Examining the Influence of Lake Water Quality on Visitor Sentiment:

Empirical Findings from 85 U.S. Lakes

Abstract:

Cultural ecosystem services (CESs) are the intangible benefits that humans receive from nature, resulting from the interplay between biotic nature and abiotic nature (Gray, 2011). And very often, the intangible dimensions hold greater significance for individuals compared to material benefits (Kobryn et al., 2018). Lakes are highly valued for CESs because they provide opportunities for recreational activities, aesthetic experiences, spiritual enrichment, and reflection (Blicharska et al., 2017). Lake CESs are intricately linked to water quality, where sedimentation, nutrients, and bacteria affect people's perceptions and interactions with the water body (Keeler et al., 2012). However, our understanding of lake water quality's influence on CESs remains limited, hindering effective management strategies to enhance the economic and social values of lakes. Water-related CESs research is also constrained at broader scales (e.g., national scale) due to measurement challenges, undermining research generalizability.

We hypothesized a connection between lake water quality and CESs, with water quality influencing CESs through recreational behaviors and visitor sentiment. To test this, we coupled big data from Google Review and National Lake Assessment to investigate water quality effects on CESs and visitor sentiment across 85 U.S. lake-front public parks. We implemented five analytical steps: (1) Collecting geographic data of 85 public-owned lakes and identifying lake-front parks using Google Place API, (2) Obtaining ~40,000 relevant text reviews with Python Crawler, (3) Filtering Google reviews related to lake CESs through content analysis with keyword labels, (4) Deriving positive, neutral, and negative sentiments from lake-related user-generated content using Google Cloud Natural Language API, and (5) Identifying associations between key water quality indicators and CESs-related visitor sentiment with a Bayesian multinomial model. Our results showed that CESs provided by lakes were significantly affected by multiple water quality indicators. Higher chlorophyll-a concentration and APHA color value increased negative sentiments toward lake CESs, while higher total phosphorus unexpectedly associated with increased positive CESs-related visitor experiences, possibly due to higher productivity. Our findings have implications for lake management, advocating targeted planning to promote harmony between human well-being and ecological health in lake ecosystems. Moreover, analysis of user-generated text contents for sentiment evaluation can yield valuable information regarding user satisfaction and memories of scenes, and thereby facilitate lake management practices aimed at fostering sustainable local economic development.

Keywords: Cultural Ecosystem Services; Visitor Sentiment; Water Quality; Content Analysis; Sentimental Analysis; Big Data.

1. Introduction

Lakes play a crucial role in human well-being in many regions by providing a diverse range of ecosystem services, e.g., water supply, irrigation, hydropower production, flood control, climate regulation, and a wide variety of cultural ecosystem services (CES) (Allan et al., 2017, Ho and Goethals, 2019, Hogeboom et al., 2018). Lakes provide ample opportunities for recreational activities like boating, swimming, fishing, and bird watching, among others (Allan et al., 2015, Angradi et al., 2016, Angradi et al., 2018b, Ghermandi and Fichtman, 2015, Sterner et al., 2020). Furthermore, they offer aesthetic experiences, educational opportunities, as well as inspirational, spiritual, or symbolic benefits (Figueroa-Alfaro and Tang, 2017, Hossu et al., 2019, Schirpke et al., 2021a). Although water quality cannot be viewed exclusively as an ultimate ecosystem service (Keeler et al., 2012), it constitutes a crucial component of numerous ecosystem services, especially CESs such as leisure and human wellness. However, existing research focused more on the lakes' ecological ecosystem services related to water quality, while CES were less studied. This is because research on lakes' CES, is limited due to the difficulties in measuring CES across different lake settings (Allan et al., 2015, Sterner et al., 2020). Also, research on lakes' CES at regional or broader scales was limited due to the difficulties in CES measurement across broader scales. To address this gap, this study established a direct linkage between water quality and lake CES at national scale.

In recent years, researchers focusing on CES have explored the potential of social media data as a tool for measuring the level of recreational use of natural areas (Heikinheimo et al., 2017, Levin et al., 2017, Sessions et al., 2016, Sinclair et al., 2020b, Sinclair et al., 2020a, Tenkanen et al., 2017, Wood et al., 2013). By analyzing the spatio-temporal distribution of social media data, they have been able to identify patterns in people's preferences for specific landscapes, landscape features, and ecosystem attributes such as water clarity (Oteros-Rozas et al., 2018, Sonter et al., 2016, Tieskens et al., 2018). This quantitative approach has been found to be effective in providing a broad understanding of the level of recreational use of natural areas. Furthermore, qualitative analysis of social media data has also been used to gain deeper insights into CES.

By examining user-generated metadata and visual content such as photographs, researchers have been able to identify the specific features and values that people associate with natural areas. For instance, the presence of certain species of plants or animals, or the quality of the landscape or water body, may be particularly meaningful to users. These approaches are based on the assumption that people post photographs of things they value, which can reveal their preferences and emotions related to CES (Egarter Vigl et al., 2021, Hausmann et al., 2018, Hausmann et al., 2020, Langemeyer et al., 2018, Moreno-Llorca et al., 2020, Oteros-Rozas et al., 2018, Pickering et al., 2020, Richards and Friess, 2015). This has enabled researchers to go beyond simple measures of use and explore the more complex cultural values associated with natural areas. As a result, these methods have become increasingly popular in recent years, providing new ways to assess and understand the value of CES and to inform management and planning decisions.

Google reviews have been utilized in many fields, such as to examine airport service quality, restaurant service and customer's eating experience, etc. Google Map user reviews have also been used for branding tourist destinations and to predict public perceptions of visiting places (Huang, et al., 2022). However, the application of social media review data in landscape and urban planning research is still preliminary. Roberts, Sadler, and Chapman (2017) use Twitter data to assess the variations in physical activity between different seasons.

Kovács-Győri, Ristea, Havas, Resch, and Cabrera-Barona (2018) explore the public emotions of events like the 2012 London Olympics by analyzing Twitter data. Zhou and Li (2018) investigate the spatial distributions of Flicker photos and hashtags to understand travel behaviors and resident-visitor interactions. Chen et al., 2018, Song and Zhang, 2020 assess landscape values using Instagram photos. However, little research exists to measure sentiments related to CES from online review data, using the measurements to guide urban planning and design practices. (Song, et al., 2022).

Our hypothesis is that lake ecosystem services (CES) and their ecological characteristics are interrelated. In particular, we posit that poor water quality in lakes can adversely affect CES by increasing the likelihood of visitors experiencing negative emotions during their visits, while good water quality can conversely increase the possibility of visitors gaining comparatively positive sentiments. With rich semantic and sentimental information from textual contents on social media, we could quantify how people value and perceive CESs. We coupled google review data and National Lake Assessment (NLA) data to examine the effects of water quality on recreational services in 182 lake-front public parks across the US.

2. Methodology

2.1 Data collection

The dataset used in this study is the Google review from 182 near-lake parks in the United States from Google Map. The geographic information of the public-owned and small/medium size lakes are collected through EPA (U.S. Environmental Protection Agency). Then the near-lake parks are extracted from Google Place API. The information of parks within 5 km from the lake center are collected, then verified manually through ArcGIS. Google Place Review of selected parks is collected with crawler based on park id. Each record in the data will contain the following information: the reviewer’s name, review time (rough time, such as “a month ago”, “a year ago”), review text, rating (1-5), park id, and nearby lake id. It is assumed that only one lake is accessible in one park, and only parks with sufficient information are selected.

Three variables: review text, rating, and whether the review includes a water-related activity are used in this study. There are around 36500 selected text reviews, 17000 of them mention activities related to water. For the reviews in languages other than English, they are translated into English. The text data is then cleaned into lowercase English characters only, with the duplicate and null dropped.

2.2 Data classification

Text reviews were labeled whether it related to lake CES, and we filtered out reviews relevant to CES with Google Natural language API and regular expression based keywords matching. The water-related activities keywords are collected from The Compendium of Physical Activities, and the activities are classified into 5 categories (boating, surfing, swimming, fishing, seen) based on how close they are related to water.

The review distribution is biased, as the majority of the visitors' perceptions are positive, and the review numbers of each near-lake park are different, which varies from less than 100 to over 1000. In order to ensure the classification performance of all the categories, resampling methods from sklearn resample library is applied to the training set.

2.3 Sentiment analysis

The study utilized sentimental analysis to extract information related to aesthetic experiences from textual content. To filter out unimportant and less meaningful text, we removed independent numerals and emoticons. We also discarded semantically meaningless stop words, and used two widely-used stop word lexicons - the Natural Language Toolkit and the Language Technology Platform - for further initialization.

To determine the sentiments associated with user-generated textual content, a lexicon-based approach is commonly used. This approach summarizes words into positive, neutral, and negative categories and records expressions for verbs, things, intensifiers, and modifiers. The proportion of these categories is approximately 4:2:1, with a greater emphasis on positive words. Machine learning methods are often used in conjunction with large sets of lexicons to train sentiment classifiers. The study utilized the NLTK library to determine the sentiment associated with user-generated text. The sentiment extent was captured in two areas: the sentiment score and magnitude score. The sentiment score quantified the overall emotional learning of the textual content of individuals within each study unit and was categorized into three groups: negative (-1), neutral (0), and positive (1).

2.4 Regression analysis of water quality and visitors' sentiments

We obtained detailed results on critical water quality indicators and related sentiments through a Bayesian multinomial model. Our model for the multinomial analysis is shown below:

Likelihood: $\text{motivators}_i \sim \text{Multinomial}\ (\text{Ppos}_i, \text{Pneu}_i, \text{Pneg}_i,; 3)$

Process model:

$$\text{Log}\ (\text{Ppos}_i) = \alpha + \beta_1\text{CHLA}_i + \beta_2\text{COLOR}_i + \beta_3\text{NTL}_i + \beta_4\text{PTL}_i + \beta_5\text{TURB}_i$$

$$\text{Log}\ (\text{Pneu}_i) = \gamma + \theta_1\text{CHLA}_i + \theta_2\text{COLOR}_i + \theta_3\text{NTL}_i + \theta_4\text{PTL}_i + \theta_5\text{TURB}_i$$

$$\text{Log}\ (\text{Pneg}_i) = \varepsilon + \lambda_1\text{CHLA}_i + \lambda_2\text{COLOR}_i + \lambda_3\text{NTL}_i + \lambda_4\text{PTL}_i + \lambda_5\text{TURB}_i$$

$$\text{Ppos}_i + \text{Pneu}_i + \text{Pneg}_i = 1$$

Where the probability of each type of response, Pposi, Pneui, and Pnegi , at each lake i, is estimated as a function of these predictors: Chlorophyll-a, color, total nitrogen, total phosphorus and turbidity represent the average Chlorophyll-a concentration in each lake, the APHA color of each lake, the total nitrogen, total phosphorus, and the total suspended solids in each lake. All these predictors are significant indicators of water quality (Egan et al., 2009, Keeler et al., 2015)). In addition, given that this study examined a large sample of 85 lakes on a national scale, there were significant challenges associated with collecting water quality data, and the predictors obtained represent the most comprehensive and reliable available for this study. We also set non-informative priors for each parameter in the model:

$\alpha_{*}, \gamma_{*}, \varepsilon_{*} \sim \log N\ (1, 100)$

$\beta_{*}, \theta_{*}, \lambda_{*}, \sim N\ (0, 10000)$

We assessed that the significance of a variable is deduced by examining whether the confidence interval of each coefficient included zero. For categorical variables we assessed statistically significance by comparing their 95%CIs, if they did not overlap we considered them to be different. The implementation of our model was carried out utilizing the RStudio software (1.4.1717).

3. Results

The correlation between water quality indicators and visitors' sentiments varied across different sentiment categories. Specifically, positive and negative sentiments showed significant correlations with some of the water quality indicators, while neutral sentiment did not exhibit any significant relationship (Table 1).

Table 1. Coefficient estimates of each parameter from Bayesian multinomial model.

	Sentiment category	**Parameters**	**2.5%**	**25%**	**50%**	**75%**	**97.5%**
1	**Positive**	Intercept α	-19.76	-6.45	0.12	6.88	20.10
		Chlorophyll-a β_1	-182.95	-78.87	-12.70	64.34	198.41
		Color β_2	-12.71	21.77	55.75	103.35	211.21
		Total nitrogen β_3	-195.97	-70.78	2.17	64.40	193.42
		Total phosphorus β_4	3.71	46.54	81.26	124.96	234.45

		Turbidity β_5	-17317	-44.54	81.26	124.96	234.45
2	**Neutral**	Intercept γ	-35.13	-25.33	-20.21	-15.37	-5.60
		Chlorophyll-a θ_1	-0.29	0.55	1.48	4.67	25.29
		Color θ_2	-0.19	0.28	0.51	0.84	2.07
		Total nitrogen θ_3	-261.05	-97.03	-39.00	-14.69	0.85
		Total phosphorus θ_4	-3.72	-0.12	0.14	0.56	3.59
		Turbidity θ_5	-18.54	0.38	3.33	9.07	43.47
3	**Negative**	Intercept ε	-20.50	-7.40	-0.61	6.00	18.33
		Chlorophyll-a λ_1	-12.98	9.66	34.16	60.77	107.36
		Color λ_2	43.33	73.31	105.91	130.54	167.08
		Total nitrogen λ_3	-61.93	46.86	88.60	133.99	223.57
		Total phosphorus λ_4	-74.81	-56.03	-43.16	-33.33	-21.72
		Turbidity λ_5	21.24	72.49	109.68	153.72	236.54

Note of Table 1: Significant water quality indicators were highlighted with pink color in this table. Different letters in parameters within each sentiment category indicated statistical differences from each other.

Among positive sentiments, only total phosphorus had a significant positive effect on the likelihood of visitors experiencing positive emotions (Fig.1). This finding was contrary to our initial hypothesis that good water quality of lakes can increase the possibility of visitors gaining comparatively positive sentiments. As such, higher levels of total phosphorus were associated with an increased likelihood of positive visitor experiences, indicating that elevated levels of total phosphorus in lakes can enhance visitors' satisfaction with lake CES.

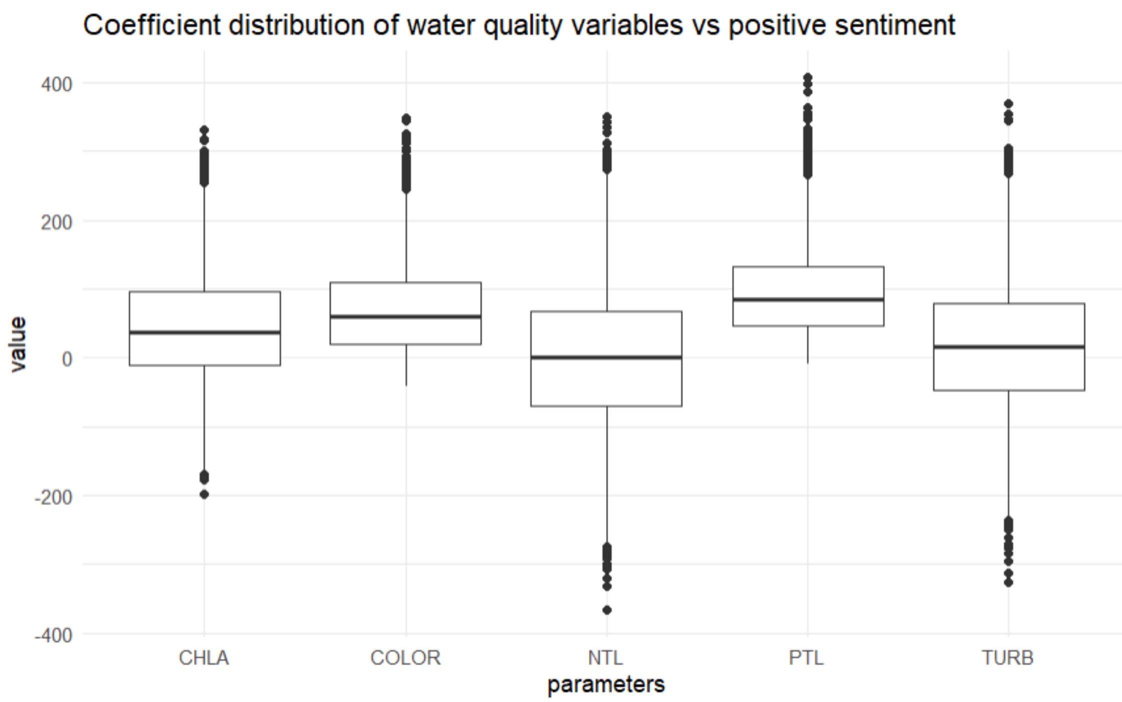

Fig.1 Distribution of coefficients for water quality variables and possibility of gaining positive sentiments. *We extracted all coefficients from 5,000 iterations of our Bayesian multinomial model. The confidence interval of our result was 95.0%.*

Besides, our analysis of negative sentiments revealed that Chlorophyll-a concentration had significantly positive effects on the likelihood of visitors experiencing negative emotions (Fig.2). This result supported our initial hypothesis that poor water quality would increase the likelihood of negative emotional responses related to CES. Specifically, our findings indicate that higher level of Chlorophyll-a concentration in lakes can reduce people's satisfaction with lake CES.

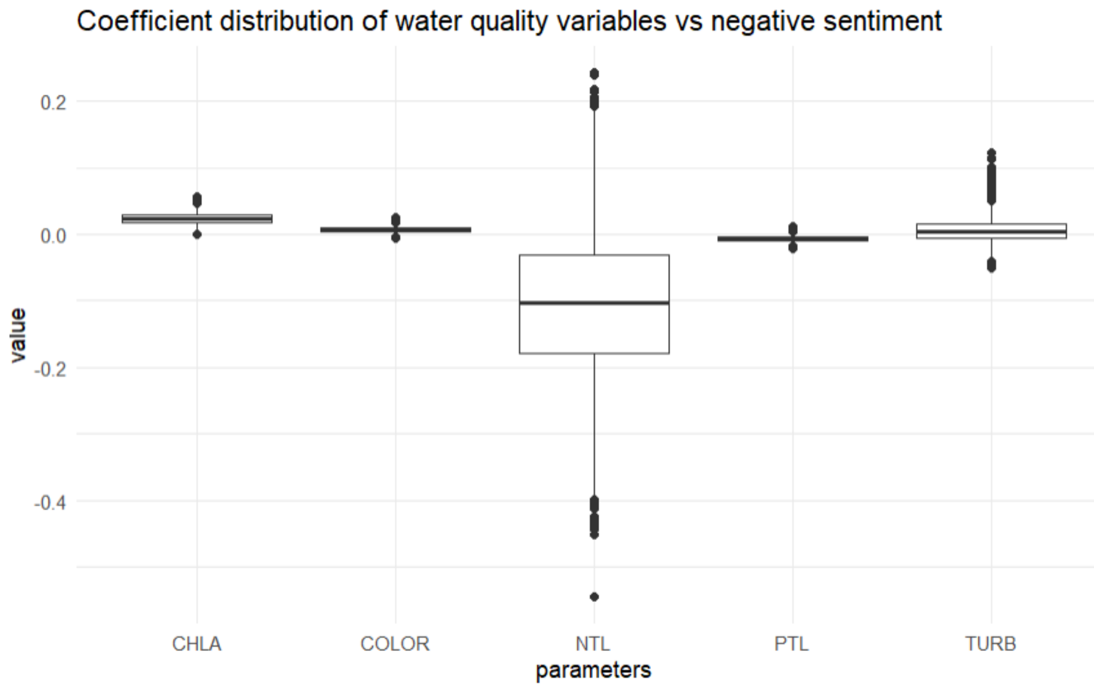

Fig.2 Distribution of coefficients for water quality variables and possibility of gaining negative sentiments. *We extracted all coefficients from 5,000 iterations of our Bayesian multinomial model. The confidence interval of our result was 95.0%.*

4. Discussion

4.1 The Importance of Water Quality in Shaping People's Perception of Lake CES

Our study found that the color and turbidity of a water body are highly valued by people and are perceived as important indicators of water quality, which is consistent with previous research. For example, Mumbi et al. (2020) found that people rely on sensory factors, such as water color and odor, to predict water quality and the risk of pollution. Similarly, a study by Keeler et al. (2015) found that, the color and clarity of water are among the top factors that influence people's perception of water quality and can significantly affect the lake visitation. People generally prefer clear and transparent water, which they associate with purity and cleanliness. On the other hand, water that appears discolored, murky, or cloudy is often perceived as dirty and polluted, which can negatively affect people's evaluation of the CES that lakes provide. However, perceived water quality can be quite different from actual water quality, leading to a discrepancy between visitors' sentiments and the real water quality (Mumbi et al., 2020, Ochoo et al., 2017). For example, people may perceive the water in a lake as clean and clear based on its color and turbidity, but the lake may contain high levels of nutrients or pollutants that can harm aquatic life and affect the quality of the water. As a result, people's emotional response to the lake may not accurately reflect the actual ecological condition of the lake.

In addition to water quality, many other factors can potentially affect people's satisfaction with lake CES. For example, the availability and quality of recreational amenities, such as boating facilities, picnic areas, and hiking trails, can greatly enhance the visitors' experience and increase their willingness to return to the lake. The accessibility of the lake, such as its proximity to urban areas and the ease of getting there, is also an important factor that can affect its use and popularity among visitors. According to a study by Cao et al. (2022), coastal CES can be categorized into six categories: recreational services, aesthetic services, amenity services, knowledge services, naturalistic services, and spiritual services. Although the study focused on coastal CES, the findings can be applied to lakes as well. Our study of CES is more focused on its recreational services, aesthetic services, naturalistic services, and spiritual services but lacks the analysis of amenity services and knowledge services. Therefore, a future step of our study could involve more lake characteristics, such as the number and type of amenities provided by lake-shore parks, the quality evaluation of lake shore landscapes, and the accessibility to the lake. By expanding our scope to include these factors, we can better understand the impact of water quality on lake CES within a larger framework.

In conclusion, water quality, especially its color and turbidity, is an important factor that can affect people's perception of lake CES. However, perceived water quality may not accurately reflect the actual ecological condition of the lake. Many other factors, such as recreational amenities, accessibility, and other CES categories, can also affect people's satisfaction with the lake. Therefore, future studies should consider these factors to provide a more comprehensive understanding of the role of water quality in shaping people's evaluation of lake CES.

4.2 Implications for Lake Management and Visitor Experience

Lake CES and its social and economic benefits. Water quality is the core of lake CES, but these services are not captured in markets and have proven difficult to quantify (Brauman et al. 2007; Keeler et al. 2012). Lack of information about the value of water-quality benefits can complicate justifying major spending on improved water quality (Egan et al., 2015). Our findings supported the argument that water quality can significantly affect visitors' satisfaction on the lakes or lake-shore parks and therefore affecting the visitation and evaluation, which could directly affect local economic development and social and cultural construction. Besides, poor water quality has higher health risk of waterborne disease for visitors who have primary contact (e.g., swimming, boating, surfing, and etc.) and secondary contact (e.g., wading, fishing, and etc.) with water. E.coli and other pollutants in water body could cause several waterborne disease such as inflammation, GI, and etc. In this way, to improve local economic and social benefits, and protect visitors from health risk, lake management strategies targeting water quality are vital. Besides, local governments should take more efforts on water quality regulation policies to protect the lake CES. All these together would not only benefit the nature and human heath, but also improve the local tourism and eventually benefit the local economic development.

Lack of information about the value of water-quality benefits can complicate justifying major spending on improved water quality.

As a natural resource, lake with its CES, provides huge social and economic benefits to human beings. However, these benefits are highly dependent on the water quality of the lake, which is difficult to be measured economically and is not captured in markets (Brauman et al. 2007; Keeler et al. 2012). Poor water quality not only affects visitors' satisfaction with the lake but also poses health risks to those who have primary contact (e.g., swimming, boating, surfing, and etc.) and secondary contact (e.g., wading, fishing, and etc.) with water (Egan et al., 2015, Kundu et al., 2013). Waterborne diseases, such as gastrointestinal, respiratory, ear and ocular and skin or wound infection, can be caused by pollutants such as coxsackieviruses, adenoviruses, echoviruses, hepatitis A virus, astroviruses and noroviruses in the lake (Sinclair et al., 2009).

To protect the social and economic benefits of Lake CES and ensure the health and safety of visitors, lake management strategies targeting water quality are essential. Some recommended management strategies include reducing nutrient pollution and controlling sediment erosion (Barnett et al., 2019). Governments should also take a more active role in regulating water quality to protect the lake and its benefits. One effective way to improve water quality is through the implementation of water-related policies. For example, the Clean Water Act (CWA) has been successful in reducing pollution in lakes and other water bodies across the United States (USEPA, 2020). Besides, the implementation of Best Management Practices (BMPs) has shown promising results in improving water quality of many lakes (EPA, 2018). By implementing effective policies and BMPs, governments can help ensure that Lake CES remains a valuable natural resource for generations to come. With these regulatory policies and management strategies, we can improve the quality of life for local communities, preserve the natural environment, and promote sustainable economic development.

4.3 Limitation and future study

This study has certain limitations that should be addressed in future research. One limitation is the lack of weather data, which could have influenced visitors' sentiment towards the lake CES. However, since there was no precise date in the textual reviews obtained through crawler, we were unable to match the exact weather data with the corresponding reviews, which limited the accuracy of our results. Additionally, it was challenging to define and explicitly explain a "neutral" sentiment, which limited our further explanation on our results. Besides, categorizing sentiments into three categories (positive, neutral, and negative) may not fully capture the complexity of visitors' experiences and perceptions. Future research could explore more nuanced ways of categorizing sentiment. To address these limitations, future studies could incorporate weather data and explore alternative ways of defining sentiments. Furthermore, adding variables such as lake and lakeshore characteristics, as well as water-related recreational behaviors, could enhance the model's predictive power.

For the data mining methodology, our next steps are to introduce different technologies in the natural language processing of the text data. We will also work on exploring how water quality impacts visitor’s perception through water-related activities by multiple mediation models.

5. Conclusion

Based on the analysis of 85 lake-front public parks in the US, this study established the linkage between water quality, CES, and visitors' sentiments. Our results provided evidence that certain water quality indicators, such as APHA color value, turbidity, and total phosphorus, have significant effects on visitors' sentiments related to CES. While existing research focused more on the ecological ecosystem services related to water quality, this study emphasized the importance of the impact of water quality on CES, which provides essential and intangible benefits for individuals and has economic importance. Through this study, we emphasized the importance of water quality in visitors' perception and evaluation of lake CES and highlighted the potential of user-generated content analysis for CES related sentiment analysis and lake management practices, which can facilitate sustainable local economic development. Overall, this research provides evidence to support sustainable planning strategies and water quality regulations to enhance the CES provided by lakes.

Reference

Allan, J. D., Smith, S. D., McIntyre, P. B., Joseph, C. A., Dickinson, C. E., Marino, A. L., ... & Adeyemo, A. O. (2015). Using cultural ecosystem services to inform restoration priorities in the Laurentian Great Lakes. *Frontiers in Ecology and the Environment*, *13*(8), 418-424.

Allan, J. D., Manning, N. F., Smith, S. D., Dickinson, C. E., Joseph, C. A., & Pearsall, D. R. (2017). Ecosystem services of Lake Erie: Spatial distribution and concordance of multiple services. *Journal of Great Lakes Research*, *43*(4), 678-688.

Angradi, T. R., Launspach, J. J., Bolgrien, D. W., Bellinger, B. J., Starry, M. A., Hoffman, J. C., ... & Hollenhorst, T. P. (2016). Mapping ecosystem service indicators in a Great Lakes estuarine Area of Concern. *Journal of Great Lakes Research*, *42*(3), 717-727.

Angradi, T. R., Launspach, J. J., & Debbout, R. (2018). Determining preferences for ecosystem benefits in Great Lakes Areas of Concern from photographs posted to social media. *Journal of Great Lakes Research*, *44*(2), 340-351.

Chen, X., Chen, Y., Shimizu, T., Niu, J., Nakagami, K. I., Qian, X., ... & Li, J. (2017). Water resources management in the urban agglomeration of the Lake Biwa region, Japan: An ecosystem services-based sustainability assessment. *Science of the Total Environment*, *586*, 174-187.

Egan, K. J., Herriges, J. A., Kling, C. L., & Downing, J. A. (2009). Valuing water quality as a function of water quality measures. American Journal of Agricultural Economics, 91(1), 106-123.

Egarter Vigl, L., Marsoner, T., Giombini, V., Pecher, C., Simion, H., Stemle, E., ... & Depellegrin, D. (2021). Harnessing artificial intelligence technology and social media data to support Cultural Ecosystem Service assessments. *People and Nature*, *3*(3), 673-685.

Environmental Protection Agency (EPA). (2018). Best management practices for lakes. Retrieved from https://www.epa.gov/lakes/best-management-practices-lakes

Figueroa-Alfaro, R. W., & Tang, Z. (2017). Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. *Journal of environmental planning and management*, *60*(2), 266-281.

Fu, B., Xu, P., Wang, Y., Yan, K., & Chaudhary, S. (2018). Assessment of the ecosystem services provided by ponds in hilly areas. *Science of the total environment*, *642*, 979-987.

Ghermandi, A., & Fichtman, E. (2015). Cultural ecosystem services of multifunctional constructed treatment wetlands and waste stabilization ponds: Time to enter the mainstream?. *Ecological Engineering*, *84*, 615-623.

Guan, J., Wang, R., Van Berkel, D., & Liang, Z. (2023). How spatial patterns affect urban green space equity at different equity levels: A Bayesian quantile regression approach. *Landscape and Urban Planning*, 233, 104709.

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), e12343.

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., Di Minin, E., 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.

Ho, L. T., & Goethals, P. L. (2019). Opportunities and challenges for the sustainability of lakes and reservoirs in relation to the Sustainable Development Goals (SDGs). *Water*, *11*(07), 1462.

Hogeboom, R. J., Knook, L., & Hoekstra, A. Y. (2018). The blue water footprint of the world's artificial reservoirs for hydroelectricity, irrigation, residential and industrial water supply, flood protection, fishing and recreation. *Advances in water resources*, *113*, 285-294.

Hossu, C. A., Iojă, I. C., Onose, D. A., Niță, M. R., Popa, A. M., Talabă, O., & Inostroza, L. (2019). Ecosystem services appreciation of urban lakes in Romania. Synergies and trade-offs between multiple users. *Ecosystem Services*, *37*, 100937.

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

Kundu, A., McBride, G., & Wuertz, S. (2013). Adenovirus-associated health risks for recreational activities in a multi-use coastal watershed based on site-specific quantitative microbial risk assessment. Water Research, 47(16), 6309-6325. doi:https://doi.org/10.1016/j.watres.2013.08.002

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.

Levin, N., Lechner, A. M., & Brown, G. (2017). An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. *Applied geography*, *79*, 115-126.

Moreno-Llorca, R., Méndez, P. F., Ros-Candeira, A., Alcaraz-Segura, D., Santamaría, L., Ramos-Ridao, Á. F., ... & Vaz, A. S. (2020). Evaluating tourist profiles and nature-based experiences in Biosphere Reserves using Flickr: Matches and mismatches between online social surveys and photo content analysis. *Science of the Total Environment*, *737*, 140067.

Mumbi, A. W., & Watanabe, T. (2020). Differences in risk perception of water quality and its influencing factors between lay people and factory workers for water management in River Sosiani, Eldoret Municipality Kenya. Water, 12(8), 2248.

Ochoo, B., Valcour, J., & Sarkar, A. (2017). Association between perceptions of public drinking water quality and actual drinking water quality: A community-based exploratory study in Newfoundland (Canada). Environmental research, 159, 435-443.

Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*, *94*, 74-86.

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*, 100252.

Reynaud, A., & Lanzanova, D. (2017). A global meta-analysis of the value of ecosystem services provided by lakes. *Ecological Economics*, *137*, 184-194.

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.

Schirpke, U., Scolozzi, R., & Tappeiner, U. (2021). "A Gem among the Rocks"—Identifying and Measuring Visual Preferences for Mountain Lakes. *Water*, *13*(9), 1151.

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

Sinclair, R. G., Jones, E. L., & Gerba, C. P. (2009). Viruses in recreational water‐borne disease outbreaks: a review. Journal of applied microbiology, 107(6), 1769-1780.

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

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*, 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), e0162372.

Sterner, R. W., Keeler, B., Polasky, S., Poudel, R., Rhude, K., & Rogers, M. (2020). Ecosystem services of Earth's largest freshwater lakes. *Ecosystem Services*, *41*, 101046.

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), 17615.

Tieskens, K. F., Van Zanten, B. T., Schulp, C. J., & Verburg, P. H. (2018). Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landscape and urban planning*, *177*, 128-137.

United States Environmental Protection Agency (USEPA). (2020). Clean water act: Progress and challenges. Retrieved from https://www.epa.gov/laws-regulations/summary-clean-water-act

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.

Compendium of physical activities. Compendium of Physical Activities. (n.d.). https://pacompendium.com/