Minireview Article

Conservation Culturomics: Potential Applications in South Asia

ABSTRACT

Unsustainable human behavior has significantly contributed to the contemporary global biodiversity crisis. In order to mitigate this, it is crucial to study the underlying factors that have resulted in such behavior. Conservation culturomics is an emerging research method that uses digitally available data to gain insights on human-nature relationships at relevant spatial and temporal scales. This method also provides a cost-effective mechanism to inform people-centric conservation policy. Depending on the research question, the data collection and analysis can be tailored, taking advantage of recent advances in information technology. While it is most often applied to global studies, conservation culturomics can also be applied at a regional scale to gather vast volumes of data that represent context-specific human values towards nature. Examination of case studies from South Asia demonstrate the range of potential subjects that culturomics studies can examine, including marine megafauna (whales), avifauna (vultures), and habitats (wetlands). Conservation outcomes of culturomics research are broad, and can include:ecosystem services valuation to motivate habitat restoration, spatial and temporal variation in whale sightings to inform ecotourism guidelines and management, and public perceptions of vultures to inform awareness and outreach initiatives. The major impediments to conservation culturomics in regions such as South Asia are socio-economic biases among internet users and the need for sophisticated technological knowledge; however, with the increasing accessibility of the internet, this is set to change in the future.

Keywords: Culturomics, Conservation science, Human-nature relationship, South Asia

1. INTRODUCTION

Anthropogenic pressures such as land-use change, pollution, climate change, overexploitation of natural resources, introduction of invasive species and others have resulted in a global biodiversity crisis [1]. In recognition of the anthropogenic influence on biodiversity, the field of conservation science has evolved to incorporate human dimensions to inform improved conservation practice and policy [2, 3]. A major factor influencing the effectiveness of conservation initiatives is public attention and interest in biodiversity, as this determines the extent to which people will support and comply with the initiative [4]. In the contemporary context, however, people are increasingly alienated from nature, as a result of which there is a loss of public recognition of its importance and a consequent lack of public support for conservation initiatives [5]. An understanding of the current status, trends, and future directions of human perspectives towards nature can, thus, advance conservation outcomes by revealing underlying reasons for behaviors that affect biodiversity, as well as potential conservation practices and policies that would be acceptable to stakeholders [6]. However, considering the paucity of funding available for conservation research as well as the resource-intensive nature of traditional social science research methods, data collection on human-nature interactions has been limited [3].

Conservation culturomics is an emerging research method that harnesses digitally available data to study human-nature interactions and public attitudes, beliefs, and perceptions towards nature at relevant spatial and temporal scales. The contemporary 'information age' has ushered in an expansion of connectivity and online interactions which leave digital

traces of individual users' interests and attitudes, thereby allowing human culture to be captured in the virtual realm [3]. As a result, the digital database contains vast volumes of user-generated information with the potential to reveal insights into human attitudes and behaviors towards nature and conservation. Through passive crowdsourcing of data, such as web search frequency, number of page visits, engagement with social media content, geo-tagged images, videos, and others, culturomics allows researchers to extract information on human values and their impact on conservation outcomes with limited investment of resources [3, 4]. While this area of research is still in its nascent stages, there has been a rise in the number of academic publications that mine data from social media platforms, web pages, online news media, search engines, and other sources to quantify and analyze societal interest in conservation-related topics [7]. Recent advances in artificial intelligence (AI) and machine learning have allowed for automation of the collection and analysis of large volumes of digital data, thereby easing the way for future research and applications of conservation culturomics [8]. This review will provide an outline of conservation culturomics as an emerging research method, with a focus on the processes involved, benefits and drawbacks, and case studies. While most of the studies that employ this method take advantage of the global reach of the internet, this review specifically highlights its potential applications at a regional scale, using examples from South Asian countries.

2. APPLICATION OF CONSERVATION CULTUROMICS

Culturomics can be used to study a wide range of conservation-related topics, including species awareness, preferences for nature-based recreation, trends and dynamics of conservation interest, perspectives on wildlife trade, cultural salience of specific species or regions, and reactions to conservation-related policies [7, 8, 9]. While it is important to define the scope of the research topic, culturomics remains a highly iterative process which requires frequent revisions to the research scope and methodology. Correia et al. [8] summarize the main stages in the decision-making process for determining the research framework of a conservation culturomics study to be: definition of scope, identification of data type, selection of data sources, and selection of an appropriate analytical approach(Figure 1).

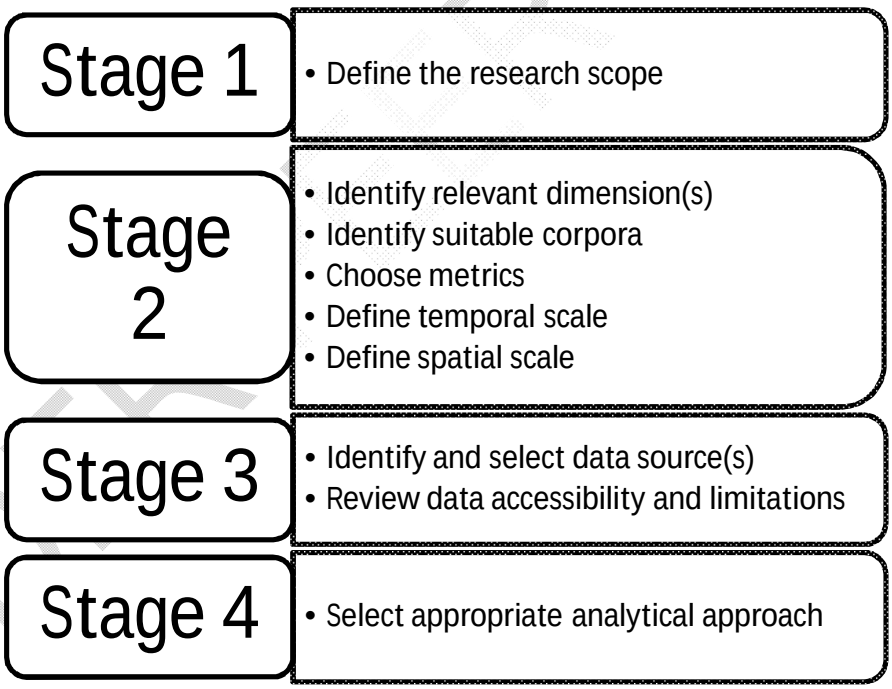

Figure 1: Key stages in the conservation culturomics research framework. Culturomics research is an iterative process, and decisions at each stage can be revisited. (Adapted from [8])

The research design of this method is based on using collections of digital data (corpora) to generate structured data sets that can be analyzed to address the research question [8]. The two main types of data analyzed in conservation culturomics are 1) content, and 2) engagement with content. Data can be extracted from a variety of corpora, such as web

pages, digitized books, news media, digital encyclopedias (like Wikipedia), social media, and video-sharing platforms. This further varies by spatial and temporal scale. Data can be spatially segregated based on internet protocol (IP) locations, geolocation tags, publisher country, and other indicators. It can be temporally segregated based on time of creation or publishing for content-based research, and by time of search, comment, share, like or review for engagement-based research [8].

The raw data can be collected from data aggregation platforms or from multiple individual platforms. This data can be made accessible through dedicated application programming interfaces (APIs) or online interfaces for data access. While some sources provide free, pre-analyzed data, others may charge a fee for access. Another option is web-scraping (automatically extracting data directly from websites); however, this requires familiarity with the terms of services as well as detailed understand of web architecture and programming languages. Having collected the raw data, it is important for the researcher/s to use filtering algorithms or other methods to remove irrelevant entries, such as data points collected for homonyms of the search term. For the duration of the study, adequate and secure data storage must be ensured to account for possible changes in accessibility of the online data source [8].

While engagement-related data can often be mined in a pre-analyzed format, content-related analysis requires the use of specifically tailored software to extract suitable data [9]. Deep learning provides tools to streamline the analysis process by employing artificial neural networks that can be taught (through programming) to perform certain tasks on input data to produce desired output data. In the case of textual data, natural language processing (NLP) methods can be used to extract useful information from the raw data through techniques such as automatic language identification, sentiment analysis (classifying text language as negative, neutral, or positive) or name recognition [8, 10, 11]. Visual data can also be analyzed using computer vision, which can be applied to automatically identify specific features, images, and patterns in images, or generate descriptions of the images that can be analyzed using textual methods [3, 8, 12].

Following this, statistical analysis can be conducted to draw inferences from the selected dataset. The main features of the data can be summarized through descriptive statistics and presented in graphical forms [8, 13]. Network analysis can be used to explore how ideas, opinions, and/or values disseminate within social networks and how online communities form around specific topics or positions [3, 8]. Furthermore, inferential statistics such as regression analysis can be used to study how the conservation-related metric under consideration relates to other variables of interest, including cultural, geographic, and social factors [14]. Temporal analysis can be conducted by visualizing time series data and exploring patterns in the same. Software packages like R allow for the decomposition of time series data into its different elements (long term trend, short term seasonal and cyclical components, and random variations), thereby providing for enhanced analysis [8]. Temporal trends can also be analyzed in the form of patterns of interest in a topic with reference to a particular event or time frame. Spatial analysis can be conducted with the use of geographical information systems (GIS) to visualize the distribution of data points. Point location data can be extracted in the form of IP addresses or geo-tagged map locations which can be plotted directly on a map to analyze spatial characteristics of the data. In cases where country or region level data is available, the attribute in consideration (e.g., frequency of searches) can be visualized by assigning values to each region based on the data collected [15].

The findings from the selected method can then be studied in the socio-cultural context of the region and time period under consideration, to further enhance the understanding of attitudes towards the topic. In order to illustrate the potential applications and value of conservation culturomics, the following section examines case studies from diverse taxa and habitats in South Asia.There has been limited use of the method in this region, and the selected studies depict examples of recent researchthat adopts uniquely tailored approaches to address their research objectives.

3. CASE STUDIES

3.1 Wetland visitation in India

Sinclair et al. [16, 17] employed conservation culturomics to study public attitudes towards the Vembanad Lake and adjacent backwaters in the state of Kerala. They used geo-tagged images posted by users of the image and video hosting service Flickr to study spatial patterns in nature-based recreation and perform an ecosystem services (ES) valuation of the lake. With this information, they also assessed the impact of changes in water quality of the lake on recreational use, using data produced by the Kerala State Pollution Control Board (KSPCB). The raw data for this study was extracted as metadata for photographs published by Flickr users who visited the lake. These were selected from the publicly available Flickr photographs geotagged within a 60-meter buffer area of Kerala's wetlands and extracted from Flickr's API using a code developed for this specific application in the R programming language. Using this data, the spatial patterns of recreational usage of the lake were mapped and visualized via GIS, and the Optimized Hot Spot Analysis tool was applied to identify the areas of interest. For the ES valuation of the lake, the researchers employed the travel cost method which estimated the recreational value of the lake based on the cost of travel for average visitors. They determined the home

location of each user who visited the lake either through the publicly disclosed location on their user profile, or through approximation based on the geotags on all their photographs. Through spatial analysis, the researchers assigned unique home locations to each user based on activity within their assumed home district. Following this, travel distance and travel time was estimated, which allowed for calculation of the value of recreational benefits of the lake and surrounding wetlands. The third component of the study involved temporal analysis to investigate the impact of changes in water quality on visitation rates. This analysis, combined with the aforementioned ES valuation, allowed them to determine the potential increase in consumer surplus or value to visitors that could result from improvements in water quality as well as lake restoration.By highlighting the economic benefits of improved water quality, the findings from this study canprovide empirically-backed motivation for authorities to invest in lake restoration measures with benefits for biodiversity.

3.2Whale watching in Sri Lanka

Bandara and Bandara [18] used tools of conservation culturomics to study images related to whale-watching related tourism in Sri Lanka. The data collected for this study included photographs and metadata from the Flickr API that was extracted using a Python script written by the researchers. They mined data by filtering with the keyword “whale” and the place identifier “Sri Lanka”. Following this, the data was interpreted through content analysis, mapping of geo-tagged whale photographs, temporal distribution of photographs, demographic data of users, and analysis of social-tags of photographs. The content analysis was manually conducted to classify the photographs based on categories such as human activity, animals, accommodation, natural phenomena, and others. Researchers also selected photographs including whales to plot the spatial distribution of whale sightings. The temporal, demographic, and social tag analysis was also conducted using custom Python scripts to arrive at descriptive statistics that summarized the data. Through these methods, the researchers found the popular social-tags among photographers and determined their demographic characteristics from the user metadata. Researchers also identified the main categories of attractions that interested whale-watching tourists and found Mirissa (in the Southern Province) to be the hotspot for whale sightings in Sri Lanka. In addition to this, the study described the time and month during which most of the photographs with whales were posted, indicating the ideal time for sightings.These findings can, thus, be used by boat tour operators and tourism policy makers to decide the most appropriate time and area to conduct activities for tourists.Despite representing a wide spatial and temporal range, this studywas conducted with limited investment of resources, illustrating the potential for conservation culturomics to inform improved management of the ecotourism industry.They could also extract spatial patterns for the distribution of whales in the region [18]. In addition to conservation culturomics, this study is an example of iEcology- as it reveals insights into ecological knowledge based on user-generated data [19].

3.3Vulture-related content in Nepal

Ghimire et al. [20] used photographs posted on Facebook, the online social media and networking service, to analyze spatial and temporal variations in vulture-related content. Considering the cultural importance of vultures as a symbol of power and insight in the South Asian context, they mined and analyzed photographs posted on Facebook within a specific temporal frame and location (Nepal) using “vulture” and the names of each species as keywords. The researchers manually filtered the dataset to identify relevant data points and identify vulture species, and conducted sentiment analysis by rating the captions on each post as positive, neutral, or negative. They also depicted the spatial distribution of vulture-related content using GIS to visualize the districts from which vulture photographs were posted and those from which they were not. This revealed that all districts towards the western region of the country logged vulture posts while some districts in the east did not, reflecting the spatial variations in vulture density. Researchers also identified the most widespread species based on its frequency of presence in photographs (another example of iEcology). All the captions were either rated positive or neutral in sentiment. In addition, the study found a temporal increase in the number of vulture-related posts however this could be attributed to the overall rise of internet users in Nepal over time [20].By presenting spatial and temporal trends in vulture-related content along with analysis of sentiments, the results from this study can be used to gauge public perceptions about vultures in different regions, along with changes over time. This has the potential to inform vulture conservation awareness groups about public interest in vulture, thus, allowing them to design more effective campaigns.

4. DISCUSSION

While a large proportion of conservation culturomics studies have been conducted at a global scale or at a country-level in the USA, Australia, Brazil and others, the case studies above [16, 17, 18, 20] demonstrate the applicability of this research method in a South Asian context. Research design can be specifically tailored to suit the study area by using geotag filters or by using keywords in regional languages [21]. The major benefit of the method is that it allows for research to be conducted over a vast spatio-temporal scale without the investment of time and resources that would be required for such a study using traditional social and natural sciences research methods. For example, in their study of the impact of water quality on recreational visitation, Sinclair et al. [16, 17] collected digital data from 81 wetlands across

Kerala between 2008 and 2016. An equivalent study using traditional methods would have required the researcher to visit the sites during the timeframe to document visitation rates. Another benefit is that the method allows large volumes of big data to be mined with the help of AI and machine learning. This significantly simplifies the process and allows for automation. It also allows for replicability as the same study can be repeated by different authors using the digitally available data, which would further verify the findings. Additionally, there is potential for real-time monitoring of human-nature interaction by taking advantage of technological advances in data mining. This would allow for the implementation of more proactive conservation measures. The metadata recorded along with user-generated data also allows for near-accurate analysis of spatial trends. As illustrated by Bandara & Bandara [18] and Ghimire et al. [20], culturomics studies can also reveal insights into the ecology of study species based on digitally available data (a field known as iEcology). This is particularly valuable in locations where field access or funding to conduct conventional field research may be limited.

Despite the advantages of conservation culturomics, there are significant disadvantages that may hinder its wide application in South Asia at present. One is the low level of internet penetration in the region. Only 58.8% of the global population has access to internet connectivity and the value for the South Asian region is significantly lower [7]. Despite the high number of individual internet users in the region, the percentage of the total population with access to the internet in South Asia remains low. This is further skewed by socio-economic factors that determine internet access [22]. Low internet accessibility would likely result in conservation culturomics data only reflecting the values of a small proportion of the population and obscuring the attitudes of mostly marginalized sections of society towards nature. This would exacerbate biased policies that reflect the values of only a specific section of society. Another drawback of this research method is that manual extraction of large volumes of data is time-consuming and tedious. While automation makes the process more efficient, it can lead to accuracy-related problem as it is difficult to verify the results. In cases where keywords have homonyms and synonyms, for example, care must be taken to ensure that this is accounted for [23]. Finally, research in the field is still in its incipient stage so there remains a lack of understanding on the biases associated with each digital corpus and methodology, implying that the accuracy of these studies is still questionable [24].

As culturomics extracts data produced by users without their explicit informed consent, there are important ethical considerations. Most importantly, the privacy of users must be secured as much as possible. While all social media consumers agree to the terms and conditions of each platform that they contribute content to, not all users are aware of these. Further, the public availability of this data does not imply that the users have given informed consent for the inclusion of their data for research [25, 26]. Thus, it is important for the researcher to minimize the data collected, anonymize data, and ensure strict data storage and management protocol to prevent any harm to the users [25]. For example, Sinclair et al. [16, 17] used geotagged image analysis to estimate the home locations of Flickr users. Anyone with access to the data could, potentially, locate the user and place them at risk. In order to prevent such threats to individuals, personal identifying features must be removed from images and the data must be stored securely. In addition to ethical behavior towards users, the researchers must also ensure that they are following copyright and database laws of the selected digital corpus. Essentially, ethical conservation culturomics research should be conducted as non-deceptive, covert observations that ensures the privacy of individuals, complies with related laws, and does not misrepresent results [27].

4. CONCLUSION

In view of the anthropogenic causes of the on-going global biodiversity crisis, conservation culturomics is a promising research method to inform more effective initiatives and policy-measures to mitigate the damage caused by human activities to biodiversity and natural resources. By revealing human perspectives and behaviors towards nature as is observable from user-generated digital data, it can provide insights from a vast spatio-temporal scale with limited investment of time and resources. Furthermore, the capacity for automation makes this method more efficient. The case studies summarized in this paper also indicate the potential for its application in the South Asian context. While it carries limitations for regions with poor internet connectivity and may require sophisticated hardware and software, culturomics can be particularly useful in situations where funding and field access may be limited.

REFERENCES

1. Maxell SL, Fuller RA, Brooks TM, Watson JE. Biodiversity: The ravages of guns, nets and bulldozers. Nature.2016;536(7615):143-145.DOI: 10.1038/536143a
2. Bennett NJ, Roth R, Klain SC, Chan KM, Clark DA, Cullman G, Epstein G, Nelson MP, Stedman R, Teel TL, Thomas REW, Wyborn C, Curran D, Greenberg A, Sandlos J, Veríssimo D. Mainstreaming the social sciences in conservation. Conservation Biology.2017;31(1):56-66.DOI: 10.1111/cobi.12788

3. Toivonen T, Heikinheimo V, Fink C, Hausmann A, Hiippala T, Järv O, Tenkanen H, Minin ED. Social media data for conservation science: A methodological overview. Biological Conservation.2019;233:298-315.DOI: 10.1016/j.biocon.2019.01.023
4. Jarić I, Bellard C, Courchamp F, Kalinkat G, Meinard Y, Roberts DL, Correia RA. Societal attention toward extinction threats: A comparison between climate change and biological invasions. Scientific Reports.2020;10(1):1-9.DOI: 10.1038/s41598-020-67931-5
5. Soga M, Gaston KJ. Extinction of experience: The loss of human–nature interactions. Frontiers in Ecology and the Environment. 2016;14(2):94-101.DOI: 10.1002/fee.1225
6. Minin ED, Correia RA, Toivonen T. Quantitative conservation geography. Trends in Ecology & Evolution.2022;37(1):42-52.DOI: 10.1016/j.tree.2021.08.009
7. Correia RA, Ladle R, Roll U.Special section: Advancing conservation culturomics – Introduction. Conservation Biology.2021;35(2):395-397.DOI: 10.1111/cobi.13700
8. Correia RA, Ladle R, Jarić I, Malhado ACM, Mittermeier JC, Roll U, Soriano-Redondo A, Veríssimo D, Fink C, Hausmann A, Guedes-Santos J, Vardi R, Minin ED. Digital data sources and methods for conservation culturomics. Conservation Biology.2021;35(2):398-411.DOI: 10.1111/cobi.13706
9. Ladle RJ, Correia RA, Do Y, Joo G, Malhado ACM, Proulx R, Roberge J, Jepson P. Conservation culturomics. Frontiers in Ecology and the Environment.2016;14(5):269-275.DOI: 10.1002/fee.1260
10. Hausmann A, Toivonen T, Fink C, Heikinheimo V, Kulkarni R, Tenkanen H, Minin ED. Understanding sentiment of national park visitors from social media data. People and Nature.2020;2(3): 750-760.DOI: 10.1002/pan3.10130
11. Kulkarni R, Minin ED. Automated retrieval of information on threatened species from online sources using machine learning. Methods in Ecology and Evolution.2021;12(7):1226-1239.DOI: 10.1111/2041-210X.13608
12. Ghermandi A, Depietri Y, Sinclair M. In the AI of the beholder: A comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green space. Landscape and Urban Planning.2022;217:e104261.DOI: 10.1016/j.landurbplan.2021.104261
13. Vardi R, Mittermeier JC, Roll U. Combining culturomic sources to uncover trends in popularity and seasonal interest in plants. Conservation Biology.2021;35(2):460-471.DOI: 10.1111/cobi.13705
14. Troumbis AY. The time and timing components of conservation culturomics cycles and scenarios of public interest in the Google era. Biodiversity & Conservation.2019;28:1717-1727.DOI: 10.1007/s10531-019-01750-7
15. Burivalova Z, Butler RA, Wilcove DS. Analyzing Google search data to debunk myths about the public's interest in conservation. Frontiers in Ecology and the Environment.2018;16(9):509-514.DOI: 10.1002/fee.1962
16. Sinclair M, Ghermandi A, Sheela AM. A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. Science of the Total Environment.2018;642:356-365.DOI: 10.1016/j.scitotenv.2018.06.056
17. Sinclair M, Ghermandi A, Moses SA, Joseph S. Recreation and environmental quality of tropical wetlands: A social media based spatial analysis. Tourism Management.2019;71:179-186.DOI: 10.1016/j.tourman.2018.10.018
18. Bandara T, Bandara TP. Whale watching in Sri Lanka: Understanding the metadata of crowd-sourced photographs on Flickr™ social media platform. Sri Lanka Journal of Aquatic Sciences.2019;24(2):41-52.DOI: 10.4038/sljas.v24i2.7566
19. Jarić I, Correia RA, Brook BW, Buettel JC, Courchamp F, Minin ED, Firth JA, Gaston KJ, Jepson P, Kalinkat G, Ladle R, Soriano-Redondo A, Souza AT, Roll U. iEcology: Harnessing large online resources to generate ecological insights. Trends in Ecology & Evolution.2020;35(7):630-639. DOI: 10.1016/j.tree.2020.03.003
20. Ghimire P, Thakuri R, Basnet A, Pandey N, Bist BS, Sharma B, Bhusal KP. Spatial-temporal analysis of vulture related content on social media. Vulture Bulletin.2020;9:8-12.
21. Correia RA, Jepson PR, Malhado AC, Ladle RJ. Familiarity breeds content: assessing bird species popularity with culturomics. PeerJ. 2016;4:e1728. DOI: 10.7717/peerj.1728
22. Zhou Y, Singh N, Kaushik PD. The digital divide in rural South Asia: Survey evidence from Bangladesh, Nepal and Sri Lanka. IIMB Management Review.2011;23(1):15-29.DOI: 10.1016/j.iimb.2010.12.002
23. Correia RA, Jarić I, Jepson P, Malhado ACM, Alves JA, Ladle RJ. Nomenclature instability in species culturomic assessments: Why synonyms matter. Ecological Indicators.2018;90:74-78.DOI: 10.1016/j.ecolind.2018.02.059

24. Ladle RJ, Jepson P, Correia RA, Malhado AC. The power and the promise of culturomics. Frontiers in Ecology and the Environment.2017;15(6):290-291.DOI: 10.1002/fee.1506
25. Minin ED, Fink C, Hausmann A, Kremer J, Kulkarni R. How to address data privacy concerns when using social media data in conservation science. Conservation Biology.2021;35(2):437-446.DOI: 10.1016/j.tree.2021.08.009
26. Zimmer M.“But the data is already public”: on the ethics of research in Facebook. The Ethics of Information Technologies.2010;12:313-325.DOI: 10.1007/s10676-010-9227-5
27. Thompson RM, Hall J, Morrison C, Palmer NR, Roberts DL. Ethics and governance for internet-based conservation science research. Conservation Biology.2021;35:1747-1754.DOI: 10.1111/cobi.13778