems influence then types of knowledge systems (in both traditional and formal) developed by different cultures.
Educational value	Ecosystems and their components and processes provide the basis for both formal and informal education in many societies.

Since the publication of the framework of ecosystem services by Costanza et al. (1997), Daily (1997), and MEA (2005), the number of studies to examine CES has substantially increased (Cheng et al., 2019; Milcu et al., 2013; Zhang et al., 2020). Most studies followed the classification

of MEA (2005) in classifying and elaborating CES values in multiscalar case studies (Bark et al., 2016; Fish et al., 2016; Hale et al., 2019), but some utilized other terms such as landscape values or social values (Table 1-2). For instance, Van Zanten et al. (2016) and Yoshimura and Hiura (2016) utilized the term landscape values to indicate aesthetic and recreational values only, but Scolozzi et al. (2015) and Chen et al. (2018) referred landscape values to demonstrate the site-specific values that do not fit into the framework of CES in MEA (2005), such as quietness, sense of home, and community attachment. Meanwhile, Brown and Brabyn (2012) defined a framework of landscape values to elaborate both material and non-material values embedded in landscapes, resulting to broaden the typologies of CES from MEA (2005). The following studies of Muñoz et al. (2019) and Brown, Reed, and Raymond, (2020) further enumerated to enlist more variety of material and non-material values as place-based values, rephrasing the cognitive process of provisioning or regularing services as intangible place values such as the appreciation of clean water and air, hunting and fishing. Social value is another terminology that has been widely used in environmental psychology and the framework of Social Values for Ecosystem Services (SolVES) (Sherrouse et al. 2014; Van Riper et al., 2017), yet the classification of social values largely aligns with CES in MEA (2005).

Table 1-2. Typologies of landscape values and cultural ecosystem services

	Frameworks		MEA (2005)			SolVES framework	Landscape values	Place-based values
	MEA (2005)	TEEB (2005)	Milcu et al. (2013)	Bark et al. (2016)	Hale et al. (2019)	Sherrouse et al. (2014)	Brown and Brabyn (2012)	Muñoz et al. (2019)
Aesthetic	Aesthetic	Aesthetic appreciation and inspiration	Aesthetic	Aesthetic	Aesthetic	Aesthetic	Scenic/aesthetic	Scenic landscape
Inspirational	Inspirational		Inspirational	Inspiration	Inspirational	-	-	-
Recreational and ecotourism	Recreational and ecotourism	Recreational and mental and physical wellbeing; Tourism	Recreation and ecotourism	Activity	Recreation	Recreational	Recreation (non-facility-based, facility-based)	Recreation
Therapeutic	-	-	-	-	-	Therapeutic		Therapeutic
Social	Social relations	-	Social relations	Social capital and cohesion	-	-	Social	Social
Spiritual and religious	Spiritual and religious	Spiritual experience and sense of place	Spiritual and religious	Spiritual	Spiritual	Spiritual	-	Spiritual
Senses of place	Senses of place		Sense of place	Identity	Sense of place	-	-	Special place
Cultural heritage	Cultural heritage; Cultural diversity	-	Cultural heritage values; cultural diversity	Place/heritage	cultural heritage values; cultural diversity	Cultural and historic	Historic/cultural	Cultural value
Knowledge system	Knowledge system	-	Knowledge system	Knowledge; Aspiration	knowledge system	-	-	-
Educational	Educational	-	Educational	-	Educational	Learning	-	-
Scientific	-	-	-	-	-	Scientific	-	-
Biological	-	-	-	-	-	Biological diversity	Native wildlife, Native vegetation, Wilderness	Biological diversity; Wilderness
Intrinsic	-	-	Bequest, intrinsic, existence	Existence/bequest; option	Existence	Intrinsic	-	-
Life-sustaining	-	-	-	-	-	Life-sustaining	Ecological/life sustaining	Clean water/air
Future	-	-	-	-	-	Future	-	-
Sustenance	-	-	-	Material	-	Subsistence	-	Gathering; Hunting/fishing
Economic	-	-	-	Employment	-	Economic	Economic	-

Current challenges of place-based conservation: national park context

National parks are one of the most prevalent PAs to promote public recreation and tourism. The Organic Act of 1916 clarifies the dual mandate of national parks to 'protect the wilderness landscape while providing the benefit and enjoyment of the people for this and future generations' (NPS, 2000). The national park systems aim to protect the aesthetic layout of geophysical components such as big rocks, caves, rivers, scenic views, geological features, as well as biological distinctions such as primary forests, engendered species, and fragile habitats.

One of the key challenges in applying place-based conservation in national parks is the lack of sufficient knowledge to understand a variety of CES values. Visitors to national parks seek for awe-inspiring experiences to appreciate natural beauty while enjoying a variety of recreational activities such as hiking, walking, horse riding, and camping, resulting in obtaining aesthetic and recreational values. In addition, national parks enshrine historic remnants and cultural landscapes, playing to promote historic and cultural education (Mcdonnell, 2008). For instance, the Great Smoky Mountains National Park in the U.S. holds rich biological diversity and important habitat for endangered species to be acknowledged as World Heritage Site for their biodiversity. In addition, the park is a home of Native American culture and heritage (i.e., Cherokees) and Scottish and Irish immigrants who resided in the 16th century. During the visit, therefore, visitors experience not only aesthetic and recreational activities but also educational opportunities from rich biodiversity and cultural display in the park (Derek et al., 2019; Johns, Shipley, and Ul-Hasan, 2017). In a contemporary context, the park fosters a cultural distinction called the Appalachian Identity, based on the place attachment and identity from the backpackers and thru-hikers on the Appalachian Trail (Kyle et al., 2004). To summarize, national parks are the living inventories of multiple values that have been assigned as a result of nature-human interactions. Yet, current

management of PAs and national parks remained in tangible features such as the geological features, habitat for endangered species, and historic buildings, resulting to overlook the values people appreciate and assign in association with the landscape.

Another concern is the lack of clarity in understanding the role of individuals and the environmental settings in the formation of CES values. The perception of intangible CES values requires two agents: human and nature. People go through different psychological and cognitive processes as a result of external environmental stimuli. Therefore, the perception of CES values is subject to the environment settings as well as individual characteristics. Considering the role of environmental settings with the perception of CES values, people enjoy aesthetic values from stunning landscapes and biological values from the environmental setting with plentiful animal species. People are also likely to appreciate the importance of culture and history when they visit a region with a rich cultural heritage and historic architecture. Yet, it should be noted that human perceptions of experiences and CES values might differ even under identical environmental settings. In that sense, tourism marketing and management are one of the most prevalent disciplines to address the influence of individual characteristics in having different decisions, experiences, and perceptions (Derek et al., 2019; Li et al., 2008). One of the key segments in classifying the cohorts is the demographic factors, such as age, gender, income, and education level, to demonstrate different travel choices of physical activities, revealed preferences, and cognitive processes (Derek et al., 2019). Trip-related characteristics (i.e., tripographics), such as past visit history, travel distance, and group size, are another factor to categorize the cohorts and their coherent travel behaviors (Li et al., 2008). Furthermore, empirical research has shown that visitors' origin (typically classified as local and non-local, or local, domestic, and international tourists) and nationality may influence diverse behaviors and experiences in tourism destinations,

because such visitor cohorts may have different cultural backgrounds, prior knowledge, and expectations (Muñoz et al. 2020; Sinclair et al. 2020). Nonetheless, there is a scarcity of research to disentangle the role of personal characteristics and environmental factors in the interpretation of landscape values and CES.

GEOTAGGED USER-GENERATED CONTENT

One of the most critical challenges to consider landscape values and CES in conservation strategies has been the lack of data to distinguish the intangible values (Daniel et al., 2012). Indeed, in contrast to tangible ecosystem services such as carbon storage or water supply, CES entail substantial uncertainty to identify and quantify the values (Chan et al., 2012). As a result, previous studies had operated qualitative research methods, such as surveys, open-ended questions, interviews, and workshops (Bieling, 2014; Pascua et al., 2017). While qualitative analysis methods shed light on underrepresented community values associated with the landscape, yet the analysis remained in smaller scopes with a fraction of the population due to the demanding data collection method (Bieling, 2014; Pascua et al., 2017).

With the advent of Web 2.0, geotagged user-generated content (UGC) has been widely applied in PA management (Barros et al. 2021; Teles da Mota and Pickering, 2020). Early studies with UGC focused on validating the UGC from the comparisons to the actual number of visitors in PAs (Sessions et al., 2016; Wood et al., 2013) or across different UGCs (Muñoz et al. 2020; Tenkanen et al. 2017). The following studies then utilized geolocations of such content as a proxy for aesthetic preference or recreation activities to map spatial patterns and hotspots in PAs (Sonter et al., 2016; Tenerelli et al. 2016; van Zanten et al., 2016; Walden-Schreiner et al., 2018). In addition, Havinga et al. (2010) retrieved CES values from eclectic geotagged UGC platforms, such

as aesthetic values from Flickr and recreational values from Strava (i.e., outdoor activity log platform) to analyze CES-specific spatial distributions. Some studies further utilized the metadata such as home locations (Sinclair et al. 2020) and timestamps (Pickering et al. 2020; Schirpke et al. 2018) to distinguish spatial variations of CES.

As big data analytics mature, relevant studies have advanced to content analysis for the interpretation of place experiences and CES values. Content analysis has been a popular netnographic approach, which employs qualitative analytic techniques using computer-mediated web material and has been widely used in tourism research (Tavakoli and Mura, 20118). Chen et al. (2018) utilized Instagram hashtags to analyze six CES in a riverine landscape in Canada. Hale et al. (2019) utilized Flickr photo tags to identify 11 CES in a river ecosystem in Idaho, USA. Ghermandi et al. (2020) also utilized title and photo tags from Flickr to analyze CES in a coastal region of Mexico. Wartmann, Acheson, and Purves (2018) compared the landscape values from the data obtained from UGCs and survey data, and Chesnokova and Purves (2018) analyzed aural experiences using georeferenced image tags from a crowdsourced platform.

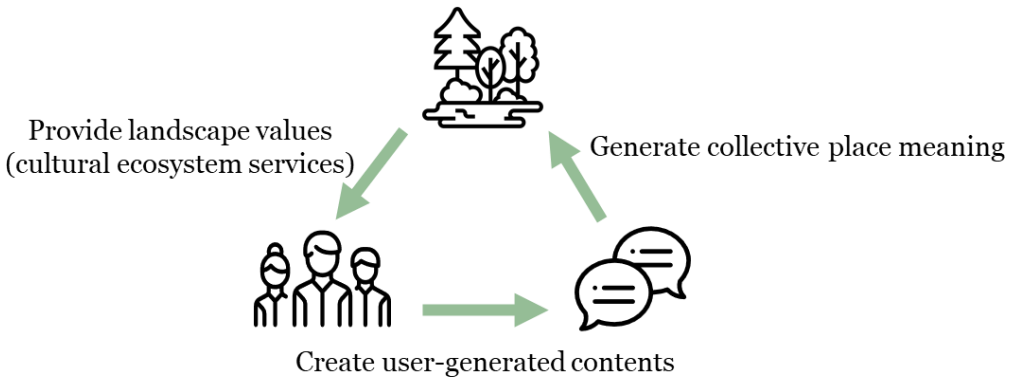


Figure 1-1. The relationship among national parks, visitors, and user-generated content

OVERARCHING GOALS AND RESEARCH OBJECTIVES

The overarching goal of the dissertation is to promote place-based conservation by utilizing geotagged user-generated content to better understand place experiences. As two proxies for place experiences, touristic preference and CES values are applied. In addition, I investigate the role of people (i.e., visitors) and PAs (i.e., national parks) in shaping place experiences and CES values. I define key research objectives as follows:

1) To examine different spatial preferences depending on visitor characteristics in a national park

a) To identify key tourism attractions based on the spatial distribution of geotagged social media content

b) To distinguish and compare key tourism attractions depending on the visitors' travel-related characteristics (i.e., Tripographics)

2) To identify the role of visitor characteristics in perceiving different CES values – single destination analysis

a) To develop a crowdsourced phrasal lexicon – a lexicon consisting of phrasal expressions (instead of single words) annotated with a crowdsourced survey (instead of expert-based annotations) using text-rich online tourism reviews

b) To identify CES values based on text-rich online tourism reviews

c) To compare the frequency of CES values and the frequent phrasal expressions depending on the language as a proxy for visitor characteristics

3) To analyze the role of national parks in providing different CES values – multi-destination analysis

a) To expand the applicability of the lexicon using text-rich online tourism reviews from diverse environmental contexts

b) To apply the lexicon and compare CES values across multiple national parks

4) To evaluate the performance of nature-based destinations in support of place-based conservation

a) To measure the gravity of national parks as a proxy to attract visitors from varying origins – visitor characteristics

b) To measure the place affinity of national parks as a proxy to provide quality tourism experiences – national park characteristics

c) To operate cross-comparisons of destination performance in national parks

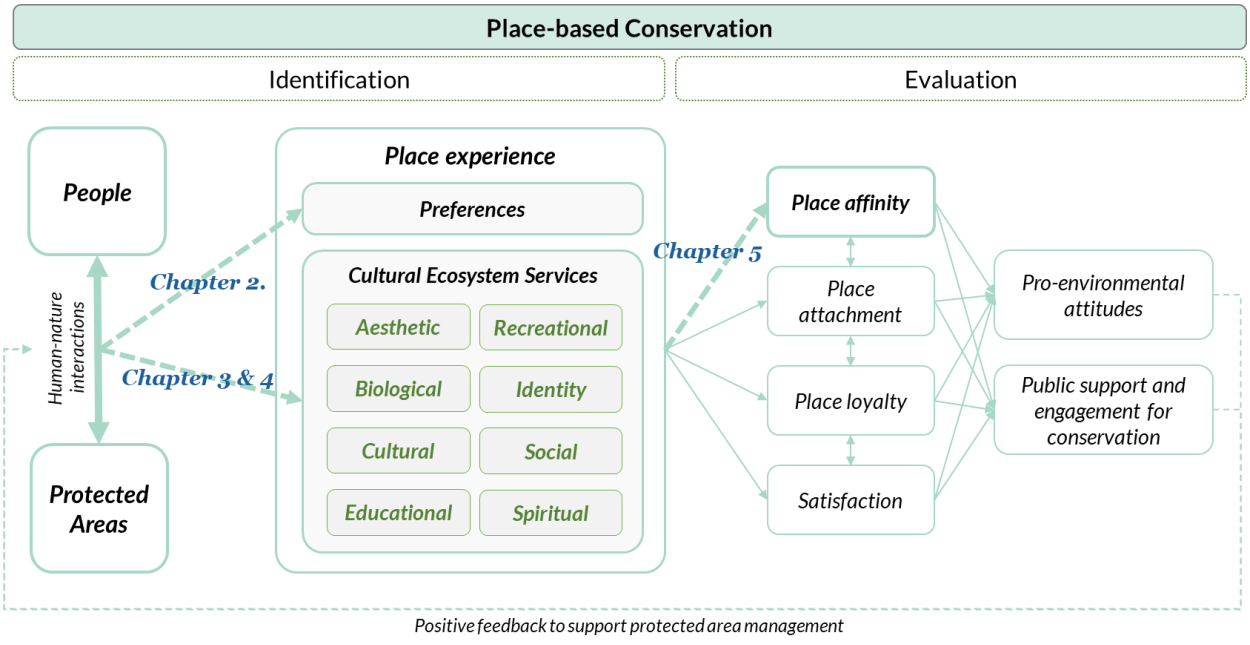


Figure 1-2. Flow of the study within the mechanism of place-based conservation

To be specific, the following four chapters – from Chapter 2 to Chapter 5, exemplify the case studies associated with research objectives. Chapter 2 applies geolocations from Flickr photographs to reveal visitor preferences that indicate aesthetic and recreational values in the Great Smoky Mountains National Park in the United States. I also distinguish preferences for attractions depending on visitors' travel-related (i.e., tripographic) characteristics retrieved from the metadata.

Chapters 3 and 4 delve into text analysis to identify CES values in support of place-based conservation. The case study presented in Chapter 3 applies two online tourism reviews (i.e., TripAdvisor and Google Maps) to identify CES values in El Cajas National Park, Ecuador, which treasures spiritual and religious heritage that has been shrined in the high mountain community, in addition to the exceptional biodiversity and scenic landscape (Sarmiento, 2016). I first develop a crowdsourced phrasal lexicon as a reference dictionary to match relevant CES values to the phrasal expresssions. Then the lexicon is applied to count CES values from the online tourim reviews. Lastly, I examine linguistic distinctions between English and Spanish content as a proxy for visitor characteristics from cultural background.

Chapter 4 further expands text analysis to operate cross-comparisons of CES values across multiple National Parks in the United States. Using TripAdvisor reviews from 48 U.S. National Parks, the crowdsourced phrasal lexicon is expanded for the applicability. After counting the frequency of CES values, national parks are grouped depending on their prevalent CES values. In addition, spatial distribution and frequent phrasal expressions of national parks are plotted by the cluster groups.

Chapter 5 focuses on measuring the destination competitiveness across U.S. National Parks using TripAdvisor. The competitive national park is capable of providing quality tourism

experiences while exerting strong gravity to attract distant visitors. The gravity of national park is measured with the number of visitors depending on the travel distance, which is measured with the distance between self-claimed home locations and destinations. The place affinity of national park is measured with textual concreteness from psycholinguistics, which indicates psychological closeness after quality experiences as external stimuli. As a result, national parks with strong gravity and place affinity are featured as competitive tourism destinations.

As a concluding remark, Chapter 6 summarizes key findings from case studies and discusses policy and scientific implications, limitations, and further directions.

REFERENCES

Adams, V. M., Pressey, R. L., & Naidoo, R. (2010). Opportunity costs: Who really pays for conservation? *Biological Conservation*, *143*(2), 439–448.

Bark, R. H., Robinson, C. J., & Flessa, K. W. (2016). Tracking cultural ecosystem services: Water chasing the Colorado River restoration pulse flow. *Ecological Economics*, *127*, 165–172.

Barros, C., Gutiérrez, J., & García-Palomares, J. (2021). Geotagged data from social media in visitor monitoring of protected areas; a scoping review. *Current Issues in Tourism*, *0*(0), 1–17.

Bennett, N. J., Roth, R., Klain, S. C., Chan, K., Christie, P., Clark, D. A., … Wyborn, C. (2017). Conservation social science: Understanding and integrating human dimensions to improve conservation. *Biological Conservation*, *205*, 93–108.

Bieling, C. (2014). Cultural ecosystem services as revealed through short stories from residents of the Swabian Alb (Germany). *Ecosystem Services*, *8*, 207–215.

Brown, G., & Brabyn, L. (2012). An analysis of the relationships between multiple values and physical landscapes at a regional scale using public participation GIS and landscape character classification. *Landscape and Urban Planning*, *107*(3), 317–331.

Brown, G., & Weber, D. (2013). A place-based approach to conservation management using public participation GIS (PPGIS). *Journal of Environmental Planning and Management*, *56*(4), 455–473.

Brown, G., Reed, P., & Raymond, C. M. (2020). Mapping place values: 10 lessons from two decades of public participation GIS empirical research. *Applied Geography*, *116*, 102156.

Buta, N., Holland, S. M., & Kaplanidou, K. (2014). Local communities and protected areas: The mediating role of place attachment for pro-environmental civic engagement. *Journal of Outdoor Recreation and Tourism*, *5–6*, 1–10.

Chan, K. M. A., Satterfield, T., & Goldstein, J. (2012). Rethinking ecosystem services to better address and navigate cultural values. *Ecological Economics*, *74*, 8–18.

Chan, K. M. A., Shaw, M. R., Cameron, D. R., Underwood, E. C., & Daily, G. C. (2006). Conservation planning for ecosystem services. *PLoS Biology*, *4*(11), 2138–2152.

Chen, Y., Parkins, J. R., & Sherren, K. (2018). Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. *Landscape and Urban Planning*, *170*(December 2016), 283–292.

Cheng, X., Van Damme, S., Li, L., & Uyttenhove, P. (2019). Evaluation of cultural ecosystem services: A review of methods. *Ecosystem Services*, *37*, 100925.

Chesnokova, O., & Purves, R. S. (2018). From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text. *Applied Geography*, *93*(February), 103–111.

Cord, A. F., Brauman, K. A., Chaplin-Kramer, R., Huth, A., Ziv, G., & Seppelt, R. (2017). Priorities to Advance Monitoring of Ecosystem Services Using Earth Observation. *Trends in Ecology and Evolution*, *32*(6), 416–428.

Costanza, R., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., … Van Den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*(6630), 253–260.

Daily, G.C. (Ed.), 1997. Nature's Services: Societal Dependence on Natural Ecosystems. Island Press, Washington DC.

Daniel, T. C., Muhar, A., Arnberger, A., Aznar, O., Boyd, J. W., Chan, K. M. A., … Von Der Dunk, A. (2012). Contributions of cultural services to the ecosystem services agenda. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 8812–8819.

Derek, M., Woźniak, E., & Kulczyk, S. (2019). Clustering nature-based tourists by activity. Social, economic and spatial dimensions. *Tourism Management*, *75*, 509–521.

Dudley, N. (2008). Guidelines for Applying Protected Area Management Categories. Gland, Switzerland: IUCN.

Egoh, B., Rouget, M., Reyers, B., Knight, A., Cowling, R., van Jaarsveld, A., Welz, A. (2007). Integrating ecosystem services into conservation assessments: a review. *Ecological Economics*, *63*, 714–721.

Fish, R., Church, A., Willis, C., Winter, M., Tratalos, J. A., Haines-Young, R., & Potschin, M. (2016). Making space for cultural ecosystem services: Insights from a study of the UK nature improvement initiative. *Ecosystem Services*, *21*, 329–343.

Ghermandi, A., Camacho-Valdez, V., & Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism Management*, *77*(September 2019), 104002.

Haines-Young, R., & Potschin, M. (2010). The links between biodiversity, ecosystem services and well-being. In D. Raffaelli, & C. Frid (Eds.). Ecosystem ecology, a new synthesis. BES Ecological Reviews SeriesCambridge, UK: Cambridge University Press

Hale, R. L., Cook, E. M., & Beltrán, B. J. (2019). Cultural ecosystem services provided by rivers across diverse social-ecological landscapes: A social media analysis. *Ecological Indicators*, *107*(July), 105580.

Havinga, I., Bogaart, P. W., Hein, L., & Tuia, D. (2020). Defining and spatially modelling cultural ecosystem services using crowdsourced data. *Ecosystem Services*, *43*(March), 101091.

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). (2016). *The methodological assessment report on scenarios and models of biodiversity and ecosystem services. ISBN: 978-92-807-3569-7*.

International Union for the Conservation of Nature (IUCN). (1994). Guidelines for Protected Area Categories. IUCN, World Conservation Monitoring Center, Gland, Cambridge, UK.

Jones, C., Shipley, N., & Ul-Hasan, S. (2017). Bringing parks back to the people: Revisiting the dual mandate and core values of the National Park Service. *The George Wright Forum*, *34*(1), 45–52.

Kaveira, P., Tallis, H., Ricketts, T. H., Gretchen, G. C., and Polasky, S. (eds) (2011). Natural Capital: Theory and Practice of Mapping Ecosystem Services.

Kim, J. H., Ritchie, J. R. B., & McCormick, B. (2012). Development of a scale to measure memorable tourism experiences. *Journal of Travel Research*, *51*(1), 12–25.

Kyle, G., Graefe, A., Manning, R., & Bacon, J. (2004). Effects of place attachment on users' perceptions of social and environmental conditions in a natural setting. *Journal of Environmental Psychology*, *24*(2), 213–225.

Lessa, T., Jepson, P., Bragagnolo, C., Campos-Silva, J. V., Barros, E., Gomes, F., … Ladle, R. J. (2021). Revealing the hidden value of protected areas. *Land Use Policy*, (February), 105733.

Li, X. (Robert), Cheng, C. K., Kim, H., & Petrick, J. F. (2008). A systematic comparison of first-time and repeat visitors via a two-phase online survey. *Tourism Management*, *29*(2), 278–293.

Maxwell, S. L., Cazalis, V., Dudley, N., Hoffmann, M., Rodrigues, A. S. L., Stolton, S., … Watson, J. E. M. (2020). Area-based conservation in the twenty-first century. *Nature*, *586*(7828), 217–227.

Mcdonnell, J. A. (2008). Reassessing the National Park Service and the National Park System. *The George Wright Forum*, *25*(2), 6–14.

Milcu, A. I., Hanspach, J., Abson, D., & Fischer, J. (2013). Cultural ecosystem services: A literature review and prospects for future research. *Ecology and Society*, *18*(3), 44–88.

Millennium Ecosystem Assessment (MEA). (2005). Millennium Ecosystem Assessment Synthesis Report.

Moore, S. A., Rodger, K., & Taplin, R. (2015). Moving beyond visitor satisfaction to loyalty in nature-based tourism: a review and research agenda. *Current Issues in Tourism*, *18*(7), 667–683.

Muñoz, L., Hausner, V. H., Runge, C., Brown, G., & Daigle, R. (2020). Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. *People and Nature*, *2*(2), 437–449.

Muñoz, L., Hausner, V., Brown, G., Runge, C., & Fauchald, P. (2019). Identifying spatial overlap in the values of locals, domestic- and international tourists to protected areas. *Tourism Management*, *71*(July 2017), 259–271.

National Park Service (NPS). (2000). America's national park system: The critical documents. Retrieved from https://www.nps.gov/parkhistory/online_books/anps/anps_1i.htm

Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D. R., … Shaw, M. R. (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment*, *7*(1), 4–11.

Pascua, P., McMillen, H., Ticktin, T., Vaughan, M., & Winter, K. B. (2017). Beyond services: A process and framework to incorporate cultural, genealogical, place-based, and indigenous relationships in ecosystem service assessments. *Ecosystem Services*, *26*, 465–475.

Phillips, A. (2003). Turning Ideas on Their Head. *The George Wright Forum*, *20*(Article 2), 8–32.

Pickering, C., Walden-Schreiner, C., Barros, A., & Rossi, S. D. (2020). Using social media images and text to examine how tourists view and value the highest mountain in Australia. *Journal of Outdoor Recreation and Tourism*, *29*(September 2019).

Pine, J., & Gilmore, J. H. (1998). Welcome to the experience economy. *Harvard Business Review*, *76*(4), 97–105.

Reed, M. S., Allen, K., Attlee, A., Dougill, A. J., Evans, K. L., Kenter, J. O., … Whittingham, M. J. (2017). A place-based approach to payments for ecosystem services. *Global Environmental Change*, *43*, 92–106.

Relph, E. (1976). Place and placelessness. London: Pion Limited.

Sarmiento, F. O. & Hitchner, S. (eds). (2017). Indigeneity and the Sacred: Indigenous Revival and the Conservation of Sacred Natural Sites in the Americas, Berghahn Books, New York.

Sarmiento, F.O. (2016). Neotropical Mountains Beyond Water Supply: Environmental Services as a Trifecta of Sustainable Mountain Development, in: Developments in Earth Surface Processes. pp. 309–324.

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

Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. M. (2016). Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management*, *183*, 703–711.

Shafer, C. L. (2015). Cautionary thoughts on IUCN protected area management categories V-VI. *Global Ecology and Conservation*, *3*, 331–348.

Sharma, P., & Nayak, J. K. (2018). Testing the role of tourists' emotional experiences in predicting destination image, satisfaction, and behavioral intentions: A case of wellness tourism. *Tourism Management Perspectives*, *28*, 41–52.

Sherrouse, B. C., Semmens, D. J., & Clement, J. M. (2014). An application of Social Values for Ecosystem Services (SolVES) to three national forests in Colorado and Wyoming. *Ecological Indicators*, *36*, 68–79.

Sinclair, M., Mayer, M., Woltering, M., & Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management*, *263*(October 2019), 110418.

Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS ONE*, *11*(9), 1–16.

Tavakoli, R., & Mura, P. (2018). Netnography in tourism – Beyond Web 2.0. *Annals of Tourism Research*, *73*, 190–192.

Taylor, K., & Lennon, J. (2011). Cultural landscapes: A bridge between culture and nature? *International Journal of Heritage Studies*, *17*(6), 537–554.

Teles da Mota, V., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism*, *30*(April), 100295.

Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, *64*, 237–248.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & 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.

The Economics of Ecosystems and Biodiversity (TEEB). (2010). The Economics of Ecosystems and Biodiversity: ecological and economic foundation. Earthscan, Cambridge.

Tuan, Y-F. (1974). Topophilia. In: a study of environmental perception, attitudes and values. Prentice-Hall, Englewood Cliffs.

Tuan, Y-F. (1977) Space and place: The perspective of experience U of Minnesota Press.

Vada, S., Prentice, C., & Hsiao, A. (2019). The influence of tourism experience and well-being on place attachment. *Journal of Retailing and Consumer Services*, *47*(December 2018), 322–330.

van Riper, C. J., Kyle, G. T., Sherrouse, B. C., Bagstad, K. J., & Sutton, S. G. (2017). Toward an integrated understanding of perceived biodiversity values and environmental conditions in a national park. *Ecological Indicators*, *72*, 278–287.

Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & 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.

Walden-Schreiner, C., Rossi, S. D., Barros, A., Pickering, C., & Leung, Y. F. (2018). Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. *Ambio*, *47*(7), 781–793.

Wartmann, F. M., Acheson, E., & Purves, R. S. (2018). Describing and comparing landscapes using tags, texts, and free lists: an interdisciplinary approach. *International Journal of Geographical Information Science*, *32*(8), 1572–1592.

Weaver, D. B., & Lawton, L. J. (2017). A new visitation paradigm for protected areas. *Tourism Management*, *60*, 140–146.

Williams, D., Steward, W. P., & Kruger, L. E. (2013). The emergence of place-based conservation. In Place-based conservation: perspectives from the social sciences, pp. 1–17.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*(1), 2976.

Worboys, G. L., Lockwood, M., Kothari, A., Feary, S., & Pulsford, I. (eds). (2015). Protected areas governance and management.

Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services*, *24*, 68–78.

Zhang, H., Huang, R., Zhang, Y., & Buhalis, D. (2020). Cultural ecosystem services evaluation using geolocated social media data: a review. *Tourism Geographies*, *0*(0), 1–23.

CHAPTER 2

USING GEOTAGGED SOCIAL MEDIA TO ANALYZE VISITOR PREFERENCES ASSOCIATED WITH MULTIVARIATE TRIPOGRAPHIC SEGMENTS[1]

[1] Kong, I., Woosnam, K.M., Sarmiento, F.O., & Yao, X.A. Submitted to *Journal of Destination Marketing & Management*, August 28, 2021.

ABSTRACT

Distinguishing visitation patterns depending on visitor segments is crucial for destination management, but it has been hampered by a lack of relevant visitor monitoring data, particularly in large nature-based tourism destinations. Taking the Great Smoky Mountains National Park in the United States as a case study, we utilize Flickr geolocations and metadata for the last ten years (2010 – 2019) to compare attraction preferences by ten visitor segments from three tripographic variables: the origin of visitors, visitation frequency, and seasonality. The findings reveal similar preference patterns for attractions for one-time, domestic, and fall visitors exhibit towards easily accessible popular attractions. Meanwhile, in-state and frequent visitors avoid such attractions to visit trail-based attractions distant from the main roads. Attraction popularity in Winter showed substantial disparity to other tripographic cohorts due to the restrictions for accessibility. Overall, the findings confirm the usefulness of geotagged social media with a tripographic-based approach to provide data-driven and site-specific evidence for visitor monitoring and destination management.

INTRODUCTION

National parks are popular nature-based tourism destinations that offer public recreation opportunities within protected area systems (Leung et al., 2018; Worboys et al., 2015). Visitors to national parks seek iconic natural attractions, such as forests, mountain peaks, large lakes, waterfalls, and scenic overlooks, along with cultural and historical heritage (Leask, 2016; Moyle et al., 2017; Schirpke et al., 2013). Diverse recreational opportunities including hiking, camping, and wildlife watching also draw huge visitor interests (Derek et al., 2019; Hausmann et al., 2018; Mancini et al., 2018). Yet, visitors in a single national park may have a variety of tourism interests.

Therefore, the demand for comprehensive visitor monitoring to examine the visitor characteristics and their corresponding tourism interests and behaviors has been increasing (Papadogiannaki et al., 2008; Pettebone & Meldrum, 2018; Weaver & Lawton, 2007). While conventional visitor monitoring focused on classifying visitor profiles based on their demogrphic characteristics, such as age, gender, income, and education level, a recent paradigm in tourism margeting and management called *persona segmentation* reflects the growing attention to define visitor segments based on revealed personal choices and behaviors in (Kang, 2020). In this regard, it is noteworthy to revisit tripographics-based visitor segmentation that classifies visitor cohorts based on travel-related characteristics (Hu & Morrison 2001). Previous studies have proved the successful segmentation of visitor interests and behaviors depending on tripographic visitor segments such as previous visits (Li et al., 2008), destination use patterns (Hu & Morrison, 2002), travel distances (i.e., the origin of visitors) (Liu et al., 2018), and tourism interests for attractions and activities (Derek et al., 2019; Fieger et al., 2019).

With the advent of Web 2.0, geotagged social media (GSM) has been emerging to supplement visitor monitoring in tourism destinations (Barros et al. 2021; Hausmann et al., 2018; Heikinheimo et al., 2017; Teles da Mota & Pickering, 2020). Based on the early studies to confirm statistical significance between the actual number of visitors and the size of the geotagged social media (Sessions et al., 2016; Tenkanen et al. 2017; Wood et al., 2013), the following studies have utilized geolocations in GSM to map spatial visitation patterns and hotspots (Sonter et al., 2016; Tenerelli et al. 2016; van Zanten et al., 2016; Walden-Schreiner et al., 2018). The metadata from GSM have also allowed to distinguish visitor segments such as visitor origins (Muñoz et al. 2020; Sinclair et al. 2020) and seasonality (Pickering et al. 2020; Schirpke et al. 2018). However, fews

studies have conducted cross-examinations of attraction preferences across multivariate tripographic segments.

This study aims to examine the role of GSM in supplementing visitor monitoring through the lens of a tripographic-based visitor segmentation and conduct cross-comparisons of tourism attractions depending on the segments. The Great Smoky Mountains National Park (GSMNP), the most visited national park in the United States (NPS, 2020), was chosen as the study site.

LITERATURE REVIEW:

TRIPOGRAPHIC VARIABLES IN NATURE-BASED TOURISM

Tripographics indicate travel-related characteristics, such as travel distances (i.e., the origin of visitors) and travel motives (Hu & Morrison, 2001; Li et al., 2008), that have received considerable attention to distinguish visitor cohorts depending on their revealed visitor choices and behaviors for destination marketing and management.

The origin of visitors has been widely discussed as a key tripographic variable to indicate travel distance, cultural distance, and emotional proximity, which resulted in different travel motives, travel choices, and visitation patterns (Liu et al., 2018; Nyaupane & Graefe, 2008). Buta et al. (2014) observed stronger place attachment of local visitors than non-local visitors in a Romanian protected area, and Mutanga et al. (2017) distinguished different travel motives between local and non-local visitors to find that local visitors were drawn to biodiversity and educational activities, while foreign visitors appreciated visual enjoyment in national parks, Zimbabwe. From the meta-analysis, Liu et al. (2018) confirmed that the origin of visitors had a positive correlation with cultural distances, which influenced the international destination choices. Supak et al. (2015) showed spatial disparities of tourism patterns of local, regional, and national visitors in federal

recreational destinations in the US. In addition, Using geotagged social media, Sinclair et al. (2020) identified visitation patterns and preferred attractions among local, domestic, and international visitors using Flickr in a German national park. Muñoz et al. (2020) also applied Flickr data to demonstrate different tourism interests depending on the origin of visitors to determine that domestic visitors preferred mountain tops, while international visitors preferred to visit trails in Norwegian national parks.

Another tripographic variable includes the previous visits, since first-time and repeat visitors have been claimed to exhibit different tourism interests and travel patterns (Hu & Morrison, 2001; Lau & McKercher, 2004). Li et al. (2008) confirmed that repeat visitors enjoy recreation, relaxation, and familiarity at the destination, while first-time visitors seek novel experiences. Using GPS trajectories of tourists in Hong Kong, McKercher et al. (2012) found that first-time visitors travel numerous attractions across dispersed areas, but repeat visitors remained in fewer locations. Using geotagged social media, Domènech et al. (2020) utilized Flickr geolocations to compare travel trajectories between first-time and repeat visitors and found different preferences of attraction types in a historic Spanish city. However, relevant studies examining the impact of previous visits on visitation patterns remain few in nature-based tourism destinations.

Lastly, the temporal dimensions of tripographics, namely the time of the visitation such as hour, month, or weekday-weekend, can also indicate different visitor interests (Lee & Tsou, 2018; Sinclair et al., 2020). Seasonality, in particular, has been regarded as a useful variable for visitor segmentation, as it may be used to categorize visitors based on their interest in seasonal attractions (Pickering et al., 2020). Previously studies have applied GPS devices or GPS-assisted apps to distinguish seasonal visitation patterns (Kim et al., 2018), but recent studies have been taking

advantage of timestamps and geolocations in GSM. Schirpke et al. (2018) utilized Flicker to compare seasonal visitation patterns and hotspots by outdoor recreationists in European Alps. Mancini et al. (2018) also applied Flickr to analyze visitation patterns of bird watchers throughout different seasons and years in Scotland. Walden-Schreiner et al. (2018) overlaid geotagged Flickr photos on the tourism facilities to contrast visitation patterns between summer and winter in two mountain protected areas in Argentina and Australia. Furthermore, Gosal et al. (2019) applied text analysis to classify visitor motives based on Flickr photos and compared their seasonal visitation patterns in a coastal area, France. Pickering et al. (2020) also compared visitation patterns and tourism interests between warm and cold seasons at the highest mountain in Australia. Despite the recent advances in GSM-assisted studies, few studies have identified multivariate visitor segments to investigate differences in their visitation patterns.

STUDY SITE AND METHODS

Study site

Located within the states of Tennessee and North Carolina in the United States, the Great Smoky Mountains National Park (GSMNP) is the largest forested national park east of the Mississippi River. The GSMNP is designated as a UNESCO Biosphere Reserve and a World Heritage Site for biological diversity and cultural heritage. Adjacent to the GSMNP, several gateway towns provide entertainment and accommodations for park visitors.

Visitor monitoring in the GSMNP involves counting the monthly visitors from the Visitor Use Statistics program and the occasional visitor surveys to gather in-depth visitor information such as age, hometowns, and satisfaction level, operated by the Comprehensive Surveys of the American Public (CSAP) and Visitor Studies Projects (VSP) (Pettebone & Meldrum, 2018;

Roemer & Vaske, 2014). However, CSAP and VSP have irregular iteration cycles due to the financial and personnel constraints associated with operating large-scale in-person or mail-based surveys (Pettebone & Meldrum, 2018). For instance, the latest VSP report for the GSMNP was published in 2009 (Papadogiannaki et al., 2009). Meanwhile, the GSMNP has been ranked as the most visited national park in the US with 12.5 million annual visitors in 2019 (NPS, 2020). Besides, the number of visitors has been steadily increasing (NPS, 2020), which may result in exacerbating over-tourism issues, such as traffic congestion, facility maintenance, and ecological damage, compounded with diverging visitor interests (Hadwen et al., 2007; Pickering et al., 2018).

Data collection

We collected data from Flickr, one of the most popular photo-sharing geotagged social media to be utilized in previous studies (Barros et al., 2021; Gosal et al., 2019; Mancini et al., 2018; Pickering et al., 2020; Sonter et al., 2016; Teles da Mota & Pickering, 2020; Walden-Schreiner et al., 2018). Data collection was operated in the R environment using Flickr application programming interfaces (API) (Flickr, 2018; R Core Team, 2018). We retrieved photo IDs containing the keyword 'smoky' in either photo tags, title, or description within a buffered boundary of the GSMNP. Then, the corresponding metadata such as timestamps, user IDs, and the self-claimed hometowns, were retrieved to the pertinent photo IDs. Using the timestamps, we chose the photos taken between January 1st, 2010, and December 31st, 2019 (10 years), considering a long-term visitor trend before the pandemic. To avoid multiple photos taken by same users on a same date, 2 km^2 hexagon grids were generated and overlayed on the spatial boundary of GSMNP to apply photo-user-days (PUDs) (Barros et al., 2021). Aside from Flickr data collection, we obtained GIS shapefiles indicating popular attractions and facilities managed by the National Park

Service (NPS) (NPS IRMA, 2019). Overall, Flickr geolocations and GIS shapefiles were mapped in ArcMap 10.7, using the North American Datum 1983 UTM Zone 17 N projection.

Defining tripographic variables

Using Flickr metadata, we defined three tripographic variables – the origin of visitors, visitation frequency, and seasonality. First, we filtered the self-claimed hometowns by user IDs to define the origin of visitors. Since the data was given as place names (e.g., 'Atlanta, GA'), we ran the R package 'ggmap' (Kahle et al., 2019) to retrieve geolocations. The geolocations of hometowns were overlaid to the U.S. state and country boundaries in ArcMap 10.7. As a result, user IDs with self-claimed hometowns were coded as in-state visitors (when their home geolocations were found within the state borders of Tennessee and North Carolina), domestic visitors (others in the U.S.), and international visitors (outside of the U.S.). Then the labels for the origin of visitors were assigned to the photo IDs corresponding to the user IDs.

Visitation frequency as a second tripographic variable was defined with Flickr timestamps. To count the monthly visits in a 10-year time frame, the timestamps of photo IDs were aggregated by user IDs to tally monthly visits. Consequently, visitation frequency in a 10-year time frame was divided into one time, two or three times, and four times and more, and labeled as one-time, a couple-times, and frequent visitors. Then the labels of user IDs were assigned to the photo IDs accordingly.

Lastly, seasonality was chosen as a third tripographic variable to represent different visitor interests in the GSMNP. We directly labeled four seasons to photo IDs based on the month of the timestamps. Photo IDs were therefore coded as spring (i.e., March, April, May), summer (i.e., June, July, August), fall (i.e., September, October, November), and winter (i.e., December,

January, February), according to the four seasons in the temperate climate zone of the Northern Hemisphere (Kim et al., 2018).

Identifying key attractions

To highlight the points of tourism interests, we identified key attractions among numerous amenities and facilities in the GSMNP. Using the entire geolocations of photo IDs, we generated point clusters using a hierarchical density-based clustering algorithm (HDBSCAN), which can detect both line-based clusters and point clusters (Campello et al., 2013). After optimizing input values considering the size of data as well as the scale of the study site, we set the minimum number of points to generate a cluster as fifty. Then we overlaid the point clusters to the GIS shapefiles from the NPS Geodatabase to define referential geolocations and place names.

Additionally, popularity rankings of key attractions were determined with the kernel density estimation (Silverman, 1986). Using the geolocations of the entire photo ID, we calculated the kernel density values which represent the spatial density of point data. After extracting and averaging kernel density values at the point clusters of key attractions, we determined the popularity rankings based on the kernel density values. The overall results from point clusters, kernel density estimation, and popularity rankings were mapped in ArcMap 10.7.

Visitor preferences by tripographic variables

For the comparisons of key attractions depending on tripographic segments, we generated kernel density maps using the geolocations pertinent to each tripographic subset variable and extracted the values at the key attraction locations. Then the kernel density values for each subset variable were rescaled between 0 and 1 using min-max normalization. Based on the results, two

statistical analyses were applied. First, the Pearson chi-square test was applied to examine significant liking or dislikes in particular key attractions depending on the tripographic subset variables. We filtered the residuals from the tests and created a bubble matrix for visual comparisons. The second analysis was Spearman's rank correlation test to examine the similarity of popularity rankings across the tripographic subset variables. The correlations between a pair of variables were then presented in a matrix for visualization. The analyses were conducted with the 'Hmisc' package (Harrell et al., 2021) and visualized with the 'corrplot' package (Wei & Simko, 2021) in R.

RESULTS

Description of collected data

We obtained 4,269 geotagged photo IDs in the GSMNP. The number of photo IDs, user IDs, and the average number of photo IDs per user ID based on the tripographic subset variables were described in Table 2-1. More than half of the photos were uploaded by domestic visitors (65.0 %), followed by in-state (27.0 %) and international visitors (8.0 %). Photos taken by visitors who visited the GSMNP more than three times in the last ten years consisted of 38.5%, yet the number was comparable to the photos taken by the visitors who visited the GSMNP once (36.0%). The most popular season to share photos was the fall with the greatest number of photo IDs (36.4 %), and winter was the least popular season (11.0%).

Table 2-1. Details of Flickr data based on tripographic subset variables.

Tripographic variables	Subsets	Number of photo IDs (%)
Origin of visitors	In-state visitors	1,153 (27.0%)
	Domestic visitors	2,775 (65.0%)
	International visitors	341 (8.0%)
Visitation frequency (in a 10-year frame)	Once	1,537 (36.0%)
	Couple times	1,087 (25.5%)
	Frequent (>3)	1,645 (38.5%)
Seasonality	Spring	1,249 (29.3%)
	Summer	996 (23.3%)
	Fall	1,556 (36.4%)
	Winter	468 (11.0%)

Key attractions in the GSMNP

The kernel density map and the point clusters from the HDBSCAN showed a spatial distribution of Flickr photos in the GSMNP (Figure 2-1). As a result, we identified 26 key attractions. Details of the key attractions such as the popularity rankings and attraction types were described in Table 2-2. The most popular attractions were Forney Ridge Overlook at Clingmans Dome and Clingmans Dome Overlook, which are all renowned iconic attractions in the GSMNP. In addition, we found that numerous popular attractions, including the most popular ones, were located along the major roads including the US 441 Highway.

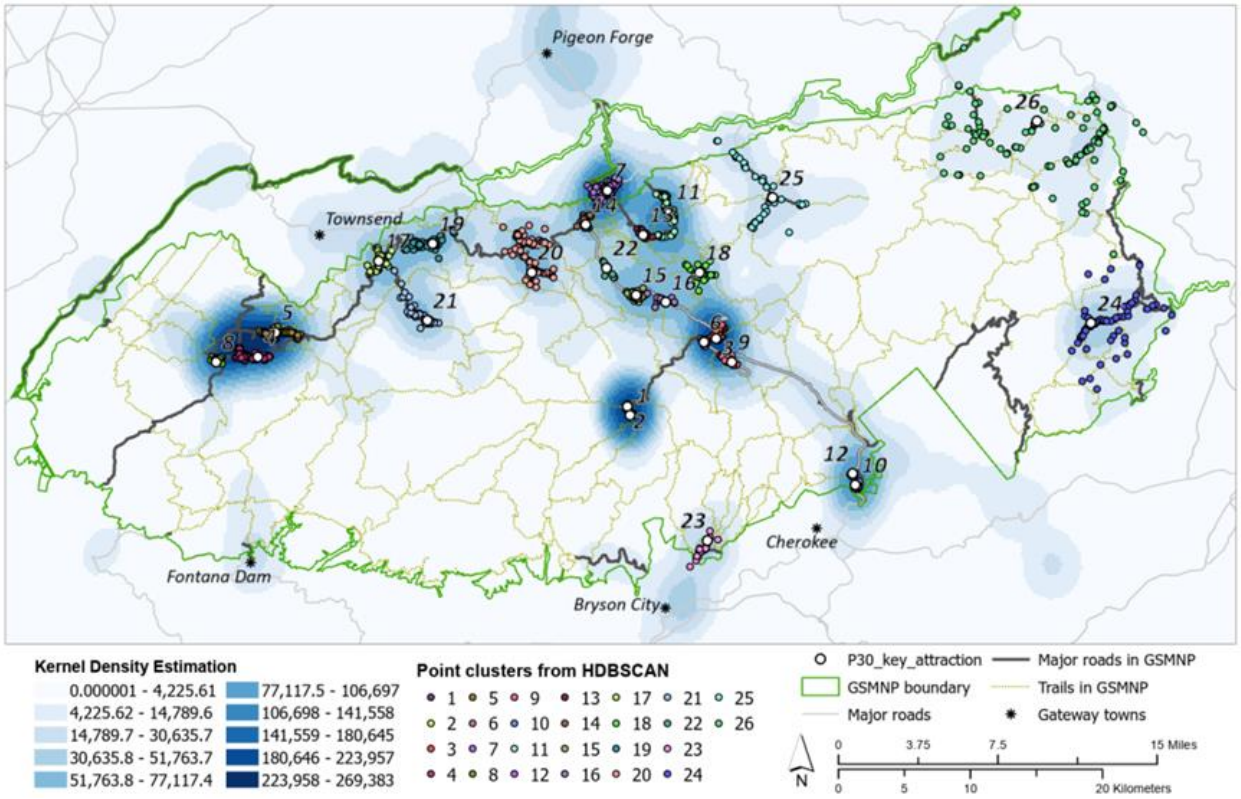


Figure 2-1. The locations of 26 top attractions in the GSMNP. They were overlayed on the kernel density map and the point clusters from the HDBSCAN analysis. The numeric labels next to the key attractions denoted the popularity rankings.

Table 2-2. Details of the 26 key attractions in the GSMNP

Popularity rank	Attraction name	Normalized kernel density values
1	Clingmans Dome Parking Area, Forney Ridge Overlook	1.00
2	Clingmans Dome Overlook	0.98
3	Rockefeller Overlook (Newfound Gap) on Newfound Gap Road	0.93
4	Cades Cove South	0.93
5	Cades Cove Entrance	0.91
6	Oconaluftee (Luftee) Overlook on Newfound Gap Road	0.79
7	Gatlinburg	0.79
8	Cades Cove Visitor Center	0.73

9	Short Beech Ridge Overlook/Swinging Bridges Overlook on Newfound Gap Road	0.57
10	Oconaluftee Visitor Center	0.51
11	Roaring Fork Motor Nature Trail	0.51
12	Mingus Mill/Mingus Trail	0.50
13	Cherokee Orchard Road / Noah Cabin	0.37
14	Sugarland Visitor Center	0.36
15	Chimney Tops Overlook on Newfound Gap Road North (Cove Hardwood Nature Trail)	0.34
16	Chimney Tops Trailhead on Newfound Gap Road (Road Prong Overlook)	0.32
17	Little River Gorge Road - West	0.29
18	Mt LeConte, Overlook	0.28
19	Little River Gorge Road - East, to Meigs Creek Trail	0.28
20	Elkmont Cabin/Campground	0.27
21	Upper Tremont Road to Middle Prong Trail	0.23
22	Carlos Campbell (North) Overlook on Newfound Gap Road North	0.20
23	Deep Creek Falls, Indian Creek Trails	0.15
24	Cataloochee Horse Trail, Cataloochee Valley Overlook	0.14
25	Greenbrier Road, Porters Creek Trail	0.08
26	Mount Cammerer Trail/Cosby Campground/Big Creek Trail	0.05

Statistical analysis of key attractions based on tripographic variables

The residuals from the Pearson chi-square test explained whether the attraction is popular (positive values) or not (negative values) across tripographic subset variables (Figure 2-3). Clingmans Dome Overlook was chosen as the most popular destination by most tripographic subset variables. Yet, in-state, couple-times, and winter visitors showed negative residuals for Clingmans Dome area.

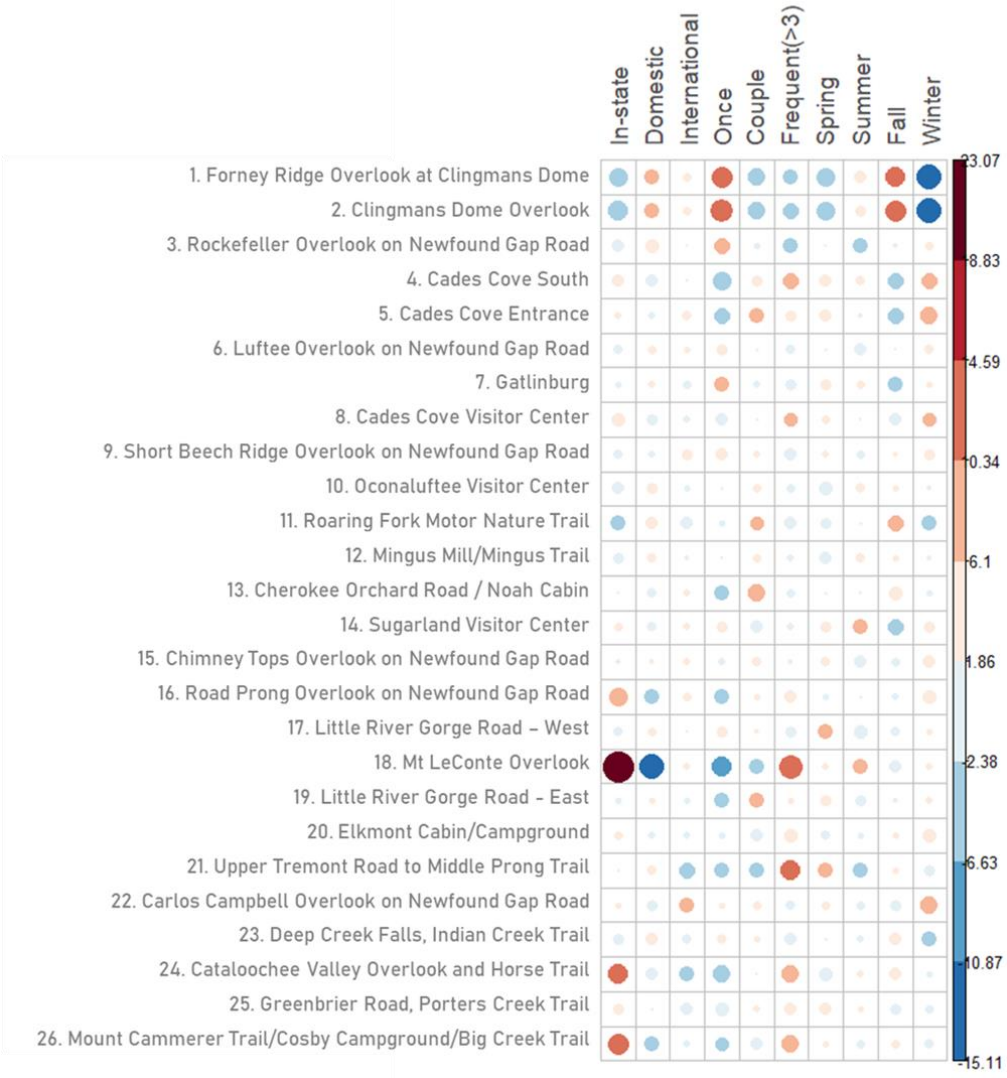


Figure 2-3. A bubble matrix showed the residuals from Pearson's chi-square test. The colored value scale indicated the positive (red) and negative (blue) residuals.

Popularity rankings of key attractions

The results of the Spearman's rho rank correlation test summarized the similarity in popularity ranking across tripographic subsets (Figure 2-4). The correlations for the popularity rankings across the tripographic subset groups were high (>0.90), indicating similar preference for touristic attractions in the GSMNP. Among the origin of visitors, domestic and international

visitors exhibited similar preferences ($p = 0.93^{***}$), but touristic interest between in-state visitors and domestic visitors less similar ($p = 0.78^{***}$). Among the visitation frequency subsets, one-time visitors showed similar interests for attractions to those of a couple-time visitors ($p = 0.90^{***}$), but less similar to those of frequent visitors ($p = 0.83^{***}$). Among the seasonality subsets, all seasons but winter showed high correlations ($p > 0.90$), indicating that people have significantly different tourist interests when visiting in winter.

When comparing the correlations across different tripographic variables, strong correlations were found between in-state and frequent visitors ($p = 0.96^{***}$), domestic and one-time visitors ($p = 0.98^{***}$). In addition, fall visitors were highly correlated with domestic ($p = 0.98^{***}$) and one-time visitors ($p = 0.95^{***}$). However, winter visitors showed weak correlation with the domestic ($p = 0.54^{***}$) and one-time visitors ($p = 0.54^{**}$), explaining their different tourism attraction preferences.

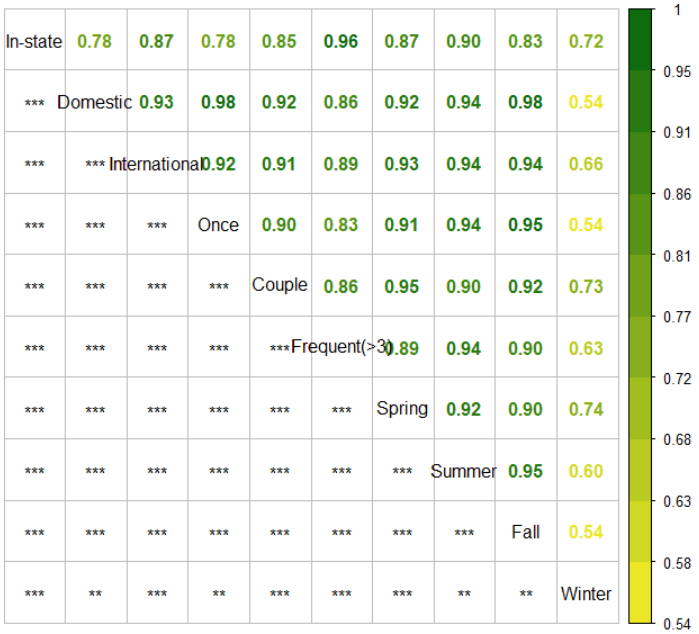


Figure 2-4. A matrix showed the results from Spearman's rank correlation across the tripographic subset variables. The numeric values in the upper matrix showed the correlations between popularity rankings while the asterisks in the bottom matrix indicated the p-values for statistical significance (*p < 0.05; ** p < 0.01; *** p < 0.001).

DISCUSSION

Utilizing geotagged social media to assist visitor monitoring

This study explored the applicability of GSM to assist comprehensive visitor monitoring with visitor segmentation and visitation pattern analysis. Geotagged social media can provide useful data to fill the managerial knowledge gap in data-lacking tourism destinations like large national parks, by bringing the substantial size of ready-to-use content generated by a large population (Richards and Friess, 2015; Sessions et al., 2016; Sonter et al., 2016).

Despite the benefits, the practical application of GSM needs to consider the innate limitations of social media content. Although the geolocations from social media have been widely accepted as a proxy for actual visitation or preference in previous studies (Gosal et al., 2019; Hausmann et al., 2018; Mancini et al., 2018; Pickering et al., 2020; Sonter et al., 2016; Tenerelli et al. 2016; Walden-Schreiner et al., 2018), the discussion for potential bias depending on the particular objectives of social media platforms is still deficient. For instance, geolocations from the photo-sharing platforms may result in the overrepresentation of scenic locations with easy accessibility. From a study to compare Flickr and Public Participatory GIS (PPGIS), Muñoz et al. (2020) noted that the spatial hotspots from Flickr were clustered near the motorized roads while PPGIS results highlighted mountaintops and trails as hotspots. In addition, Moyle et al. (2017) found that when people were given cameras to take pictures of their favorite attractions, trails were more popular than scenic highways. Likewise, the most popular attraction in this study was Clingmans Dome Overlook which is located near major roads, but the VSP report in the GSMNP rated Clingmans Dome Overlook as the fifth visited attraction (Papadogiannaki et al., 2009). Yet, tourist preferences for particular attractions can be highly influenced by the site characteristics, the

niche of destinations, as well as individual visitors, so further analysis is required to disentangle the influence on the result depending on the social media platforms.

Another potential shortcoming is the user bias in social media platforms. Despite the social media platforms have been acclaimed to have a large user population (Heikinheimo et al., 2017), Flickr is known for the user bias towards residents in North America and Europe (van Zanten et al., 2016; Yang et al., 2017). Such bias, therefore, can result in favor of specific visitor subsets or geographic locations (van Zanten et al., 2016). However, the influence of user bias has not been resolved yet. Previous studies using geotagged social media in North America, including this study, have found that local or short-haul visitors make the most frequent visits to the nearby parks (Sonter et al., 2016). Nonetheless, a similar study with Flickr in a German national park showed that domestic visitors were the most frequent visitors followed by local and international visitors (Sinclair et al., 2020). As a result, the application of geotagged social media analysis for destination management requires solid validation using secondary data such as in-person surveys (Heikinheimo et al., 2017), semi-structured online surveys or PPGIS (Muñoz et al., 2020), and multiple social media platforms (Bubalo et al. 2019; Tenkanen et al., 2017).

The influence of tripographics on visitor preference for attractions

In this study, we examined tourism attractions across ten subset variables across three tripographic variables. While the findings examined the similarity and dissimilarity of tourism preferences, it was also noteworthy to define visitor cohorts more than a binary classification such as first-time and repeat visitors or local and non-local visitors. Accordingly, the diversified visitor segments enriched the analysis, resulting in a variety of visitation patterns with statistical significance.

The results showed that one-time, domestic, and international visitors favored iconic and popular natural attractions located along main roads, but in-state and frequent visitors shunned such attractions. The findings aligned with previous studies to distinguish tourism interests between first-time and repeat visitors (Li et al., 2008; Mckercher et al., 2012), exemplified as repeat visitors enjoying recreation, relaxation, and familiarity at the destination, whereas first-time visitors seeking novel experiences. Furthermore, the findings in this study also confirmed the negative correlations between the visitation frequency and travel distances (i.e., the origin of visitors) from previous studies (Nyaupane & Graefe, 2008).

Nonetheless, it should be noted that visitation patterns can be highly dependent on the site characteristics as well as tripographic variables. In this study, frequent visitors exhibited the most dispersed visitation patterns while one-time visitors preferred easily accessible attractions along the major roads. In Hong Kong, McKercher et al. (2012) discovered that first-time visitors had broader and more dispersed visitation patterns, while repeat visitors checked in at fewer attractions. Therefore, further study is needed to explain the coherent relationship between previous visits and visitation patterns.

Concerning seasonality, the highest numbers of photo contents in summer and fall corresponded with the two peak seasons of the GSMNP from the NPS visitor statistics (NPS, 2020). Also, key attractions of summer and fall visitors corresponded with those of first-time, repeat, and domestic visitors, implying major visitor segments during the peak seasons in the GSMNP. Spatial analysis for summer and fall showed the most dispersed patterns with the greatest number of hotspots, indicating the wide-ranging visitor interests during the peak seasons. Meanwhile, winter had the fewest and the least dispersed hotspots in spatial analysis. Such findings supported the previous studies to show broader and diverse visitor interests during favorable

seasons in the nature-based tourism destinations (Gosal et al., 2019; Kim et al., 2018; Lee & Tsou, 2018; Mancini et al., 2018; Pickering et al., 2020). Mountain landscapes, in particular, showed a significant seasonal variance in visitation, as severe weather conditions can limit tourist mobility due to restricted access to hazardous places (Schirpke et al., 2018; Walden-Schreiner et al., 2018).

Lastly, the findings confirmed tourist bias towards easily accessible roadside attractions. In the GSMNP, trail-based natural attractions outnumbered roadside natural attractions, but the roadside natural attractions dominated the highest-ranked attractions. Since the role of accessibility has been discussed in regards to the visitor experience and destination evaluation (Tverijonaite et al., 2018), further analysis is required to examine the quality of experiences at the tourism destinations.

Managerial implications

Utilizing geotagged social media with a multivariate tripographic approach has managerial implications for park rangers, planners, volunteers, and policymakers in nature-based tourism destinations. First, geotagged social media can assist visitor segmentation, which is critical in destination marketing and management to identify key customers and the niche of the destination (Nyaupane & Graefe, 2008). Utilizing several metadata in GSM, we presented a multivariate visitor segmentation based on tripographic characteristics, which can explain homogeneous visitor cohorts that have similarities in revealed behavior, attraction choices, and tourism interest (Hu & Morrison, 2001). Though, classifying visitor segments from the metadata of social media can be versatile, which can be re-classified or modified depending on the aim of study or the destination characteristics.

Second, spatial analysis associated with the geolocations of social media enables to fill the managerial gap in large natural destinations. Park managers can identify visitation hotspots and assign proper resources or regulations to reduce negative pressure on attractions and facilities. As a result, the visitor pressure on attractions can be controlled within their carrying capacity, which will improve visitor satisfaction by avoiding negative experiences like traffic jams or poorly managed attractions and recreational facilities (Hadwen et al., 2007). Also, longitudinal spatial analytics (Mancini et al., 2018) can assist to develop long-term management plans to preserve the integrity of destinations.

Still, the locations of where the pictures were taken may have significant disparity of where the attractive features are. In this study, we found several scenic overlooks along the major roads as key popular attractions in the GSMNP, yet the spatial locations of such overlooks does not necessarily indicate the locations of beautiful features but the locations to observe the beautiful sceneries. To fill such gaps, several studies have implemented content analysis to analyze the dominant features appearing in geotagged photos (Richards and Friess, 2015; Richards and Tunçer, 2018). Furthermore, Yoshimura and Hiura (2017) and Van Berkel et al. (2018) suggested a viewshed analysis of geotagged photos based on a probabilistic estimation or the overlay with high-resolution topography data (i.e., LiDAR) to retrieve the directions of photos and distinguish the key attractions. Therefore, analyzing where people were gazing when they are taking pictures can provide in-depth understanding of key attractions in tourism destinations.

Overall, GSM-assisted visitor monitoring can support visitor-oriented management, which aims to engage visitors in destination management (Weaver & Lawton, 2017). While traditional visitor management largely involved visitor regulations and restrictions to safeguard the integrity of the destinations, visitor-oriented management aims to provide quality experiences in pursuit of

bringing mutual benefits to both visitors and destinations (Mason, 2005; Pettebone & Meldrum, 2018; Weaver & Lawton, 2017). The shifting paradigm is supported from the previous studies that providing quality experiences can maximize visitor satisfaction and foster strong place attachment, affinity, and loyalty to the destination (Arnberger et al., 2012; Kyle et al., 2004; Moore et al., 2015). Consequently, strong positive experiences and psychological closeness to the destination will lead to strong public support and engagement for the destination management (Buta et al., 2014; Weaver & Lawton, 2017). Therefore, comprehensive visitor monitoring from GSM-assisted analytics can offer key insights to tailor tourism experiences and enhance visitor satisfaction, which will bring substantial public support to buttress destination management. The implementation of visitor-oriented management is expected to benefit nature-based tourism destinations the most, as they confront budgetary constraints and a shortage of employees to run administrative operations across large territories.

CONCLUSION

The study highlighted the applicability of GSM to facilitate data-driven and site-specific visitor management in a large-size national park. Using the metadata and geolocations from Flickr, we presented managerial insights to implement multivariate visitor segmentation and visitation pattern analysis. The findings confirmed similar preference patterns for attractions for one-time, domestic, and fall visitors exhibit towards easily accessible popular attractions. Meanwhile, in-state and frequent visitors avoid such attractions to visit trail-based attractions distant from the main roads. Despite the innate limitations of GSM, the outcomes of this study demonstrated the feasible and efficient role of implementing GSM to assist visitor-oriented management, in pursuit of mutual benefits between visitors and destinations.

REFERENCES

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, *27(2)*, 93–115.

Arnberger, A., Eder, R., Allex, B., Sterl, P., & Burns, R. C. (2012). Relationships between national-park affinity and attitudes towards protected area management of visitors to the Gesaeuse National Park, Austria. *Forest Policy and Economics*, *19*, 48–55.

Barros, C., Gutiérrez, J., & García-Palomares, J. (2021). Geotagged data from social media in visitor monitoring of protected areas; a scoping review. *Current Issues in Tourism*, 1–17.

Bubalo, M., van Zanten, B. T., & Verburg, P. H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landscape and Urban Planning*, *184*(January), 101–111.

Buta, N., Holland, S. M., & Kaplanidou, K. (2014). Local communities and protected areas: The mediating role of place attachment for pro-environmental civic engagement. *Journal of Outdoor Recreation and Tourism*, *5–6*, 1–10.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7819 LNAI*(PART 2), 160–172.

Derek, M., Woźniak, E., & Kulczyk, S. (2019). Clustering nature-based tourists by activity. Social, economic and spatial dimensions. *Tourism Management*, *75*, 509–521.

Domènech, A., Mohino, I., & Moya-Gómez, B. (2020). Using Flickr Geotagged Photos to Estimate Visitor Trajectories in World Heritage Cities. *ISPRS International Journal of Geo-Information*, *9*(11), 646.

Eagles, P. F. J. (2014). Research priorities in park tourism. *Journal of Sustainable Tourism*, *22*(4), 528–549.

Fieger, P., Prayag, G., & Bruwer, J. (2019). 'Pull' motivation: an activity-based typology of international visitors to New Zealand. *Current Issues in Tourism*, *22*(2), 173–196.

Flickr. (2018). Flickr Terms of Use. Available from https://www.flickr.com/help/terms (last accessed July 16, 2021)

Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., & Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem Services*, *38*(June 2018), 100958.

Hadwen, W. L., Hill, W., & Pickering, C. M. (2007). Icons under threat: Why monitoring visitors and their ecological impacts in protected areas matters. *Ecological Management and Restoration*, *8*(3), 177–181.

Harrell, F.E. Jr., with contributions from C. Dupont and many others. (2021). Hmisc: Harrell Miscellaneous. R package version 4.5-0. Available from https://CRAN.R-project.org/package=Hmisc (last accessed July 16, 2021)

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & 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), 1–10.

Heikinheimo, V., Minin, E. Di, Tenkanen, H., Hausmann, A., Erkkonen, J., & 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.

Hu, B., & Morrison, A. M. (2002). Tripography: Can destination use patterns enhance understanding of the VFR market? *Journal of Vacation Marketing*, *8*(3), 201–220.

Jenkins, A., Croitoru, A., Crooks, A. T., & Stefanidis, A. (2016). Crowdsourcing a collective sense of place. *PLoS ONE*, *11*(4), 1–20.

Kahle, D., Wickham, H., Jackson, S., Korpela, M. (2019). R package 'ggmap' (version 3.0.0) Available from https://github.com/dkahle/ggmap (last accessed July 16, 2021)

Kang, R. (2020). Using logistic regression for persona segmentation in tourism: A case study. *Social Behavior and Personality*, *48*(4), 1–17.

Kim, J., Thapa, B., Jang, S., & Yang, E. (2018). Seasonal spatial activity patterns of visitors with a mobile exercise application at Seoraksan National Park, South Korea. *Sustainability (Switzerland)*, *10*(7).

Kulczyk, S., Woźniak, E., & Derek, M. (2018). Landscape, facilities and visitors: An integrated model of recreational ecosystem services. *Ecosystem Services*, *31*, 491–501.

Kyle, G., Graefe, A., Manning, R., & Bacon, J. (2004). Effects of place attachment on users' perceptions of social and environmental conditions in a natural setting. *Journal of Environmental Psychology*, *24*(2), 213–225.

Lau, A. L. S., & McKercher, B. (2004). Exploration versus acquisition: A comparison of first-time and repeat visitors. *Journal of Travel Research*, *42*(3), 279–285.

Leask, A. (2016). Visitor attraction management: A critical review of research 2009–2014. *Tourism Management*, *57*, 334–361.

Lee, J. Y., & Tsou, M. H. (2018). Mapping spatiotemporal tourist behaviors and hotspots through location-based photo-sharing service (Flickr) data. *Lecture Notes in Geoinformation and Cartography*, (208669), 315–334.

Leung, Y., Spenceley, A., Hvenegaard, G., Buckley, R., & Groves, C. (2018). *Tourism and visitor management in protected areas : guidelines for sustainability*. *Tourism and visitor management in protected areas : guidelines for sustainability*.

Li, X. (Robert), Cheng, C. K., Kim, H., & Petrick, J. F. (2008). A systematic comparison of first-time and repeat visitors via a two-phase online survey. *Tourism Management*, *29*(2), 278–293.

Liu, H., Li, X. (Robert), Cárdenas, D. A., & Yang, Y. (2018). Perceived cultural distance and international destination choice: The role of destination familiarity, geographic distance, and cultural motivation. *Journal of Destination Marketing and Management*, *9*(February), 300–309.

Mancini, F., Coghill, G. M., & Lusseau, D. (2018). Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. *PLoS ONE*, *13*(7), 1–19.

Mason, P. (2005). Visitor management in protected areas: From "hard" to "soft" approaches? *Current Issues in Tourism*, *8*(2–3), 181–194.

McKercher, B., Shoval, N., Ng, E., & Birenboim, A. (2012). First and Repeat Visitor Behaviour: GPS Tracking and GIS Analysis in Hong Kong. *Tourism Geographies*, *14*(1), 147–161.

Moore, S. A., Rodger, K., & Taplin, R. (2015). Moving beyond visitor satisfaction to loyalty in nature-based tourism: a review and research agenda. *Current Issues in Tourism*, *18*(7), 667–683.

Moyle, B. D., Scherrer, P., Weiler, B., Wilson, E., Caldicott, R., & Nielsen, N. (2017). Assessing preferences of potential visitors for nature-based experiences in protected areas. *Tourism Management*, *62*, 29–41.

Muñoz, L., Hausner, V. H., Runge, C., Brown, G., & Daigle, R. (2020). Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. *People and Nature*, *2*(2), 437–449.

Mutanga, C. N., Vengesayi, S., Chikuta, O., Muboko, N., & Gandiwa, E. (2017). Travel motivation and tourist satisfaction with wildlife tourism experiences in Gonarezhou and Matusadona National Parks, Zimbabwe. *Journal of Outdoor Recreation and Tourism*, *20*(June), 1–18.

National Park Service Integrated Resource Management Applications (NPS IRMA). (2019) https://grsm-nps.opendata.arcgis.com/ (last accessed July 16, 2021)

National Park Service. (2020). National Park Service Visitation Numbers. https://www.nps.gov/aboutus/visitation-numbers.htm (last accessed July 16, 2021)

Nyaupane, G. P., & Graefe, A. R. (2008). Travel distance: A tool for nature-based tourism market segmentation. *Journal of Travel and Tourism Marketing*, *25*(3–4), 355–366.

Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management*, *33*(3), 672–682.

Papadogiannaki, E., Braak, A., Holmes, N., Eury, D., & Hollenhorst, S. J. (2008). *Great Smoky Mountains National Park Visitor Study*. Social Science Program, National Park Service, U.S., Department of the Interior.

Pearce, J., & Dowling, R. (2019). Monitoring the quality of the visitor experience: An evolutionary journey. *Journal of Outdoor Recreation and Tourism*, *25*(November 2017), 87–90.

Pettebone, D., & Meldrum, B. (2018). The Need for a Comprehensive Socioeconomic Research Program for the National Park Service. *The George Wright Forum*, *35*(1), 22–31.

Pickering, C., Rossi, S. D., Hernando, A., & Barros, A. (2018). Current knowledge and future research directions for the monitoring and management of visitors in recreational and protected areas. *Journal of Outdoor Recreation and Tourism*, *21*(June 2017), 10–18.

Pickering, C., Walden-Schreiner, C., Barros, A., & Rossi, S. D. (2020). Using social media images and text to examine how tourists view and value the highest mountain in Australia. *Journal of Outdoor Recreation and Tourism*, *29*(September 2019).

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/. (last accessed July 16, 2021)

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

Roemer, J. M., & Vaske, J. J. (2014). National Park Service Visitor Satisfaction: A Comparative Analysis. *Journal of Park and Recreation Administration Winter*, *32*(4), 35–51.

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

Schirpke, U., Tasser, E., & Tappeiner, U. (2013). Predicting scenic beauty of mountain regions. *Landscape and Urban Planning*, *111*(1), 1–12.

Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. M. (2016). Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management*, *183*, 703–711.

Silverman, B. (1986). Density estimation for statistics and data analysis, in monographs on statistics and applied probability. London, UK: Chapman and Hall.

Sinclair, M., Mayer, M., Woltering, M., & Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management*, *263*(October 2019), 110418.

Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS ONE*, *11*(9), 1–16.

Stamberger, L., van Riper, C. J., Keller, R., Brownlee, M., & Rose, J. (2018). A GPS tracking study of recreationists in an Alaskan protected area. *Applied Geography*, *93*, 92–102. https://doi.org/10.1016/j.apgeog.2018.02.011

Supak, S., Brothers, G., Bohnenstiehl, D. W., & Devine, H. (2015). Geospatial analytics for federally managed tourism destinations and their demand markets. *Journal of Destination Marketing and Management*, *4*(3), 173–186.

Teles da Mota, V., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism*, *30*(April), 100295.

Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, *64*, 237–248.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & 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.

Tverijonaite, E., Ólafsdóttir, R., & Thorsteinsson, T. (2018). Accessibility of protected areas and visitor behaviour: A case study from Iceland. *Journal of Outdoor Recreation and Tourism*, *24*(September), 1–10.

Van Berkel, D. B., Tabrizian, P., Dorning, M. A., Smart, L., Newcomb, D., Mehaffey, M., … Meentemeyer, R. K. (2018). Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. *Ecosystem Services*, *31*, 326–335.

Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & 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.

Walden-Schreiner, C., Rossi, S. D., Barros, A., Pickering, C., & Leung, Y. F. (2018). Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. *Ambio*, *47*(7), 781–793.

Weaver, D. B., & Lawton, L. J. (2007). Twenty years on: The state of contemporary ecotourism research. *Tourism Management*, *28*(5), 1168–1179.

Weaver, D. B., & Lawton, L. J. (2017). A new visitation paradigm for protected areas. *Tourism Management*, *60*, 140–146.

Wei, T. & Simko, V. (2021). R package "corrplot": Visualization of a Correlation Matrix (Version 0.89). Available from https://github.com/taiyun/corrplot (last accessed July 16, 2021)

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*(1), 2976.

Worboys, G. L., M. Lockwood, A. Kothari, S. F. and I. P. (eds). (2015). *Protected areas governance and management*. ANU Press.

Yang, L., Wu, L., Liu, Y., & Kang, C. (2017). Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from flickr. *ISPRS International Journal of Geo-Information*, *6*(11), 1–18.

Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services*, *24*, 68–78.

CHAPTER 3

ONLINE REVIEWS AND CROWDSOURCED PHRASAL LEXICON TO IDENTIFY CULTURAL ECOSYSTEM SERVICES: A CASE STUDY OF EL CAJAS NATIONAL PARK, ECUADOR[2]

[2] Kong, I. & Sarmiento, F.O. Submitted to *Ecosystem Services*, September 10, 2021.

ABSTRACT

Social media has been utilized to identify cultural ecosystem services (CES) that encapsulate place-dependent non-material values in protected areas, yet related studies remained in single-word lexicons using short text data such as photo tags. Taking El Cajas National Park in Ecuador as a case study, we propose a novel approach to utilize text-rich online reviews and a phrasal lexicon comprised of noun and verb phrases to identify CES. Also, we suggested a crowdsourced survey to gather public consensus in pairing the phrases with CES for the lexicon. The final lexicon was then applied to count CES-positive online reviews, and the results were further compared across different languages (English vs. Spanish), platforms (TripAdvisor vs. Google Maps), and phrasal types (noun vs. verb). Additionally, frequent phrases for CES in English and Spanish reviews were juxtaposed to highlight the linguistic distinctions in describing CES. Overall, we confirmed the usefulness of online reviews and the phrasal lexicon approach to improving the accuracy of interpreting site-specific CES, with high-context and concrete phrasal expressions. The approach, therefore, can help identify CES in data-lacking protected areas, harnessing CES-inclusive conservation strategies.

INTRODUCTION

Protected areas are a key component of global conservation strategy for preserving biological and cultural diversity, while also providing public recreation and tourism opportunities (Leung et al., 2018; Worboys et al., 2015). From this perspective, protected areas offer comprehensive cultural ecosystem services (CES), ranging from aesthetic and recreational values to biological and educational values (MEA, 2005). In addition, protected areas serve as a symbolic place to embody local culture and identity that has been shared within the communities, promoting

historic awareness, traditional knowledge system, and social engagement (Bieling, 2014; Cebrián-Piqueras et al., 2020; Pascua et al., 2017). Thus, CES values play a crucial role to support the physical and mental wellbeing of the community (Kosanic and Petzold, 2020). However, many protected areas have been facing managerial challenges, particularly in countries undergoing radical socioeconomic changes involving massive land transitions for agriculture and urban expansion (Lambon et al., 2001; Ross et al., 2017). The following socio-cultural transformation compounded with radical environmental degradation can weaken the human-nature interactions, resulting in the marginalization of CES values (Fisher and Eastwood, 2016; Pascua et al., 2017). For the last two decades, numerous studies have examined site-specific CES throughout the world to highlight the significance of such values and incorporate them in conservation practices, yet there has been criticism for the previous studies having a strong country bias towards high-income countries in North America and Western Europe (Englund et al., 2017; Kosanic and Petzold, 2020; Milcu et al., 2013). Indeed, several conservation policies based on the framework of ecosystem services (i.e., payment for ecosystem services) have been widely implemented in middle to low-income countries of South America, Asia, and Africa, yet CES values have been largely overlooked in such programs (Balvanera et al., 2012; Calvet-Mir et al., 2015; Grima et al., 2016).

One of the most critical challenges to consider CES in current conservation policy has been the lack of data and tools to distinguish intangible CES embedded in the landscape and community (Chan et al., 2006; Daniel et al., 2012). In contrast to tangible ecosystem services such as carbon storage or water supply, CES entail substantial uncertainty to identify and quantify the values (Chan et al., 2012). Furthermore, CES values are found to be dependent on socio-environmental context, requiring case-by-case examinations (Bieling, 2014; Pascua et al., 2017). Early CES studies have taken environmental variables, such as land cover and accessibility to the roads, to

reveal the spatial distribution CES over large areas (Chan et al., 2006; Nahuelhual et al., 2013). However, indirect approaches have been limited to assessing aesthetic and recreational values (Hernández-Morcillo et al., 2013; Milcu et al., 2013). Meanwhile, qualitative approaches such as free listing surveys or interviews have been used to obtain a wider range of CES directly from the public, but the costly and time-consuming process has constrained such analysis to local scales (Bieling, 2014; Pascua et al., 2017). Also, the survey respondents often found it challenging to elaborate CES, due to the immaterial and immeasurable nature of CES (Stålhammar and Pedersen, 2017). When analyzing the results, many have encountered challenges with the bundled CES that could not be split into a single value as well as the site-dependent CES that did not correspond to the ecosystem services frameworks such as MEA (2005) or CICES (Haines-Young and Potschin, 2010) (Bieling, 2014; Bullock et al., 2018; Pascua et al., 2017).

The application of social media in CES studies, therefore, has been acclaimed to alleviate the shortcomings of previous methods, with a vast size of ready-to-use content created by large populations (Richards and Friess, 2015; Teles da Mota and Pickering, 2020). While several studies have utilized the geolocations of social media to exhibit the distribution of aesthetic or recreational values (Gosal et al., 2019; Langemeyer et al., 2018), recent studies have incorporated text analysis to distinguish diverse CES from social media contents (Chen et al., 2018; Hale et al., 2019; Ghermandi et al., 2020; Rossi et al., 2020). Chen et al. (2018) utilized Instagram hashtags to analyze six CES in a riverine landscape in Canada, and Hale et al. (2019) utilized Flickr photo tags to identify 11 CES in a river ecosystem in Idaho, USA. Ghermandi et al. (2020) also utilized title and photo tags from Flickr to analyze CES in a coastal region of Mexico. Several studies focused on the photo contents to manually annotate the objects on the photos and identified CES in mangrove forests in Singapore (Richards and Friess, 2015), a coastal area of Brazil (Retka et al.,

2019), and a protected in Argentina (Rossi et al., 2020). Automatic image recognition techniques in recent years have been applied to expedite the photo annotation to analyze CES in Singapore (Richards and Tunçer, 2018), marine protected areas (Ruiz-Frau et al., 2020), and northern Italy (Egarter Vigl et al., 2021).

Despite significant advances in CES research using social media, most studies have relied primarily on short text contents and single-word lexicons, namely a single-word-based dictionary (Chen et al. 2018; Egarter Vigl et al. 2021; Ghermandi et al., 2020; Hale et al. 2019; Richards and Friess, 2015). For instance, Chen et al. (2018) defined Instagram posts containing '*beauty*' and '*beautiful*' to indicate aesthetic value. Likewise, Ghermandi et al. (2020) identified Flickr posts with the word '*landscape*' to imply aesthetic value. Yet, the single-word lexicon has several limitations. First, it may misinterpret the meaning of a word in a specific context, resulting in a false-positive annotation of CES. This problem is the most noticeable when it comes to homonyms (i.e., terms that share spelling but differ in meaning) (Roll et al., 2018). For instance, the term '*spring*' can be found in phrases such as '*enjoy spring season*' and '*see hot spring*', thus creating a lexicon for '*spring*' based on the former (i.e., season) may result in a misleading result for the latter (i.e., geological feature). Second, a single word lexicon can overestimate the frequency of CES by counting every word in the texts. For instance, '*Beautiful!*' or '*It was amazing*' barely indicate particular objects or actions to associate with CES but are likely to be counted as CES with a single word lexicon.

Moreover, previous lexicons have been developed from a handful of experts involved in the study (Chen et al. 2018; Hale et al. 2019; Richards and Friess, 2015). The expert-based lexicon is convenient to reach consensus in pairing the words with CES by exclusively engaging with scholars who are familiar and well-versed with CES. However, the reliability of such lexicons

might be questioned because it involves the viewpoint of a small number of experts, which may contradict the public opinions as well as the researchers in other case studies. For instance, Hale et al. (2019) defined '*beauty*' as recreational and existential values, whereas several other lexicons applied in Chen et al. (2018), Chen et al. (2020), and Richards and Friess (2015), have defined the word as aesthetic value.

In this study, we propose to identify CES based on online reviews, which have been largely underutilized despite the text-rich and concrete linguistic expressions (Cong et al., 2014; Spalding and Parrett, 2019). Using the text-mining techniques for the unstructured texts of online reviews, we develop a lexicon comprised of phrasal expressions. Also, we suggest a crowdsourced method to annotate the phrases to relevant CES as a novel strategy to minimizing expert bias. Using the lexicon, we analyze CES frequencies from online reviews and compare the results across languages, review platforms, and phrase types to untangle the influence of the individual factors on the identification of CES. We further compare key phrases from different languages to highlight linguistic distinctions in describing CES. El Cajas National Park in Ecuador was chosen as a study site.

METHODS

Study site

We chose El Cajas National Park in Ecuador, to be one of the most underrepresented geographic locations in current CES discourses (Englund et al., 2017; Kosanic and Petzold, 2020; Milcu et al., 2013; Rossi et al., 2020), despite the rich biological and cultural distinctions. El Cajas National Park (CNP) is a protected area located in the southern Andes of Ecuador, near the city of Cuenca (Figure 3-1). The high altitude of the park, ranging from 3,100 to 4,500 m.a.s.l., creates a

cool and wet climate that distinguishes the site from typical subtropical locations. Accordingly, the climate shapes a unique ecosystem called páramo, a high-altitude grassland native to South America, alongside thousands of endemic species (Sarmiento, 2016). The omnipresent ancient glacial lakes also consist of unique landscape features of CNP, providing scenic beauty and year-round water resources to Cuenca (Sarmiento, 2016; van Colen et al., 2016). CNP was declared as a national park in 1996, followed by international recognition as a Ramsar Wetland in 2002 (van Colen et al., 2016). CNP is also a UNESCO Biosphere Reserve for biocultural heritage embedded in the landscape to represent Andean highland culture (Sarmiento, 2016). Residents in the Andean highlands have a strong connection to the mountains and venerate the mountain goddess Pachamama, as the creator of all life on Earth (Sarmiento, 2016).

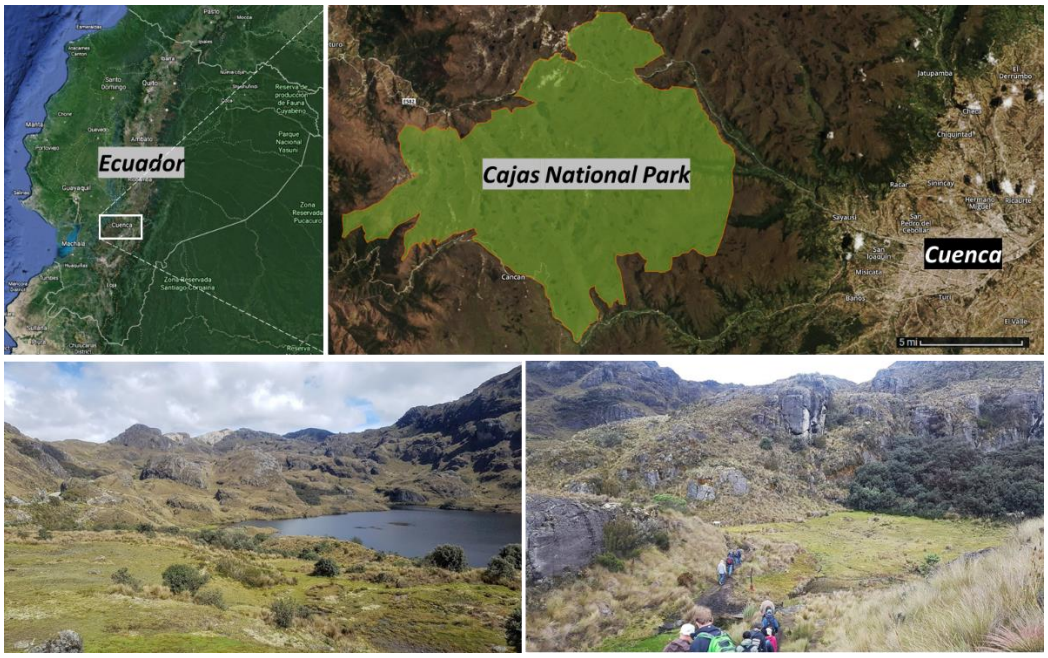


Figure 3-1. El Cajas National Park in Ecuador as a study site.

The photos below were taken by the first author.

Accordingly, CNP has been a flagship tourist destination in Ecuador, with over 80,000 visitors in 2019 (Ministerio del Ambiente, 2019). The number of tourists in CNP has doubled since 2001, and the percentage of international visitors has increased from 29 percent in 2001 to 39 percent in 2019 (Ministerio del Ambiente, 2019). Nonetheless, the peripheral páramo environment, which serves as a buffer zone for CNP, has been threatened by land transition pressures such as land burning and cattle grazing (Donoso and Sarmiento, 2021; Ross et al., 2017). Furthermore, the concurrent sociocultural shifts may result in the devaluation or marginalization of indigenous culture and heritage associated with CNP (Sarmiento, 2016). Since 2008, CNP has been implementing payment for ecosystem services to protect the park and its surroundings, yet the policy has been exclusively addressing tangible water resources (Bremer et al., 2016), overlooking CES values (Farley and Bremer, 2017; Grima et al., 2016).

Data acquisition

We collected all available online tourism reviews from TripAdvisor and Google Maps as of February 2021. TripAdvisor was chosen as the most popular online review platform in tourism research (Cong et al., 2014; Spalding and Parrett, 2019). Using the *rvest* (Wickham, 2020) and the *Rselenium* (Harrison, 2020) packages, we wrote the R scripts to automate the web scraping for TripAdvisor online reviews. Once the script opened the TripAdvisor page for El Cajas National Park and proceeded to the Review section, each page displayed five online reviews. Then the script activated to click the 'More...' buttons in the Review section and scraped the full-length 'review' as well as the ‘title’ (hereafter, the reviews). The scraping process was iterated until the last page was reached. Subsequently, we changed the language setting from English to Spanish and followed the same procedure as mentioned above, to operate cross-linguistic analysis. The Spanish reviews

were then translated into English using Google Translate, which has proven to be a reliable translator in cross-linguistic research (Barbro et al., 2020; Mathayomchan and Sripanidkulchai, 2019).

We also obtained online reviews from Google Maps, which has been known to represent a broader range of user demographics particularly in non-English-speaking countries (Mathayomchan and Sripanidkulchai, 2019). From this dual-platform approach, we could reduce the commonly known bias of TripAdvisor to overrepresent the users and destinations in North America and Western Europe (Kladou and Mavragani, 2015; Taecharungroj and Mathayomchan, 2019). We manually scraped online reviews from Google Maps since the platform was built on a dynamic JAVA environment so that automatic web scraping was restricted. In addition, Google Maps did not distinguish the language but attached automated Google translations when the original contents were not in English. To determine the original languages, we used two language detection packages, the *textcat* (Hornik et al., 2013) and the *cld2* (Ooms, 2020) in R Studio. Since the results from the two packages showed substantial inconsistency among Catalan, Basque, Galician, and Spanish, we classified the review to be Spanish when either outcome was Spanish. The reviews without auto-translations were classified as English.

Developing a crowdsourced phrasal lexicon for CES

We utilized text mining methods to extract phrasal expressions from online reviews. First, we cleaned the unstructured text of online reviews using natural language processing, which included tokenization, lemmatization, part-of-speech (POS) tagging, and the removal of punctuation, emojis, and selective determinators (i.e., a, an, the). Named entity recognition was

also applied to detect place names and merge them into a chunk (e.g., 'El-Cajas-National-Park'). The process was run in R Studio using the *spacyr* package (Benoit and Matsuo, 2020).

Then, we utilized the phrase parsing (i.e., phrase chunking) technique to extract noun phrases (i.e., phrases that embellish a noun) and verb phrases (i.e., phrases that begin with a verb), which play a crucial role in sentences to indicate objects and actions (Rudra et al., 2018). Noun phrases can also supplement verb phrases with auxiliary verbs like 'be' or 'have,' which are often absent in informal writing (e.g., *'[It was a] Beautiful landscape!'*). We chose a POS-based phrase parsing algorithm that detects text chunks based on the POS tags associated with the lemmas (Wijffels, 2019). Unlike vector-based methods that rely on statistical cooccurrence of the phrases, the POS-based parsing method allows the researchers to customize the phrasal structures of interests (Wijffels, 2019). After evaluating feasible POS combinations from the entire review, we built a POS pattern library for noun and verb phrases, which can be found in Appendix A. For instance, '(VERB*(ADP|ADJ|NOUN)*NOUN$)' indicated a POS pattern of verb phrases that begin with a verb, followed by any combination of adposition, adjective, noun, and always conclude with a noun. The exemplary phrases from this pattern included '*eat trout* (VERB-NOUN)', '*hike around trail* (VERB-ADP-NOUN)', and '*enjoy nice view* (VERB-ADJ-NOUN)'. The phrase parsing was operated using the *udpipe* in R Studio (Wijffels, 2019).

Following the retrieval of thousands of crude phrases from the phrasal parsing, two researchers skimmed the crude phrases and selected viable phrases to indicate CES. We removed fragmented sentences that were less likely to suggest CES values (i.e., '*go to*', '*stop at*', '*be in*') as well as the phrases involving advice, warning, recommendations, and complaints (i.e., *'take bus', 'wear warm clothes'*). The phrases that appeared only a few times were mostly removed, except for highly specific and relevant expressions to describe site-specific values. Lastly, we manually

eliminated certain phrases that were highly context-dependent. For instance, the phrase '*good walking*' can indicate a recreational activity, but when it is found in '*bring good walking shoes*', the meaning wights on advice rather than a recreational value. The skimming process was drafted by the first author and proofread by the second author, who had both been to CNP.

We pre-defined CES categories as a common procedure to create lexicons in CES studies (Chen et al., 2018; Ghermandi et al., 2020; Hale et al., 2019). After examining CES categories in MEA (2005), TEEB (2010), and CICES (Haines-Young et al., 2010) as well as review papers on CES case studies (Englund et al., 2017; Hernández-Morcillo et al., 2013; Kosanic and Petzold, 2020; Milcu et al., 2013), we defined eight CES as follows: aesthetic, biological, cultural (i.e., cultural heritage), educational, identity (i.e., sense of place), recreational (i.e., recreational and ecotourism), social (i.e., social relations), and spiritual (i.e., spiritual and religious) values. Here, we defined biological value to reflect biological distinctions in protected areas although the value has been co-founded in existential value (Hale et al., 2019) or recreational value (MEA, 2005). Subsequently, we built a survey comprised of a set of phrases to be labeled with eight CES categories and operated a crowdsourced survey using Amazon Mechanical Turk (MTurk), an Amazon-operated survey platform that has a large pool of survey participants (Paolacci and Chandler, 2014). The survey obtained the IRB exemption from the University of Georgia IRB office. The detailed procedure of the survey as well as the definitions of eight CES can be found in Appendix C. Once the survey hit 15 responses, it was automatically closed for a further influx of survey participants. As post-processing, we compiled the responses and determined the most dominant CES to be the most relevant CES for phrases. Some phrases were attached with more than one CES due to the comparable number of responses. For the cases with considerable inconsistency among similar phrases or unreasonable CES being attached to the phrase, the

researchers had to make some adjustments, but these changes were kept to a minimum. As a result, we created a crowdsourced phrasal lexicon for eight CES.

Analyzing CES frequency

The crowdsourced phrasal lexicon was then applied in online reviews to identify CES. When a set of lemmas in a review matched with a phrase from the lexicon in the same sequence, the review was assigned the CES corresponding to the phrase. When a review had multiple phrases indicating the same CES, the review was reduced to be 'positive' for the CES. Therefore, one review can have up to eight CES. The analysis was carried out in R Studio using the *udpipe* package (Wijffels, 2019). The results were aggregated and tallied to count the frequency of CES by language subgroups (i.e., English and Spanish), platform subgroups (i.e., TripAdvisor and Google Maps), and phrasal types (i.e., verb phrases and noun phrases) as well as the entire online reviews.

Comparing key phrases

To highlight linguistical distinctions in describing CES, we filtered 30 of the most frequently occurring phrases from the subgroups of language-CES pairs. When the frequencies of many phrases were tied, the phrase with the least n-grams (i.e., the number of lemmas in a phrase) was prioritized for selection. Then, we applied a network analysis that utilizes a graph theory to assess the relationship of a system of connected elements (Ruiz-Frau et al., 2020). We divided the 30 key phrases into lemmas and tallied the occurrences of the lemmas throughout the 30 phrases, resulting to generate a matrix for vertices (i.e., nodes). The POSs of the lemmas were then attached to the vertex matrix. Subsequently, edges (i.e., links) were created from filtering the consecutive

pairs of lemmas in the phrases. When the n-gram of a phrase was more than two, multiple edges were generated from the phrase by moving the bigram window within the phrase. For instance, the verb phrase '*breathe fresh air*' generated two edges: '*breathe (originating) - fresh (ending)*' and '*fresh (originating) - air (ending)*'. Then the frequency of the phrases was attached to the edges to indicate degree centrality (i.e., edge thickness). We also kept the directions of the edges so that it can specify in-degree centrality (i.e., the edge originating from a vertex) and out-degree centrality (i.e., the edge ending to a vertex). For the comparisons of the networks across language-CES pairs, the size of vertices and degree centrality of edges were normalized between 0 and 1. The analysis was operated using the *igraph* package in R Studio (Csardi and Nepusz, 2006).

RESULTS

Data summary

We collected 3,201 online reviews from TripAdvisor and Google Maps (Table 3-1). TripAdvisor had a comparable number of reviews between English and Spanish (1,103 in English and 953 in Spanish). However, Google Maps showed a dominant share of Spanish reviews than English reviews (194 in English and 951 in Spanish). The median word count for online reviews revealed that the reviews written in English were lengthier than those written in Spanish for both platforms. Also, online reviews from TripAdvisor were lengthier than Google Maps in both languages.

Table 3-1. The details of collected online reviews by languages and platforms

	TripAdvisor		Google Maps	
The number of online reviews	2,056		1,145	
	English	Spanish	English	Spanish
	1,103	953	194	951
Median word counts per review (in English)	60	40	14	9

Summary of the crowdsourced phrasal lexicon for CES

The phrase parsing resulted in 6,552 crude noun phrases and 10,832 crude verb phrases. Following the skimming process, we selected 872 noun phrases and 651 verb phrases for the viable phrases in the crowdsourcing survey. The phrases included the expressions indicating objects (e.g., *'see llama', 'paper tree forest'*), activities (e.g., *'do day hike', 'sport fishing'*), as well as metaphorical and symbolic expressions (e.g., *'national treasure', 'hidden gem', 'heritage of humanity'*). Table 3-2 summarized the number of phrases in the final lexicon by CES, language, and phrasal type. The full lexicon can be found on https://github.com/ihKong/phrasal_lexicon_CES.

The final lexicon had the phrases for recreational value the most, followed by aesthetic and biological values. Cultural and educational values, on the other hand, had the fewest phrases. Depending on the languages, the number of phrases for English reviews outnumbered Spanish reviews, except for social, spiritual, and cultural values. Meanwhile, noun phrases were the dominating phrasal type for aesthetic, biological, and identity values, but verb phrases prevailed for recreational, social, and spiritual values.

Table 3-2. The number of viable phrases by CES categories, languages, and phrasal types

		Aesthetic	Biological	Cultural	Educational	Identity	Recreational	Social	Spiritual
Total [a]		**361**	**289**	**17**	**22**	**167**	**446**	**87**	**169**
ENG		**255**	**218**	**9**	**18**	**113**	**361**	**31**	**51**
	VP	52	46	2	14	20	226	20	30
	NP	203	172	7	4	93	135	11	21
SPN		**247**	**156**	**9**	**11**	**100**	**203**	**66**	**141**
	VP	57	36	2	7	20	112	43	87
	NP	190	120	7	4	80	91	23	54

[a] *Due to duplication of phrases between languages, the sum of phrases from two languages surpassed the provided total number of phrases. The sum of English and Spanish phrases for aesthetic value, for example, was 255+247=502, although the total number of viable phrases for aesthetic value was 361.*

CES frequency

The results showed the frequency of online reviews including relevant CES. We first presented the distribution of CES frequency across the language as well as CES categories (Figure 3-2). Aesthetic and recreational values were the two most prevalent CES in online reviews written in both English and Spanish. The linguistic difference in aesthetic value was minimal, with more than half of both English (52.7%) and Spanish (55.1%) reviews mentioning the value. However, recreational value showed a significant language gap that over a half of English reviews (55.3%) described phrases associated with recreational value, whereas less than a quarter of Spanish reviews (24.8%) did. In addition, biological (26.7%) and identity values (24.1%) ranked the third and fourth most common CES in English reviews, but Spanish reviews ranked spiritual value (17.2%) as the third most common CES, followed by biological (15.4%) and identity values (13.8%). The remaining CES were found less than five percent of the entire reviews for both English and Spanish reviews, except for social value in Spanish reviews (5.4%). Educational and cultural values were the least prevalent CES for both languages, accounting for less than two percent of online reviews.

Furthermore, we examined the CES frequency based on phrasal types (Figure 3-2). The result described the performance of each phrasal type in identifying relevant CES from online reviews. The sum of the results from noun-phrase-only and verb-phrase-only was not equal to the result from the entire lexicon, due to the overlapping phrases in online reviews. Throughout most CES categories, noun phrases outperformed verb phrases by a wide margin to retrieve CES-related phrases. Also, the difference between noun phrase-only and the entire lexicon (i.e., both noun and verb phrases) remained marginal throughout most CES. For instance, the result from noun-phrase-only for aesthetic value was almost identical with the result from the entire lexicon, found in over half of English and Spanish reviews. Yet, there were a few exceptions. Recreational value demonstrated a significant role of verb phrases that were as frequent as the result from noun-phrase-exclusive. Also, the gap between the noun phrase and the entire lexicon was the greatest, highlighting the role of verb phrases in identifying recreational value.

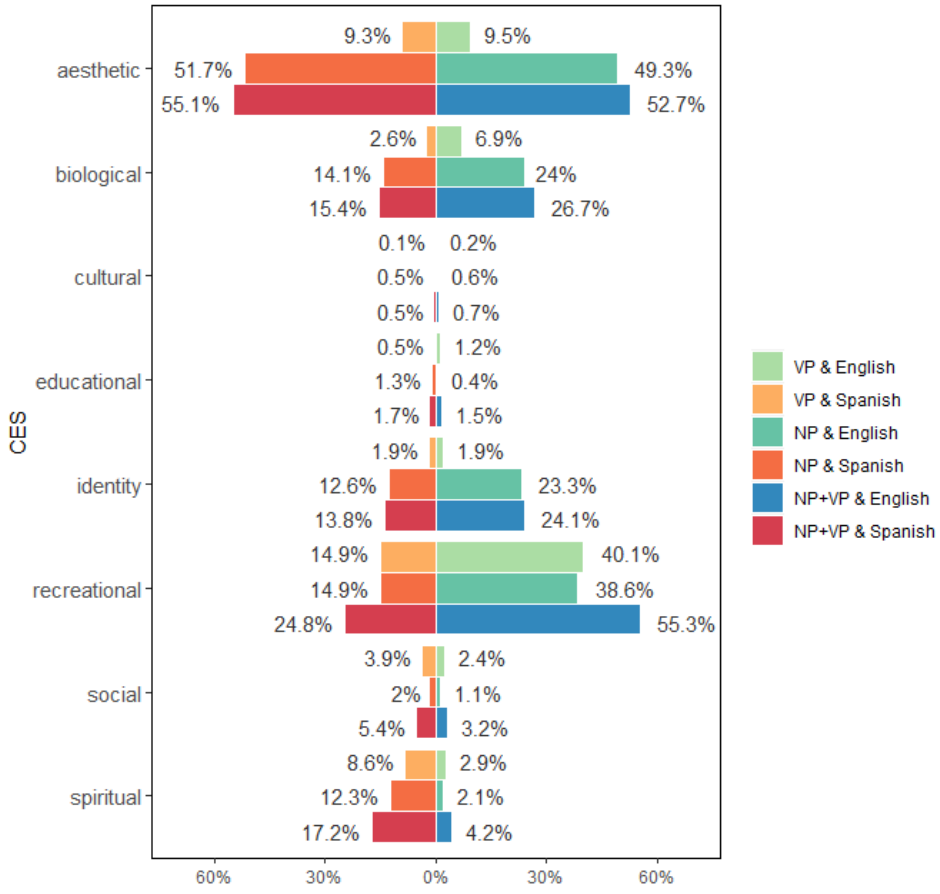


Figure 3-2. CES frequencies by CES and the phrasal types. For each CES category, the bars in red and blue indicate the results from Spanish (left) and English (right), while showing the results from verb phrases only (VP, on top), noun phrases only (NP, in the middle), and the comprehensive lexicon of both verb and noun phrases (NP+VP, on the bottom).

The following result in Figure 3-3 illustrated the role of platforms in the frequencies of CES. TripAdvisor outperformed Google Maps across all CES categories for both languages. Furthermore, TripAdvisor alone has proven to be more effective than utilizing both TripAdvisor and Google Maps. For instance, recreational values were found in 37.5% of Spanish reviews from TripAdvisor but were found in 24.8% of online reviews from both platforms. In other words, online reviews from Google Maps were less likely to contain phrases that imply CES values. The findings

can be ascribed to the average length of online reviews, as longer texts are more likely to contain more CES-related phrases.

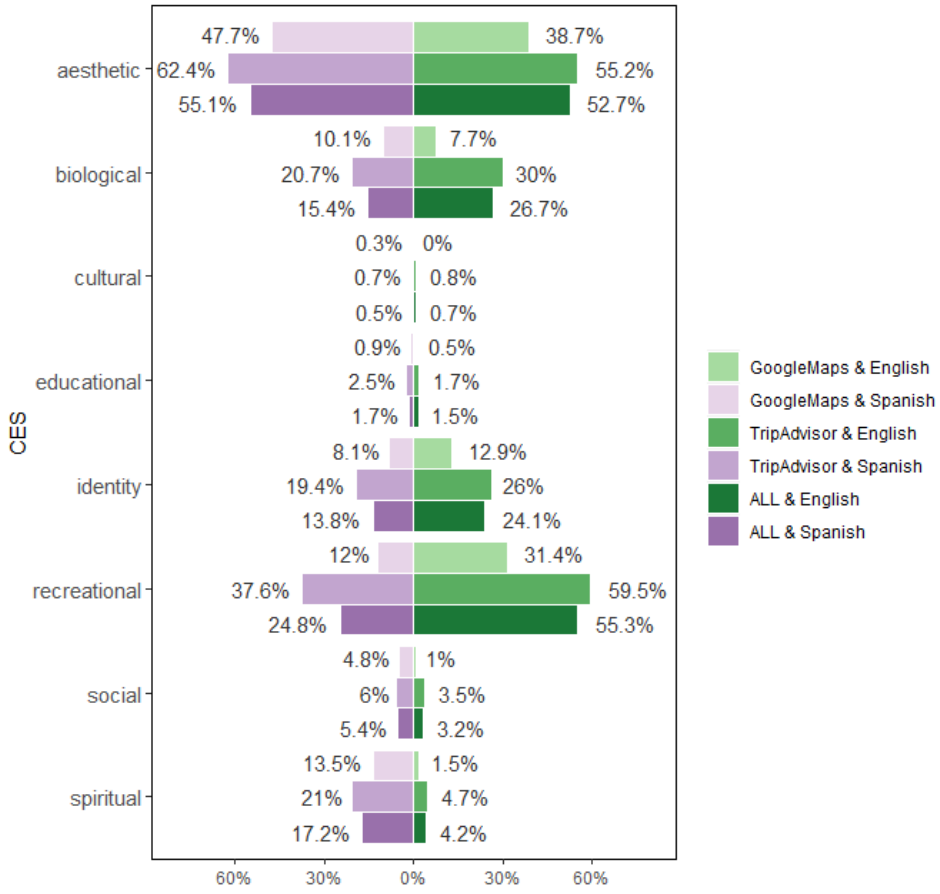


Figure 3-3. CES frequencies by CES and the platforms. For a CES category, the bars in purple and green indicate the results from Spanish (left) and English (right), while showing the results from Google Maps only (top), TripAdvisor only (middle), and both platforms (bottom).

Text network of frequent phrases by CES

The results from network analysis (Figure 3-4) reveal the interconnectedness of the 30 most frequent phrases in the CES-language pairs. Throughout the networks, we highlight the differences in key vertices, edges, and network connectivity between English and Spanish reviews.

- Aesthetic value

The text networks for aesthetic value exhibited a predominance of adjective and noun vertices, demonstrating the importance of noun phrases in describing aesthetic experiences for both English and Spanish reviews. Particularly, the substantial size of the adjective vertex '*beautiful*' reflected the frequent occurrence of the word throughout the phrases. Furthermore, the vertex '*beautiful*' served as a hub connecting multiple noun vertices with great out-degree centralities, notably with '*place*', '*landscape*', and '*park*'. To emphasize language distinctions, online reviews in English tended to use the terms '*scenery*' and '*view*' with various adjectives whereas Spanish reviews preferred to use the word '*landscape*' with various adjectives. The clusters in both networks were well-connected, with a few fragments like '*take photo*' or '*natural beauty*'.

- Biological value

The networks of biological value for English and Spanish reviews included numerous large vertices connected with significant degree centrality, indicating that a few phrases exhibited exceptionally higher frequency among the selected phrases. Noun phrases were the most common in both languages, but English reviews had a verb vertex '*see*' connected with several nouns. The reviews in English and Spanish both had '*flora and fauna*' as the most frequent phrase, but the remaining phrases showed significant linguistic differences. English reviews mentioned '*see*

llama', '*wild llama*', '*see bird*', '*plant life*', '*paper tree*' with comparable vertex sizes and degree centrality. Particularly, English reviews appreciated the encounter with '*llama*' while the reviews in Spanish rarely mentioned llama. Meanwhile, Spanish reviews mentioned '*many lagoon*' and '*paper tree*' but the degree centrality was overall weak in the entire network, indicating the strong domination of '*flora and fauna*'. In addition, the network for Spanish reviews was more fragmented than English reviews, which can be explained with more eclectic expressions in Spanish reviews. Indeed, Spanish reviews tended to elaborate biological features more concretely, such as '*polylepis forest*', '*andean paramo*', and '*endemic species*', whereas English reviews remained to mention '*interesting bird*' or '*small flower*'.

- Cultural value

Since the lexicon contained less than 30 phrases for cultural values, the key phrases from the entire review were also less than 30. The phrases for cultural value showed highly divergent interests depending on the language. The most prevalent phrases in English reviews included '*history of area*', '*old brewery*', and '*coca tea*' as shown in large vertices and strong degree centrality, but Spanish reviews preferred the phrase '*heritage of humanity*' the most. Meanwhile, both languages exhibited extremely fragmented networks, suggesting that few phrases shared common lemmas. Such results can be ascribed to the modest number of phrases in the lexicon.

- Educational value

Similar to cultural value, the number of phrases for educational value was less than 30 so that the network was constructed with fewer phrases than other CES. From the networks, we also found distinctive expressions between two languages in describing educational values. To

elaborate educational values, online reviews written in English commonly included a verb phrase beginning with '*learn about*,' such as '*learn about history*.' In addition, the phrase '*learn about*' played an important role in connecting varied expressions, resulting in the formation of a robust network around the phrase. However, the phrase '*learn about*' did not appear in the key phrases for Spanish reviews. Meanwhile, both languages appreciated to '*reconnect with nature*' or of '*pure nature*' as shown in the network with great degree centrality as well as large vertex sizes. The remaining edges, on the other hand, showed marginal degree centrality for both language networks, indicating a considerable skewness in key phrase frequencies.

- Identity value

The most common phrasal type of identity value was noun phrases containing adjective and noun vertices. The networks, however, revealed substantial differences in key phrases between English and Spanish reviews. The most popular phrase for English reviews was '*high altitude*' with great degree centrality and large vertex sizes, but the reviews in Spanish barely mentioned '*high altitude*' but preferred '*unique place*' the most. Both languages strongly praised the omnipresent glacial lakes in CNP, as shown in the great sizes of relevant vertices, '*lake*' or '*lagoon*' ('laguna' in Spanish), yet in-degree centrality for those vertices remained liminal. Such findings revealed that a large number of key phrases shared the lemma 'lake' or 'lagoon' in various adjective combinations. The network for English reviews exhibited more disconnected patterns than the network of Spanish reviews, with several fragmented bigram phrases (i.e., '*continental divide*', *'cold weather*', '*thin air*').

- Recreational value

The predominant usage of verb phrases was one of the most distinguishing features of recreational value. The verb vertices such as '*hike*' and '*walk*' had the largest vertices despite the absence of overpowering degree centrality associated with these vertices, demonstrating the predominance of verb phrases comprising '*hike*' and '*walk*'. Nonetheless, English reviews indicated a substantial preference for 'hike' activities such as '*hike around lake*' and '*day hike*,' but Spanish reviewers favored '*walk*.' In terms of the most common phrase (i.e., the edge with the greatest degree centrality), English reviews cited '*day trip*' the most, while Spanish reviews mentioned '*sport fishing*' the most, suggesting different interests for recreational activities. The overall network connectivity demonstrated that Spanish reviews included eclectic phrasal expressions to describe a variety of recreational activities, such as '*place to camp'*, '*spend night*', and '*bird watching*', whereas English reviews largely remained in describing hiking or walking activities.

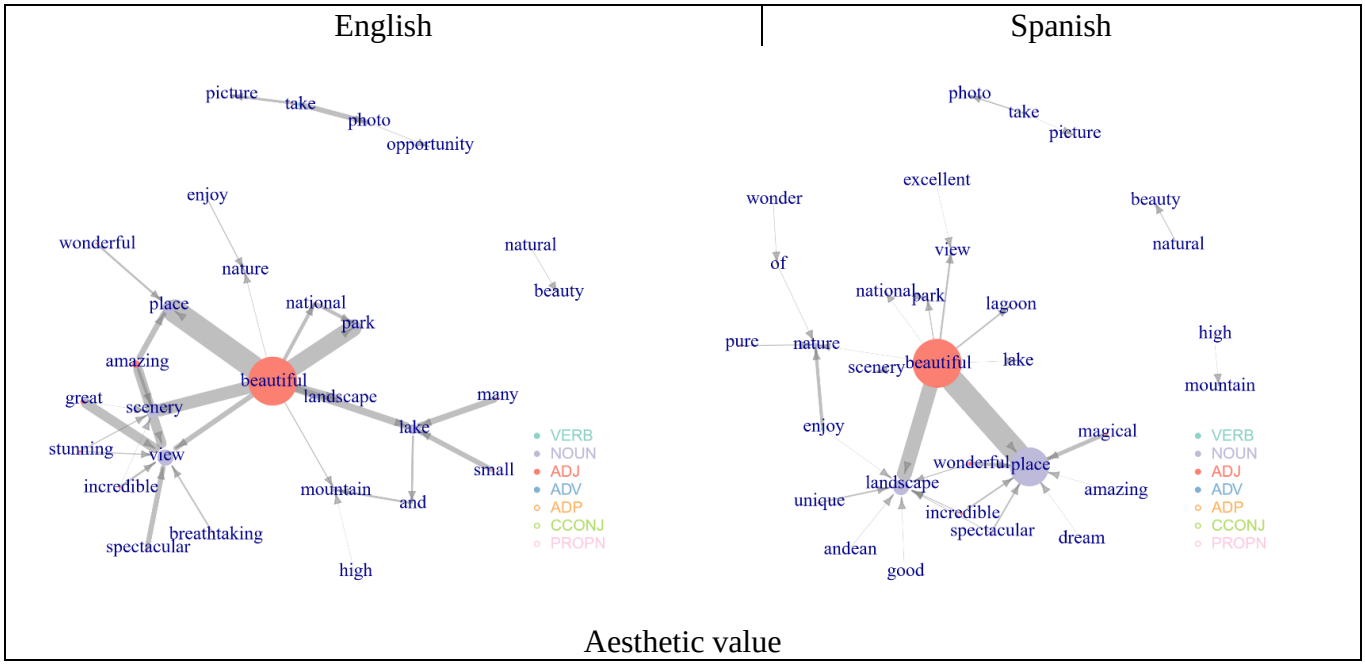
- Social value

The phrasal types for social value showed the dominance of noun phrases with a few verb phrases. Both languages shared a phrase '*have lunch*' with great degree centrality, while also sharing several phrases including '*friend*' and '*family*'. Specifically, English reviews mentioned doing a variety of activities '*with friend*' whereas Spanish reviews preferred to do '*with family*'. In addition, English reviews exhibited moderate to weak degree centrality in the network whereas Spanish reviews exhibited diverse phrasal expressions with moderate to great degree centrality across several phrases (i.e., '*spend with family*', '*family and friend*'). Such results could pertain to the similar frequency of numerous key phrases in Spanish reviews or the extremely pervasive usage of a few key phrases in English reviews. In terms of network connectivity, both languages

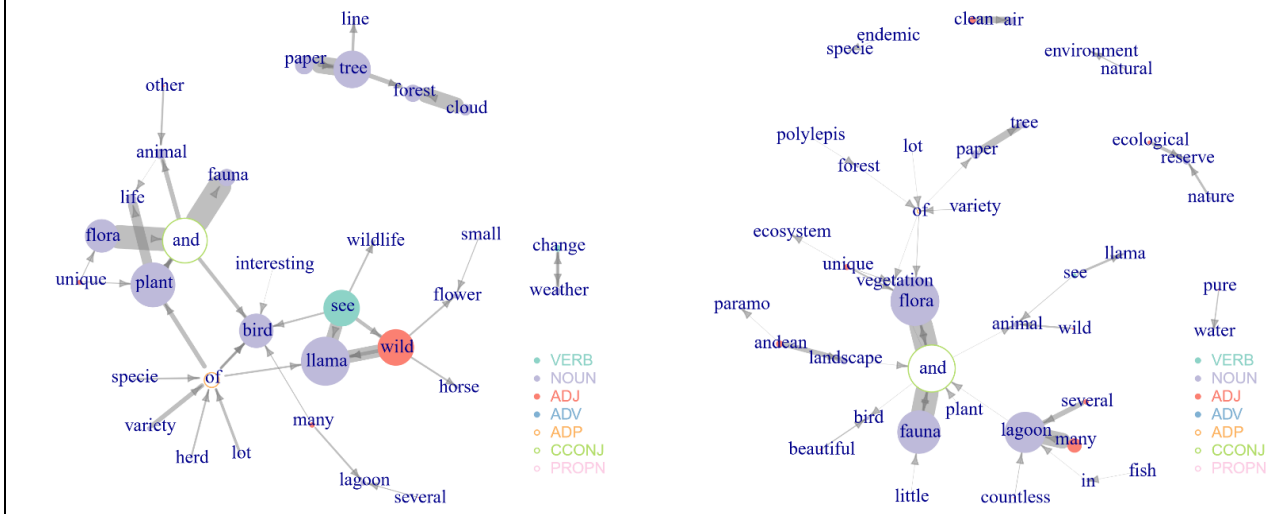
resulted in well-connected network clusters, suggesting significant overlaps of lemmas throughout the key phrases.

- Spiritual value

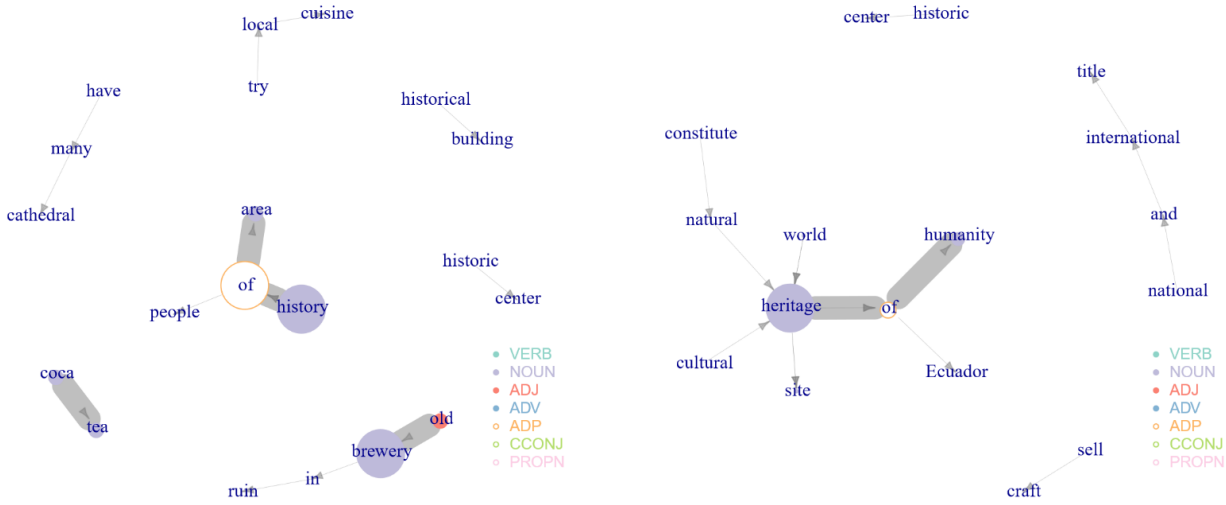
While both languages showed dominant noun phrases in the key phrases, the network analysis resulted in meaningful cross-linguistical distinctions. We found that two networks barely shared common phrases except for '*magical place*'. Specifically, English reviews elaborated '*place*', such as '*magical place*' and '*peaceful place*'. Spanish reviews also contained '*magical place*' as a key phrase, yet more emphasis was laid on describing activities '*with nature*', such as '*contact with nature*', '*connect with nature*', and '*get in touch with nature*.' English reviews commonly described to '*feel altitude*' while Spanish reviews rarely elaborated such experience. Overall, Spanish reviews exhibited a well-connected network than English reviews, featuring substantial overlaps of lemmas among the key phrases.



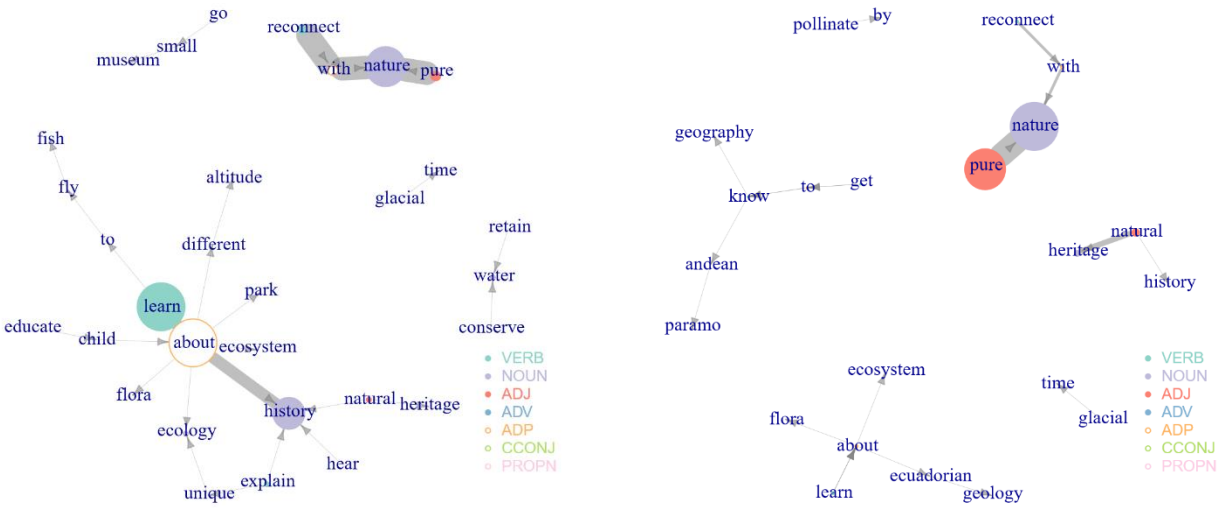
Aesthetic value



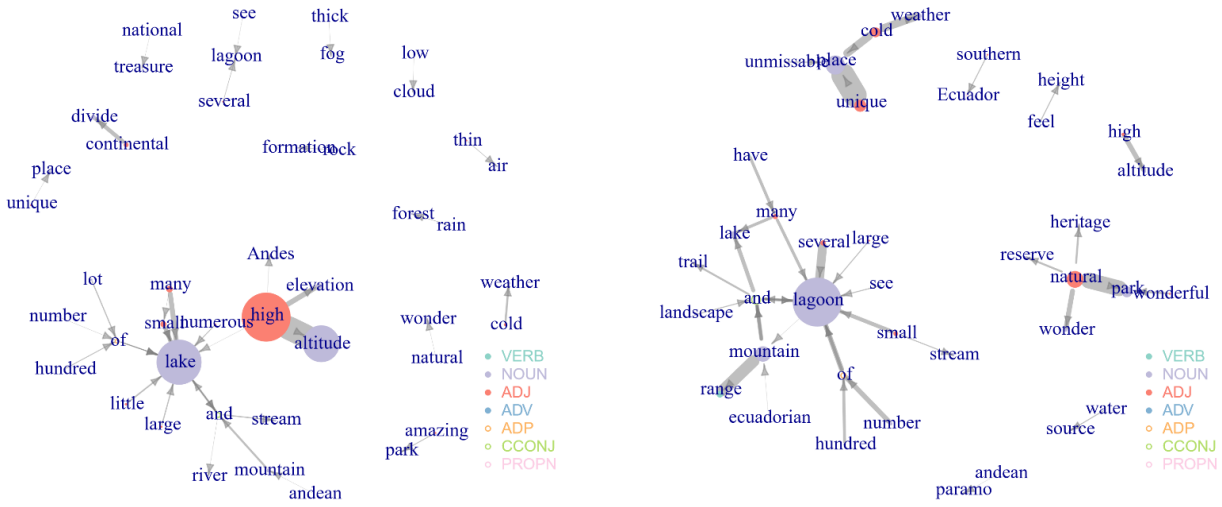
Biological value



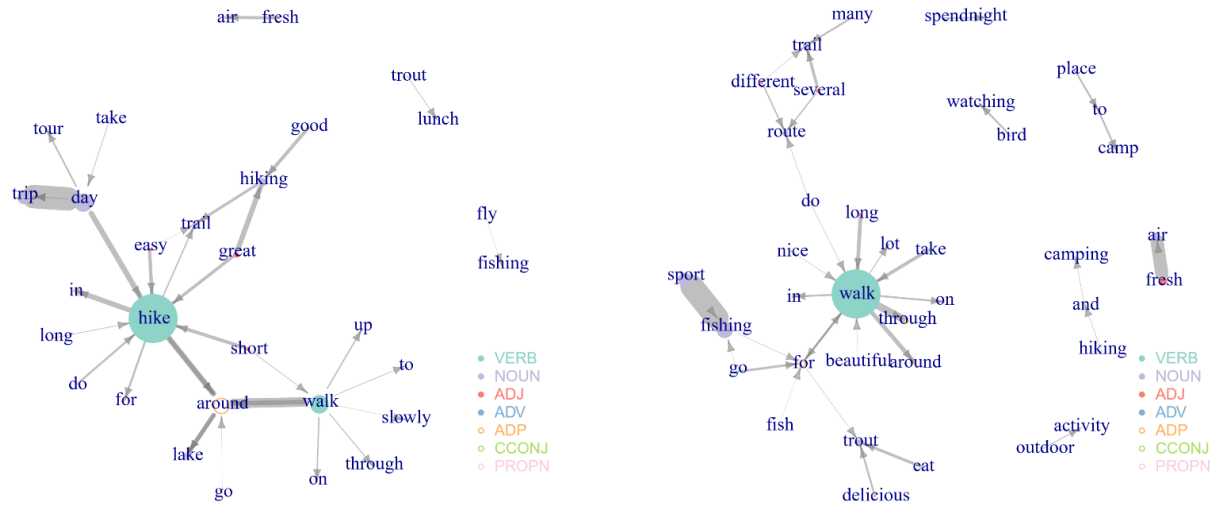
Cultural value



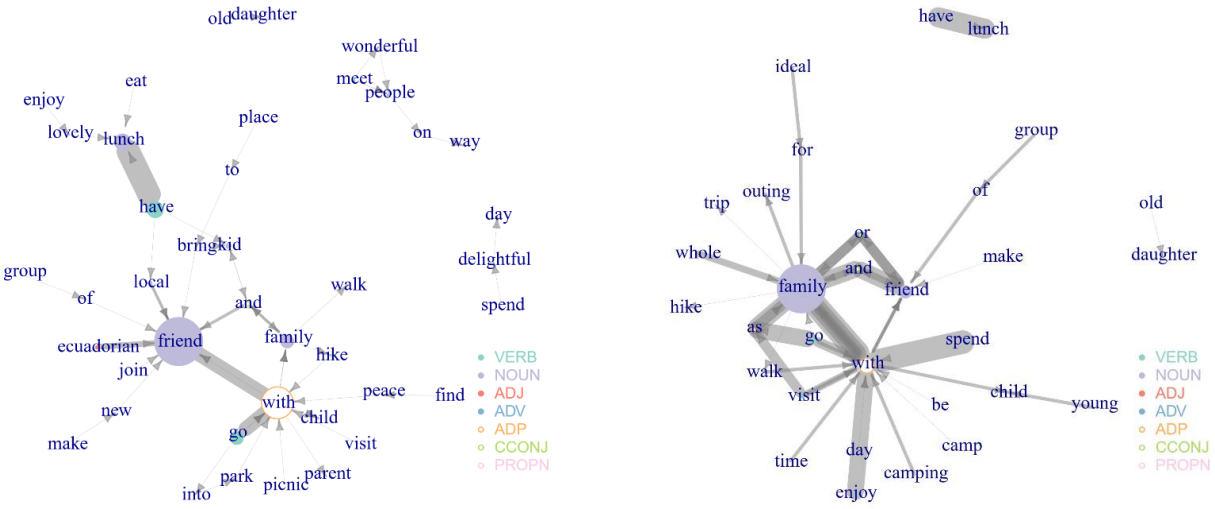
Educational value



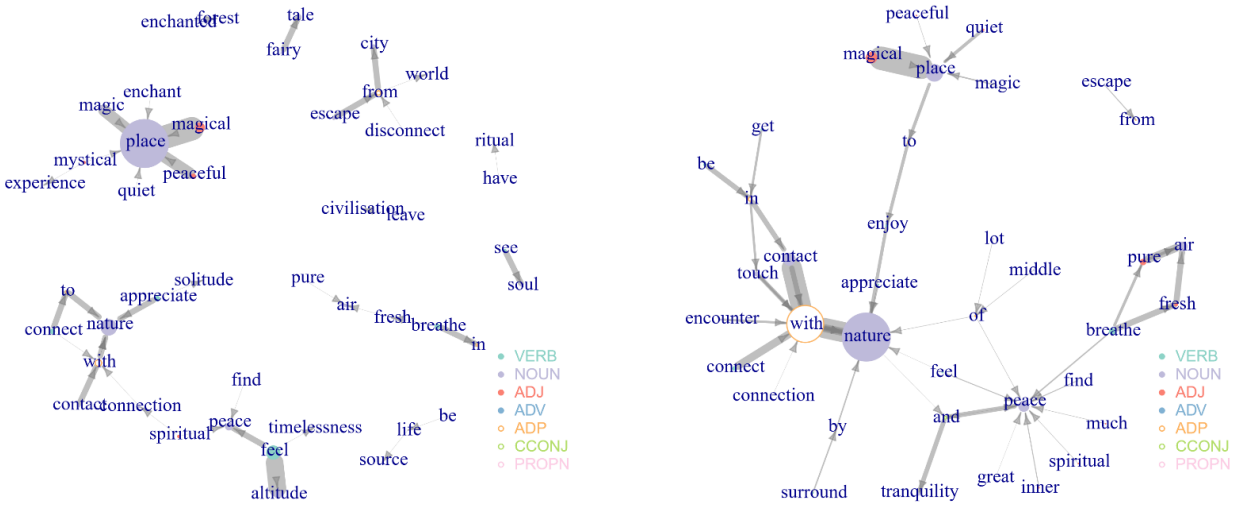
Identity value



Recreational value



Social value



Spiritual value

Figure 3-4. Networks of the 30 most frequent phrases in describing CES between English and Spanish reviews

DISCUSSION

Online reviews as a data source to identify CES

This study explored the usefulness of online reviews in analyzing CES. Aside from the benefit of a significant amount of user-generated content as one of the social media, online reviews can provide lengthy text contents with eloquent and articulate statements (Kladou and Mavragani, 2015; Taecharungroj and Mathayomchan, 2019). Therefore, phrasal expressions derived from online reviews provide high-context information that can improve the accuracy and reliability of CES identification. In addition, we discovered that TripAdvisor, which offers longer reviews than Google Maps, outperformed in bringing more relevant phrases, demonstrating the importance of the volume of the texts in CES analysis.

Yet, we noticed that online reviews resulted in the prominence of aesthetic and recreational values that had been commonly observed in previous studies using social media, such as geotagged

photographs (Egarter Vigl et al., 2021; Ghermandi et al., 2020). Although the dominance of aesthetic or recreational value can be accurate depending on the study site, the studies with qualitative research approaches in natural settings have revealed contradictory findings (Bieling, 2014; Pascua et al., 2017), questioning the bias derived from social media content. From an interview-based study at a natural reserve in Germany, Bieling (2014) observed that identity value was the most commonly mentioned CES, followed by recreational, aesthetic, and heritage values. Pascua et al. (2017) also addressed that aesthetic and recreational values were not the most important values among indigenous communities in Hawai'i. When we consider that early studies have taken the geolocations of social media contents as a proxy for aesthetic and recreational values (Langemeyer et al., 2018; Nahuelhual et al., 2013), the contested bias is not surprising. Online reviews are no exception, as users of online review platforms are primarily concerned with sharing tourism-related experiences, advice, and recommendations (Taecharungroj and Mathayomchan, 2019). Since Wartmann et al. (2018) confirmed the divergence of key findings of landscape values across travel blogs, Flickr tags, and free-listing surveys, it is highly encouraged to consider a variety of data sources to identify CES in the future.

Crowdsourced phrasal lexicon to identify CES

A single-word lexicon has been the most common method to assess CES in previous studies with social media (Chen et al., 2018; Ghermandi et al., 2020; Hale et al., 2019; Richards and Friess, 2015). In this study, we suggested a phrase-based lexicon approach, which involved phrasal expressions to annotate CES. Accordingly, we could distinguish aesthetic values based on the phrases like '*beautiful landscape*' and '*beautiful lake*', instead of '*beautiful*'. We further confirmed the efficacy of the crowdsourced survey to harness the lexicon by bringing public

consensus. The responses from the survey turned out to be satisfactory, allowing to keep post-processing work minimum. Overall, future research can take advantage of this lexicon to minimize time and resources in the analytic process. Also, we expect to improve the reproducibility of the lexicon by accumulating site-dependent CES-phrase pairs from the case studies. Considering the single-word lexicons in different environmental contexts, such as coastal (Ghermandi et al., 2020), riverain (Chen et al., 2018; Hale et al. 2019), and marine environments (Retka et al., 2019; Ruiz-Frau et al., 2020), can be also helpful to fill the contextual gap of the lexicon. Eventually, the lexicon can be applied to the spatial analysis by engaging with geotagged text content, such as Instagram posts or Twitter, allowing researchers to explore spatial human-environment dynamics of CES values.

Nonetheless, the lexicon has several limitations to address. During the screening process, we deleted thousands of crude phrases due to the lack of frequency, yet they may have contained site-specific CES. Indeed, longer phrases tend to include concrete expressions relating to CES since they have higher context information. However, the longer the phrases were, the less frequently they appeared in online reviews. For example, the phrases '*beautiful nature*' and '*enjoy nature*' were frequently found in online reviews, whereas '*learn about ecuadorian geology*' and '*see tiny hummingbird nest*' appeared only a few times. However, it was not possible to include all of such phrases due to the limited budget for the crowdsourced survey.

In addition, the lexicon in this study was heavily weighted on phrases associated with aesthetic and recreational values, which may have influenced the results of CES frequency across online reviews. While it seems inevitable when using social media content that is specialized in tourism-related expressions (Egarter Vigl et al., 2021; Ghermandi et al., 2020), the predominance of aesthetic and recreational values in CES studies can misleadingly legitimize the prioritization

of tourism-related values and undermine less prevalent but indispensable CES during the decision-making process in determining conservation priorities. Since several studies have emphasized the significance of acknowledging wide-ranging CES values to safeguard human wellbeing (Kosanic and Petzold, 2020), it is required to explore novel approaches as well as new data sources to complement current limitations. For instance, text-based big data whose key agendas are not exclusive to tourism, such as Twitter and Reddit (Fox et al., 2021), as well as hybrid methods, such as public participatory mapping (PPGIS), may fill the knowledge gaps in articulating understated but essential CES values.

Linguistic distinctions of CES

We examined linguistic distinctions between English and Spanish reviews in describing CES values, focusing on cross-linguistic comparisons of the frequent phrases. In cross-linguistic studies, language has been used as a proxy to define cultural cohorts that share mindset, values, belief systems, and preferences (Mathayomchan and Sripanidkulchai, 2019). Yet, related studies to examine how different language can influence on the perception on CES has been rarely discussed (Spalding and Parrett, 2019; Wartmann and Purves, 2018). In that sense, this study was the first to investigate such distinctions focusing on phrasal expressions.

The findings included several insights concerning the language and CES. First, English reviews tended to run longer than Spanish reviews in both TripAdvisor and Google Maps, suggesting the language-specific verbosity. Also, TripAdvisor reviews had lengthier texts compared to Google Maps, highlighting the platform-specific verbosity. Second, the lexicon revealed that English reviews had more phrases in biological and recreational values whereas Spanish reviews had more social and spiritual values. The number of phrases largely corresponded

with the result of the CES frequency. For instance, English reviews commented on recreational value as frequently as aesthetic value, yet Spanish reviews mentioned recreational values far less than English reviews while outperforming on social and spiritual values. The following analysis to analyze cross-linguistic distinctions of key phrases further revealed how different languages described CES values. While we did not find linguistic distinctions aesthetic value, the remaining CES exhibited meaningful differences. To summarize, English reviews prefer to '*hike*' (i.e., recreational value), '*with friend*' (i.e., social value), feeling '*high altitude*' (i.e., identity value), whereas Spanish reviews prefer to '*walk*' or do '*sport fishing*' (i.e., recreational value), '*with family'* (i.e., social value), thinking of CNP as a '*unique place*' (i.e., identity value). Such distinctions may reflect the role of language in influencing perceptions of CES as well as tourism interests, but further research is needed to clarify the consistency of such relationships across broad socioenvironmental settings.

CONCLUSION

Identifying CES is a primary step to safeguard CES in the landscape, yet the process has been lagged due to the lack of viable data. This study presents a novel approach to identify CES using online tourism reviews and a crowdsourced phrasal lexicon in El Cajas National Park, Ecuador. The findings in this study demonstrate the originality of the phrasal lexicon method and online reviews to provide high-context information, which can improve the accuracy and accountability of CES analysis. Also, we highlighted the linguistic distinctions in describing CES values between English and Spanish reviews, with network analysis from the key phrasal expressions. Overall, utilizing the online reviews along with the crowdsourced phrasal lexicon can

help leverage data-driven analysis for protected area management to implement CES-inclusive conservation strategies.

ACKNOWLEDGEMENT

The crowdsourced survey in this study was supported by the UGA Summer Research Grant endowed to the first author.

REFERENCES

Balvanera, P., Uriarte, M., Almeida-Leñero, L., Altesor, A., DeClerck, F., Gardner, T., Hall, J., Lara, A., Laterra, P., Peña-Claros, M., Silva Matos, D.M., Vogl, A.L., Romero-Duque, L.P., Arreola, L.F., Caro-Borrero, Á.P., Gallego, F., Jain, M., Little, C., de Oliveira Xavier, R., Paruelo, J.M., Peinado, J.E., Poorter, L., Ascarrunz, N., Correa, F., Cunha-Santino, M.B., Hernández-Sánchez, A.P., Vallejos, M., 2012. Ecosystem services research in Latin America: The state of the art. Ecosyst. Serv. 2, 56–70.

Barbro, P.A., Mudambi, S.M., Schuff, D., 2020. Do Country and Culture Influence Online Reviews? An Analysis of a Multinational Retailer's Country-Specific Sites. J. Int. Consum. Mark. 32, 1–14.

Benoit, K., Matsuo, A., 2020. spacyr: Wrapper to the 'spaCy' 'NLP' Library. R package version 1.2.1. https://CRAN.R-project.org/package=spacyr

Bieling, C., 2014. Cultural ecosystem services as revealed through short stories from residents of the Swabian Alb (Germany). Ecosyst. Serv. 8, 207–215.

Bremer, L.L., Auerbach, D.A., Goldstein, J.H., Vogl, A.L., Shemie, D., Kroeger, T., Nelson, J.L., Benítez, S.P., Calvache, A., Guimarães, J., Herron, C., Higgins, J., Klemz, C., León, J., Sebastián Lozano, J., Moreno, P.H., Nuñez, F., Veiga, F., Tiepolo, G., 2016. One size does not fit all: Natural infrastructure investments within the Latin American Water Funds Partnership. Ecosyst. Serv. 17, 217–236.

Brown, G., Reed, P., Raymond, C.M., 2020. Mapping place values: 10 lessons from two decades of public participation GIS empirical research. Appl. Geogr.

Bullock, C., Joyce, D., Collier, M., 2018. An exploration of the relationships between cultural ecosystem services, socio-cultural values and well-being. Ecosyst. Serv. 31, 142–152.

Calvet-Mir, L., Corbera, E., Martin, A., Fisher, J., Gross-Camp, N., 2015. Payments for ecosystem services in the tropics: A closer look at effectiveness and equity. Curr. Opin. Environ. Sustain. 14, 150–162.

Cebrián-Piqueras, M.A., Filyushkina, A., Johnson, D.N., Lo, V.B., López-Rodríguez, M.D., March, H., Oteros-Rozas, E., Peppler-Lisbach, C., Quintas-Soriano, C., Raymond, C.M., Ruiz-Mallén, I., van Riper, C.J., Zinngrebe, Y., Plieninger, T., 2020. Scientific and local ecological knowledge, shaping perceptions towards protected areas and related ecosystem services. Landsc. Ecol. 35, 2549–2567.

Chan, K.M.A., Satterfield, T., Goldstein, J., 2012. Rethinking ecosystem services to better address and navigate cultural values. Ecol. Econ. 74, 8–18.

Chan, K.M.A., Shaw, M.R., Cameron, D.R., Underwood, E.C., Daily, G.C., 2006. Conservation planning for ecosystem services. PLoS Biol. 4, 2138–2152.

Chen, Y., Caesemaecker, C., Rahman, H.T., Sherren, K., 2020. Comparing cultural ecosystem service delivery in dykelands and marshes using Instagram: A case of the Cornwallis (Jijuktu'kwejk) River, Nova Scotia, Canada. Ocean Coast. Manag. 193, 105254.

Chen, Y., Parkins, J.R., Sherren, K., 2018. Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. Landsc. Urban Plan. 170, 283–292.

Cong, L., Wu, B., Morrison, A.M., Shu, H., Wang, M., 2014. Analysis of wildlife tourism experiences with endangered species: An exploratory study of encounters with giant pandas in Chengdu, China. Tour. Manag. 40, 300–310.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research, InterJournal, Complex Systems 1695.

Daniel, T.C., Muhar, A., Arnberger, A., Aznar, O., Boyd, J.W., Chan, K.M.A., Costanza, R., Elmqvist, T., Flint, C.G., Gobster, P.H., Grêt-Regamey, A., Lave, R., Muhar, S., Penker, M., Ribe, R.G., Schauppenlehner, T., Sikor, T., Soloviy, I., Spierenburg, M., Taczanowska, K., Tam, J., Von Der Dunk, A., 2012. Contributions of cultural services to the ecosystem services agenda. Proc. Natl. Acad. Sci. U. S. A. 109, 8812–8819.

Donoso, M.E., Sarmiento, F.O., 2021. Changing mountain farmscapes: vulnerability and migration drivers in the Paute River watershed, Southern Ecuador. J. Mt. Sci. 18, 1902–1919.

Egarter Vigl, L., Marsoner, T., Giombini, V., Pecher, C., Simion, H., Stemle, E., Tasser, E., Depellegrin, D., 2021. Harnessing artificial intelligence technology and social media data to support Cultural Ecosystem Service assessments. People Nat. 1–13.

Englund, O., Berndes, G., Cederberg, C., 2017. How to analyse ecosystem services in landscapes—A systematic review. Ecol. Indic. 73, 492–504.

Farley, K.A., Bremer, L.L., 2017. "Water Is Life": Local Perceptions of Páramo Grasslands and Land Management Strategies Associated with Payment for Ecosystem Services. Ann. Am. Assoc. Geogr. 107, 371–381.

Fischer, A., Eastwood, A., 2016. Coproduction of ecosystem services as human-nature interactions-An analytical framework. Land use policy 52, 41–50.

Fox, N., Graham, L.J., Eigenbrod, F., Bullock, J.M., Parks, K.E., 2021. Reddit: A novel data source for cultural ecosystem service studies. Ecosyst. Serv. 50, 101331.

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

Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B., Ziv, G., 2019. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosyst. Serv. 38, 100958.

Grima, N., Singh, S.J., Smetschka, B., Ringhofer, L., 2016. Payment for Ecosystem Services (PES) in Latin America: Analysing the performance of 40 case studies. Ecosyst. Serv. 17, 24–32.

Haines-Young, R., Potschin, M., Raffaelli, D.G., Frid, C.L.J., 2010. In: Ecosystem Ecology: A New Synthesis. Cambridge University Press, Cambridge, pp. 110–139.

Hale, R.L., Cook, E.M., Beltrán, B.J., 2019. Cultural ecosystem services provided by rivers across diverse social-ecological landscapes: A social media analysis. Ecol. Indic. 107, 105580.

Harrison, J., 2020. RSelenium: R Bindings for 'Selenium WebDriver'. R package version 1.7.7. https://CRAN.R-project.org/package=RSelenium

Hernández-Morcillo, M., Plieninger, T., Bieling, C., 2013. An empirical review of cultural ecosystem service indicators. Ecol. Indic. 29, 434–444.

Hitchner, S., Schelhas, J., Brosius, J.P., Nibbelink, N., 2019. Thru-hiking the John Muir Trail as a modern pilgrimage: implications for natural resource management. J. Ecotourism 18, 82–99.

Hornik, K., Mair, P., Rauch, J., Geiger, W., Buchta, C., Feinerer, I., 2013. The textcat Package for n-Gram Based Text Categorization https://CRAN.R-project.org/package=cld2

Johnson, M.L., Campbell, L.K., Svendsen, E.S., McMillen, H.L., 2019. Mapping urban park cultural ecosystem services: A comparison of twitter and semi-structured interview methods. Sustain. 11.

Kladou, S., Mavragani, E., 2015. Assessing destination image: An online marketing approach and the case of TripAdvisor. J. Destin. Mark. Manag. 4, 187–193.

Kosanic, A., Petzold, J., 2020. A systematic review of cultural ecosystem services and human wellbeing. Ecosyst. Serv. 45, 101168.

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., 2001. The causes of land-use and land-cover change: moving beyond the myths. GLOBAL ENVIRONMENTAL CHANGE-HUMAN AND POLICY DIMENSIONS. Glob. Environ. Chang. 11, 261–269.

Langemeyer, J., Calcagni, F., Baró, F., 2018. Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. Land use policy 77, 542–552.

Leung, Y., Spenceley, A., Hvenegaard, G., Buckley, R., Groves, C., 2018. Tourism and visitor management in protected areas : guidelines for sustainability.

Mathayomchan, B., Sripanidkulchai, K., 2019. Utilizing Google Translated Reviews from Google Maps in Sentiment Analysis for Phuket Tourist Attractions. JCSSE 2019 - 16th Int. Jt. Conf. Comput. Sci. Softw. Eng. Knowl. Evol. Towar. Singul. Man-Machine Intell. 260–265.

Milcu, A.I., Hanspach, J., Abson, D., Fischer, J., 2013. Cultural ecosystem services: A literature review and prospects for future research. Ecol. Soc. 18, 44–88.

Millennial Ecosystem Assessment (MEA), 2005. Ecosystems and Human Well-being: Synthesis, Washington D.C.: Island Press.

Ministerio del Ambiente. 2019. National system of protected areas: Reportes Históricos de Visitas. http://areasprotegidas.ambiente.gob.ec/en/reporte-de-visitas

Nahuelhual, L., Carmona, A., Lozada, P., Jaramillo, A., Aguayo, M., 2013. Mapping recreation and ecotourism as a cultural ecosystem service: An application at the local level in Southern Chile. Appl. Geogr. 40, 71–82.

Ooms, J., 2020. cld2: Google's Compact Language Detector 2. R package version 1.2.1.

Paolacci, G., Chandler, J., 2014. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. Curr. Dir. Psychol. Sci. 23, 184–188.

Pascua, P., McMillen, H., Ticktin, T., Vaughan, M., Winter, K.B., 2017. Beyond services: A process and framework to incorporate cultural, genealogical, place-based, and indigenous relationships in ecosystem service assessments. Ecosyst. Serv. 26, 465–475.

Retka, J., Jepson, P., Ladle, R.J., Malhado, A.C.M., Vieira, F.A.S., Normande, I.C., Souza, C.N., Bragagnolo, C., Correia, R.A., 2019. Assessing cultural ecosystem services of a large marine protected area through social media photographs. Ocean Coast. Manag. 176, 40–48.

Richards, D.R., Friess, D.A., 2015. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. Ecol. Indic. 53, 187–195.

Roll, U., Correia, R.A., Berger-Tal, O., 2018. Using machine learning to disentangle homonyms in large text corpora. Conserv. Biol. 32, 716–724.

Ross, C., Fildes, S., Millington, A.C., 2017. Land-use and land-cover change in the páramo of South-Central Ecuador, 1979-2014. Land 6, 46.

Rossi, S.D., Barros, A., Walden-Schreiner, C., Pickering, C., 2020. Using social media images to assess ecosystem services in a remote protected area in the Argentinean Andes. Ambio 49, 1146–1160.

Rudra, K., Goyal, P., Ganguly, N., Mitra, P., Imran, M., 2018. Identifying sub-events and summarizing disaster-related information from microblogs. 41st Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR 2018 265–274.

Ruiz-Frau, A., Ospina-Alvarez, A., Villasante, S., Pita, P., Maya-Jariego, I., de Juan, S., 2020. Using graph theory and social media data to assess cultural ecosystem services in coastal areas: Method development and application. Ecosyst. Serv. 45, 101176.

Sarmiento, F.O., 2002. Anthropogenic change in the landscapes of highland Ecuador. Geogr. Rev. 92, 213–234.

Sarmiento, F.O., 2016. Neotropical Mountains Beyond Water Supply: Environmental Services as a Trifecta of Sustainable Mountain Development, in: Developments in Earth Surface Processes. pp. 309–324.

Spalding, M., Parrett, C.L., 2019. Global patterns in mangrove recreation and tourism. Mar. Policy 110, 103540.

Stålhammar, S., Pedersen, E., 2017. Recreational cultural ecosystem services: How do people describe the value? Ecosyst. Serv. 26, 1–9.

Taecharungroj, V., Mathayomchan, B., 2019. Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tour. Manag. 75, 550–568.

TEEB, 2010. The Economics of Ecosystems and Biodiversity: ecological and economic foundation. Earthscan, Cambridge.

Teles da Mota, V., Pickering, C., 2020. Using social media to assess nature-based tourism: Current research and future trends. J. Outdoor Recreat. Tour. 30, 100295.

Van Colen, W.R., Mosquera, P., Vanderstukken, M., Goiris, K., Carrasco, M.C., Decaestecker, E., Alonso, M., León-Tamariz, F., Muylaert, K., 2017. Limnology and trophic status of glacial lakes in the tropical Andes (Cajas National Park, Ecuador). Freshw. Biol. 62, 458–473.

Wartmann, F.M., Acheson, E., Purves, R.S., 2018. Describing and comparing landscapes using tags, texts, and free lists: an interdisciplinary approach. Int. J. Geogr. Inf. Sci. 32, 1572–1592.

Wartmann, F.M., Purves, R.S., 2018. Investigating sense of place as a cultural ecosystem service in different landscapes through the lens of language. Landsc. Urban Plan. 175, 169–183.

Wickham, H., 2020. rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.6. https://CRAN.R-project.org/package=rvest

Wijffels, J., 2019. udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit. R package version 0.8.3. https://CRAN.R-project.org/package=udpipe

Worboys, G. L., Lockwood, M., Kothari, A., Feary, S., and Pulsford, I. (eds), 2015. Protected areas governance and management. ANU Press, Canberra.

CHAPTER 4

CROWDSOURCED ANALYSIS OF CULTURAL ECOSYSTEM SERVICES ACROSS THE U.S. NATIONAL PARKS[3]

[3] Kong, I., Mu, L., & Sarmiento, F.O. To be submitted to *The Professional Geographers*, December 2021.

ABSTRACT

Understanding site-specific cultural ecosystem services (CES) values is crucial for comprehensive protected area management that considers both biophysical and human dimension of landscape values. Yet, few research have conducted cross-comparisons of CES values across different landscapes. Using online tourism reviews from TripAdvisor, we develop and apply a crowdsourced phrasal lexicon to identify eight CES throughout 48 U.S. National Parks. The national parks are then classified into four groups using a Self-Organizing Map (SOM) based on their predominant CES values. Key phrasal expressions of national parks that represent four SOM groups are shown in text networks. Furthermore, the geographic locations of national parks are mapped to examine geospatial bias based on the four groups. Overall, the findings assert the significance of site-specific management strategies due to the variety of CES values among national parks despite their same managerial system and designation rationales.

INTRODUCTION

National park is a type of protected areas for the protection of ecological processes and cultural heritage while providing tourism and recreational opportunities (Dudley, 2008). During the visit to national parks, people perceive a variety of cultural ecosystem services (CES) that indicate intangible benefits people perceive from human-nature interactions (Chan et al., 2012; MEA, 2005). Visitors appreciate natural beauty and enjoy recreational opportunities, such as hiking and camping. National parks with diverse wildlife and cultural heritage can offer biological, cultural, and educational values to the visitors (Mallarach, 2008).

Understanding a variety of CES values is therefore important not only for visitor and destination management but also for the preservation of multidimensional human values that

consist of place integrity and human wellbeing (Bieling et al., 2014; Bullock et al., 2018; Willis, 2015). Nonetheless, integrating CES into the conservation policy of national park management is in a nascent stage, largely ascribed to the lack of relevant data to reveal immaterial and intangible CES values (Chan et al., 2012; Milcu et al., 2013). Numerous studies have operated visitor surveys or interviews to identify CES values (Bieling, 2014; Cheng et al., 2019), but qualitative approaches are resource-intensive and often limited to small spatial scopes and a small percentage of populations.

Recent application of user-generated content (UGC) in understanding CES has been acclaimed to fill the data gap with the massive size of the content generated from a large size of individuals (Richards and Friess, 2015; Zhang et al., 2020). Moreover, the creation of UGC is associated with the memorable experience and perceived enjoyment (Oliveira et al., 2020). Therefore, the geolocations of UGC have been considered as a proxy for the spatial distribution of visitor preference for aesthetic and recreational values in natural landscapes (Sinclair et al., 2019; Mancini et al., 2018; Yoshimura and Hiura, 2017). Content analysis has been also applied to distinguish multifaceted CES using texts and photo tags from UGC (Hale et al., 2019; Richards and Friess, 2015). The common approach in the content analysis involved the development of a lexicon that plays as a dictionary to match CES values to the relevant texts, and such lexicons consisted of various CES, such as aesthetic, recreational, social, historic, spiritual, and educational values (Chen et al. 2018; Hale et al., 2019; Richards and Friess, 2015). Richards and Friess (2015) classified eight CES using Flickr tags in mangrove forests and Hale et al. (2019) analyzed 11 CES from Flickr tags in a riverain landscape in the U.S. Chen et al. (2018) utilized Instagram posts to analyze CES values in Canada.

Despite the advances, content analysis for CES values has remained in single-word lexicons that may involve confusions and ambiguity with little contextual information. This is largely ascribed to the limited text content available from photograph-oriented social media such as Flickr and Instagram. Additionally, the lexicons had been developed by a small number of experts, which may generate significant cognitive disparity between the content generator (i.e., the public) and the content interpreter (i.e., researchers). Furthermore, there has been a lack of analysis to operate cross-comparisons of CES, which can help distinguish unique and site-specific values for park management (Dai et al., 2019). For instance, national parks are designated for their humanistic distinctions for natural and cultural values and have been managed under the same categorical system, not all national parks provide identical CES.

In this study, we aim to utilize text-rich online tourism reviews to analyze CES values and compare them across multiple national parks. First, we focus on developing a lexicon that contains higher context phrasal expressions. The lexicon aims to aggregate the public consensus for the annotation of CES to alleviate potential expert bias. As a result, we apply the lexicon to identify CES values for national parks. For the post-analysis, we categorize national parks depending on the prevalent CES values and examine dominant expressions from key groups.

METHODS

Data collection

We obtained online tourism reviews (OTRs) from TripAdvisor to be one of the most popular online platforms to share tourism experiences (Taecharungroj and Mathayomchan, 2019; Xiang et al. 2017). Of 62 national parks within the U.S. National Park System, we concentrated on national parks situated inside and near to the contiguous United States to reduce deviations in

national park experiences. Gateway Arch National Park and New River Gorge National Park were also omitted due to their urban setting and a small number of reviews (i.e., less than 100). Finally, TripAdvisor considers Sequoia National Park to be a subset of Kings Canyon National Park, therefore the two parks have been integrated. As a result, 48 national parks were chosen for the analysis.

To scrap TripAdvisor reviews, we first identified the URLs for the front pages of 48 national parks (e.g., "Yosemite National Park"). Then we wrote R scripts to obtain the directories of subordinate attractions (e.g., "Half Dome"), up to 30 for each national park. It may have less than 30 attractions depending on the national parks, even a single attraction for a national park. Then the R scripts navigated to each attraction and scraped all available OTRs, including the title, full-text review, and the Date of Experience, until the last page. The data collection was undertaken using the *rvest* for HTML web scraping (Wickham, 2020) and *RSelenium* for dynamic web browsing (e.g., clicking the next web page number) (Harrison, 2020). After the data scraping, we kept the OTRs whose Date of Experience is before March 2020 to avoid any anomalous experiences caused by managerial regulations since the breakout of COVID-19. As a result, 294,486 reviews were obtained for the study (Appendix B).

Retrieving CES values from texts

The unstructured texts from TripAdvisor reviews were cleaned with Natural Language Process (NLP) that includes tokenization, lowercasing, lemmatization, and the part-of-speech (POS) tagging. Named Entity Recognition was also operated to detect place names and exclude them from further analysis. Selective determinators (i.e., a, an, the) were also removed. The entire NLP was applied with *spacyr* (Benoit and Matsuo, 2020).

Subsequently, we developed a lexicon, as a dictionary to annotate CES to the texts. Instead of single words that have been widely applied in previous CES lexicons, we focused on phrasal expressions that can deliver higher context information, alleviating confusions or ambiguity in matching the CES values. Taking the POS pattern library from Kong & Sarmiento (under review)(Appendix A), we extracted noun phrases (i.e., phrases that embellish a noun) and verb phrases (i.e., phrases that begin with a verb) using the *udpipe* package (Wijffels, 2019). After the retrieval of thousands of crude phrases, we skimmed and filtered viable phrases to contain CES values. In the meantime, we defined eight CES categories – aesthetic, biological, cultural, educational, identity, recreational, social, and spiritual values – considering the national park experiences as well as the ecosystem service frameworks such as MEA (2005).

Then we operated a crowdsourced survey to match the phrases with relevant CES. Compared to previous lexicons that had been developed by a handful of researchers, a crowdsourced method can gather public perception of CES in association with phrasal expressions. The survey was operated on Amazon Mechanical Turk and the survey obtained the IRB exemption. At the beginning of the survey, we put a brief definition of eight CES along with exemplary phrases (Table 4-1). The detailed procedure of the survey can be found in Appendix C. Upon the completion of the survey, we tallied the response and chose the most prominent CES for the phrase. When the survey results showed a comparable number of different CES values for one phrase, multiple CES values were allowed. As a result, we created a crowdsourced phrasal lexicon for eight CES.

Table 4-1. The definitions and exemplary phrases for eight CES

CES	Definition	Exemplary phrases
Aesthetic value	People value the beauty of national parks.	beautiful waterfall, take picture of majestic scenery
Biological value	People value various plants and animals in national parks.	spot wolf, encounter black bear, exotic plant
Cultural value	People value cultural heritage preserved in national parks.	see petroglyph, enjoy the historic tour, ancient rock art
Educational value	People value the educational opportunities to learn something in national parks.	learn a lot about geology, join junior ranger program, educational display
Identity value	People value the unique characteristics of national parks.	national treasure, unique rock formation
Recreational value	People value the recreational activities in national parks.	hike up mountain, take boat across lake, stargazing sky
Social value	People value social interactions in national parks.	family vacation, make friend, hike with friend
Spiritual value	People value spiritual or therapeutic experiences in national parks.	enjoy solitude, soothe your soul

We identified CES values from TripAdvisor reviews using the crowdsourced phrasal lexicon. When a lexical phrase matched one in the review, we assigned the relevant CES values to the review using the *udpipe* package in R (Wijffels, 2019). Although a review can contain many phrases that correspond to one CES value, we simplified the outcome to a binary value based on whether the review mentions certain CES. Therefore, a single review can contain up to eight CES. As a consequence, we aggregated the frequency of eight CES values at each national park level and calculated a correlation across 48 national parks.

Clustering national parks depending on prevalent CES values

We ran a clustering analysis to detect prevalent compositions of CES values. For the clustering method, we applied a self-organizing map (SOM), an unsupervised clustering method based on a neural network approach by Kohonen (1982) (Wehrens and Kruisselbrink, 2018). Compared to other clustering methods, SOM excels in converting complex multi-dimensional data into low-dimensional 'map units' (i.e., clusters in other methods) with a powerful visualization ability (Wehrens and Kruisselbrink, 2018). In addition, SOM does not require *a-priori* knowledge for the input data, enabling it to be a data-driven clustering. Given n-dimensional data (i.e., an n-variable data matrix), SOM calculates the nearest neighbor distance between the objects and identifies the best matching unit. Then, over the unsupervised iterations, it modifies distance weights to reorganize the objects and aggregates the nearby objects to be clustered. In this study, we prepared a data matrix consisting of eight CES frequencies as variables and 48 national parks as objects. Using the *kohonen* package in R (Wehrens and Kruisselbrink, 2018), we trained the data to classify the national parks into four map units.

Analyzing national parks based on key CES clusters

The map units were then assigned to the corresponding national parks to operate two comparative analyses. First, the top 30 most common phrases from each national park were selected to generate a text cloud. The size of the phrase indicated the frequency and the color of the phrase indicated CES values. The same phrase may appear more than once in different colors when the phrase was tagged with multiple CES values. Therefore, we can validate SOM results with actual phrases from TripAdvisor reviews. The word clouds were visualized with the R package *wordcloud* (Fellows, 2018). Second, we examined the spatial distribution of national

parks, categorized into four map units. Based on the locations, we generated ellipses depicting the directional distribution of national parks with the range of one standard deviation (~68 percent). The spatial analysis was operated in ESRI ArcGIS Pro 2.8.0.

RESULTS

Frequency of CES values

The crowdsourced phrasal lexicon contained more than five thousand phrasal expressions. The entire lexicon can be found at *https://github.com/ihKong/phrasal_lexicon_CES*. The aggregated frequency of eight CES values across 48 U.S. National Parks was listed in Appendix D and summarized in Table 4-2. The two most frequent CES values from 48 national parks were recreational and aesthetic values, to be mentioned in 54.5 percent and 41.0 percent of TripAdvisor reviews on average. Still, the frequency of CES values varied depending on national parks. Aesthetic value was found in 63.3 percent from North Cascades NP and 14.7 percent from Mammoth Cave NP. Recreational value also showed a wide-ranging disparity, with 73.0 percent in Pinnacles NP and 33.6 percent in Yellowstone NP. Biological value displayed the greatest standard deviation (i.e., 11.5), explaining the wide disparity between Everglades NP (53.0 percent) and Canyonlands NP and Hot Springs NP (2.3 percent). Cultural, educational, and spiritual values were less prevalent throughout national parks, rating 3 to 5 percent on average.

Table 4-2. The frequency of CES values across 48 U.S. National Parks

CES	Min (%)	Mean (%) [sd]	Max (%)
Aesthetic value	14.7 (Mammoth Cave)	41.0 [11.9]	63.3 (North Cascades)
Biological value	2.3 (Canyonlands, Hot Springs)	15.4 [11.5]	53.0 (Everglades)
Cultural value	0.3 (North Cascades)	3.6 [4.0]	21.9 (Mesa Verde)
Educational value	1.3 (Acadia)	4.8 [2.9]	14.0 (Mammoth Cave)
Identity value	2.8 (Everglades)	13.1 [6.8]	42.2 (White Sands)
Recreational value	33.6 (Yellowstone)	54.5 [9.2]	73.0 (Pinnacles)
Social value	2.0 (Death Valley)	6.5 [2.8]	13.8 (Mammoth Cave)
Spiritual value	1.0 (Biscayne)	3.1 [1.4]	8.1 (Redwood)

We determined the correlations between different CES values by calculating the Pearson correlation coefficient across 48 national parks (Table 4-3). Aesthetic value was negative correlated with most CES values, showing statistically significant negative relationships with cultural, educational, and social values (p = -0.39**, p = -0.42**, p = -0.50***). Cultural value was positively correlated with educational value (p = 0.34*), and social value was positively correlated with identity (p = 0.36**) and recreational value (p = 0.39**). Biological value was positively correlated with spiritual value (p = 0.50***).

Table 4-3. Correlation matrix of eight CES values

	Aesthetic	Biological	Cultural	Educational	Identity	Recreational	Social	Spiritual
Aesthetic	1.00	-0.08	-0.39**	-0.42**	-0.09*	-0.11	-0.50***	0.07***
Biological		1.00	-0.26	0.23**	-0.31	-0.01	-0.11	**0.50**
Cultural			1.00	**0.34***	-0.05	-0.07	0.20	-0.18
Educational				1.00	-0.11	0.01	0.25**	0.14
Identity					1.00	0.10	**0.36**	-0.01
Recreational						1.00	**0.39****	0.08
Social							1.00	0.03
Spiritual								1.00

(n = 48; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; statistically significant and positive correlations are bolded)

Classification of national parks depending on dominant CES values

The SOM clustering resulted in four map units. The compositions of eight CES values for each map unit were illustrated in a normalized value range (Figure 4-1). Based on the prevalent CES values, the map units were labeled as follows: (a) Aesthetic, (b) Biological/Spiritual, (c) Recreational/Identity, and (d) Cultural/Educational/Social.

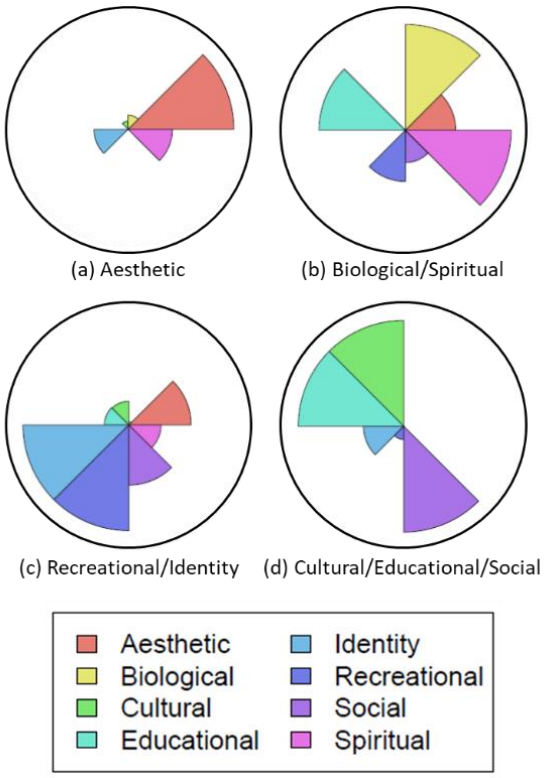


Figure 4-1. Four map units from SOM clustering

Key phrase analysis

The word cloud maps illustrated the frequent phrases from four national parks that represent each of four map units (Figure 4-2). For (a) Aesthetic category, North Cascades NP depicted a dominant share of phrases that correspond to aesthetic values. For (b) Biological/Spiritual category, Everglades NP showed that most phrases correspond to biological values. For (c) Recreational/Identity category, Pinnacles NP was shown to represent prevalent recreation-related phrases. For (d) Cultural/Educational/Social category, Mammoth Cave NP was chosen to show the phrasal combination of cultural, educational, and social values.

Figure 4-2. Word clouds of frequent phrasal expressions from national parks that represent four SOM map units

Spatial distribution analysis

Geographic locations of national parks depending on their SOM map units were highlighted with spatial ellipses of one standard deviation (i.e., 68 percent probability) (Figure 4-3). All ellipses for four SOM map units were tilted toward the western U.S. due to the region's

dominance of national parks, but the elliptical boundaries revealed regional biases. Aesthetic national parks were distributed along the Pacific West and Intermountain regions, including numerous popular national parks such as Yellowstone NP, Yosemite NP, Grand Canyon NP, and Great Smoky Mountains NP. Recreational/Identity national parks overlapped with the regional boundaries of Aesthetic national parks but stretched out to the southern part of these areas, including Pinnacles NP, Joshua Tree NP, White Sands NP, and Big Bend NP. Biological/Spiritual national parks were spotted at the coastal regions of the U.S., such as Channel Islands NP and Everglades NP. Meanwhile, Cultural/Educational/Social national parks were discovered in the inland of the U.S. continent, including Mammoth Cave NP, Mesa Verde NP, and Wind Cave NP.

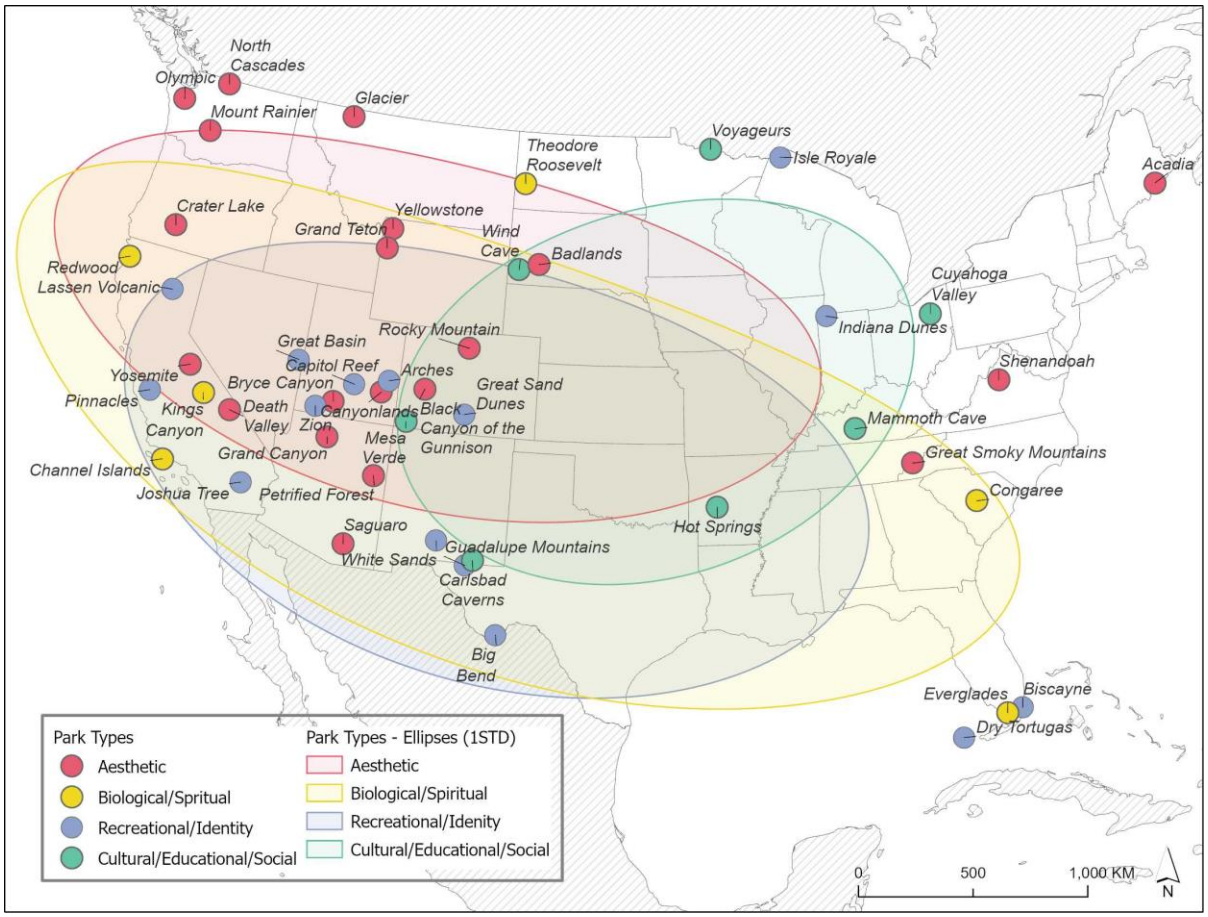


Figure 4-3. Spatial distribution of national parks associated with four SOM map units and probabilistic ellipses

DISCUSSION

Understanding human-nature interactions and the subsequent perception of various CES is imperative to achieve comprehensive conservation planning in protected areas. While the efforts for CES-oriented landscape management have been hampered by the lack of relevant public data as well as the labor-intensive qualitative data collection process, the advent of user-generated content has been acclaimed to alleviate the obstacles for data accessibility with large size of online content.

The findings revealed the diversity of CES values across the U.S. National Parks, supporting the diversity of national park values. Still, the most dominant CES values were recreational and aesthetic values, similar to the results from other studies (Sultana and Selim, 2021; Milcu et al., 2013). Furthermore, four types of prevalent CES types – aesthetic, recreational, cultural, and biological values – resonate with the mission of the national park to support public recreation while preserving biological and cultural distinctions (NPS, 2000).

The correlation matrix that explained the co-occurrence of CES values across 48 national parks aligned with the composition of four map units from SOM analysis. The negative correlations between aesthetic value and other CES values corresponded with the SOM plot of (a) Aesthetic type showing the substantial size of the pie for aesthetic value. In addition, strong correlations between biological and spiritual value, and cultural and educational value were supported by the SOM plots of (b) Biological/Spiritual and (d) Cultural/Educational/Social types. However, recreational and identity values did not show a strong correlation but were classified together in (c) Recreational/Identity type. Such a finding can be explained by a high correlation between social and recreational values, as well as identity values. Overall, the finding in this study largely resonates with common sense, but we found that the outcome may be exclusive to national

park experiences in the U.S. since similar studies to compare CES values in urban parks in China (Dai et al., 2019) and Bangladesh (Sultana and Selim, 2021) resulted in different correlations across CES values. Dai et al. (2019), for example, verified the positive correlations between aesthetic, recreational, and spiritual values. Sultana and Selim (2019) identified positive relationships between natural awareness, cultural heritage, and religious and spiritual values, as well as negative correlations between natural awareness and recreational value, and aesthetic value and spiritual value. Such differences can be driven by the source data (Chinese tourism check-in websites in Dai et al. (2019) and in-person survey in Sultana and Selim (2019)) as well as the typologies of CES values. As a result, it is necessary to take a cautious approach when interpreting and comparing CES results across different study sites. Furthermore, additional studies are required to have full knowledge of the correlations between CES values.

Spatial locations of 48 National Parks divided by four SOM map units further demonstrated the relationship between geophysical attributes and the perception of CES values. While aesthetic national parks were spotted in the mid- to the northern region of Pacific West and Intermountain regions, recreational national parks were found in the mid- to the southern part of Pacific West and Intermountain regions. National parks with cultural and educational distinctions were mostly located in the inland, while national parks with biological and spiritual values were found along the coastlines and lowland areas. Such findings confirmed a range of place experiences and CES values in national parks, although they are classified and managed in the same institutional and managerial system.

CONCLUSION

Identifying CES values is critical for developing a holistic conservation strategy that considers landscape values that people perceive in protected areas. We used extensive text analysis from TripAdvisor reviews to uncover CES values and compare them across 48 U.S. National Parks. The studies found a difference in CES values amongst national parks, not only the frequency of various CES values but key phrasal usage and spatial distribution depending on the prevalent CES values. Overall, identifying site-specific CES values may help managers develop management methods for customizing visitor experiences while also making further efforts to protect such values. It may also help with tourism destination marketing and management by leveraging key visitor niches that differentiate the park from rival destinations.

REFERENCES

Benoit, K., & Matsuo, A. (2020). spacyr: Wrapper to the 'spaCy' 'NLP' Library. R package version 1.2.1. https://CRAN.R-project.org/package=spacyr

Bieling, C. (2014). Cultural ecosystem services as revealed through short stories from residents of the Swabian Alb (Germany). *Ecosystem Services*, *8*, 207–215.

Bieling, C., Plieninger, T., Pirker, H., & Vogl, C. R. (2014). Linkages between landscapes and human well-being: An empirical exploration with short interviews. *Ecological Economics*, *105*, 19–30.

Bullock, C., Joyce, D., & Collier, M. (2018). An exploration of the relationships between cultural ecosystem services, socio-cultural values and well-being. *Ecosystem Services*, *31*, 142–152.

Chan, K. M. A., Satterfield, T., & Goldstein, J. (2012). Rethinking ecosystem services to better address and navigate cultural values. *Ecological Economics*, *74*, 8–18.

Chen, Y., Parkins, J. R., & Sherren, K. (2018). Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. *Landscape and Urban Planning*, *170*(December 2016), 283–292.

Cheng, X., Van Damme, S., Li, L., & Uyttenhove, P. (2019). Evaluation of cultural ecosystem services: A review of methods. *Ecosystem Services*, *37*, 100925.

Dai, P., Zhang, S., Chen, Z., Gong, Y., & Hou, H. (2019). Perceptions of cultural ecosystem services in urban parks based on social network data. *Sustainability (Switzerland)*, *11*(19).

Dudley, N. (2008). Guidelines for Applying Protected Area Management Categories. Gland, Switzerland: IUCN.

Fellows, I. (2018). wordcloud: Word Clouds. R package version 2.6. https://CRAN.R-project.org/package=wordcloud

Hale, R. L., Cook, E. M., & Beltrán, B. J. (2019). Cultural ecosystem services provided by rivers across diverse social-ecological landscapes: A social media analysis. *Ecological Indicators*, *107*(July), 105580.

Harrison, J. (2020). RSelenium: R Bindings for 'Selenium WebDriver'. R package version 1.7.7. https://CRAN.R-project.org/package=RSelenium

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69.

Mallarach, J.-M. (2008). Protected Landscapes and Cultural and Spiritual Values. Volume 2 in the series Values of Protected Landscapes and Seascapes, IUCN, GTZ and Obra Social de Caixa Catalunya. Kasparek Verlag, Heidelberg.

Mancini, F., Coghill, G. M., & Lusseau, D. (2018). Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. *PLoS ONE*, *13*(7), 1–19.

Milcu, A. I., Hanspach, J., Abson, D., & Fischer, J. (2013). Cultural ecosystem services: A literature review and prospects for future research. *Ecology and Society*, *18*(3), 44–88.

Millennium Ecosystem Assessment (MEA). (2005). Millennium Ecosystem Assessment Synthesis Report.

National Park Service (NPS). (2000). America's national park system: The critical documents. Retrieved from https://www.nps.gov/parkhistory/online_books/anps/anps_1i.htm

National Park Service. (2020). National Park Service Visitation Numbers. https://www.nps.gov/aboutus/visitation-numbers.htm

Oliveira, T., Araujo, B., & Tam, C. (2020). Why do people share their travel experiences on social media? *Tourism Management*, *78*(November 2019), 104041.

Richards, D. R., & 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.

Sinclair, M., Ghermandi, A., Moses, S. A., & Joseph, S. (2019). Recreation and environmental quality of tropical wetlands: A social media based spatial analysis. *Tourism Management*, *71*(2751), 179–186.

Sultana, R., & Selim, S. A. (2021). Residents' perceptions of the role and management of green spaces to provide cultural ecosystem services in Dhaka, Bangladesh. *Ecology and Society*, *26*(4).

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*(July), 550–568.

Wehrens, R. & Kruisselbrink, J. (2018). Flexible Self-Organizing Maps in kohonen 3.0. *Journal of Statistical Software*, 87(7), 1-18.

Wickham, H. (2020). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.6. https://CRAN.R-project.org/package=rvest

Wijffels, J. (2019). udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit. R package version 0.8.3. https://CRAN.R-project.org/package=udpipe

Willis, C. (2015). The contribution of cultural ecosystem services to understanding the tourism-nature-wellbeing nexus. *Journal of Outdoor Recreation and Tourism*, *10*(June), 38–43.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65.

Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services*, *24*, 68–78.

Zhang, H., Huang, R., Zhang, Y., & Buhalis, D. (2020). Cultural ecosystem services evaluation using geolocated social media data: a review. *Tourism Geographies*, *0*(0), 1–23.

CHAPTER 5

ANALYZING DESTINATION COMPETITIVENESS CONSIDERING TOBLER'S FIRST LAW OF GEOGRAPHY: A CROSS-COMPARISONS OF U.S. NATIONAL PARKS[4]

[4] Kong, I., Mu, L., Woosnam, K.M., & Sarmiento, F.O. Submitted to *Tourism Management Perspectives*, October 29, 2021.

ABSTRACT

Tobler's first law of geography explains how geographic proximity between visitors and tourism destinations promotes visitation and engaging experiences. Competitive tourism destinations, however, may deviate from the law by exerting strong gravity to attract distant visitors and offering quality tourism experiences to visitors regardless of their origins. In this study, we utilize TripAdvisor to assess the competitiveness of 48 U.S. National Parks, focused on destination gravity and experience quality. The plots for destination gravity and experience quality based on travel distances confirmed the Tobler's first law of geography, yet the patterns were distinctive to national parks, allowing them to be classified into four distinct clusters. As a result, we distinguished the most competitive national parks that exert stronger gravity and provide quality experiences. Overall, assessing destination competitiveness based on geographic perspective can assist in optimization of destination marketing and management.

INTRODUCTION

National parks provide competitive tourism opportunities based on nationally distinguished biodiversity and cultural distinctions (Leung et al., 2018). With increasing public interest in tourism and leisure activities, national park tourism has grown rapidly as such a trend is expected to continue (Clark et al., 2019; Rice et al., 2019). Prior to the COVID-19 global pandemic, more than 300 million people visited the U.S. National Parks in 2019 (NPS, 2020). However, not all national parks exhibit the same level of destination competitiveness despite falling within the U.S. Department of the Interior’s National Park System.

Destination competitiveness has been applied in tourism studies to measure the ability to “increasingly attract visitors while providing them with satisfying, memorable experiences and to

do so in a profitable way" (Ritchie & Crouch, 2003). A competitive tourism destination can appeal to a large number of people across diverse markets while providing quality tourism experiences so that it can secure economic sustainability. Previous research has proposed a variety of frameworks for measuring destination competitiveness, ranging from the number of visitors to diverse proxy variables such as destination image, attractiveness, and satisfaction (Jensen, Li, & Uysal, 2017; Kladou & Mavragani, 2015; Kozak & Rimmington, 1999; Mutanga et al., 2017). In addition, multiple criteria decision-making frameworks have been devised to evaluate wide-ranging dimensions of destination competitiveness, as shown in the Travel and Tourism Competitiveness Index that considers policy efficacy and infrastructure capacity, as well as natural and cultural resources (WEF, 2019).

According to the definition of competitive destination by Ritchie and Crouch (2003) which mentioned the ability to attract visitors, national parks have varying degrees of gravity: more competitive destinations have a greater number of visitors. According to the U.S. National Park visitor statistics (NPS, 2020), over four million people visited Yosemite National Park in 2019, whereas around 600 thousand people visited Kings Canyon National Park, which is only 150 kilometers (i.e., 100 miles) away. In addition to the number of visitors, competitive tourism destinations exert gravity to the farther markets, encouraging distant visitors to make long-distance trips. Generally, distance has been considered to be an obstacle in making travel decisions (McKercher, 2018; McKercher & Mak, 2019), as conceptualized in the First Law of Geography by Tobler (1970), stating that "everything is related to everything else, but near things are more related than distant things" (p. 236). However, the emergence of low-cost airlines in the last decade has reduced the role of physical distance as a key constraint in making tourism decisions, weighing more on the pull factors of tourism destinations (Olipra, 2012). Simultaneously, the growing

number of empirical works has suggested that competitive tourism destinations now are better positioned to draw more long-distance visitors from remote markets (Neuvonen et al., 2010; Norman & Pickering, 2019; Shen, Huang, & Zhang, 2019). Big data analysis has also revealed that popular national parks in the United States are welcoming more overseas visitors (VISA, n.d.). People also express a stronger willingness to make long-distance travels to visit iconic tourism destinations seeking once-in-a-lifetime experiences (Choy & Prineas, 2006; Eagles et al., 2015; Pinkus et al., 2016), even paying for premiums (Mandić & Petrić, 2021).

Another aspect of competitive tourism destinations is the ability to provide enjoyable and satisfying tourism experiences (Ritchie & Crouch, 2003). Since the declaration of *experience economy* by Pine and Gilmore (1999), the managerial agenda of tourism destinations has shifted toward implementing memorable tourism experiences, which aim to offer novel and unforgettable tourism experiences (Kim, Ritchie, & McCormick, 2012). By operating memorable tourism experiences, tourism destinations can boost destination competitiveness to attract more visitors while enhancing visitor satisfaction, place attachment, and destination loyalty (Sharma & Nayak, 2018; Pinkus et al., 2016; Vada, Prentice, & Hsiao, 2019). However, measuring the quality of the tourism experience has been a major challenge as it involves the psychological process that occurs as a result of onsite experience at destinations. As a result, numerous studies have proposed and implemented indicator-based analysis to reduce the dimension of experience psychology, suggesting visitor satisfaction, place attachment, and loyalty (Moore et al., 2015; Vada, Prentice, & Hsiao, 2019). With improvements in measuring the quality of tourism experiences, the following question delves into the subjectivity of tourism experiences. Several previous studies have proven that the origin of visitors determines the cultural distance and place familiarity with the destinations (Bi & Lehto, 2018; McKercher & Chow, 2001), resulting in noted differences

across tourism interests, activities, and satisfaction levels between local and non-local visitors (Mutanga et al., 2017; Nyaupane & Graefe, 2008). However, most studies have compared visitors based on categorical classifications, namely locals versus non-locals, mostly at a single destination, due to the constraints from the conventional research methods that require significant time and budgeting. Indeed, nature-based tourism destinations including national parks have been dealing with the lack of relevant data to operate visitor and destination management (Eagles, 2014).

The advent of user-generated content (UGC), therefore, has garnered significant attention in capturing massive amounts of public content which can assist nature-based tourism destinations with management strategies (Teles da Mota & Pickering, 2020; Wilkins, Wood, & Smith, 2020). Accordingly, the contents and metadata from UGC have facilitated analysis to support visitor monitoring and destination assessment, while validating the data integrity using official statistics and field survey data (Heikimeimo et al., 2017; Sessions et al., 2016). Still, the application of UGCs in harnessing destination management remains underutilized, notably for cross-comparisons of destination competitiveness. Notably, current knowledge about national park competitiveness remains limited, particularly in terms of how gravity works to attract visitors from different source markets and how destinations provide quality experiences to visitors at each destination level. Above all, assessing and comparing the destination competitiveness is critical to deal with increasing and diversifying tourist demand, which has posed significant challenges (Eagles, 2014; Leung et al., 2018; Taylor, Frost, & Laing, 2017).

In this study, we aim to assess national park competitiveness focusing on two factors associated with Tobler's first law of geography. One is the gravity of destinations to attract visitors from varying travel distances and another is the quality of tourism experiences perceived by visitors of varying origins. When a destination is competitive, it will exert a stronger gravity to

attract more long-haul visitors and provide a quality experience to them regardless of visitor origin. As a case study, we chose national parks in the contiguous U.S. for this research.

LITERATURE REVIEW

The gravity of tourism destinations

Tobler's First Law of Geography explains the diminishing interactions with an increase of distance (Tobler, 1970). The relationships between visitors and tourism destinations are not an exception, showing that visit frequency diminishes with an increase in travel distance (Cheung, 1972; Chung, Dietz, & Liu, 2018; Eagles et al. 2015; Rossi, Byrne, & Pickering, 2015; Schuett & Hollenhorst, 2010; Shen, Huang, & Zhang, 2019). The classical pattern that explains such a relationship is the negative linear or negative exponential growth with a peak, explaining that visitation reaches a pinnacle at the close-distanced market and diminishes rapidly as distance increases (McKercher, 2018).

However, the classic pattern does not apply to all destinations. People are willing to take a long-haul trip to iconic destinations in search of novelty and a once-in-a-lifetime experience (Eagles et al., 2015; Pinkus et al., 2016). Among nature-based tourism destinations, Choy and Prineas (2006) found that national parks had a slowly declining curve with a fat-tail while local and regional recreational sites exhibited a radical decay of visit frequency as travel distance increased. From the cross-comparisons of 35 national parks in Finland, Neuvonen et al. (2010) identified that national parks with more attractive attributes such as natural distinctions, recreational facilities, and tourism services showed a higher number of visitations. Using geotagged social media, Norman and Pickering (2019) further demonstrated that visitors made longer trips to national parks in search of lengthier recreational trails or rugged terrains where they

could enjoy non-daily activities. To summarize, competitive destinations have stronger gravity to attract visitors from remote distances, illustrating multiple peaks (i.e., one big peak at close distance and other peaks at distant locations) or a plateau (i.e., flat peak throughout certain distance ranges) in plotting the relationship between travel distance and visit frequency.

Still, we have limited knowledge in understanding the pattern of destination gravity on a continuous geographic scale. From a case study for five provincial parks in Canada, Eagles et al. (2015) exemplified a variety of patterns including multiple peaks as well as the classic exponential curve, suggesting the diversity of gravity depending on destinations. Using geotagged Flickr, Sessions et al. (2016) also compared the proportion of visitors depending on travel distances at 38 national parks in the western U.S, to find different patterns of Tobler's First Law of Geography. Despite such findings, a dearth of studies exists to distinguish the power of destination gravity across multiple destinations falling within the same managing agency (i.e., U.S. National Park Service).

Quality of tourism experiences

To measure the quality of tourism experiences, one of the most prevalent approaches in nature-based tourism research focuses on destination satisfaction (Cong et al., 2014; Mutanga et al., 2017; Roemer & Vaske, 2014). Using empirical research, Cong et al. (2014) and Mutanga et al. (2017) confirmed that the quality of wildlife tourism, particularly when accompanied with interpretation and interactions, can result in higher visitor satisfaction in wildlife tourism. The U.S. National Park Service also monitors visitor satisfaction at the destination level, to compare satisfaction levels based on geographic regions and park designations (Roemer & Vaske, 2014). In addition to the satisfaction level, Moore et al. (2015) has proposed destination loyalty as a

subsequent indicator of satisfaction. The following studies, therefore, have verified the efficacy of satisfaction and loyalty in explaining positive word-of-mouth, revisit intention, and willingness to pay more at nature-based tourism destinations (Kwenye & Freimund, 2016; Pearce & Dowling, 2019; Pinkus et al., 2016).

Recent studies further explore psychological approaches in measuring the quality of tourism experiences, such as sentiments (Hausmann et al., 2020; Mangachena & Pickering, 2021), sensory perceptions (Lv, Li, & McCabe, 2020), and long-term memories (Agapito, Pinto, & Mendes, 2017; Jorgenson et al., 2019). Similarly, psychological distance is a metric that has been applied to assess the tourism experience based on construal level theory, to measure the conceptual distance between a subject and an object (Trope & Liberman, 2010). As the psychological distance can be shortened with meaningful stimuli, it is believed that memorable tourism experiences can reduce the psychological distance of visitors toward the destinations (Massara & Severino, 2013). Nonetheless, psychological measurement in assessing the quality of tourism experiences is in a nascent stage (Hausmann et al., 2020; Mangachena & Pickering, 2021).

Another key concern with assessing the quality of experience is how visitors perceive the experience differently. It has long been considered that visitor origin or travel distance influences the cultural distance and psychological familiarity with the destination (Bi & Lehto, 2018; McKercher & Chow, 2001), shaping different tourism interests, motivations, and experiences (Mutanga et al., 2017; Nyaupane & Graefe, 2008). Specifically, Nyaupane and Graefe (2008) found that local visitors engaged more in camping and overnight stays than non-local visitors within U.S. national parks, revealing stronger place attachment. Mutanga et al. (2017) also distinguished different tourism interests between local and non-local visitors, demonstrating that the former expressed stronger place attachment and satisfaction following safari tourism

experiences. Similarly, Joo et al. (2017) found stronger place attachment among local visitors in nature-based tourism destinations in Texas. as Xiao et al. (2021) validated the role of distance in shaping place authenticity and identity from mountain tourism in China.

Nonetheless, the questions remain concerning the salient factors in determining quality of experience between visitor characteristics and destinations. Although numerous studies have demonstrated that the variance of tourism experiences depends on visitor origin, the significance of tourism destinations in affecting quality of tourism experiences has not been explored very much. A handful of studies have argued that the quality tourism experiences from the destination itself can be more deterministic in shaping visitor satisfaction and place attachment than the residential proximity or visit frequency (Jensen, Li, & Uysal, 2017; Vada, Prentice, & Hsiao, 2019). Thus, assessing perceptions of quality experiences at destinations, as well as how such perceptions vary based on visitor origin can help decipher the interplay of visitors and destinations.

DATA AND METHODS

Study site and data collection

We obtained online tourism reviews (OTRs) along with the date and self-claimed home locations from TripAdvisor, which has been utilized in several tourism studies (Taecharungroj & Mathayomchan, 2019; Xiang et al. 2017; Zhou, Wang, & Li, 2017). Of 62 national parks within the U.S. National Park Service System, we filtered national parks in the contiguous U.S. continent, excluding those in Alaska or remote islands (e.g., American Samoa) to minimize distance variance as potential outliers. Gateway Arch National Park was also excluded given its urban setting, which may result in a distinctive tourism experience and accessibility. Lastly, Sequoia National Park was dropped as it was subordinate to Kings Canyon National Park in TripAdvisor. As a result, we

defined 48 national parks as the study sites. The locations of study sites were illustrated with the number of visitors in 2019 (NPS, 2020) and the NPS managerial regions in Figure 5-1.

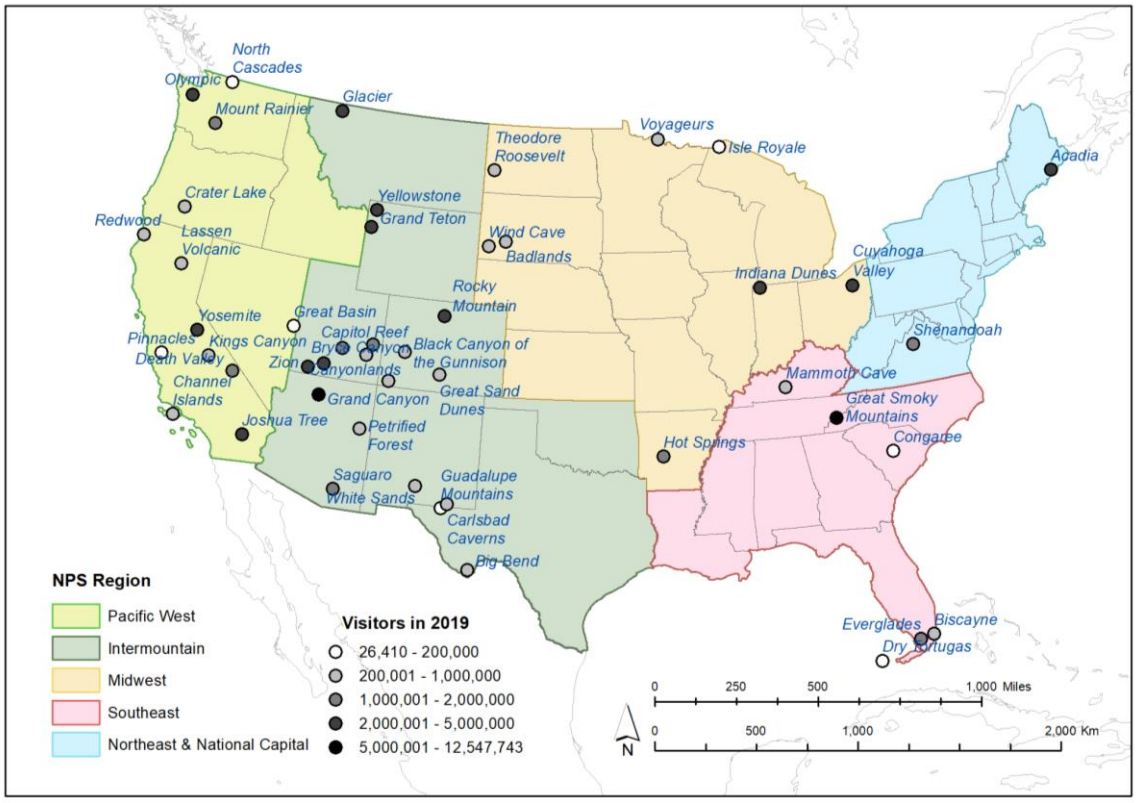


Figure 5-1. The study sites of 48 U.S. National Parks

To operate web scraping of TripAdvisor, we developed the scripts in R Studio. As one national park has multiple attractions and the reviews are accumulated under each attraction, we first identified the URLs for the front pages of 48 national parks (e.g., “Yosemite National Park”) and let the R scripts acquire the directories of subordinate attractions (e.g., “Half Dome”). We set to acquire up to 30 subordinate attractions, yet some national parks had fewer attractions or consisted of a single attraction (see Appendix B). Then the scripts navigated to each attraction page and collect all available OTRs, along with user ID, date of experience, and the self-reported

home locations. The data collection was undertaken using the *rvest* for HTML web scraping (Wickham, 2020) and *RSelenium* for dynamic web browsing (e.g., clicking the next web page number, clicking '…More' button to expand the review) (Harrison, 2020). As a result, we obtained 292,306 reviews from the 48 national parks.

Data processing

Before cleaning the data, we only kept the reviews written before March 2020 to avoid unusual experiences after the pandemic and consequent managerial regulations (e.g., park closure, modified park programs). Then, we counted the number of visitors by pairing user IDs and the date of experience, given that one visitor can post several reviews across multiple attractions. Then, the self-reported home locations (e.g., 'Nashville, Tennessee') from unique pairs were geocoded to obtain geographic coordinates using the R package *ggmap* (Kahle & Wickham, 2013). When the pairs returned invalid geolocations due to the inaccurate home locations (e.g., 'United States', 'nomad'), we removed the pairs. The number of valid users were then applied to analyze the gravity of destinations (Chapter 3.3). The valid user information was also applied to filter corresponding reviews for psychological distance analysis (Chapter 3.4). The data processing flow is illustrated in Figure 5-2. Details of the dataset can be found in Appendix B.

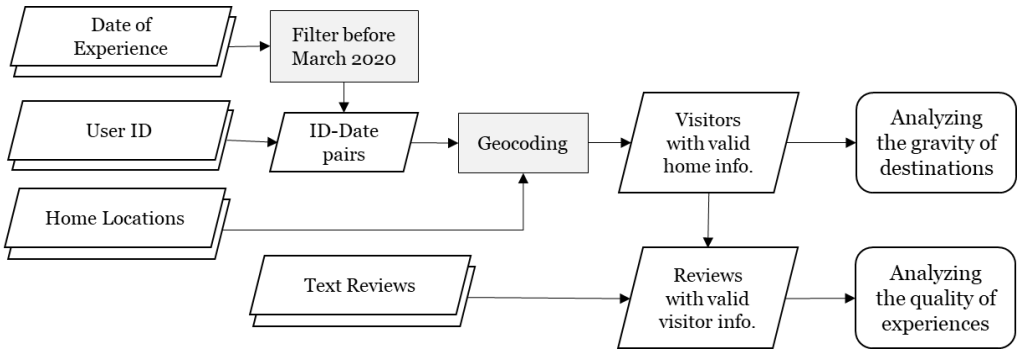


Figure 5-2. Flow of data processing

Assessing the gravity of destination

To plot the gravity of national parks that represent the relationship between travel distance and visit frequency, we first calculated the travel distances of unique visitors based on the origin-destination model. Due to the minuscule distances across subordinate attractions in a national park, we defined one representative geolocation for each national park as the centroid of park boundary (NPS, 2019). Then the R package *geosphere* was applied to measure the distance between the national park and the home locations of unique visitors, using the Haversine method (Hijmans, 2019). As a result, we generated a density plot with the smoothed density estimates to show the relationship between the travel distance (x-axis) and the share of visitors (y-axis). The x-axis ranged between 0 and roughly 18,000 km. The visualization was operated with the *ggplot2* package (Wickham, 2016).

Then, we clustered the plots based on the patterns. We extracted density values (i.e., the share of visitors) from the constant intervals of the x-axis (i.e., travel distance) and created a matrix table summarizing the density values across 48 national parks. Then we applied a distance matrix computation (DMC) to calculate numeric distances between the rows of a data matrix. Using the results from DMC, the hierarchical clustering method with Ward's minimum variance was applied to group the patterns of destination gravity. As a result, we distinguished four different patterns

based on the best-fitting clustering hierarchy. The process was operated with the *stats* package (R Core Team, 2021).

Assessing the quality of experiences

The quality of tourism experiences was estimated with the textual concreteness of TripAdvisor reviews. According to construal level theory in psycholinguistics, meaningful stimuli can lower the level of construal, reducing the psychological distance between the subject and the object (Trope & Liberman, 2010). When the psychological distance is close, people are likely to use more concrete expressions instead of abstract expressions (Snefjella & Kuperman, 2015; Yeomans, 2021).

Before estimating the textual concreteness, we first applied Natural Language Process (NLP) to the unstructured texts of TripAdvisor reviews. The process included tokenization (i.e., dissecting the sentences), lowercasing the text, lemmatization (i.e., returning the words into the base forms), and part-of-speech tagging (Wijffels, 2019). Named Entity Recognition (NER) was also operated to detect place names. For instance, when NER finds the phrase 'White Sands National Park' as a place name, it recognizes the chunk of lemmas as an aggregated phrase and skips calculating 'white' or 'sand' in the following analysis. Among thousands of results from NER, we applied an 80% threshold of the cumulative frequency to keep frequently used place names. The entire NLP was applied with *udpipe* based on American English ('English-ewt') (Wijffels, 2019).

Accordingly, we applied the concreteness ratings from Brysbaert, Warriner, & Kuperman (2014). The ratings included an array of words (lemmas) and concreteness scores ranging from 0.0 (abstract) to 5.0 (concrete). For instance, the concreteness scores for 'tree,' 'facility,' and 'happy'

were 5.0, 3.58, and 2.56 respectively. After assigning the concreteness scores to the corresponding lemmas in each OTR, we calculated the average score as a proxy for experience quality perceived by the individuals. Then we aggregated the concreteness scores of OTRs by each national park and generated scatter plots with a linear model, showing the relationship between the concreteness scores (y-axis) and the travel distance (x-axis). Subsequently, we grouped the patterns based on *k*-means clustering, using two parameters of the linear models – the slope and y-intercept. As a result, we obtained four different patterns of the quality of tourism experiences in association with travel distance for the 48 national parks. The analysis was processed with *stats* package (R Core Team, 2021).

Cross-comparisons of national park competitiveness

To compare the competitiveness of national parks, the travel distances and concreteness scores were aggregated at each national park to calculate the average. Then, the average travel distance and average concreteness scores of 48 national parks were shown in a scatter plot. Then we colored the plot by the clusters of destination gravity and experience quality respectively, with ellipses showing 95% confidence. The visualization was operated with the *ggplot2* in R (Wickham, 2016). For additional analysis, we mapped the spatial distribution of national parks based on the clusters of destination gravity and experience quality, to distinguish spatial distinctions related to the clusters. The map was generated using ArcMap 10.7.1.

RESULTS

Four clusters of destination gravity

We plotted individual patterns of destination gravity for 48 national parks (see Appendix E-1). The results revealed that most national parks followed Tobler's first law of geography, a declining visit frequency as travel distance increases. Then we grouped the patterns into four clusters, to name the clusters as Classic (G1), Plateau (G2), Three Peaks (G3), and Local Peak (G4) (Figure 5-3). National parks classified as Classic (G-1) and Local Peak (G4) showed the classical pattern of Tobler's first law of geography, with the decaying gravity to attract visitors as travel distance increases. Yet, national parks of Local Peak (G4) exhibited radical declines in exerting gravity to attract faraway visitors, hosting most visitors from nearby areas. The Great Smoky Mountains National Park and Cuyahoga National Parks fell into this group. National parks of the Classic pattern (G1) also had most visitors from the closest market, yet the peak was not as significant as the patterns of Local Peak (G4). The Classic pattern (G-1) was shown from Acadia National Park and Rocky Mountains National Park.

Meanwhile, some national parks exhibited extended gravity to attract more faraway visitors, as the patterns of Plateau (G2) and Three Peaks (G3) demonstrated. National parks with the Plateau pattern (G2) attracted a steady number of visitors within a 3,500-kilometer radius, with an extra peak of international visitors from 8,000 kilometers in distance. Grand Canyon National Park and Yellowstone National Park were two destinations with the Plateau pattern (G2). National parks with the Three Peaks (G3) showed three peaks of visitor distances, indicating short-haul domestic (i.e., Western U.S.), long-haul domestic (i.e., Eastern U.S., and international visitors (i.e., Europe) based on the geographic scales. Yosemite National Park and Olympic National Park showed the Three Peaks (G3). To summarize, the national parks with the Plateau (G2) and Three

Peaks (G3) were competitive destinations that attracted visitors from farther distances as well as close distances.

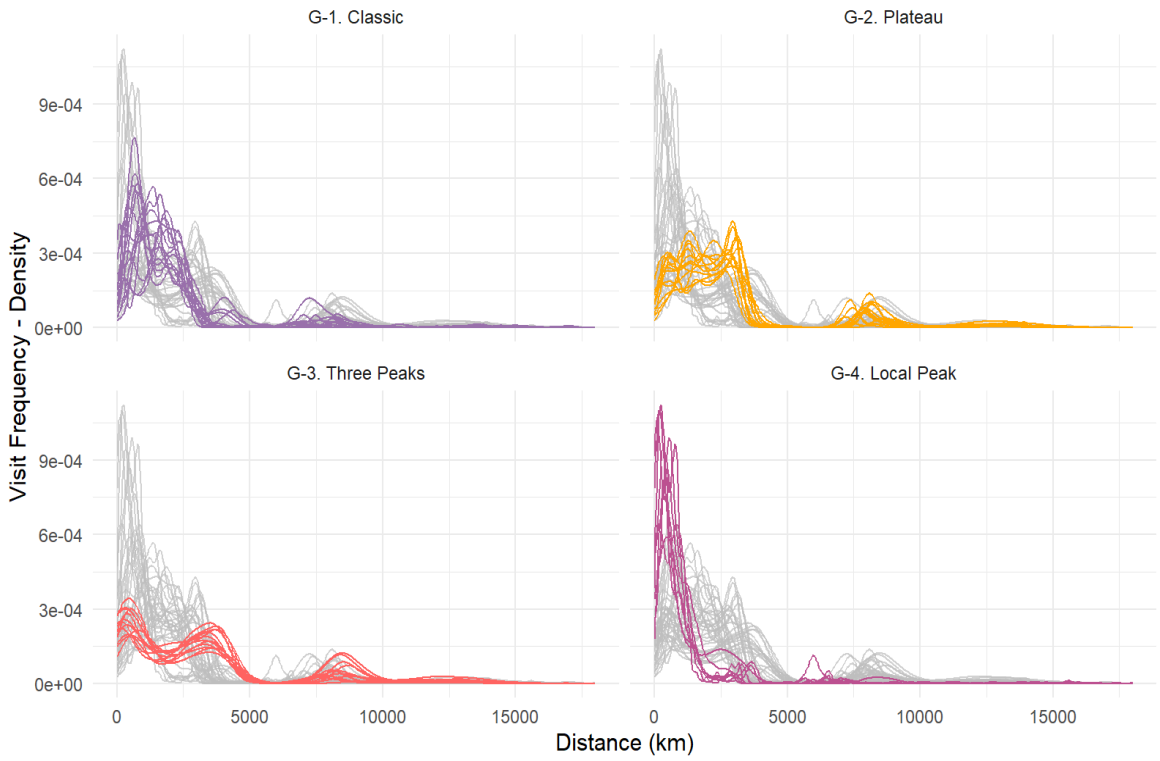


Figure 5-3. Four patterns to describe the gravity of destinations in association with travel distances. Each line indicates a national park; colored lines indicate national parks that fall into the pattern group while grey lines indicate national parks in other pattern groups.

Four clusters of experience quality

We plotted individual patterns of experience quality for the 48 national parks (see Appendix E-2). Most national parks followed Tobler's first law of geography, with a declining concreteness score as travel distance increases, with some exceptions, such as Isle Royale National Park and Biscayne National Park, to exhibit the opposite pattern. We assumed that the low number of reviews in these national parks may have influenced such patterns (see Appendix B), but kept

all data for the clustering analysis, except for Isle Royale National Park whose farthest visitor traveled roughly 4,000 kilometers to get to the park.

After the clustering, we distinguished four patterns and named them based on the y-intercept (i.e., concreteness scores to be above 2.6 as high) and the slope of the decline (i.e., constant, declining, and dropping, based on the angle). As a result, we obtained High-Constant (Q1), High-Dropping (Q2), Moderate-Declining (Q3), and High-Declining (Q4) patterns (Figure 5-4). National parks with High-Constant (Q1) patterns exhibited that visitors expressed a great quality of tourism experiences regardless of their travel distances, with the examples of Everglades National Park and Olympic National Park. National parks with the High-Dropping (Q2) patterns explained that local visitors experienced quality experiences whereas visitors from farther distances used much lower concrete expressions regarding their experiences. Examples of this included the Great Sand Dunes and Voyageurs National Park. National parks with the Moderate-Declining (Q3) patterns explained that the quality of experience ranged between moderate and low, and the quality of experiences moderately declined as the travel distance increased. Grand Canyon National Park and Yellowstone National Park were examples of Q-3. The High-Decay (Q4) patterns, which included Yosemite National Park and Great Smoky Mountains National Park, showed a moderate to high quality of experiences for local visitors and it declined slightly as visitors' travel distances increased. To summarize, the national parks with the High-Constant (Q1) were the most competitive destinations to offer quality experiences regardless of visitor origin.

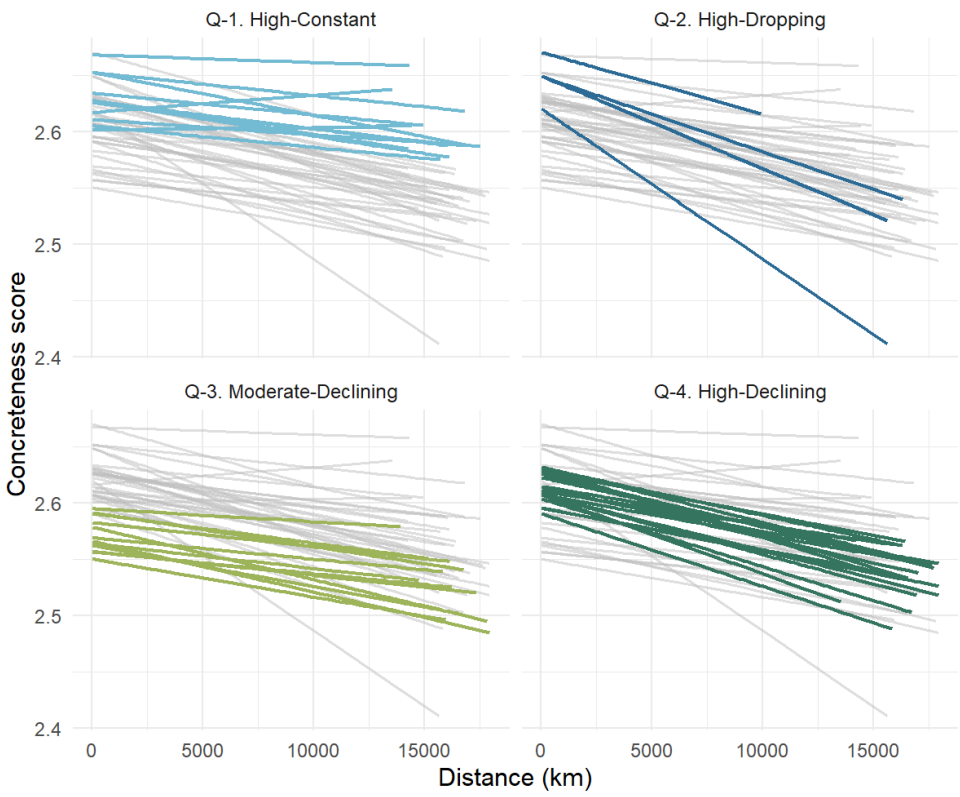


Figure 5-4. Four patterns to describe the quality of experiences in association with travel distances. Each line indicates a national park; colored lines indicate national parks that fall into the pattern group while grey lines indicate national parks in other pattern groups.

Identifying competitive national parks

Highlighted with the clusters of destination gravity (Figure 5-5) and experience quality (Figure 6), the scatter plots visualized the average travel distance and concreteness score of the 48 national parks. As shown in Figure 5, national parks in the same destination gravity clusters had comparable ranges of travel distances. Local Peak (G4) national parks had the shortest average distance among their visitors, followed by Classic pattern (G1). Plateau (G2) and Three Peaks (G3) national parks had higher ranges of average travel distances of their visitors, suggesting their greater gravity in attracting visitors from further distances. Meanwhile, the gravity of destinations

had a minimal association with experience quality (i.e., concreteness scores), since all the ellipses of gravity clusters were vertically deformed to include wide-ranging concreteness scores.

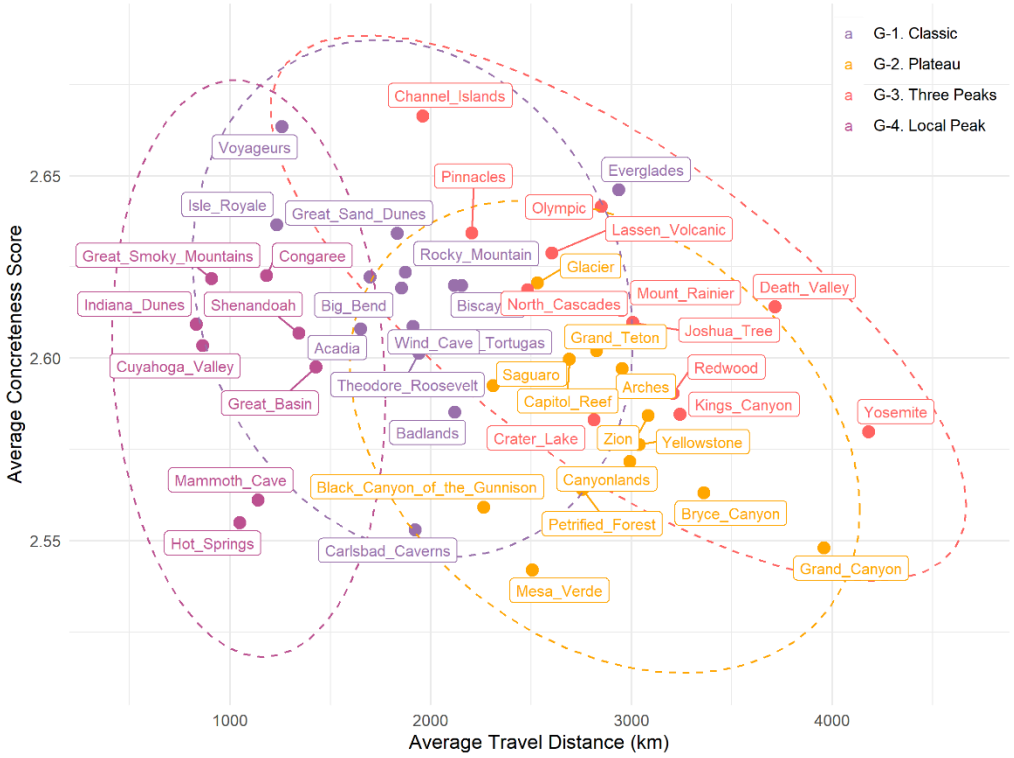


Figure 5-5. Scatter plots of destination gravity. The colors indicate the destination gravity clusters

Figure 5-6 highlighted the national parks based on the clusters of experience quality. The lowest ranges of concreteness scores were found in Moderate-Declining (Q3) national parks, followed by High-Declining (Q4) national parks. National parks of High-Constant (Q1) and High-Dropping (Q2) patterns were blended in the plot. Similar to the findings from Figure 5, the quality of experience had a minimal association with the travel distance since all the ellipses of quality clusters were horizontally deformed to include wide-ranging travel distances.

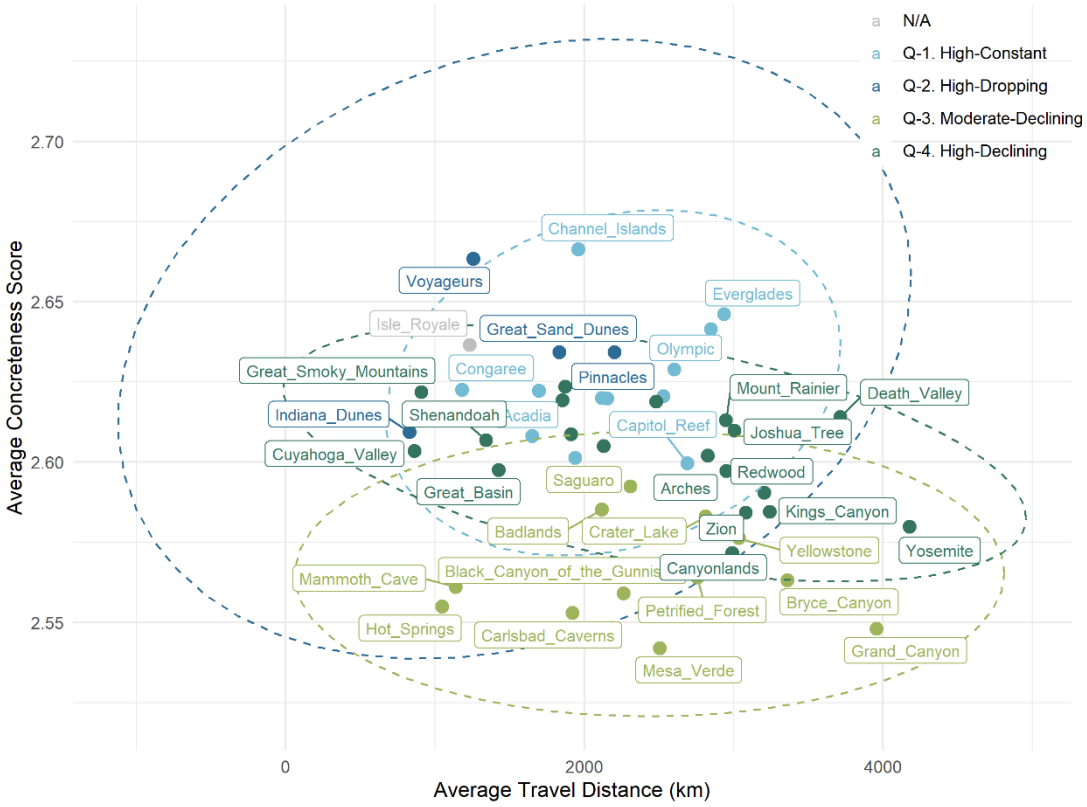


Figure 5-6. Scatter plots of experience quality. The colors indicate the experience quality clusters

The matrix map (Figure 5-**7**) illustrated the spatial distribution of national parks based on the cluster combinations of destination gravity and experience quality. The subtotal map of destination gravity clusters (i.e., the first row) revealed distinctive geographical patterns of national parks. National parks of Classic pattern (G1) exhibited the most dispersed geographic patterns, ranging from the Rocky Mountains to Florida and Maine. National parks of Three Peaks (G3) were exclusively found in the Pacific West, while national parks of Plateau (G2) were located in the Intermountain region. As Plateau (G2) and Three Peaks (G3) were two of the most competitive national parks in terms of destination gravity, the results supported the competitiveness of national parks in the Pacific West and Intermountain areas to attract visitors of varying travel distances.

Meanwhile, most of the Local Peak (G4) national parks were located in the eastern United States, implying the dominance of local and short-distanced visitors in these national parks.

The subtotal map of experience quality clusters (i.e., the first column), however, did not show distinctive geographical patterns. Still, we found that national parks of Moderate-Declining (Q3) corresponded the most with Plateau (G2), and national parks of High-Declining (Q4) corresponded the most with Three Peaks (G3). Given that the Plateau (G2) and Three Peaks (G3) were the two most competitive destinations in terms of great destination gravity, and the Moderate-Declining (Q3) was the least competitive in terms of experience quality, national parks in the G3-Q4 pair outperformed national parks in the G2-Q3 pair as competitive destinations.

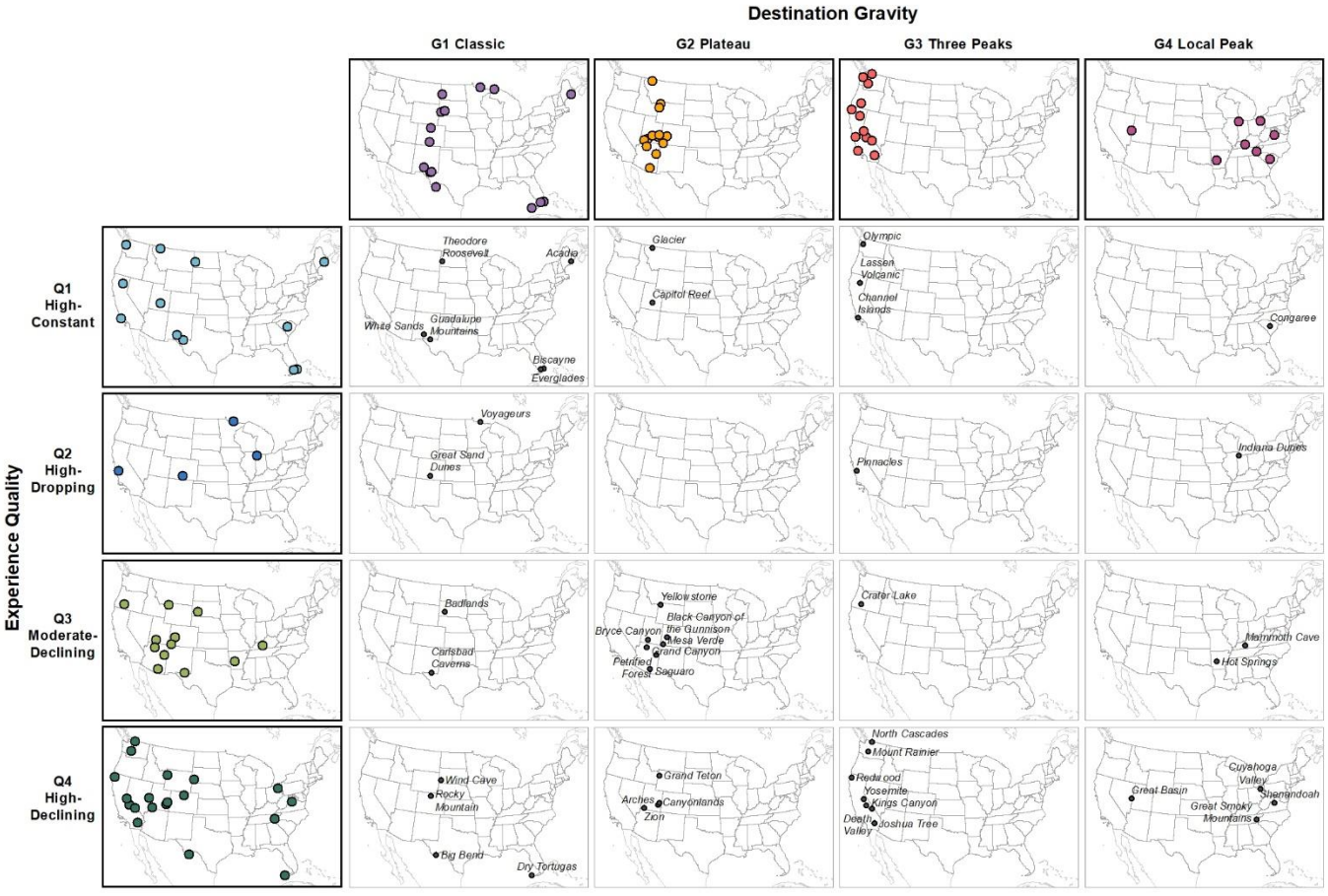


Figure 5-7. A matrix map showing the spatial distribution of 48 national parks based on four clusters of destination gravity and four clusters of experience quality

DISCUSSION

Destination gravity to indicate a competitive tourism destination

In this study, we focused on destination gravity and experience quality to assess destination competitiveness in 48 U.S. national parks, to find varying levels of competitiveness. From the first analysis to measure the destination gravity, we found that Plateau (G2) and Three Peaks (G3) had the greatest gravity to attract more distant visitors among the four clusters of destination gravity, compared to Classic (G1) and Local Peak (G4) patterns. Spatial distribution analysis revealed that the national parks that have strong gravity (G2, G3) were situated in the Pacific West and Intermountain areas, explaining the geographic distribution of the competitive national parks. Such finding is consistent with the popularity of national parks in the Pacific West and Intermountain regions, such as the Grand Canyon, Yellowstone, and Yosemite National Parks. Eventually, such national parks have the gravity to draw visitors from various origins, as well as a large number of visitors.

Additionally, the plots of destination gravity further indicated the average travel distances to visit the U.S. national parks. Previous research has determined that the travel distance threshold for urban parks in China is 4 kilometers (Liu, Chen, & Dong, 2017), provincial parks in Canada are 300 kilometers (Eagles et al., 2015), and international tourism destinations are 5,000 kilometers (McKercher & Mak, 2019). According to the findings in this study, people traveled ~1,000 km to visit national parks of Local Peak (G4), ~2,000 km to visit Classic pattern (G1), and ~3,000 km to visit Plateau (G2) and Three Peak (G3) national parks, on average.

Still, some may question the role of accessibility to plot such results. Admittedly, previous studies have claimed the role of proximate cities in bringing a higher number of visitors to protected areas (Chung, Dietz, & Liu, 2018; Neuvonen et al., 2010; Norman & Pickering, 2019).

Likewise, the number of visitors to the U.S. national parks was related to the existence of nearby cities (Burkett et al. 2010; Shen, Huang, & Zhang, 2019; Schuett & Hollenhorst, 2010). Nonetheless, the findings of this study indicated that proximity to large cities is not the single most significant factor in determining the gravity of tourism destinations. This is because it did not explain why national parks on the East Coast had Local Peaks while national parks on the West Coast had Three Peaks. If all national parks had the same gravitational pull, national parks in the eastern U.S. should have had a secondary peak about 4,000 kilometers away, indicating visits from the West Coast. Eventually, the result explained the deterministic role of destination competitiveness to exert gravity in pulling distant visitors.

Quality of tourism experiences to indicate a competitive tourism destination

In this study, we proposed textual concreteness to measure the quality of tourism experience, based on the construal level theory (Trope & Liberman, 2010; Yeomans, 2021). The findings showed that the perception of quality experiences in the U.S. national parks declined with an increase of travel distance of visitors, supporting Tobler's first law of geography. However, not all national parks exhibited same level of quality experience. Cross-comparisons across the 48 national parks revealed that national parks with the High-Constant pattern (Q1) had high concreteness ratings that remained consistent throughout travel distances. Meanwhile, other national parks, such as those with a Moderate-Declining (Q3) pattern, had lower-than-average concreteness at the destination and it decreased as travel distance increased. Such distinctions can harness the argument that the personal perception of quality experiences can be determined by the performance of tourism destinations as well as predetermined visitor factors such as travel distance, cultural distance, and psychological familiarity with the destination. Furthermore, when

we plotted the locations of national parks based on quality experience clusters, it showed random distributions, debunking the notion that climate or natural landscape is the most important factor in providing quality experiences. Instead, the result emphasized the role of individual destinations to provide memorable and enjoyable tourism experiences, supporting previous studies to emphasize the tourism contents (Jensen, Li, & Uysal, 2017) and memorable tourism experiences (Vada, Prentice, & Hsiao, 2019).

However, there are some technical limitations in measuring the quality of tourism experiences based on concreteness ratings. First, the concreteness ratings developed by Brysbaert, Warriner, and Kuperman (2014) were not intended to be applied in a tourism setting. As a consequence, we discovered that certain vocabulary, such as ecological words (e.g., species names), which may provide significant attributes in tourism experiences, were lacking. To improve analytic accuracy, therefore, relevant vocabulary that can reflect the contexts and domains should be adjusted (Yeomans, 2021). Second, we disregarded place names from calculating concreteness scores, although the construal level theory classifies place names as concrete expressions (Trope & Liberman, 2010). However, we concluded that mentioning place names is more of a chance rather than showing psychological closeness. Despite the limitations, we presented the usefulness of online reviews for evaluating destination competitiveness based on psychological distance and concreteness scores. Indeed, many studies in recent years have applied content analysis from UGCs to shed light on the diverse dimension of tourism experiences, including sentiment, perceptions, and memory (Agapito, Pinto, & Mendes, 2017; Hausmann et al., 2020; Jorgenson et al., 2019; Lv, Li, & McCabe, 2020; Mangachena & Pickering, 2021). Exploring diverse types of content, platforms, and analytic methods may therefore bring substantial benefits to assist visitors and destination management organizations.

Data validity and integrity of TripAdvisor

As previously stated, using UGCs for tourism analysis can help to address the data gap for destination management in natural environments (Heikinheimo et al., 2017). However, the credibility and integrity of UGC have been called into question due to biased user demographics or anonymity on online platforms. For instance, TripAdvisor is known to have more users from North America and Europe (Kladou & Mavragani, 2015; Taecharungroj & Mathayomchan, 2019). Furthermore, UGCs can be skewed toward exceptional experiences that create great delight (Oliveira, Araujo, & Tam, 2020), which may overrepresent the case of international travels or trips to popular destinations.

To address the data validity of UGCs, previous studies juxtaposed the number of UGCs with the number of visitors from the official statistics (Sessions et al., 2016), or compared the analysis from UGCs with the ground survey results (Heikinheimo et al., 2017). Tenkanen et al. (2017) incorporated multiple UGC platforms to minimize data bias from a single platform. However, validating content analysis results involves several difficulties. To begin with, few platforms exist where shared tourism experiences in natural environments can be obtained. Some studies have examined online review content from Yelp and Expedia, but these platforms are for sharing comments on restaurants and hotels (Xiang et al., 2017; Zhou, Wang, & Li, 2017). Another difficulty for content validation is a lack of comparable data. The origin of visitors or their tourism experiences is rarely monitored or gathered at the destination level. As a result, many previous studies did not apply data validation and instead relied on the quality control protocols from the platforms (Hausmann et al., 2020; Xiang et al., 2017; Zhou, Wang, & Li, 2017).

While there may be minor concerns about verifying the authenticity of contents, we discovered two inherent limitations of TripAdvisor. First, the number of TripAdvisor reviews has

been declining after reaching a peak around 2014 (see Appendix E-3). As a result, the disparity between the number of monthly visits and the number of monthly TripAdvisor reviews has been widening. Therefore, the content analysis of TripAdvisor reviews may reflect the experience of the past. Another limitation we observed is the uneven number of reviews across national parks. As Oliveira et al. (2020) mentioned, social media contents are selectively generated by individuals so that it may strengthen the popularity bias, which is also observed in the dataset in this study. For instance, both Crater Lake National Park and Biscayne National Park were visited by more than 700,000 visitors in 2019, yet the number of accumulative TripAdvisor reviews were significantly different, 5,160 and 104, respectively. Overall, by thoroughly addressing data characteristics as well as constraints, OTRs and UGCs can help to fill the knowledge gap in visitor monitoring and destination management, particularly in data-deficient, nature-based tourism destinations.

CONCLUSION

This study examined the destination competitiveness across 48 U.S. national parks using TripAdvisor. In evaluating competitiveness, we considered destination gravity to attract visitors from varying travel distances, and experience quality to provide memorable tourism experiences. The results revealed that both the destination gravity and experience quality declined as the travel distance of visitors increased, according to Tobler's first law of geography. However, the patterns were primarily unique to the national parks, indicating the different levels of destination competitiveness in attracting visitors and providing quality experiences. Overall, assessing tourism destinations from a geographic perspective can assist in evaluating destination performance and competitiveness. As a consequence, the findings of this study can help destination managers

identify unique tourism niches based on cross-comparisons, allowing them to harness the optimization of destination marketing and management.

REFERENCES

Agapito, D., Pinto, P., & Mendes, J. (2017). Tourists' memories, sensory impressions and loyalty: In loco and post-visit study in Southwest Portugal. *Tourism Management*, *58*, 108–118.

Bi, J., & Lehto, X. Y. (2018). Impact of cultural distance on international destination choices: The case of Chinese outbound travelers. *International Journal of Tourism Research*, *20*(1), 50–59.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Burkett, J., Tyrrell, T., & Virden, R. (2010). National Park Service Visitation Trends: Exploring the forces that influence visitation. *National Park Service*, (July).

Cheung, H. (1972). A day-use park visitation model, *Journal of Leisure Research, 7,* 139– 156.

Choy, D. L., & Prineas, T. (2006). Parks for People: Meeting the outdoor recreation demands of a growing regional population. *Annals of Leisure Research*, *9*(1–2), 86–109.

Chung, M. G., Dietz, T., & Liu, J. (2018). Global relationships between biodiversity and nature-based tourism in protected areas. *Ecosystem Services*, *34*(September), 11–23.

Clark, M., Wilkins, E. J., Dagan, D. T., Powell, R., Sharp, R. L., & Hillis, V. (2019). Bringing forecasting into the future: Using Google to predict visitation in U.S. national parks. *Journal of Environmental Management*, *243*(February 2019), 88–94.

Cong, L., Wu, B., Morrison, A. M., Shu, H., & Wang, M. (2014). Analysis of wildlife tourism experiences with endangered species: An exploratory study of encounters with giant pandas in Chengdu, China. *Tourism Management*, *40*, 300–310.

Eagles, P. F. J. (2014). Research priorities in park tourism. *Journal of Sustainable Tourism*, *22*(4), 528–549.

Eagles, P. F. J., Johnson, P. A., Potwarka, L. R., & Parent, C. (2015). Travel distance classes for tourism destinations: A proposal from Ontario Provincial Park camping. *Journal of Ecotourism*, *14*(1), 64–84.

Harrison, J. (2020). RSelenium: R Bindings for 'Selenium WebDriver'. R package version 1.7.7. https://CRAN.R-project.org/package=RSelenium

Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Kulkarni, R., Tenkanen, H., & Di Minin, E. (2020). Understanding sentiment of national park visitors from social media data. *People and Nature*, *2*(3), 750–760.

Heikinheimo, V., Minin, E. Di, Tenkanen, H., Hausmann, A., Erkkonen, J., & 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.

Hijmans, R.J. (2019). geosphere: Spherical Trigonometry. R package version 1.5-10. https://CRAN.R-project.org/package=geosphere

Jensen, Ø., Li, Y., & Uysal, M. (2017). Visitors' satisfaction at managed tourist attractions in Northern Norway: Do on-site factors matter? *Tourism Management*, *63*, 277–286.

Joo, D., Woosnam, K. M., Shafer, C. S., Scott, D., & An, S. (2017). Considering Tobler's first law of geography in a tourism context. *Tourism Management*, *62*, 350–359.

Jorgenson, J., Nickerson, N., Dalenberg, D., Angle, J., Metcalf, E., & Freimund, W. (2019). Measuring Visitor Experiences: Creating and Testing the Tourism Autobiographical Memory Scale. *Journal of Travel Research*, *58*(4), 566–578.

Kahle, D. & H. Wickham. (2013). ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Kim, J. H., Ritchie, J. R. B., & McCormick, B. (2012). Development of a scale to measure memorable tourism experiences. *Journal of Travel Research*, *51*(1), 12–25.

Kladou, S., & Mavragani, E. (2015). Assessing destination image: An online marketing approach and the case of TripAdvisor. *Journal of Destination Marketing and Management*, *4*(3), 187–193.

Kozak, M., & Rimmington, M. (1999). Measuring tourist destination competitiveness: Conceptual considerations and empirical findings. International Journal of Hospitality Management, 18(3), 273–284.

Kwenye, J. M., & Freimund, W. (2016). Domestic tourists' loyalty to a local natural tourist setting: Examining predictors from relational and transactional perspectives using a Zambian context. *Tourism Management Perspectives*, *20*, 161–173.

Leung, Y., Spenceley, A., Hvenegaard, G., Buckley, R., & Groves, C. (2018). *Tourism and visitor management in protected areas : guidelines for sustainability*.

Lv, X., Li, C. (Spring), & McCabe, S. (2020). Expanding theory of tourists' destination loyalty: The role of sensory impressions. *Tourism Management*, *77*(November 2019).

Mandić, A., & Petrić, L. (2021). The impacts of location and attributes of protected natural areas on hotel prices: implications for sustainable tourism development. *Environment, Development and Sustainability, 23 (1),* 833-863.

Mangachena, J. R., & Pickering, C. M. (2021). Implications of social media discourse for managing national parks in South Africa. *Journal of Environmental Management*, *285*(September 2020), 112159.

Massara, F., & Severino, F. (2013). Psychological distance in the heritage experience. *Annals of Tourism Research*, *42*, 108–129.

McKercher, B. (2018). The impact of distance on tourism: a tourism geography law. *Tourism Geographies*, *20*(5), 905–909.

McKercher, B., & Chow, B. (2001). Cultural distance and participation in cultural tourism. Pacific Tourism Review, 5(1), 23–32.

McKercher, B., & Mak, B. (2019). The impact of distance on international tourism demand. *Tourism Management Perspectives*, *31*(March), 340–347.

Moore, S. A., Rodger, K., & Taplin, R. (2015). Moving beyond visitor satisfaction to loyalty in nature-based tourism: a review and research agenda. *Current Issues in Tourism*, *18*(7), 667–683.

Mutanga, C. N., Vengesayi, S., Chikuta, O., Muboko, N., & Gandiwa, E. (2017). Travel motivation and tourist satisfaction with wildlife tourism experiences in Gonarezhou and Matusadona National Parks, Zimbabwe. *Journal of Outdoor Recreation and Tourism*, *20*(June), 1–18.

National Park Service (NPS) (2020). National Park Service Visitor Use Statistics from Integrated Resource Management Applications (IRMA). https://irma.nps.gov/STATS

National Park Service (NPS). (2019). National park boundary GIS shapefile. https://public-nps.opendata.arcgis.com/datasets/nps-boundary-1

Neuvonen, M., Pouta, E., Puustinen, J., & Sievänen, T. (2010). Visits to national parks: Effects of park characteristics and spatial demand. *Journal for Nature Conservation*, *18*(3), 224–229.

Norman, P., & Pickering, C. M. (2019). Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. *Journal of Environmental Management*, *249*(July), 109413.

Nyaupane, G. P., & Graefe, A. R. (2008). Travel distance: A tool for nature-based tourism market segmentation. *Journal of Travel and Tourism Marketing*, *25*(3–4), 355–366.

Olipra, L. (2012). The impact of low-cost carriers on tourism development in less famous destinations. Conference proceeding. *CITTASLOW: Il Valore Della Lentezza per Il Turismo Del Futuro*, 41–56.

Oliveira, T., Araujo, B., & Tam, C. (2020). Why do people share their travel experiences on social media? *Tourism Management*, *78*(November 2019), 104041.

Pearce, J., & Dowling, R. (2019). Monitoring the quality of the visitor experience: An evolutionary journey. *Journal of Outdoor Recreation and Tourism*, *25*(November 2017), 87–90.

Pine, B. J. II., & Gilmore, J. H. (1998). Experience Economy. *Harvard Business Review*, (July-Augus), 97.

Pinkus, E., Moore, S. A., Taplin, R., & Pearce, J. (2016). Re-thinking visitor loyalty at 'once in a lifetime' nature-based tourism destinations: Empirical evidence from Purnululu National Park, Australia. *Journal of Outdoor Recreation and Tourism*, *16*, 7–15.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rice, W. L., Park, S. Y., Pan, B., & Newman, P. (2019). Forecasting campground demand in US national parks. *Annals of Tourism Research*, *75*(January), 424–438.

Ritchie, J. R. B. (1993). The measurement of destination image: an empirical assessment. *Journal of Travel Research*, 3–13.

Ritchie, J.R.B.; Crouch, G.I. (2003). The Competitive Destination: A Sustainable Tourism Perspective; CABI: Wallingford, Oxfordshire, UK. ISBN 978-0-85199-664-6.

Roemer, J. M., & Vaske, J. J. (2014). National Park Service Visitor Satisfaction : A Comparative Analysis. *Journal of Park and Recreation Administration*, *32*(4), 35–51.

Rossi, S. D., Byrne, J. A., & Pickering, C. M. (2015). The role of distance in peri-urban national park use: Who visits them and how far do they travel? *Applied Geography*, *63*, 77–88.

Schuett, M. A., Le, L., & Hollenhorst, S. J. (2010). Who visits the U.S. National parks? An analysis of park visitors and visitation: 1990–2008. *World Leisure Journal*, *52*(3), 200–210.

Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. M. (2016). Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management*, *183*, 703–711.

Sharma, P., & Nayak, J. K. (2018). Testing the role of tourists' emotional experiences in predicting destination image, satisfaction, and behavioral intentions: A case of wellness tourism. *Tourism Management Perspectives*, *28*, 41–52.

Shen, X., Huang, R., & Zhang, J. (2019). U.S. national parks accessibility and visitation. *Journal of Mountain Science*, *16*(12), 2894–2906.

Snefjella, B., & Kuperman, V. (2015). Concreteness and psychological distance in natural language use. *Psychological Science*, *26*(9), 1449–1460.

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*(July), 550–568.

Taylor, P., Frost, W., & Laing, J. (2017). Meeting the challenge of managing visitor experiences at tourism attractions. *Visitor Management in Tourism Destinations*, (November 2016), 22–32.

Teles da Mota, V., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism*, *30*(April), 100295.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & 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.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234-240.

Trope, Y., & Liberman, N. (2010). Construal Theory. *Psychological Review*, *117*(2), 440–463.

Vada, S., Prentice, C., & Hsiao, A. (2019). The influence of tourism experience and well-being on place attachment. *Journal of Retailing and Consumer Services*, *47*(December 2018), 322–330.

VISA. (n.d.) Visa Business and Economic Insights; VISIT database; Haver analytics, U.S. National Park Service Visitor Use Statistics. Retrieved from https://usa.visa.com/partner-with-us/visa-consulting-analytics/western-us-national-parks-popular-with-foreign-tourists.html

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham, H. (2020). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.6. https://CRAN.R-project.org/package=rvest

Wijffels, J. (2019). udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit. R package version 0.8.3. https://CRAN.R-project.org/package=udpipe

Wilkins, E. J., Wood, S. A., & Smith, J. W. (2020). Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environmental Management*, 120–132.

World Economic Forum. (2019). *The Travel and Tourism Competitiveness Report 2019.* Retrieved from http://www3.weforum.org/docs/WEF_TTCR_2019.pdf

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65.

Xiao, X., Zhang, J., Lu, J., Li, L., Xu, Y., Zhang, H. lei, & Riungu, G. K. (2021). Distance decay of place attachment and perceived authenticity of mountain tourism destinations in China. *Journal of Mountain Science*, *18*(1), 194–204.

Yeomans, M. (2021). A concrete example of construct construction in natural language. *Organizational Behavior and Human Decision Processes*, *162*(November 2020), 81–94.

Zhou, X., Wang, M., & Li, D. (2017). From stay to play – A travel planning tool based on crowdsourcing user-generated contents. *Applied Geography*, *78*, 1–11.

CHAPTER 6

CONCLUSION

SUMMARY OF THE FINDINGS

This research explores geotagged user-generated content to support place-based conservation in protected areas. Four research objectives in this study involve disentangling the role of humans and nature in the creation of unique place experiences in national parks.

The first research objective is to examine different spatial preferences of attractions depending on visitor characteristics. Based on a general assumption that the geolocations of user-generated content indicate the preference, popularity, as well as aesthetic and recreational values in tourism destinations, the case study utilized the geolocations of Flickr data to identify key attractions in the Great Smoky Mountains National Park in the United States. Then the study compared the popularity of the attraction depending on three tripographic variables of visitors, based on Flickr metadata: the timestamp of the content, visitor logs, and the self-claimed home locations were utilized to retrieve seasonality, previous visit history, and the origin of visitors. The results validated different visitor preferences for attractions depending on tripographic variables with solid geospatial discrepancy and statistical significance. In addition, the findings revealed that local and frequent visitors exhibited similar spatial preferences avoiding popular attractions, whereas first-time, domestic, and international visitors showed a strong preference for easily accessible popular attractions.

The second and third research objectives delve into the application of content analysis to identify site-specific CES values from online tourism reviews of TripAdvisor. The second study focuses on the development and application of a crowdsourced phrasal lexicon to identify eight CES values in El Cajas National Park, Ecuador. The lexicon is innovative to bring phrasal expressions that connotate higher context information than conventional single-word lexicons in CES studies. Also, the phrase-CES pairs in the lexicon were generated from a crowdsourced survey instead of expert-based annotations, bringing public consensus in matching the expression with CES values. After applying the lexicon to retrieve phrasal expressions and CES values from TripAdvisor reviews, the results were compared to distinguish any discrepancies depending on the original language of the content – English or Spanish, which is a popular proxy for the cultural background of visitors. The findings indicated that aesthetic value was the most significant value in the study site regardless of languages, but substantial linguistic distinctions were observed in other CES values. For instance, English reviews perceived more recreational values in the park than Spanish reviews. Spanish reviews, on the other hand, perceived more spiritual values than English reviews. In addition, the result showcased different usage of key phrasal expressions to express each CES value depending on the language.

The third case study expands the applicability of the crowdsourced phrasal lexicon to incorporate phrasal expressions retrieved from TripAdvisor reviews of 48 U.S. National Parks. Accordingly, the lexicon is equipped with phrase-CES pairs from diverse environmental contexts. Upon the completion of the lexicon, the key research objective delves into the destination-oriented comparisons of CES values and phrasal expressions. The findings showcased the diversity of CES values depending on national parks although they are subordinate to the identical managerial systems. The correlation matrix unraveled the co-occurrence of CES values, and the SOM

clustering analysis allowed to categorize national parks into four groups depending on the prevalent CES types. The subsequent word clouds corroborated the prevalent CES values associated with SOM map units. Furthermore, the geospatial distribution of national parks based on four SOM map units demonstrated similar spatial patterns in national park locations.

The last research objective aims to evaluate destination performance based on the visitors and their place experiences. The framework of destination competitiveness consists of two pillars; one is the gravity of the destinations to attract visitors from varying origins, and another is the ability of the destinations to provide quality experiences. The quantification of two variables was conducted with TripAdvisor metadata and reviews, respectively: the gravity was measured with the origin of visitors and the quality of place experience was assessed with textual concreteness from the construal level theory that indicates psychological affinity. Therefore, the study examined the positive correlation between place experience and place affinity. After discerning four types of gravity variables and another four types of affinity variables, the results evaluated the performance of national parks with geographic distinction. Overall, the findings highlighted the data-driven performance of national parks based on geotagged UGC.

IMPLICATIONS, UNCERTAINTY, AND FUTURE DIRECTIONS

Scientific implications

The studies contribute to advancing scientific and technical metrics of geotagged user-generated content in exploring the human-nature relationship across national parks with an emphasis on CES values. One of the key contributions in this work is the development of a crowdsourced phrasal lexicon. Compared to the single-word lexicons in previous studies, the lexicon enabled to detection of CES values based on phrasal expressions that have higher context

information. In addition, the lexicon was developed from the crowdsourced survey that allowed to bring a public understanding of CES values, instead of the conventional expert-based annotations. Lastly, the lexicon is applicable for various environmental settings ranging from high-altitude grasslands, boreal forests, temperate forests, deserts, wetlands, and islands, since it was developed from TripAdvisor reviews from more than fifty national parks in the U.S. Therefore, the lexicon, which is now available on GitHub (https://github.com/ihKong/phrasal_lexicon_CES), is expected to play as a reference for the subsequent studies to save time and workforces in developing site-specific lexicons while allowing for the modification and customization for their lexicons.

Policy implications

Analyzing place experiences and multiple CES values from large user-generatec content in PA management can bring several benefits. First, CES-conscious management plans can benefit both visitors and destination management based on the positive feedback loop (see Figure 1-2). According to the empirical studies, memorable place experiences that involve a variety of experiences and cognitive perceptions can improve the satisfaction of visitors, further developing place attachment and destination loyalty (Moore, Rodger, and Taplin, 2015; Sharma and Nayak, 2018; Vada, Prentice, and Hsiao, 2019). Consequently, positive tourism experiences and place attachment can reduce psychological distance and improve place affinity to the place, resulting to engender strong public support and engagement for the conservation actions (Buta et al., 2014; Weaver and Lawton, 2017). In addition, acknowledging diverse CES values can raise public awareness that conservation policy is relevant to personal and community well-being (Bullock et al., 2018; Chan et al., 2016; Kosanic and Petzold, 2020; Plieninger et al., 2015). These benefits

align with the key message in Relph (1976) that strong place attachment and rootedness lead to "a sense of deep care and concern for that place (p. 37)".

Second, acknowledging multiple landscape values on top of the biodiversity can bring additional values to support the rationales for PA management (Chan et al., 2006; Lessa et al., 2021). Indeed, PA management requires substantial financial budgets for the general maintenance of the landscape integrity and the recreational facilities. The massive cost for reserving a huge landmass for the conservation of ecological and cultural heritage has been often criticized for substantial opportunity costs and trade-offs (Adams, Pressey, and Naidoo, 2010; Nelson et al., 2009). Therefore, identifying a variety of CES values that directly contribute to human well-being can corroborate the rationale for the expenditures for conservation (Lessa et al., 2021). Furthermore, the perception of such values that directly benefits the general public can boost collective support for the inward conservation investment to the PAs that are often inadequately resourced (Lessa et al., 2021). Data-driven identification and quantitation of site-specific CES values from geotagged UGC can also ease the communication between visitors, park rangers, and stakeholders for comprehensive conservation planning of PAs (Daniel et al., 2012).

Third, site-specific CES values can play as tourist niche for tourism destination management in PAs. As tourism markets become competitive, tourism destinations have been seeking unique tourism niches that can attract visitors. The emergence of the *experience economy*, therefore, highlighted the importance of engaging customer experiences for successful business models (Pine and Gilmore, 1998), and tourism destinations were not an exception. Tourism destinations have also been implementing memorable tourism experiences that provide a memorable and unforgettable experience for visitors (Kim, Ritchie, and McCormick, 2012). As prevalent nature-based tourism destinations, national parks have been introducing interactive

visitor experiences that involve wildlife watching or cultural and historic education to nurture sociocultural values in addition to aesthetic and recreational values (Pettebone and Meldrum, 2018). Overall, understanding unique CES values and place experiences can assist tourism destination marketing and management by differentiating the park from rival destinations.

Uncertainty

Despite the advances, this work is not without its shortcomings. First, the analyses have largely relied on user-generated content with few attempts for validations. This is largely due to the lack of available data to juxtapose the original data or validate the results. Particularly, compared to urban tourism experiences (i.e., restaurants or hotels) that can be obtained from Yelp or Expedia, tourism experiences in natural landscapes are extremely rare. Accordingly, the future direction of content analysis to identify landscape values and CES needs to explore hybrid approaches to obtain comparable data that comes directly from the public. For instance, public participatory GIS can be an alternative to obtain public responses with little geographic, temporal, or workforce-related limitations. In addition, it can ask questions that directly relate to research purposes, such as CES values or perceived landscape values, instead of indirect interpretations from the user-generated content.

Another limitation in this work involves the crowdsourced phrasal lexicon. When developing the crowdsourced phrasal lexicon, the crowdsourced survey to annotate relevant CES to the phrases has been exclusively operated with the residents of the U.S., due to the key participant pools from Amazon Mechanical Turk. Also, the survey has gathered fifteen responses for each phrase, which may not be enough to represent the general population. For example, two similar phrases ‘see bear’ and ‘see bear cub’ were annotated with educational value and biological

value, respectively. Therefore, to improve the universality and the reliability of the lexicon, the additional iteration of the crowdsourced survey that includes more diverse respondent backgrounds will be necessary. Second, current analysis with the POS-based phrasal retrieval can find the same phrases from TripAdvisor reviews to annotate relevant CES values from the lexicon. This may exclude some relevant expressions when the phrases in TripAdvisor reviews have slight aberrations. For instance, the current method can detect 'see bear' to annotate educational value but 'see four bear(s)' will be excluded. Therefore, the probabilistic approach in text analysis needs to be considered in the retrieval of CES values. Furthermore, although the lexicon has been developed from text content encompassing diverse environmental settings, it does not necessarily include all the site-specific expressions and values. As a result, I anticipate that the publicly accessible lexicon will serve as a reference for other researchers to modify and customize the lexicon to contribute to improving the lexicon's applicability, reproducibility, and reliability.

Future directions

For future directions of this work, I propose two suggestions. First, the theoretical and technological approaches in this study can be further applied in landscapes that are not designated in institutional conservation systems (i.e., protected areas) but have unique place values. Indeed, the global effort for landscape management has resulted in an unprecedented number of PAs, yet numerous landscapes remain overshadowed. Especially where radical economic growth and urban expansion are undergoing, such landscapes face severe threats in safeguarding highly valuable heritage that has been created, accumulated, and shared over time. Therefore, recognizing the importance of landscapes via the lens of CES can boost public awareness of the need for conservation and provide scientific and quantifiable data for decision-makers. The framework of

CES in conservation policy may also result in the adoption of an unconventional conservation strategy that does not need an institutional conservation scheme, but rather autonomous and bottom-up conservation measures driven by the community or local government.

Second, the outline of place-based conservation presents the mechanism of understanding and identifying CES values in the ultimate good for conservation planning that both benefits people and nature. Yet, this work largely focused on the identification, quantification, and comparison of CES values, while one case study was spared to examine the psychological affinity as a result of place experiences. Therefore, the subsequent mechanisms of place-based conservation need to be thoroughly examined. To be more specific, the beneficial influence of perceiving diverse CES values on enhancing satisfaction, place attachment, and willingness to support and engage in conservation actions should be examined to provide solid and trustworthy evidence for the efficacy of place-based conservation.

REFERENCES

Adams, V. M., Pressey, R. L., & Naidoo, R. (2010). Opportunity costs: Who really pays for conservation? *Biological Conservation*, *143*(2), 439–448.

Bullock, C., Joyce, D., & Collier, M. (2018). An exploration of the relationships between cultural ecosystem services, socio-cultural values and well-being. *Ecosystem Services*, *31*, 142–152.

Buta, N., Holland, S. M., & Kaplanidou, K. (2014). Local communities and protected areas: The mediating role of place attachment for pro-environmental civic engagement. *Journal of Outdoor Recreation and Tourism*, *5–6*, 1–10.

Chan, K. M. A., Balvanera, P., Benessaiah, K., Chapman, M., Díaz, S., Gómez-Baggethun, E., … Turner, N. (2016). Why protect nature? Rethinking values and the environment. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(6), 1462–1465.

Chan, K. M. A., Shaw, M. R., Cameron, D. R., Underwood, E. C., & Daily, G. C. (2006). Conservation planning for ecosystem services. *PLoS Biology*, *4*(11), 2138–2152.

Daniel, T. C., Muhar, A., Arnberger, A., Aznar, O., Boyd, J. W., Chan, K. M. A., … Von Der Dunk, A. (2012). Contributions of cultural services to the ecosystem services agenda. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 8812–8819.

Kim, J. H., Ritchie, J. R. B., & McCormick, B. (2012). Development of a scale to measure memorable tourism experiences. *Journal of Travel Research*, *51*(1), 12–25.

Kosanic, A., & Petzold, J. (2020). A systematic review of cultural ecosystem services and human wellbeing. *Ecosystem Services*, *45*(October 2019), 101168.

Lessa, T., Jepson, P., Bragagnolo, C., Campos-Silva, J. V., Barros, E., Gomes, F., … Ladle, R. J. (2021). Revealing the hidden value of protected areas. *Land Use Policy*, (February), 105733.

Moore, S. A., Rodger, K., & Taplin, R. (2015). Moving beyond visitor satisfaction to loyalty in nature-based tourism: a review and research agenda. *Current Issues in Tourism*, *18*(7), 667–683.

Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D. R., … Shaw, M. R. (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment*, *7*(1), 4–11.

Pettebone, D., & Meldrum, B. (2018). The Need for a Comprehensive Socioeconomic Research Program for the National Park Service. *The George Wright Forum*, *35*(1), 22–31.

Pine, J., & Gilmore, J. H. (1998). Welcome to the experience economy. *Harvard Business Review*, *76*(4), 97–105.

Plieninger, T., Bieling, C., Fagerholm, N., Byg, A., Hartel, T., Hurley, P., … Huntsinger, L. (2015). The role of cultural ecosystem services in landscape management and planning. *Current Opinion in Environmental Sustainability*, *14*, 28–33.

Relph, E. (1976). Place and placelessness. London: Pion Limited.

Sharma, P., & Nayak, J. K. (2018). Testing the role of tourists' emotional experiences in predicting destination image, satisfaction, and behavioral intentions: A case of wellness tourism. *Tourism Management Perspectives*, *28*, 41–52.

Vada, S., Prentice, C., & Hsiao, A. (2019). The influence of tourism experience and well-being on place attachment. *Journal of Retailing and Consumer Services*, *47*(December 2018), 322–330.

Weaver, D. B., & Lawton, L. J. (2017). A new visitation paradigm for protected areas. *Tourism Management, 60*, 140–146.

APPENDICES.

APPENDIX A. THE PART-OF-SPEECH (POS) LIBRARY TO RETRIEVE CRUDE NOUN PHRASES AND VERB PHRASES (CHAPTER 3 AND 4)

Phrasal type	Model structures in the R script	POS structures [a]	Example phrases [b]
NP	((ADJ|NOUN)*NOUN$)	ADJ ADJ NOUN	'breathe fresh air'
		ADJ ADJ NOUN NOUN	'many high mountain lagoon'
	(NOUN*ADP*NOUN$)	NOUN ADP NOUN	'herb with liquor'
	(ADJ*CCONJ*ADJ*NOUN$)	ADJ CCONJ ADJ NOUN	'unique and spectacular place'
	(NOUN*CCONJ*NOUN$)	NOUN CCONJ NOUN	'charm and mystery'
	((ADJ|NOUN)*NOUN*PART*VERB*(ADP|ADJ|NOUN)*NOUN$)	ADJ NOUN PART VERB ADJ NOUN	'spectacular opportunity to admire andean landscape'
		ADJ NOUN PART VERB ADP NOUN ADP NOUN	'ideal place to be in contact with nature'
	(ADJ*NOUN*ADP*PROPN$)	ADJ NOUN ADP PROPN	'cultural heritage of Ecuador'
	(ADJ*PROPN$)	ADJ PROPN	'beautiful Cajas'
		ADJ PROPN	'beautiful National Park'
VP	(VERB*ADV$)	VERB ADV	'walk outdoors'
	(VERB*ADV*(ADP|ADJ|NOUN)*NOUN$)	VERB ADV NOUN NOUN NOUN	'be also sport fishing activity'
		VERB ADV ADP NOUN ADP NOUN NOUN	'stand there in middle of mountain range'
	(VERB*PART*VERB*((ADP|ADJ|VERB)*NOUN))	VERB PART VERB VERB NOUN	'have to try grill trout'
		VERB PART VERB NOUN	'want to see nature'
		VERB PART VERB ADP NOUN	'have to register at park entrance'
	(VERB*(ADP|ADJ|NOUN|PRON)*NOUN$)	VERB ADP NOUN	'visit with family '
		VERB ADP ADJ ADJ NOUN	'walk in pristine natural setting'
		VERB ADJ NOUN	'enjoy beautiful landscape'
		VERB ADJ ADJ NOUN	'include regional endemic specie'
		VERB ADJ NOUN ADP NOUN	'choose short route around lake'
		VERB ADJ ADJ NOUN ADP NOUN	'see many different specie of bird'
		VERB ADJ ADJ ADJ NOUN	'have great little interpretive museum'
		VERB ADJ ADP ADJ NOUN	'know more about beautiful place'
		VERB ADJ ADP ADJ NOUN ADP ADJ NOUN	'struggle little with thin air at high altitude'
		VERB ADJ ADP NOUN	'be high in altitude'
		VERB ADJ ADP NOUN ADP ADJ NOUN	'be different from rest of ecuadorian highland'
	(VERB*(ADJ|NOUN)*CCONJ*(ADJ|NOUN)*NOUN$)	VERB ADJ CCONJ ADJ NOUN	'have relaxing and interesting stroll'
		VERB ADJ ADJ NOUN CCONJ NOUN	'be many beautiful landscape and vegetation'
		VERB ADJ NOUN CCONJ ADJ ADJ NOUN	'see beautiful orchid and many other plant'

[a] The POS structures shown in this column do not represent all of the structures that could be found with the model structure.
[b] The exemplary phrases shown in this column may not have been included in the final lexicon in this study.

APPENDIX B. THE SUMMARY OF COLLECTED DATA (CHAPTER 4 AND 5)

	National Park	TripAdvisor			Annual visitors (2019) [*1]
		Number of attractions	**Number of reviews (pre-pandemic)**	Number of valid single visitors	
1	Grand Canyon National Park	30	31,306	19179	5,974,411
2	Yellowstone National Park	30	29,616	9792	4,020,288
3	Yosemite National Park	30	24,657	10521	4,422,861
4	Great Smoky Mountains National Park	30[Tennessee], 20[North Carolina]	22,846	12982	12,547,743
5	Zion National Park	30	21,224	10335	4,488,268
6	Bryce Canyon National Park	30	13,221	7146	2,594,904
7	Death Valley National Park	30	13,080	4260	1,740,945
8	Acadia National Park	30	11,872	5484	3,437,286
9	Glacier National Park	30	10,987	4728	3,049,839
10	Rocky Mountain National Park	30	9,700	5487	4,670,053
11	Kings Canyon National Park	30	8,812	3426	632,110
-	Sequoia National Park	N/A [*4]			1,246,053
12	Grand Teton National Park	30	8,536	4429	3,405,614
13	Olympic National Park	30	8,481	3889	3,245,806
14	Badlands National Park	25	5,544	2987	970,998
15	Crater Lake National Park	30	5,153	3245	704,512
16	Arches National Park	30	4,936	2132	1,659,702
17	Carlsbad Caverns National Park	16	4,920	2886	440,691
18	Shenandoah National Park	23	4,740	3109	1,425,507
19	Mammoth Cave National Park	11	4,462	3394	551,590
20	Mesa Verde National Park	21	4,274	2153	556,203
21	Everglades National Park	30	4,197	2320	1,118,300
22	Mount Rainier National Park	30	3,925	2330	1,501,621
23	Petrified Forest National Park	18	3,902	2158	643,588
24	Capitol Reef National Park	30	3,812	1880	1,226,519
25	Big Bend National Park	30	3,765	1313	463,832
26	White Sands National Park	Single [*3]	3,570	3060	608,785
27	Joshua Tree National Park	30	2,987	1353	2,988,547
28	Canyonlands National Park	30	2,844	1304	733,996
29	Lassen Volcanic National Park	24	1,652	914	517,039
30	Black Canyon of the Gunnison National Park	12	1,632	1028	432,818
31	Redwood National Park	5	1,116	856	504,722
32	North Cascades National Park	26	1,000	546	38,208

33	Saguaro National Park	Single [*3]	879	800	1,020,226
34	Hot Springs National Park	Single [*3]	857	745	1,467,153
35	Cuyahoga Valley National Park	Single [*3]	754	661	2,237,997
36	Theodore Roosevelt National Park	11	723	417	691,658
37	Guadalupe Mountains National Park	14	657	408	188,833
38	Wind Cave National Park	5	609	534	615,350
39	Great Basin National Park	11	520	315	131,802
40	Pinnacles National Park	Single [*3]	519	474	177,224
41	Dry Tortugas National Park	4	516	375	79,200
42	Congaree National Park	Single [*3]	498	429	159,445
43	Channel Islands National Park	8	489	353	409,630
44	Indiana Dunes National Park	Single [*3]	361	316	2,134,285
45	Great Sand Dunes National Park	5	340	263	527,546
46	Voyageurs National Park	15	235	148	232,974
47	Isle Royale National Park	10	225	126	26,410
48	Biscayne National Park	3	101	80	708,522

[*1] National Park Service (NPS) Integrated Resource Management Applications (IRMA). (2020). National Park Service Visitor Use Statistics. https://irma.nps.gov/STATS

[*2] National Park Service (NPS) Integrated Resource Management Applications (IRMA). (na) National Park GIS shapefile https://irma.nps.gov/DataStore/Reference/Profile/2224545?lnv=True. Cross-validated with the official report published in Dec 2020: https://www.nps.gov/subjects/lwcf/upload/NPS-Acreage-12-31-2020.pdf

[*3] NP as a single attraction in TripAdvisor

[*4] Sequoia National Park was merged to Kings Canyon National Park in this study

APPENDIX C. THE OPERATION OF CROWDSOURCED SURVEY TO DEVELOP A CROWDSURCED PHRASAL LEXICON (CHAPTER 3 AND 4)

We operated a crowdsourced survey to match the phrases with eight CES categories, using Amazon Mechanical Turk (MTurk). The survey design was largely indebted to Brysbaert et al. (2014)[5].

The survey was designed using Qualtrics. On the front pages, we outlined the consent letter that does not require signatures, followed by the descriptions and exemplary phrases of eight CES categories (see Table 1 in the paper). The body of the survey was filled with phrases to be classified to CES. Since we had thousands of phrases, we randomly divided the phrases into 47 groups so that we can lessen the burden for survey respondents while maximizing the survey coherence. As a result, each survey had 112 phrases and 8 caliber phrases for quality control. The caliber words included 'beautiful view' for aesthetic value, 'historical building' for cultural value, 'learn about ecology' for educational value, 'national treasure' for identity value, 'alpine forest' for biological value, 'go with family' for social value, 'hike around lake' for recreational value, 'enjoy tranquility' for spiritual value. The 120 phrases were then listed in alphabetical order. Each phrase was allowed to match up with two CES, and it can be also left blank if no CES is relevant. On the last page, we described how to claim a reward in Amazon MTurk when they finish the survey on Qualtrics. The estimated time for finishing the survey was 20 minutes.

The URLs for the Qualtrics were then posted on the Amazon Mturk website. The surveys were only accessible to Mturkers whose HIT Approval Rate was greater than 95% and the number of approved HITs was more than 500. Also, all MTurkers are older than 18 years old. The surveys

[5] Brysbaert, M., Warriner, A.B., Kuperman, V., 2014. Concreteness ratings for 40 thousand generally known English word lemmas. Behav. Res. Methods 46, 904–911.

were approved to be IRB-exempt by the University of Georgia IRB office. When the survey hits 15 responses, the link is automatically closed in Amazon MTurk. Then we examined the survey responses from Qualtrics, focusing on the positive rates for the caliber phrases and the time durations to finish the survey. Also, we reviewed whether the Qualtrics survey was taken by different individuals. As a result, we selectively approved the Mturk responses, while rejecting unqualified ones. The satisfactory MTurk survey takers were then rewarded with $1.5, considering the Fair Labor Standards Act which set the federal minimum hourly wage at $7.25 as of 2021.

Overall, the final lexicon in this study is available on *https://github.com/ihKong/phrasal_lexicon_CES*. In addition to relevant CES and phrasal types (i.e., noun phrases or verb phrases), we also attached the applicability information (i.e., CNP-specific or universal) according to our discretion.

APPENDIX D. THE FREQUENCY OF EIGHT CES VALUES ACROSS 48 U.S. NATIONAL PARKS (CHAPTER 4)

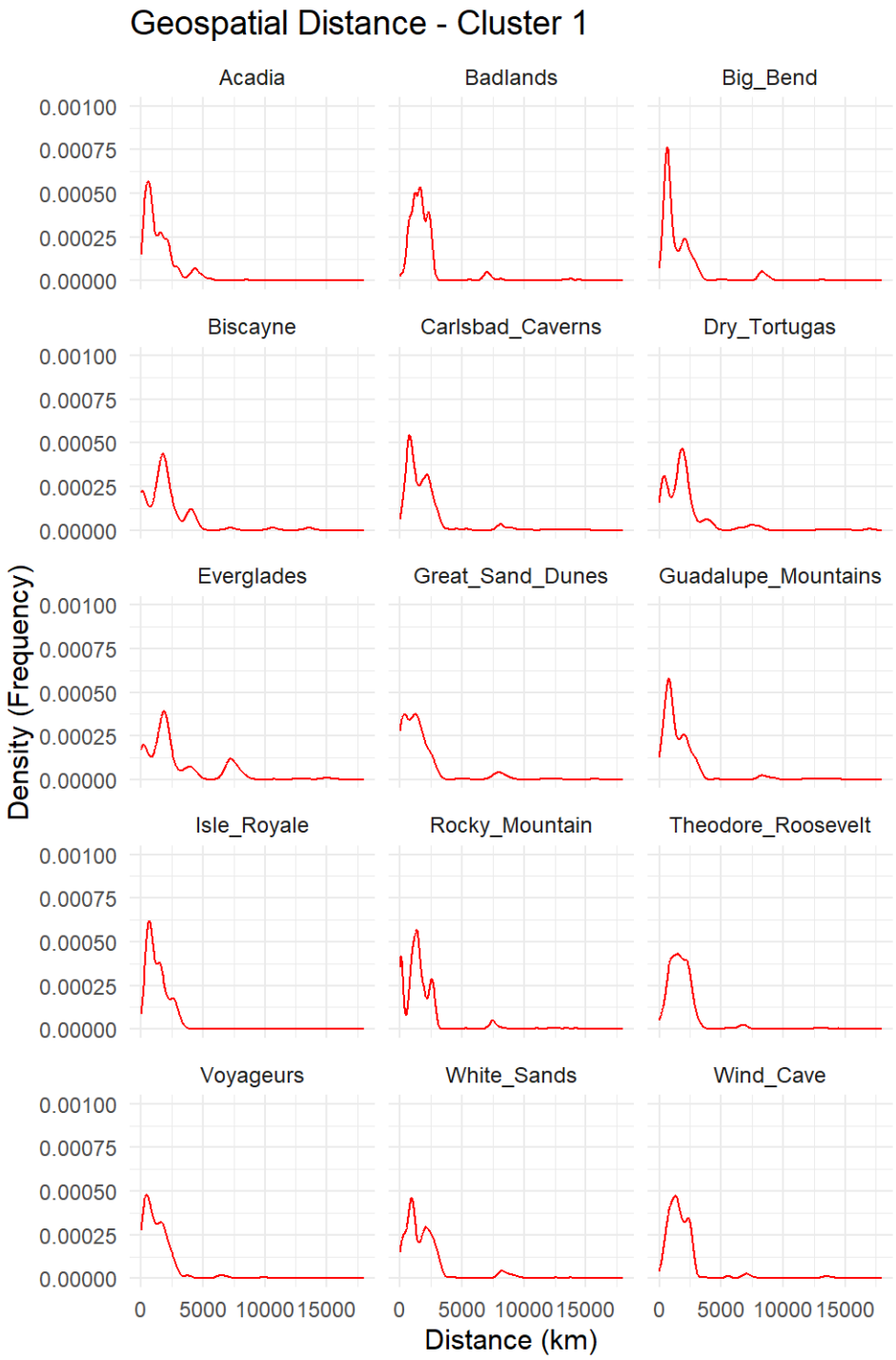
National Parks	Aesthetic	Biological	Cultural	Educational	Identity	Recreational	Social	Spiritual	Type
Acadia	0.532	0.043	0.004	0.013	0.095	0.490	0.065	0.034	(a)
Arches	0.397	0.029	0.009	0.022	0.168	0.661	0.049	0.025	(c)
Badlands	0.464	0.190	0.016	0.049	0.108	0.396	0.046	0.026	(a)
Big_Bend	0.430	0.104	0.031	0.032	0.093	0.613	0.040	0.020	(c)
Biscayne	0.356	0.129	0.040	0.040	0.099	0.673	0.089	0.010	(c)
Black_Canyon_of_the_Gunnison	0.558	0.068	0.018	0.062	0.105	0.477	0.044	0.025	(a)
Bryce_Canyon	0.519	0.051	0.012	0.027	0.158	0.532	0.036	0.035	(a)
Canyonlands	0.552	0.023	0.010	0.019	0.114	0.550	0.037	0.027	(a)
Capitol_Reef	0.393	0.048	0.078	0.067	0.161	0.533	0.042	0.025	(c)
Carlsbad_Caverns	0.215	0.129	0.013	0.064	0.127	0.422	0.088	0.037	(d)
Channel_Islands	0.329	0.350	0.025	0.053	0.092	0.728	0.119	0.053	(b)
Congaree	0.211	0.414	0.026	0.060	0.133	0.685	0.074	0.070	(b)
Crater_Lake	0.560	0.045	0.018	0.037	0.163	0.580	0.057	0.031	(a)
Cuyahoga_Valley	0.301	0.119	0.070	0.069	0.098	0.585	0.084	0.024	(d)
Death_Valley	0.384	0.024	0.021	0.036	0.170	0.394	0.020	0.022	(a)
Dry_Tortugas	0.339	0.138	0.091	0.109	0.097	0.614	0.109	0.054	(c)
Everglades	0.216	0.530	0.006	0.048	0.028	0.587	0.047	0.024	(b)
Glacier	0.520	0.206	0.008	0.024	0.106	0.569	0.057	0.028	(a)
Grand_Canyon	0.528	0.055	0.028	0.032	0.092	0.430	0.052	0.037	(a)
Grand_Teton	0.545	0.198	0.031	0.038	0.075	0.433	0.045	0.029	(a)
Great_Basin	0.298	0.164	0.015	0.060	0.192	0.548	0.073	0.025	(c)
Great_Sand_Dunes	0.350	0.032	0.018	0.050	0.306	0.553	0.112	0.024	(c)
Great_Smoky_Mountains	0.470	0.217	0.097	0.101	0.070	0.578	0.086	0.048	(a)
Guadalupe_Mountains	0.458	0.148	0.038	0.044	0.160	0.714	0.064	0.037	(c)
Hot_Springs	0.256	0.023	0.138	0.095	0.116	0.461	0.068	0.012	(d)
Indiana_Dunes	0.344	0.055	0.031	0.031	0.155	0.551	0.122	0.022	(c)
Isle_Royale	0.444	0.200	0.031	0.107	0.107	0.729	0.036	0.044	(c)
Joshua_Tree	0.379	0.114	0.014	0.019	0.184	0.567	0.050	0.019	(c)
Kings_Canyon	0.386	0.270	0.017	0.041	0.082	0.494	0.048	0.038	(b)
Lassen_Volcanic	0.438	0.101	0.082	0.068	0.200	0.617	0.064	0.035	(c)
Mammoth_Cave	0.147	0.050	0.108	0.140	0.178	0.479	0.138	0.021	(d)
Mesa_Verde	0.223	0.028	0.219	0.075	0.060	0.498	0.058	0.018	(d)
Mount_Rainier	0.526	0.189	0.020	0.035	0.105	0.579	0.064	0.029	(a)
North_Cascades	0.633	0.104	0.003	0.017	0.109	0.522	0.050	0.019	(a)
Olympic	0.480	0.221	0.007	0.016	0.185	0.540	0.061	0.042	(a)
Petrified_Forest	0.453	0.276	0.085	0.060	0.085	0.370	0.030	0.025	(a)
Pinnacles	0.378	0.208	0.021	0.027	0.268	0.730	0.091	0.033	(c)

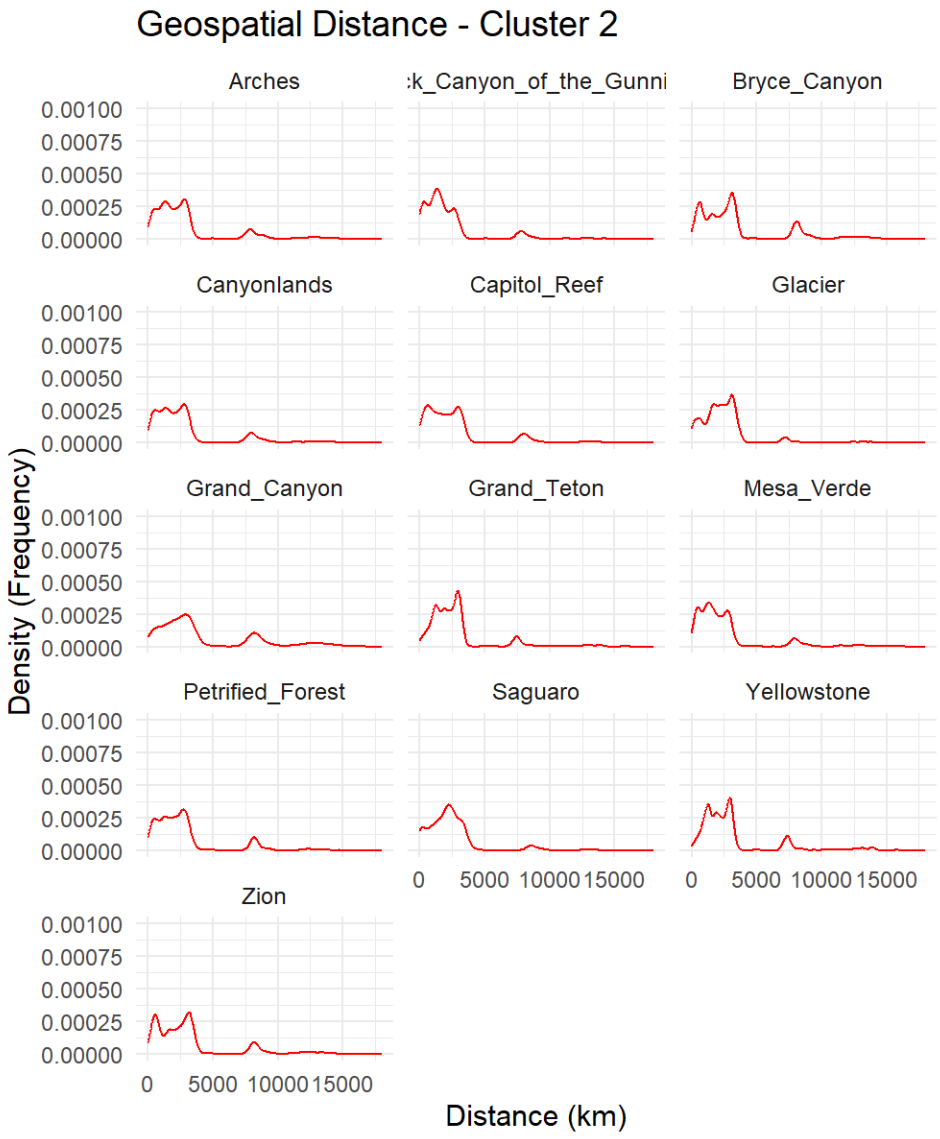
Redwood	0.359	0.404	0.009	0.016	0.078	0.487	0.057	0.081	(b)
Rocky_Mountain	0.557	0.212	0.006	0.013	0.113	0.577	0.068	0.032	(a)
Saguaro	0.382	0.126	0.035	0.042	0.063	0.517	0.036	0.030	(a)
Shenandoah	0.548	0.185	0.010	0.044	0.067	0.591	0.065	0.040	(a)
Theodore_Roosevelt	0.593	0.400	0.026	0.042	0.098	0.515	0.044	0.033	(b)
Voyageurs	0.285	0.153	0.021	0.043	0.060	0.532	0.106	0.013	(d)
White_Sands	0.422	0.120	0.036	0.040	0.422	0.561	0.120	0.047	(c)
Wind_Cave	0.166	0.177	0.036	0.097	0.182	0.522	0.099	0.016	(d)
Yellowstone	0.384	0.184	0.019	0.026	0.155	0.336	0.028	0.024	(a)
Yosemite	0.518	0.081	0.011	0.021	0.093	0.454	0.046	0.034	(a)
Zion	0.440	0.063	0.007	0.013	0.094	0.598	0.058	0.030	(c)

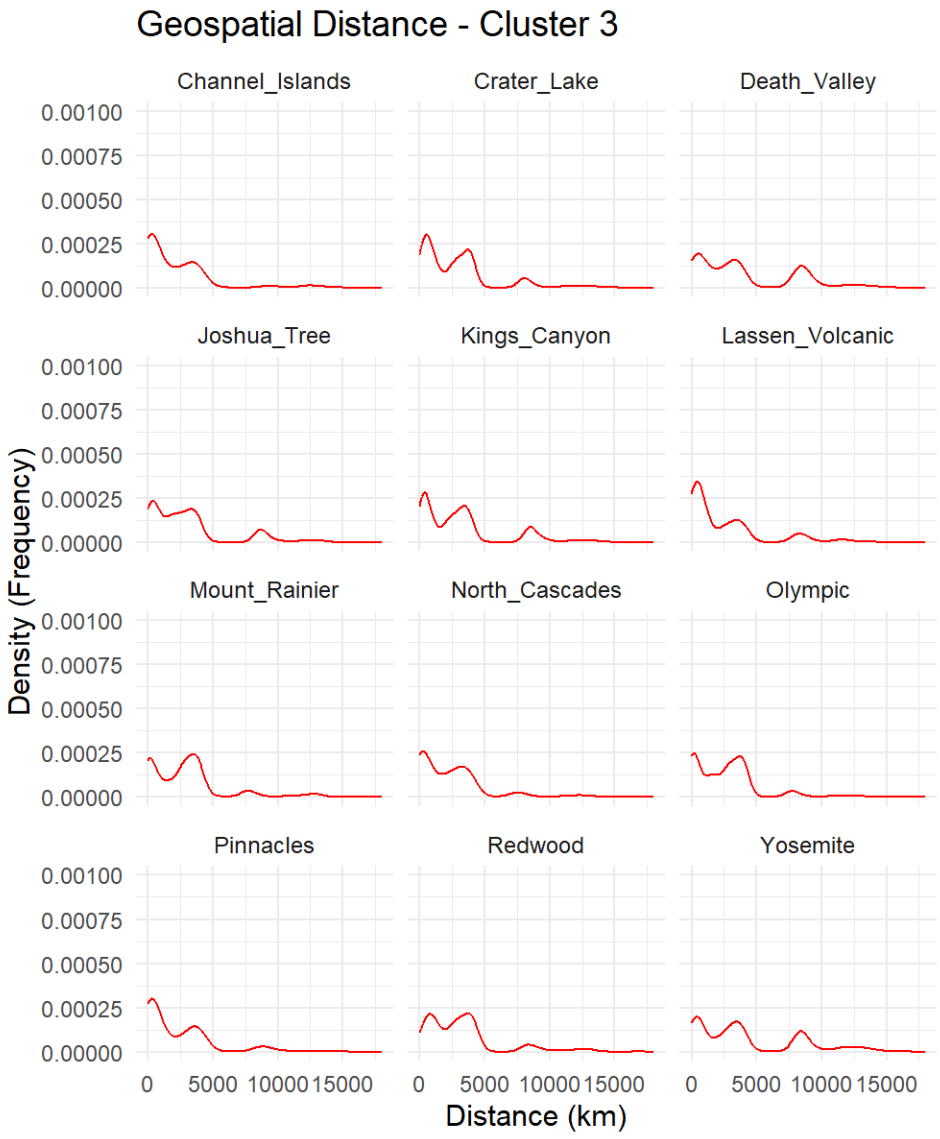
[*] *In the Type column, (a), (b), (c), and (d) indicate (a) Aesthetic, (b) Biological/Spiritual, (c) Recreational/Identity, and (d) Cultural/Educational/Social from SOM clustering analysis.*
[**] *The shaded green color indicates the normalized min-max values for each column.*

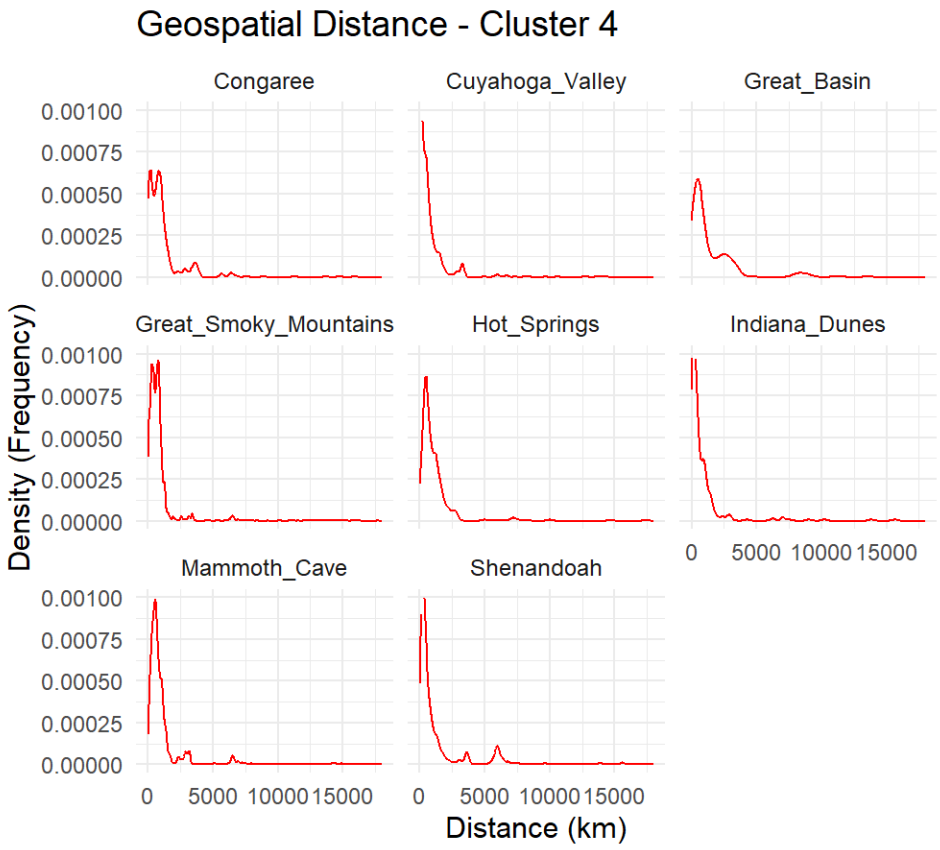
APPENDIX E. SUPPLEMENTARY MATERIALS FOR CHAPTER 5

Appendix E-1. Visitation frequency by geographic travel distance

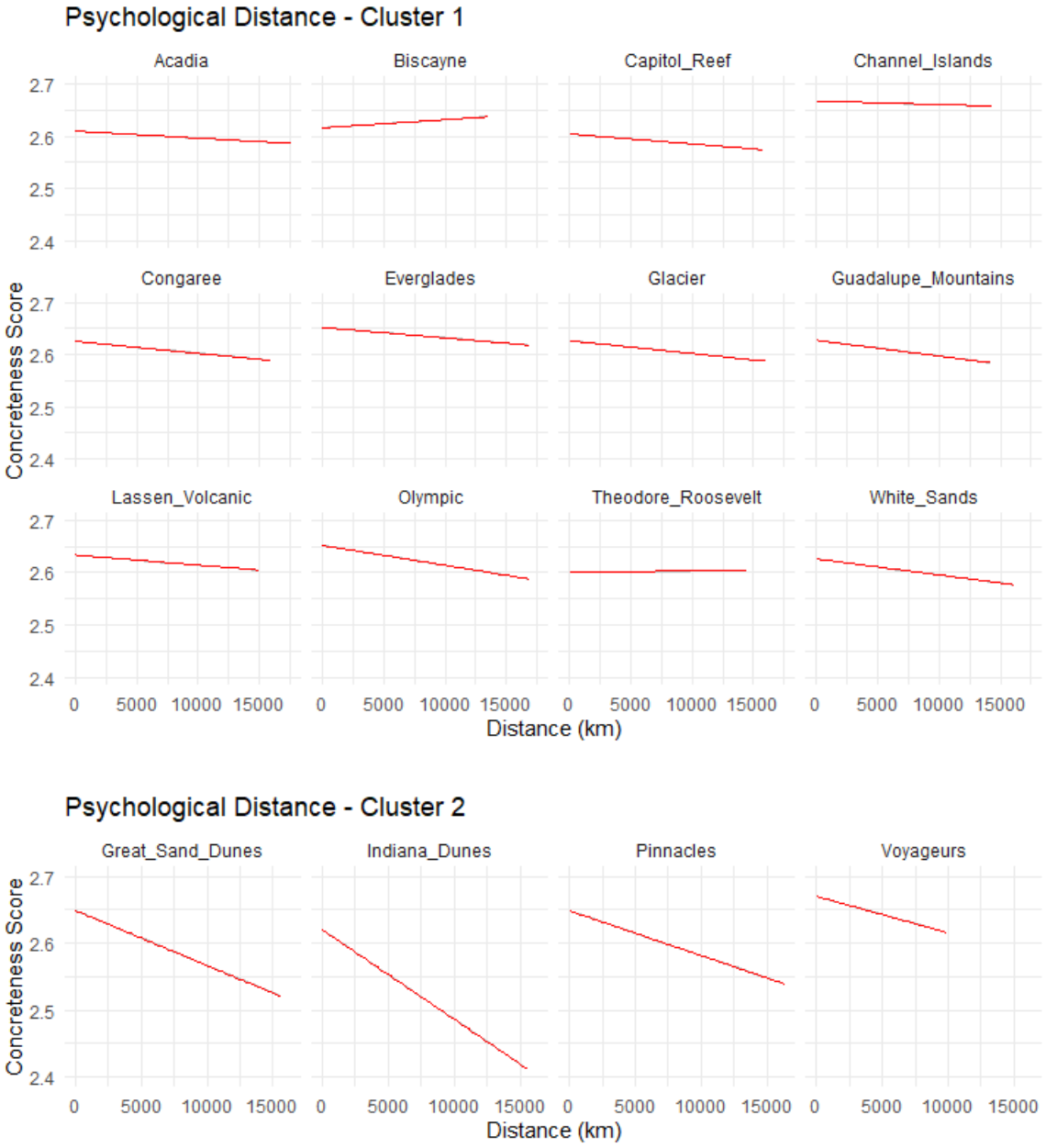


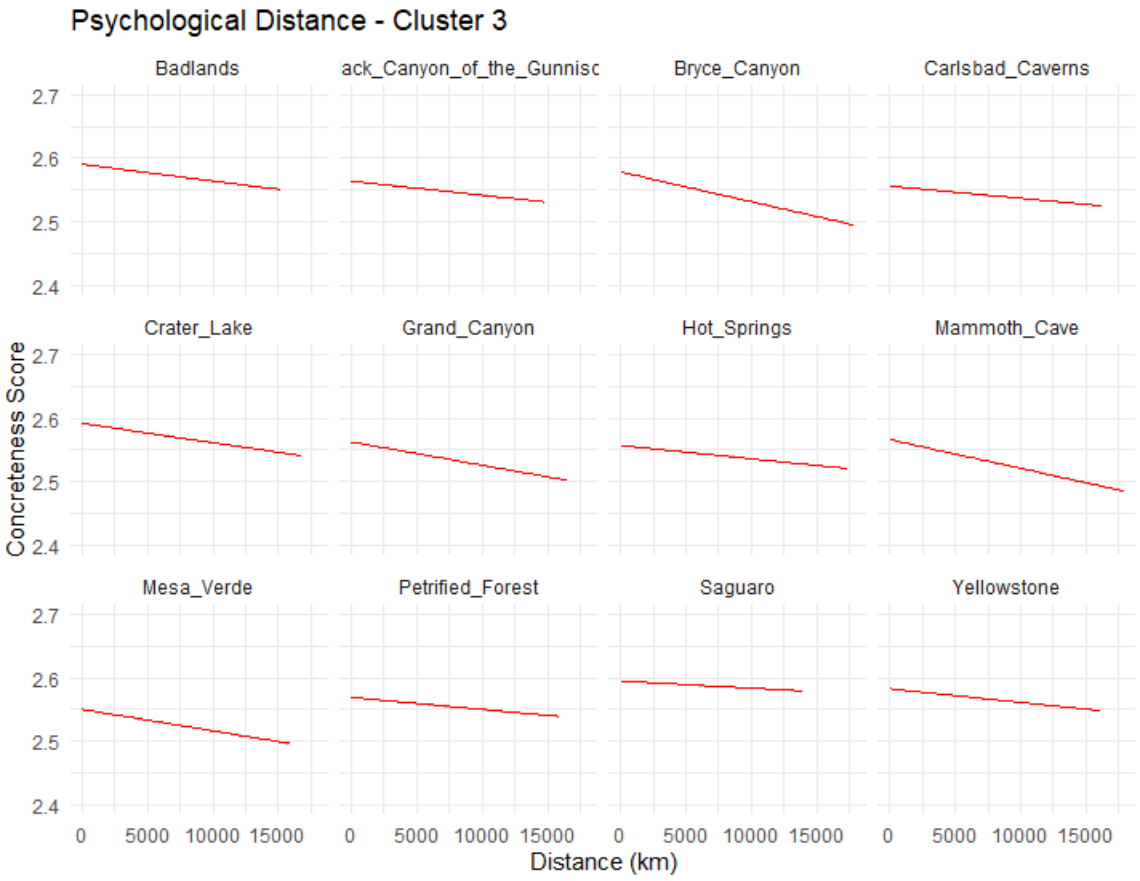

Geospatial Distance - Cluster 2
Arches
:k_Canyon_of_the_Gunni
Bryce_Canyon
Canyonlands
Capitol_Reef
Glacier
Grand_Canyon
Grand_Teton
Mesa_Verde
Petrified_Forest
Saguaro
Yellowstone
Zion
Density (Frequency)
0.00100
0.00075
0.00050
0.00025
0.00000
0
5000
10000
15000
Distance (km)

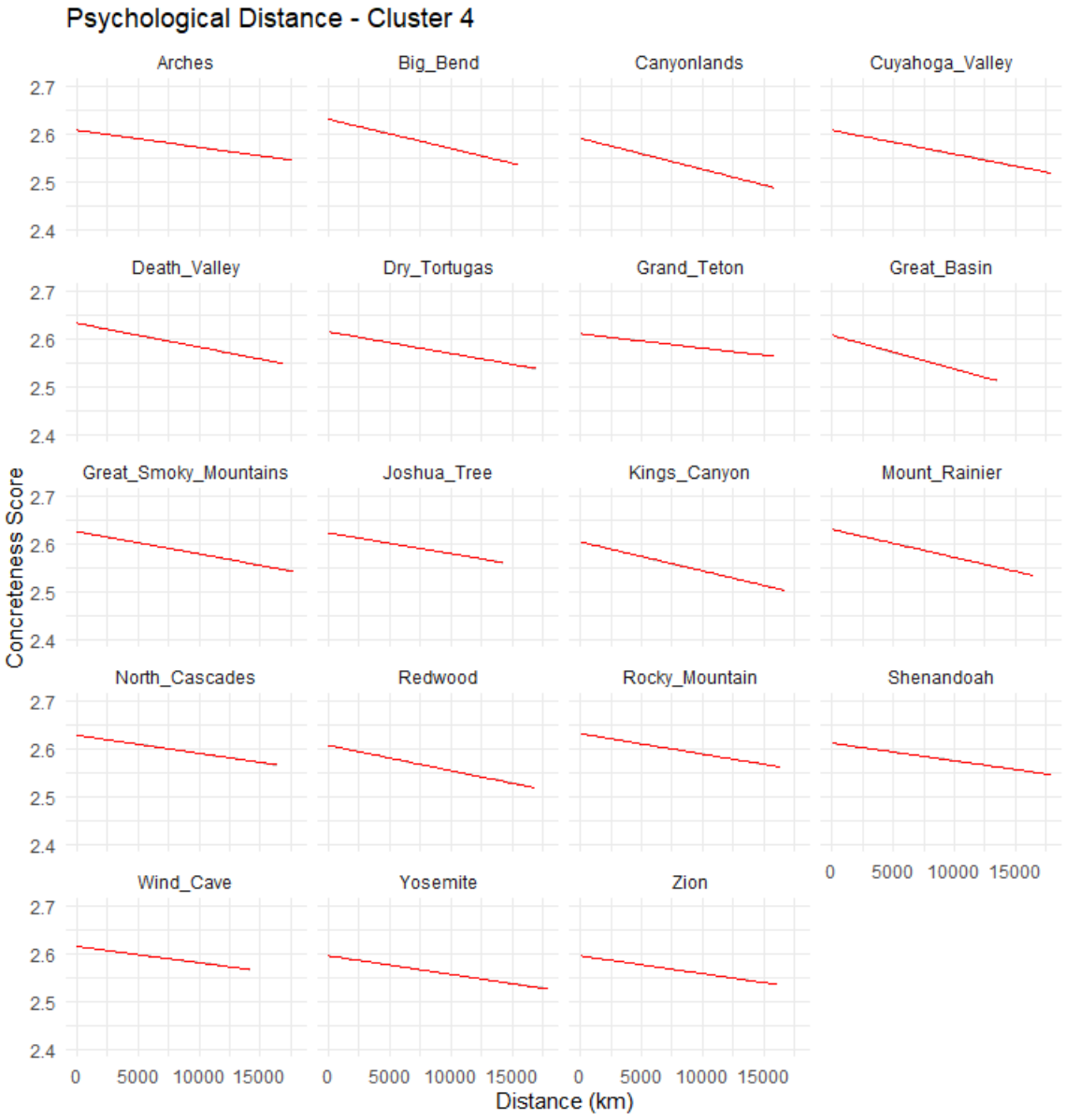

Geospatial Distance - Cluster 3
Channel_Islands
Crater_Lake
Death_Valley
Joshua_Tree
Kings_Canyon
Lassen_Volcanic
Mount_Rainier
North_Cascades
Olympic
Pinnacles
Redwood
Yosemite
Density (Frequency)
0.00100
0.00075
0.00050
0.00025
0.00000
0
5000
10000
15000
Distance (km)


Geospatial Distance - Cluster 4
Congaree
Cuyahoga_Valley
Great_Basin
Great_Smoky_Mountains
Hot_Springs
Indiana_Dunes
Mammoth_Cave
Shenandoah
Density (Frequency)
0.00100
0.00075
0.00050
0.00025
0.00000
0
5000
10000
15000
Distance (km)

Appendix E-2. Psychological distance by geographic travel distance




Psychological Distance - Cluster 3
Badlands
ack_Canyon_of_the_Gunnisc
Bryce_Canyon
Carlsbad_Caverns
Crater_Lake
Grand_Canyon
Hot_Springs
Mammoth_Cave
Mesa_Verde
Petrified_Forest
Saguaro
Yellowstone
Concreteness Score
2.7
2.6
2.5
2.4
0
5000
10000
15000
Distance (km)


Psychological Distance - Cluster 4
Arches
Big_Bend
Canyonlands
Cuyahoga_Valley
Death_Valley
Dry_Tortugas
Grand_Teton
Great_Basin
Great_Smoky_Mountains
Joshua_Tree
Kings_Canyon
Mount_Rainier
North_Cascades
Redwood
Rocky_Mountain
Shenandoah
Wind_Cave
Yosemite
Zion
Concreteness Score
2.7
2.6
2.5
2.4
0
5000
10000
15000
Distance (km)

Appendix E-3. The number of OTRs generated annually for each national park