# Can social media help us understand the impact of climate change on forests in the US?

1st Derek Van Berkel
*School for Environment & Sustainability*
*University of Michigan*
Ann Arbor, USA
dbvanber@umich.edu

2nd Nathan Fox
*School for Environment & Sustainability*
*University of Michigan*
Ann Arbor, USA
foxnat@umich.edu

3rd Stella Cousins
*School for Environment & Sustainability*
*University of Michigan*
Ann Arbor, USA
scousins@umich.edu

4th Kai Zhu
*School for Environment & Sustainability*
*University of Michigan*
Ann Arbor, USA
zhukai@umich.edu

***Abstract*—While social media data are increasingly being used in the study of pressing environmental problems, their ability to monitor environmental changes has scarcely been assessed. Understanding this viability is highly important as climate change increasingly impacts public health, and behavior. We examine social media photographs associated with wildfires in Yellowstone National Park to assess if images and content can adequately capture environmental change associated with large-scale landscape impacts - wildfires - using computer vision, natural language processing and spatiotemporal analysis. We find that social media posts associated with wildfire events rarely capture the fires themselves, while landscape impacts including burnt trees and early succession are more frequently the topic of photography. Furthermore, we find that computer vision has challenges with capturing these phenomena. While capturing wildfires proved difficult, developing multimodal analysis including natural language processing, spatial, trend and computer vision analysis at scale may open opportunities for more general understanding of social media's efficacy for monitoring environmental change.**

***Keywords*—Computer vision, Natural language processing, mapping, wildfire, climate change**

## I. Introduction

As the impact of climate change intensifies, forest ecosystems are experiencing increasing threats that undermine their vital role in regulating the atmospheric system, carbon stock stability, and recreation functions [1]. Shifting disturbance dynamics are contributing to larger, more damaging wildfires [2], drought [3], and surging pests and pathogens [4]. These impacts are especially acute in the Western U.S., where together they have recently driven unprecedented mortality [5]. Social media (SM ) may offer unique insights into social and ecological impacts from these threats from climate-driven stress and disturbance. The accumulation and accessibility of unprecedented amounts of user-generated data via social media [6; 7] offers unique opportunities to bring high fidelity socio-environmental understanding to rapidly changing forest ecosystems.

Sustainability researchers are increasingly turning to social media (SM) to investigate socio-ecological interactions, climate change discourses, explore urban sustainability, and provide novel insights into ecology and conservation science [8]. Forest-specific applications include disaster risk reduction [9], forest recreation and perceived attractiveness [10], and environmental quality monitoring [11]. However, its utility in detecting forest change and use is underexplored [12]. Questions remain about whether meaningful information can be gleaned from posts that include photographs, locations and text in forest contexts [13].

To date, large-scale detection of damage to trees and forests and their impact has been primarily conducted using remote sensing techniques, and in-person surveys. The spectral reflectance of burn scars, tree crown color, and phenological shifts are used to characterize the spatial and structural properties of a given disturbance episode, such as extent and severity [14; 15]. Determining tree decline and die-off, however, requires costly and time consuming onsite observations and expertise to determine how and why trees die (e.g. "ground-truthing"), limiting the scope and extent of validation efforts [16]. Though visually-quantifiable patterns have the potential to be detected in images captured through SM photographs,

DVB & NF contributed equally as first authors on this manuscript.

Funding for this project was provided by the US National Science Foundation (NSF) through grant 2033320 (NF and DVB), and 2244711, 2306198 (KZ), and the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship (NF).

ecologists have yet to fully tap into this novel data source. Photo-sharing platforms may offer large amounts of geo-referenceable visual and textual content for forests over time, which could add a new level of insight unavailable through traditional satellite and airphoto remote sensing [8].

Several scalable techniques may enable detection of forest changes for ground truthing and wildfire event monitoring. For example, computer vision has been used to detect phenological changes in photographs [17; 18], and wildfire detection in remotely sensed images [19]. Whether the subtle visual qualities of forest burns and wildfire can be detected in SM photography has yet to be thoroughly examined. Natural language processing (NLP), likewise, might offer enhanced detection of wildfire themed photography through analysis of text posted by SM users [12]. Visitation avoidance of forest changes inferred from SM activity, extensively used as a robust approximation of total visitation to protected areas [20], might also serve as an indicator of wildfire and its impacts. Our study aims to provide a proof of concept of how this unprecedented data might more deeply reveal the ecological and social impacts of forest change.

We aim to assess how effective disturbance photography from SM is for observing forest dynamics and environmental change, assessing whether there are sufficient photographs in locations of wildfire to detect these phenomena. We also assess if changes in forest conditions are tied to changes in visitation, and photography and might serve as a proxy of climate-induced behavioral change. This will contribute to furthering our understanding of SM's potential for monitoring landscapes and human behavior impacted by climate change.

## II. Methods

### A. Overview

To evaluate the effectiveness of SM in detecting wildfires, we conducted a comprehensive analysis of photographs sourced from the SM platform Flickr using the API and the R package photosearcher [21]. We applied our techniques to Yellowstone National Park, which has experienced multiple wildfires, and is frequented by visitors uploading their photographs and experiences of the protected areas. Flickr, a popular photo-sharing website, is frequently utilized by sustainability researchers due to its accessibility and the abundance of recreational photography shared within protected areas [8].

Our approach involved filtering the data to include only those photographs taken within a 5-kilometer radius of documented wildfires. We sourced information on these wildfires from the U.S. Geological Survey and USDA Forest Service's burn severity maps (Zhu et al., 2023). Given the visible effects of wildfires, such as smoke and significant landscape changes resulting from forest burn scars, we expected that their impact would be observable even at this distance. We analyzed and map visitation and photographic trends in Yellowstone, and conducted significance tests (i.e., t-test before and after wildfire) to identify any possible behavioral changes from wildfires. Finally, we employed computer vision and natural language processing techniques to assess whether photographers photographed or referenced wildfires in SM.

### B. Social Media, Computer Vision & NLP

We utilized the Google Vision API to classify the SM photographs for the years when Yellowstone experienced wildfires. The Google Vision API is a user-friendly image classification tool that provides a list of image tags (e.g., 10 tags) and their associated model confidence scores (e.g., 95% confidence of a tree in an image). It relies on convolutional neural networks (CNN) trained on manually classified datasets to identify various aspects of image context, such as color, edges, texture layers, and more.

While computer vision models have been developed for real-time wildfire detection in satellite imagery [19], these models are not openly accessible and have not been applied to analyzing photography. In our study, we specifically looked for tags associated with wildfires (e.g., fire, flame, smoke, forest fire, brush fire) and post-burn landscapes (e.g., burnt, scorch, singe, sear, char, blacken) within the results generated by Google Vision. We also conducted a manual visual assessment of image content to evaluate the accuracy of Google Vision in identifying wildfire-themed images.

Additionally, we applied Natural Language Processing (NLP) techniques to detect posts related to wildfires. We used the Bidirectional Encoder Representations from Transformers (BERT) topic modeler known as 'BERTopic' [22]. This approach allowed us to analyze the titles, descriptions, and tags across the entire SM dataset. BERTopic is capable of identifying semantic topic clusters within a given corpus, enabling us to distinguish meso-level topics in the text. This approach provided an efficient and scalable method for detecting whether our SM dataset contained accounts of wildfires.

To the best of the author's knowledge, this study represents a novel attempt to leverage multimodal data from social media to monitor climate-induced stressors like wildfires.

## III. Results

Between 2007 and 2023, a total of 147,469 photographs were uploaded by 4,898 Flickr users in Yellowstone National Park, averaging approximately 30.8 photographs per user. The weekly trend analysis of Flickr uploads (Fig. 1) demonstrates that the majority of visits, quantified in terms of unique photographers (Photo User Day or PUD), were observed during the summer months, coinciding with the peak tourism season (Fig. 1).

While there was a visible decline in photography in the weeks following a wildfire event (highlighted by the red line on the graph), t-test comparing these levels two weeks prior to a wildfire ($\overline{X}$ = 291.3) with those two weeks after ($\overline{X}$ = 253.9) did not yield a statistically significant difference (t = 0.36, p-value = 0.72). Similarly, visitation levels before and after a wildfire ($\overline{X}$ = 32.0 visitors pre-wildfire and 32.6 post-wildfire) were statistically indistinguishable (t = -0.06, p-value = 0.96).

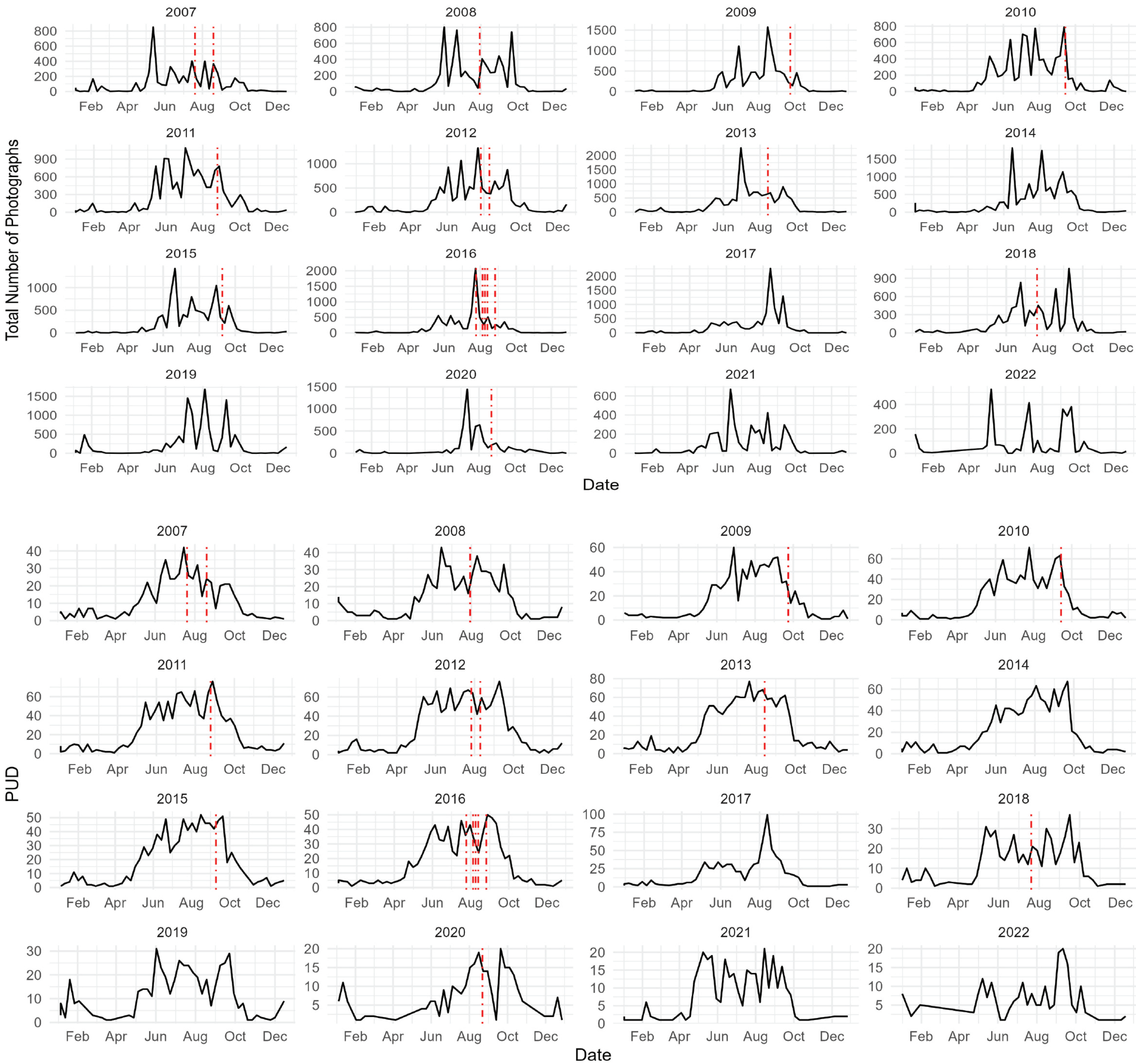


Fig. 1. Total Photo-User Days (PUD); and the cumulative count of photographs uploaded by Flickr users in Yellowstone National Park from 2007 to 2022 (bottom). Date of a wildfire is indicated with a red line.

Maps of photography taken two weeks before a wildfire and two weeks after display similar patterns, likewise, suggesting that there were no discernible behavioral responses to the fires (Fig. 2). Interestingly, the majority of Flickr photography posts are concentrated along roadways.

In the 9 years that it experienced a wildfire, 1,399 users uploaded 15,743 ($\overline{X}$ = 11.3) photographs within 5 kms of the locations of the fire events. Computer vision classified 29 photographs with the tag "fire", and 83 with "smoke". No additional tags that might indicate wildfire were detected. Manual validation of the computer vision results indicated poor identification of wildfire themed photographs. A mere 7 (6.3% of the total) photographs (3 with fire, 4 with smoke) were positively identified as wildfire. Photographs were either misclassified by the algorithm, (e.g. steam from the geological formation within the park classified as smoke; n = 74), or contained tag mixing (e.g. a campfire rather than a wildfire n = 12).

BERT topic modeling of the titles, descriptions and volunteered tags revealed a single semantic topic related to fire (n= 174). It also identified topics related to bison, elk, and the Norris Geyser Basin and "old faithful" geysers (Fig. 4).

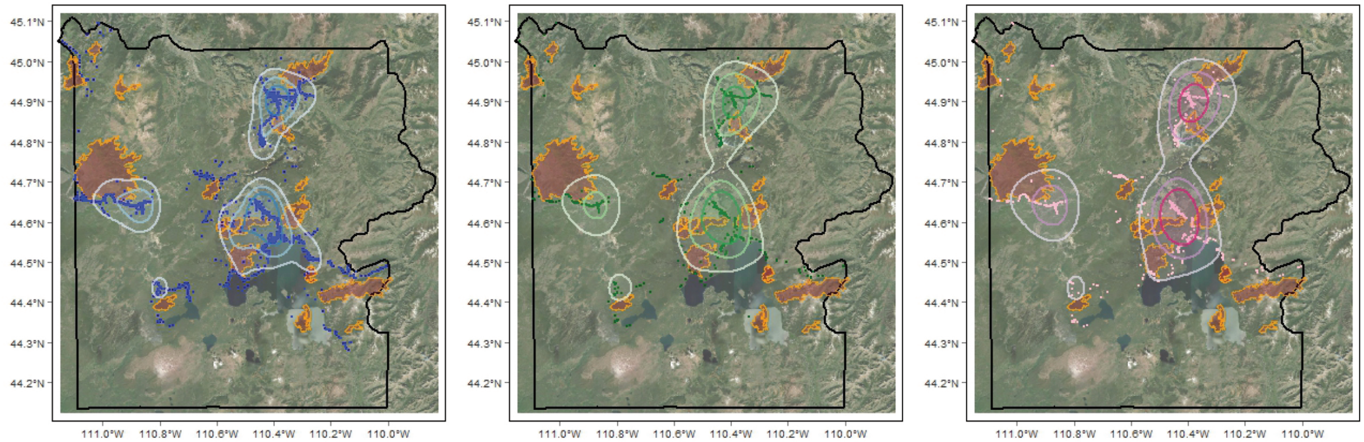


Fig. 2. Flickr photography (blue points) within 5km of a wildfire year, Yellowstone National Park (left); and total photographs (dark green) 2 weeks prior (middle), and (pink) the 2 weeks post a wildfire (right). Fires are depicted in red with orange outlines, and the isoline represents the 75 (darkest colors), 50, and 25(lightest) percentile of total points.

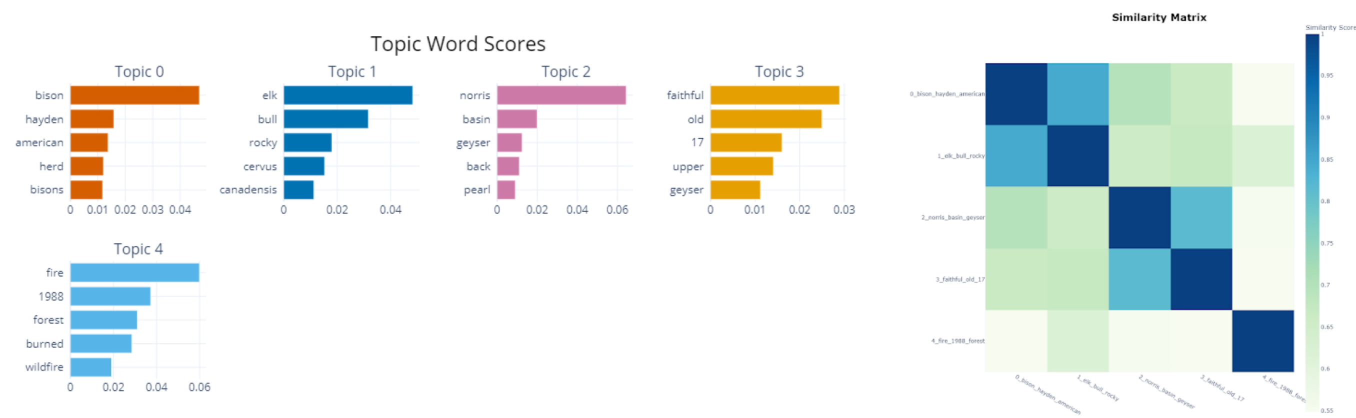


Fig. 3. The BERT topic model extracted the five most frequent topics from post descriptions. The topic similarity matrix provides a visual representation of the semantic distances between these topics. Notably, the wildfire topic (Topic 4) exhibits a substantial dissimilarity in comparison to all the other topics.

Fig. 4. Examples of wildfire photography (above); and post-wildfire impact including recent and early succession (below). All photograph attribution is available upon request.

Manual assessment of the photographs from the "fire" topic revealed that 15 (8.6%) contained characteristics indicating wildfire (e.g., smoke, fire), and 95 (55.0%) a recent wildfire or more advanced ecological succession after a burn. The other 65 (37.4%) did not include characteristics associated with wildfire or post burns from our visual inspection. Several of the photographs, 33 (18.6%), could not be conclusively determined to include visual elements associated with wildfire, while others contained geysers 11 (6.3%), campfires 6 (3.4%) and other non-applicable content 15 (8.6%). Moreover, BERT topic modeling proved more effective in identifying posts related to wildfires compared to basic keyword searches using terms such as "fire," "flame," "smoke," "forest fire," "brush fire," "burnt," "scorch," "singe," "sear," "char," and "blacken." Notably, BERT topic modeling successfully differentiated between wildfire-related content and other references, such as the popular tourist site "Firehole Falls" and the fire monitoring station in Yellowstone, which were included in the simplistic keyword search. Figure 4 showcases explicit examples of posts related to wildfires and post-wildfire photography, as identified through the combined analysis of computer vision and natural language processing techniques.

## IV. Discussion and conclusion

In this study, our objective was to assess the effectiveness of SM in gauging environmental changes. To accomplish this, we conducted tests involving Google Vision for image analysis and NLP for the automated identification of photographs depicting wildfires and their associated landscape impacts. Furthermore, we examined whether there were observable changes in SM activity in response to wildfire events. Our results indicate that wildfires are rarely captured by SM users, compared to the total SM content captured in Yellowstone. When it is the topic of photography, the actual wildfire events are rarely captured. Instead, landscape impacts including burnt trees and early succession are more frequently the topic of photography. While this might be due to restricted access to the area while the wildfires are burning (e.g. Yellowstone National Park often restricts access in wildfire events), our trend and spatial analysis did not indicate a significant decline in visitation and photography. By and large, photographers appear to document popular tourist sites, wildlife, and their own recreational activities on SM platforms [7]. Whether there is sufficient photography of wildfire to detect environmental change phenomena requires additional research at broader scales, and methods for distinguishing between recreation and events like wildfires.

Our evaluation of the effectiveness of these methodologies in detecting wildfire-related themes within SM posts yielded mixed results. The computer vision algorithm struggled to accurately distinguish wildfires and their associated impacts. In contrast, NLP analysis of descriptions provided by SM users proved to be significantly more useful in this regard. Future efforts should focus on developing specialized computer vision classifiers tailored specifically to wildfires and their ecological consequences. Such classifiers should be capable of distinguishing between wildfire images and those depicting campfires, clouds, steam, as well as burn scars and deadfall, which are indicative of post-wildfire landscapes. The vast repository of images available on various SM platforms, along with their accompanying textual descriptions, can serve as valuable resources for training and refining models aimed at detecting these phenomena. In this context, NLP is likely to remain a valuable tool for identifying topics and gauging attitudes within SM discussions related to wildfires and environmental changes.

Understanding the limitations of SM, and their associated textual, spatial and temporal metadata, will provide greater insight into the specific way scientists can monitor and understand change to human-nature interaction as a result of climate change impacts. Although the task of capturing wildfire-related data presented challenges, exploring a multimodal approach that integrates trend analysis, spatial analysis, and the scalable use of computer vision and NLP for other climate-induced stressors, such as pine beetles and drought, may unlock new avenues for gaining a broader and more comprehensive understanding of the effectiveness of social media in tracking environmental transformations related to climate change.

## References


[1] R. Seidl, D. Thom, M. Kautz, D. Martin-Benito, M. Peltoniemi, G. Vacchiano, J. Wild, D. Ascoli, M. Petr, J. Honkaniemi *et al.*, "Forest disturbances under climate change," *Nature climate change*, vol. 7, no. 6, pp. 395–402, 2017.

[2] V. Iglesias, J. K. Balch, and W. R. Travis, "Us fires became larger, more frequent, and more widespread in the 2000s," *Science advances*, vol. 8, no. 11, p. eabc0020, 2022.

[3] J. T. Overpeck and B. Udall, "Climate change and the aridification of north america," *Proceedings of the national academy of sciences*, vol. 117, no. 22, pp. 11 856–11 858, 2020.

[4] C. J. Fettig, C. Asaro, J. T. Nowak, K. J. Dodds, K. J. Gandhi, J. E. Moan, and J. Robert, "Trends in bark beetle impacts in north america during a period (2000–2020) of rapid environmental change," *Journal of Forestry*, vol. 120, no. 6, pp. 693–713, 2022.

[5] J. Axelson, J. Battles, B. Bulaon, D. Cluck, S. Cousins, L. Cox, B. Estes, C. Fettig, A. Hefty, S. Hishinuma *et al.*, "The california tree mortality data collection network—enhanced communication and collaboration among scientists and stakeholders," *California Agriculture*, vol. 73, no. 2, pp. 55–62, 2019.

[6] B. T. Van Zanten, D. B. Van Berkel, R. K. Meentemeyer, J. W. Smith, K. F. Tieskens, and P. H. Verburg, "Continental-scale quantification of landscape values using social media data," *Proceedings of the National Academy of Sciences*, vol. 113, no. 46, pp. 12 974–12 979, 2016.

[7] D. B. Van Berkel, P. Tabrizian, M. A. Dorning, L. Smart, D. Newcomb, M. Mehaffey, A. Neale, and R. K. Meen-

temeyer, “Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and lidar,” *Ecosystem services*, vol. 31, pp. 326–335, 2018.

[8] A. Ghermandi, J. Langemeyer, D. Van Berkel, F. Calcagni, Y. Depietri, L. E. Vigl, N. Fox, I. Havinga, H. Jäger, N. Kaiser *et al.*, “Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use,” *One Earth*, vol. 6, no. 3, pp. 236–250, 2023.

[9] F. C. Moore and N. Obradovich, “Using remarkability to define coastal flooding thresholds,” *Nature communications*, vol. 11, no. 1, p. 530, 2020.

[10] P. Pereira, I. Bogunovic, W. Zhao, and D. Barcelo, “Short-term effect of wildfires and prescribed fires on ecosystem services,” *Current Opinion in Environmental Science & Health*, vol. 22, p. 100266, 2021.

[11] A. S. Vaz, J. F. Gonçalves, P. Pereira, F. Santarém, J. R. Vicente, and J. P. Honrado, “Earth observation and social media: Evaluating the spatiotemporal contribution of non-native trees to cultural ecosystem services,” *Remote Sensing of Environment*, vol. 230, p. 111193, 2019.

[12] N. N. Kaiser, C. K. Feld, and S. Stoll, “Does river restoration increase ecosystem services?” *Ecosystem Services*, vol. 46, p. 101206, 2020.

[13] E. J. Wilkins, S. A. Wood, and J. W. Smith, “Uses and limitations of social media to inform visitor use management in parks and protected areas: A systematic review,” *Environmental Management*, vol. 67, no. 1, pp. 120–132, 2021.

[14] C.-y. Huang, W. R. Anderegg, and G. P. Asner, “Remote sensing of forest die-off in the anthropocene: From plant ecophysiology to canopy structure,” *Remote Sensing of Environment*, vol. 231, p. 111233, 2019.

[15] J. C. Mitchell, D. M. Kashian, X. Chen, S. Cousins, D. Flaspohler, D. S. Gruner, J. S. Johnson, T. D. Surasinghe, J. Zambrano, and B. Buma, “Forest ecosystem properties emerge from interactions of structure and disturbance,” *Frontiers in Ecology and the Environment*, vol. 21, no. 1, pp. 14–23, 2023.

[16] A. J. Das, N. L. Stephenson, and K. P. Davis, “Why do trees die? characterizing the drivers of background tree mortality,” *Ecology*, vol. 97, no. 10, pp. 2616–2627, 2016.

[17] A. S. Gosal, I. R. Geijzendorffer, T. Václavík, B. Poulin, and G. Ziv, “Using social media, machine learning and natural language processing to map multiple recreational beneficiaries,” *Ecosystem Services*, vol. 38, p. 100958, 2019.

[18] D. L. Correia, W. Bouachir, D. Gervais, D. Pureswaran, D. D. Kneeshaw, and L. De Grandpre, “Leveraging artificial intelligence for large-scale plant phenology studies from noisy time-lapse images,” *IEEE Access*, vol. 8, pp. 13 151–13 160, 2020.

[19] A. Bouguettaya, H. Zarzour, A. M. Taberkit, and A. Kechida, “A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms,” *Signal Processing*, vol. 190, p. 108309, 2022.

[20] S. A. Wood, A. D. Guerry, J. M. Silver, and M. Lacayo, “Using social media to quantify nature-based tourism and recreation,” *Scientific reports*, vol. 3, no. 1, p. 2976, 2013.

[21] N. Fox, T. August, F. Mancini, K. E. Parks, F. Eigenbrod, J. M. Bullock, L. Sutter, and L. J. Graham, ““photosearcher” package in r: An accessible and reproducible method for harvesting large datasets from flickr,” *SoftwareX*, vol. 12, p. 100624, 2020.

[22] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based tf-idf procedure,” *arXiv preprint arXiv:2203.05794*, 2022.