LETTER • OPEN ACCESS

Mapping and modeling the impact of climate change on recreational ecosystem services using machine learning and big data

To cite this article: Kyle Manley and Benis N Egoh 2022 *Environ. Res. Lett.* **17** 054025

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Mapping and modeling the impact of climate change on recreational ecosystem services using machine learning and big data

Kyle Manley[*] and Benis N Egoh

Department of Earth System Science, University of California Irvine, Irvine, CA 92697, United States of America
* Author to whom any correspondence should be addressed.

E-mail: ktmanley@uci.edu

Abstract

The use of recreational ecosystem services is highly dependent on the surrounding environmental and climate conditions. Due to this dependency, future recreational opportunities provided by nature are at risk from climate change. To understand how climate change will impact recreation we need to understand current recreational patterns, but traditional data is limited and low resolution. Fortunately, social media data presents an opportunity to overcome those data limitations and machine learning offers a tool to effectively use that big data. We use data from the social media site Flickr as a proxy for recreational visitation and random forest to model the relationships between social, environmental, and climate factors and recreation for the peak season (summer) in California. We then use the model to project how non-urban recreation will change as the climate changes. Our model shows that current patterns are exacerbated in the future under climate change, with currently popular summer recreation areas becoming more suitable and unpopular summer recreation areas becoming less suitable for recreation. Our model results have land management implications as recreation regions that see high visitation consequently experience impacts to surrounding ecosystems, ecosystem services, and infrastructure. This information can be used to include climate change impacts into land management plans to more effectively provide sustainable nature recreation opportunities for current and future generations. Furthermore, our study demonstrates that crowdsourced data and machine learning offer opportunities to better integrate socio-ecological systems into climate impacts research and more holistically understand climate change impacts to human well-being.

1. Introduction

Ecosystem services are the benefits provided to us by nature, for example, carbon sequestration, flood regulation, and recreation. Recently, many global policies and assessments have been incorporating ecosystem services in order to demonstrate the importance of properly functioning ecosystems and to justify their restoration and conservation (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) 2019, European Commission *et al* 2020, Convention on Biological Diversity 2021). In January 2021 U.S. President Biden included ecosystem services in an executive order for climate change action (The White House 2021). Such high-profile interest in ecosystem services and associated assessments rely upon mapping and modeling to effectively understand, monitor, and preserve ecosystem services and to assess impacts of climate change on ecosystem services. Ecosystem service research is a relatively young discipline, but the scientific literature is growing and methods are improving, yet major gaps and limitations still remain (Burkhard and Maes 2017, McDonough *et al* 2017, Ochoa and Urbina-Cardona 2017, Kosanic and Petzold 2020). Specifically, data availability issues force researchers to rely on expert knowledge or simple proxies (i.e. land use and land cover) that

do not fully capture the multidimensional traits and dynamics of ecosystems and their services, including the essential socio-ecological aspects (Ochoa and Urbina-Cardona 2017, Lautenbach *et al* 2019). This contributes to another major gap within the field, the lack of connecting social and biophysical (i.e. socio-ecological) aspects of ecosystem services (Lavorel *et al* 2017, Mandle *et al* 2021).

As computational technology and techniques have improved, a greater set of resources have become available, including big data and machine learning techniques. Vast amounts of big data now exist from our modern technologies including social media, mobile phones, sensors, and a plethora of other sources that can provide myriad insights into environmental attributes, including ecosystem services, and aid researchers in overcoming limitations (Havinga *et al* 2020, Xia *et al* 2020). Machine learning techniques have also quickly become an indispensable tool for many researchers, partially due to the data driven attributes (i.e. empirically modeled data with few or no *a-priori* assumptions) as well as the efficacy and efficiency when handling big data (Willcock *et al* 2018, Scowen *et al* 2021). Both resources offer opportunities to overcome the major limitations and fill gaps within ecosystem service research. For example, geolocated big data from social media gives us previously unavailable insights into cultural ecosystem services from a beneficiary perspective along with high resolution corresponding spatiotemporal information (Havinga *et al* 2020, Wood *et al* 2020). This has an even more significant use due to the high relevance for studying cultural ecosystem services, like recreation, which are currently understudied compared to other ecosystem services largely due to low data availability (Cheng *et al* 2019, Kosanic and Petzold 2020). Furthermore, machine learning techniques, like random forest, handle the nonlinear dynamics of ecosystem services and socio-ecological systems in general by allowing us to approximate relationships between model inputs and outputs directly from data rather than applying the linear constraints and *a priori* assumptions of traditional regression algorithms (Ochoa and Urbina-Cardona 2017, Lautenbach *et al* 2019, Scowen *et al* 2021).

These improved techniques and tools are essential for effectively mapping and modeling ecosystem services and are especially important as the climate continues to rapidly change. Since ecosystem services are a direct product of functional ecosystems, which are highly vulnerable to climate and environmental changes, climate change poses a huge threat and in some cases is already impacting ecosystem services (Pachauri *et al* 2014, Palomo 2017, Runting *et al* 2017, Pörtner *et al* 2021). To this point, climate impacts research has mainly focused on impacts to ecological systems (e.g. species, ecosystem structure and functioning) or impacts to human systems, but have rarely connected the two through impacts

to ecosystem services (Zommers *et al* 2014, Palomo 2017, van der Geest *et al* 2019, Brauman *et al* 2019). This is of concern as it has been shown that climate change is increasingly becoming a more prevalent direct driver in changes to ecosystem services and is consequently increasingly impacting human well-being (IPBES 2019, Pörtner *et al* 2021). Thus, it is essential to recognize that human systems and ecosystems are coupled (Berkes and Folke 1998). An understanding of climate change impacts to ecosystem services will help inform more effective policy and management plans for mitigation and adaptation in the future (UNEP 2016, van der Geest *et al* 2019, Pörtner *et al* 2021).

Recreation's close link with natural resources makes it more vulnerable to climate and environmental change than other sectors of the economy (National Academy of Sciences 1992) and evidence has shown that climate change is a significant threat to recreation and requires immediate addressing by decisionmakers and land managers (Walls *et al* 2009, Askew and Bowker 2018). Recreation contributes direct benefits to national, statewide, and surrounding local economies as well as indirect economic benefits through contributions to individual physical health, and psychological and emotional well-being (Haines-Young and Potschin 2012, Hermes *et al* 2018). Beyond material benefits, cultural ecosystem services, including recreation, also provide non-material benefits to individual well-being through good social relationships, cultural identity, and freedoms of choice and action (Shin *et al* 2019). Despite this importance cultural services have been understudied compared to other ecosystem services (Runting *et al* 2017). This understudy is even more important to note when considering cultural ecosystem services vulnerability to climate and environmental change. Thus, in this study we assess this vulnerability in the state of California. According to the latest statewide climate change assessment, California is projected to experience an increase in average maximum daily temperature of ∼5.6 °F–8.8 °F by late in the century, decreases in precipitation in the south and increases in the north, decreases in snowpack, and much more (Bedsworth *et al* 2018). Along with such drastic regional shifts in climate, it would also be expected that ecosystems and the services they provide will be impacted. To add further stress, California recreation, and thus the demand for recreational ecosystem services, is projected to increase due to the continued increase in population (Halofsky *et al* in press).

This study looks to fill knowledge gaps relating to fully integrating socio-ecological aspects of ecosystem services in order to better assess climate change impacts. Furthermore, we look to address the disparity in ecosystem service studies that have largely understudied cultural ecosystem services. We do this through a case study on the impacts of climate

change to non-urban recreational ecosystem services in California and overcome historical limitations using geolocated social media data and a machine learning algorithm. Specifically, we address these gaps by (a) mapping the current patterns in recreational ecosystem service use in California, (b) modeling the relationship between the demand for these ecosystem services and biophysical influences, and (c) modeling and mapping how future climatic changes will impact those relationships and thus, recreational ecosystem services in California in the future. Some studies have also used big data and/or machine learning to assess ecosystem services, but based on our knowledge, there have not been any studies taking advantage of these tools to assess climate change impacts on ecosystem services. Due to the regional focus, high-resolution social media recreation data, and machine learning algorithm that allows us to better connect the natural and human dimensions of ecosystem services, this study provides detailed insights that can be used by decision makers to focus land management on outcomes beneficial to human well-being under a future of environmental and climate change.

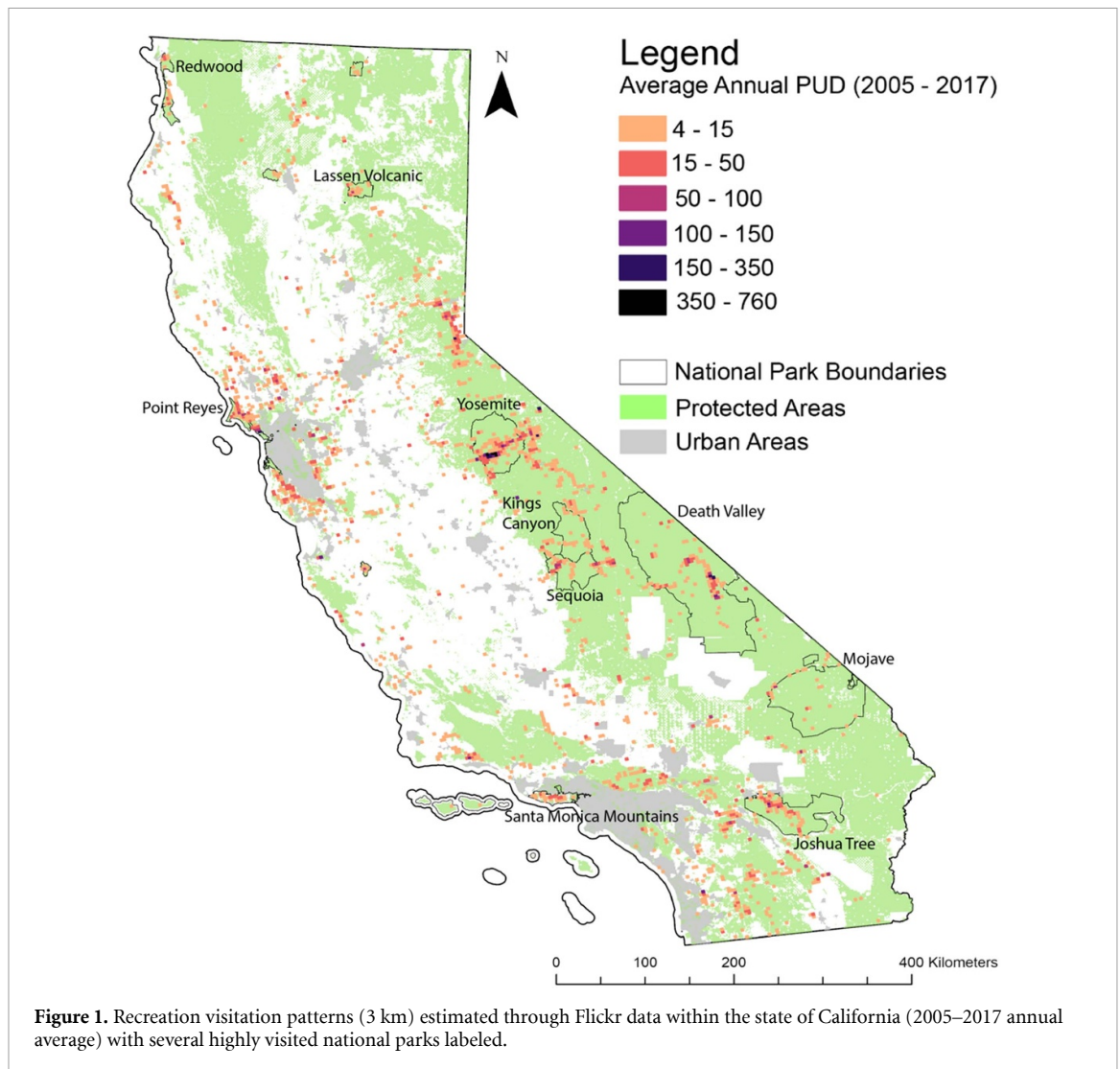## 2. Data and methods

2.1. Study region
Recreation is a major service provided by California's uniquely diverse ecosystems. According to the California Protected Areas Database, approximately 47% (∼49.5 million acres) of Californian land is protected (CPAD 2021). This land is managed by a multitude of agencies, with the US Forest Service (∼20.7 million acres), Bureau of Land Management (∼14 million acres), and National Park Service (∼7.6 million acres) managing the majority. Most of the state is within the California Floristic Province which contains exceptionally high diversity and endemism (Burge *et al* 2016). This highly diverse landscape consists of approximately 49% grassland/shrubland, 27% forest, 12% agriculture, 4% developed, and 8% other (water, barren, mining, wetlands, ice/snow) (Wilson *et al* 2015) (figure S1 available online at stacks.iop.org/ERL/17/054025/mmedia). Furthermore, the state is defined by five ecoregions, the most prominent being the Mediterranean ecoregion comprising most of the area (coastal and central), West Cordillera in the northern and Sierra Nevada regions, and warm deserts in the southeast (figure S1). According to the Bureau of Economic Analysis, some of the popular recreational activaties include boating/fishing, RVing, biking, climbing, hiking, and hunting. In 2019 recreation added $57 billion to the state economy, approximately 2% of state gross domestic product and the highest amount of any state and is the second fastest growing state recreation economy in the country (U.S. Bureau of Economic

Analysis 2019). Recreation also provided over 575000 jobs, the most of any state recreation economy in the country.

2.2. Recreation visitation and climate data
Social media data presents a great opportunity for novel insights by greatly increasing the magnitude and spatiotemporal coverage of data that is relevant to both the social and ecological aspects of cultural ecosystem services. Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) is a suite of models, developed by the Natural Capital Project, that map, value, and assess ecosystem services. The InVEST recreation model utilizes social media data to quantify the recreational ecosystem service usage rates across landscapes. Because empirical data on recreation rates is often hard to come by, the model utilizes recreational related photos from the social media site Flickr to quantify the photo-user-days (PUDs) of a defined area of interest. PUD calculates the number of unique users who took at least one photo within each grid cell in order to avoid overestimates from a single user taking and uploading many photos (Wood *et al* 2013). This PUD calculation is not an actual estimation of total recreational visitation but is a proxy for recreational patterns within a specific region (Kim *et al* 2019).

In order to map the use of recreational ecosystem services within California, we queried the collection of millions of geolocated global photographs through the InVEST recreation model for the complete available time period of 2005–2017 and for our area of interest, California. Although some recreational ecosystem services are supplied by developed regions, we focus solely on services outside of these urban areas. We use the 2010 adjusted urban area boundaries shapefile provided by the California Department of Transportation derived from the 2010 Census to mask out the urban regions, in order to best assess recreational ecosystem services provided by nature. This method allows us to filter out the unrelated noise from heavily populated areas like Los Angeles and the Bay Area that supply recreational services beyond those that are provided by ecosystems. Once developed areas are masked out, we input the shapefile to the model, specify the model to grid the output data to 3 km, and run the model. Output data is a shapefile of the PUD calculation for each grid cell within the area of interest for each year as well as for the entire time frame of 2005–2017 (figure 1). Our final dataset included an annual average of ∼34 750 Flickr photos across California, with most photos coming from central and southern California. Thus, we can incorporate recreation from a multitude of environments and their varying biophysical aspects across California, which would otherwise be difficult using traditional data collection methods such as interviews, surveys, or focus-groups. This strategy

Figure 1. Recreation visitation patterns (3 km) estimated through Flickr data within the state of California (2005–2017 annual average) with several highly visited national parks labeled.

also allows us to use empirical data on human behavior that is directly related to recreational ecosystem services to estimate the spatiotemporal preferences for the demand of recreational services in California, thus addressing some of the issues that come with quantifying cultural ecosystems services.

This study also required both past regional climate data and future downscaled climate projections. Historical 1 km monthly gridded climate data, including maximum temperature and precipitation, were acquired from the Daymet version 4 monthly climate summaries dataset (Thornton *et al* 2020) and wind speed data was acquired from the Terraclimate global monthly climate dataset (Abatzoglou *et al* 2018). Future max temperature, precipitation, and wind speed projection data was acquired from the 'Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections Archive' (Reclamation 2013), in which we specifically used data from the CMIP5 CESM1-BGC projections for representative concentration pathway (RCP) scenarios 4.5 and 8.5 (1/8 degree resolution). Major highway data was acquired from the US Census Bureau's TIGER geodatabase

and water body data was acquired from the California Department of Fish and Wildlife. Koppen-Geiger climate classifications were derived from Beck *et al* (2018), but future projections were not used due to data availability issues. Current and future land cover data was obtained from Sleeter *et al* (2017), who used population and climate projections under a business-as-usual scenario to backcast and project land use land cover under a business-as-usual scenario for California. All data was clipped to the area of interest and historical data was averaged for the timeframe of the visitation data (2005–2017) and for the peak season (July–September).

2.3. Random forest model
Even though random forest is not explicitly a spatial method, it performs comparatively as well as traditional spatial methods and is effective at solving spatial problems (Breiman 2001, Fox *et al* 2020). Using the PUD within different areas in California and the social and biophysical data within these regions, we created a random forest regression model to connect the summer demand for recreational ecosystem

services to social, environmental, and climate variables. We then project how a change in these factors under different future climate scenarios (RCP 4.5 and 8.5) could impact the suitability of future recreational services. It is important to note that the purpose of the model predictions is not to project future recreation, but the suitability for the use of recreational ecosystem services as different variables that are important for use change. This model is created by inputting the explanatory variables (social and biophysical) and the dependent variable (recreational visitation estimates) for the peak season (July–September) within a random forest model in ArcGIS. The random forest algorithm then creates an ensemble of hundreds of 'decision trees' for each pixel that each test a large and randomly selected subset of various combinations of all the input variables and a subset of the data between those variables. The model takes an average of each tree's prediction in order to create the best possible overall model for the recreational ecosystem services within the region. One big advantage of the random forest algorithm is that it adds a large amount of randomness into the regression analysis causing it to be a relatively effective method for avoiding overfitting. This is important because recreation is based upon a large set of complexly interacting variables and is thus vulnerable to overfitting. Furthermore, because random forest is a data driven model, it can undertake modeling tasks without making *a-priori* assumptions about the input data, allowing it to oftentimes more accurately deduce relationships and making it a more effective method for modeling the complex nonlinear socio-ecological relationships of ecosystems and the services they produce (Willcock *et al* 2018, Rammer and Seidl 2019, Scowen *et al* 2021).
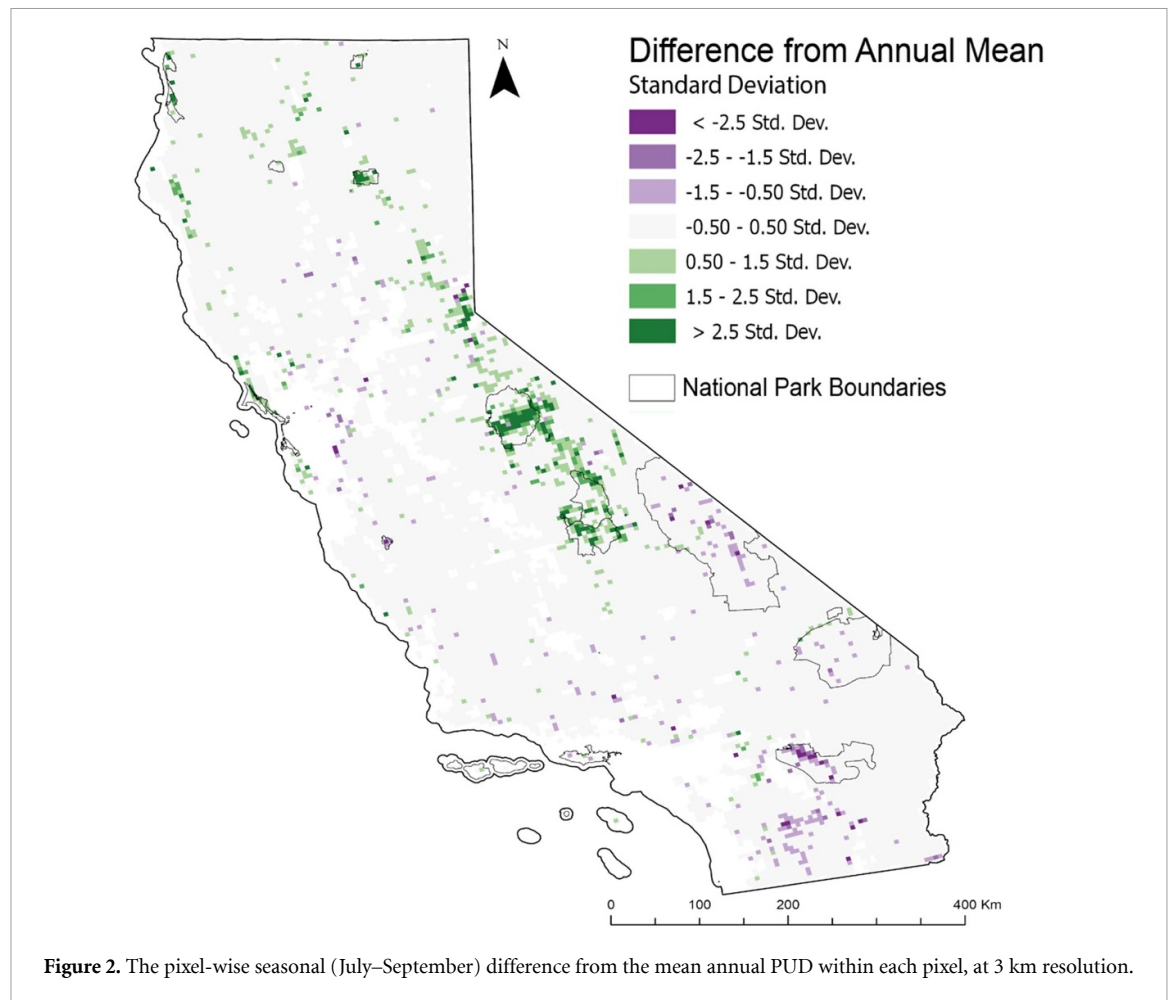
Explanatory variables that were tested for the recreation model include maximum temperature, precipitation, wind speed, Koppen-Geiger climate classifications, land cover, distance from highways, and trail density. Another benefit of using random forest is its ability to calculate the importance of each input variable to predicting the target output variable (i.e. the Gini coefficient). This allowed us to test the model on various combinations of these variables. Ultimately, median importance varied from ∼50 to over 500, we excluded variables with median importance values under 100 (land cover), in an attempt to improve the model by minimizing the amount of noise and lowering the chance of overfitting by only including relatively important variables. After testing these variables, the most effective model (as determined through out-of-bag estimates) included trail density, average maximum temperature, average wind speed, average precipitation, distance from major highways, the Koppen-Geiger climate classification, and distance from water bodies. The climate data that are used as explanatory variables are seasonal averages for the timeframe of 2005–2017. We train the recreation model on the mean difference

of seasonal recreation from annual recreation. We do this by first computing the average monthly visitation for the summer recreation season in California (July–September) and taking the difference from the average monthly visitation for the year (2005–2017) within each pixel of the study area. Using the difference from the mean for the season allows us to add a temporal component rather than just a spatial analysis, allowing for a better examination on how climate interacts spatiotemporally with the demand for recreational ecosystem services (figure S2).

We separated the input data with 80% being used for training the model and the other 20% for validation. We specified the model so that it generates 1000 separate trees to get a robust training set and 500 trees for validation purposes. We fine-tuned hyperparameters by running the model and checking out of bag errors in order to avoid overfitting and increase generalizability of the model. We also quantified the relative importance of each climatic variable for recreation. This variable importance is a measure of the impact of each individual variable on a correct estimate of recreational visitation. This assessment gives some insight into what factors are playing a role in influencing the highest annual average recreation visits in Summer and allows us to exclude less influential factors and minimize noise. Once hyperparameters are optimized and only relatively important explanatory variables are included, the model is trained and validated. We also calculate the partial dependence of each covariate to illustrate the non-linear relationships of recreational ecosystem services (figure 3). Using the trained model, we project how recreation patterns could be impacted in the future by the years 2030, 2050, and 2099 under RCP scenarios 4.5 and 8.5. Future climate data is used for each respective year and RCP scenario, while all other variables are held constant due to limitations on projection availability. We choose these two climate scenarios to provide a range of impacts in which RCP 4.5 would show impacts to recreational ecosystem services under a future in which greenhouse gas emissions are relatively rapidly mitigated versus RCP 8.5 in which little to no action is taken on mitigating greenhouse gas emissions.

3. Results

Assessing the seasonal difference of recreational patterns from the annual mean shows there is a spatiotemporal difference in recreation throughout the state (figure 2). Milder mountainous, northern, and coastal regions like those in and around Yosemite, Sequoia, Lassen Volcanic, and Point Reyes see above average recreation whereas southern and historically hotter regions like those in and around Death Valley, Mojave, and Joshua Tree see lower than average recreation during the peak season (July–September) (figure S2).

Figure 2. The pixel-wise seasonal (July–September) difference from the mean annual PUD within each pixel, at 3 km resolution.
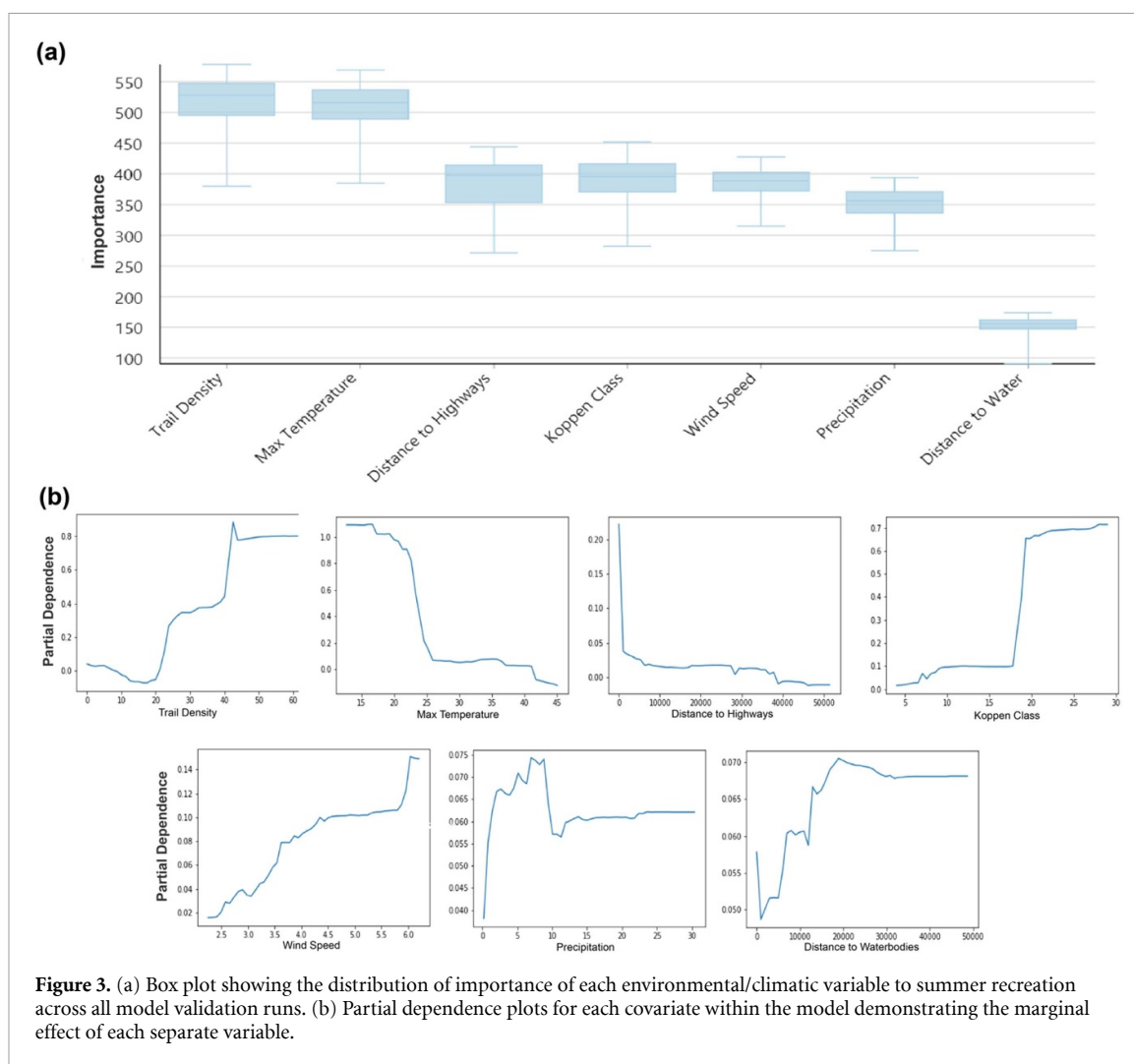
The initial training of the model using random forest allowed us to assess the relative importance of each explanatory variable to recreation in the region (figure 3(a)). The final model included seven environmental and climatic variables, with trail density and average seasonal max temperature being the most important for recreation. Furthermore, we calculated the partial dependence of each variable to assess the marginal effect of each separate variable on the model prediction of recreational ecosystem services (figure 3(b)).

Once the model was trained, we used validation runs to assess the prediction efficacy of the model (figure 4). The training runs resulted in an R^2 of 0.924, a p-value of <0.001, and a standard error of 0.002. Although the validation resulted in a lower R^2 (0.305) likely due to the inability of the model to predict similar visitation values as estimated by Flickr data, the model still is able to effectively predict overall spatial and magnitudinal patterns in recreational ecosystem services (figure 4). A calculated Pearson correlation coefficient of 0.94 confirms this spatial correlation between the empirical and modeled PUD.
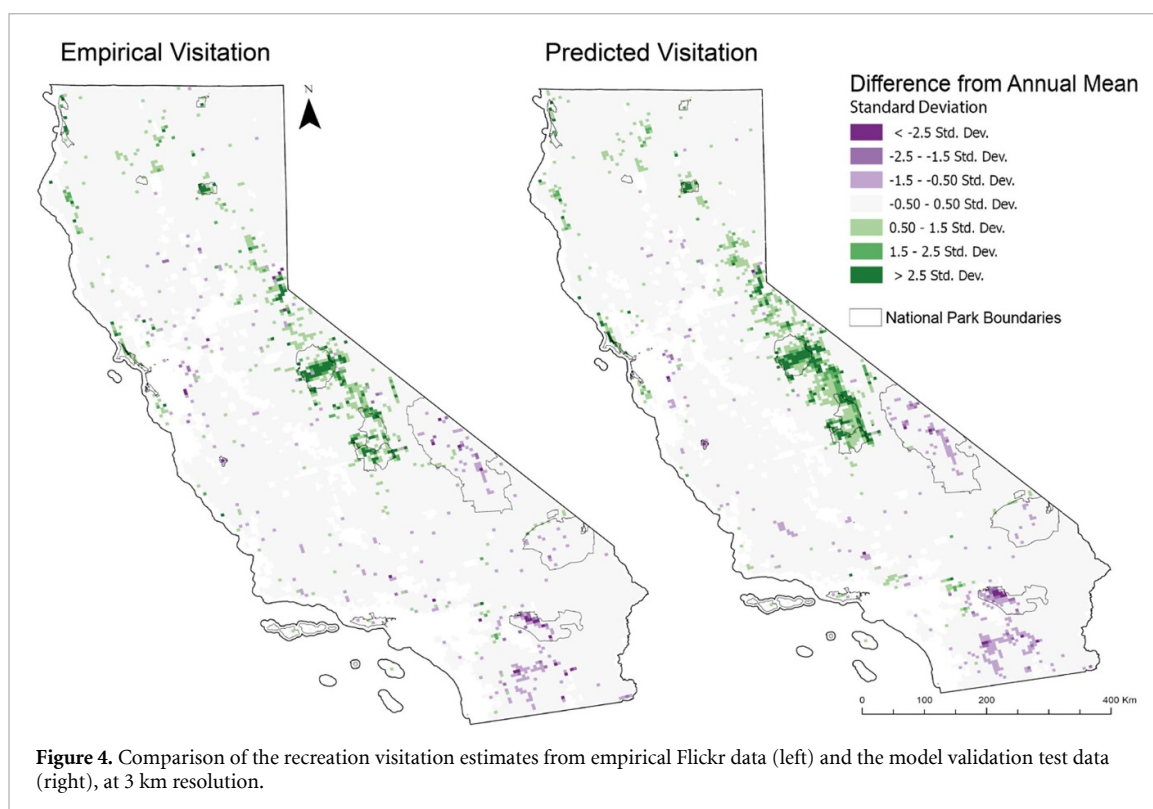
Model results for each scenario show patterns of change in the suitability for future recreational ecosystem service use based on the socio-ecological relationships that the model was able to deduce

(figure 5). Both scenarios show increases in the suitability for recreation over time in mountain, northern, and some coastal regions. Under RCP 4.5 there is a slight increase in non-suitable area within the Central Valley by 2099. This Central Valley phenomenon is seen much earlier in the RCP 8.5 scenario (by 2030) and extends throughout the entire valley by 2099. Generally, patterns seen within the Flickr data (e.g. high recreation in the Sierra Nevada) are exacerbated under a warming climate and more so under greater warming scenarios.

In order to add further context to the model results we assessed how the original recreational estimates and the model results for 2099 differed based on the corresponding current and future land cover (figure 6). We assessed the corresponding land cover class for areas of above average recreation (>0.5 standard deviations) and below average recreation (<0.5 standard deviations). Unsurprisingly, most recreation in the summer was within the forest while below average recreational activities occurred in shrublands. This pattern persists into the future based on our model projections. Overall, areas that are suitable for recreational activities drop significantly under climate change. Future predictions show that, under both scenarios, land area that is suitable for recreation is less than land area not suitable for recreational

Figure 3. (a) Box plot showing the distribution of importance of each environmental/climatic variable to summer recreation across all model validation runs. (b) Partial dependence plots for each covariate within the model demonstrating the marginal effect of each separate variable.
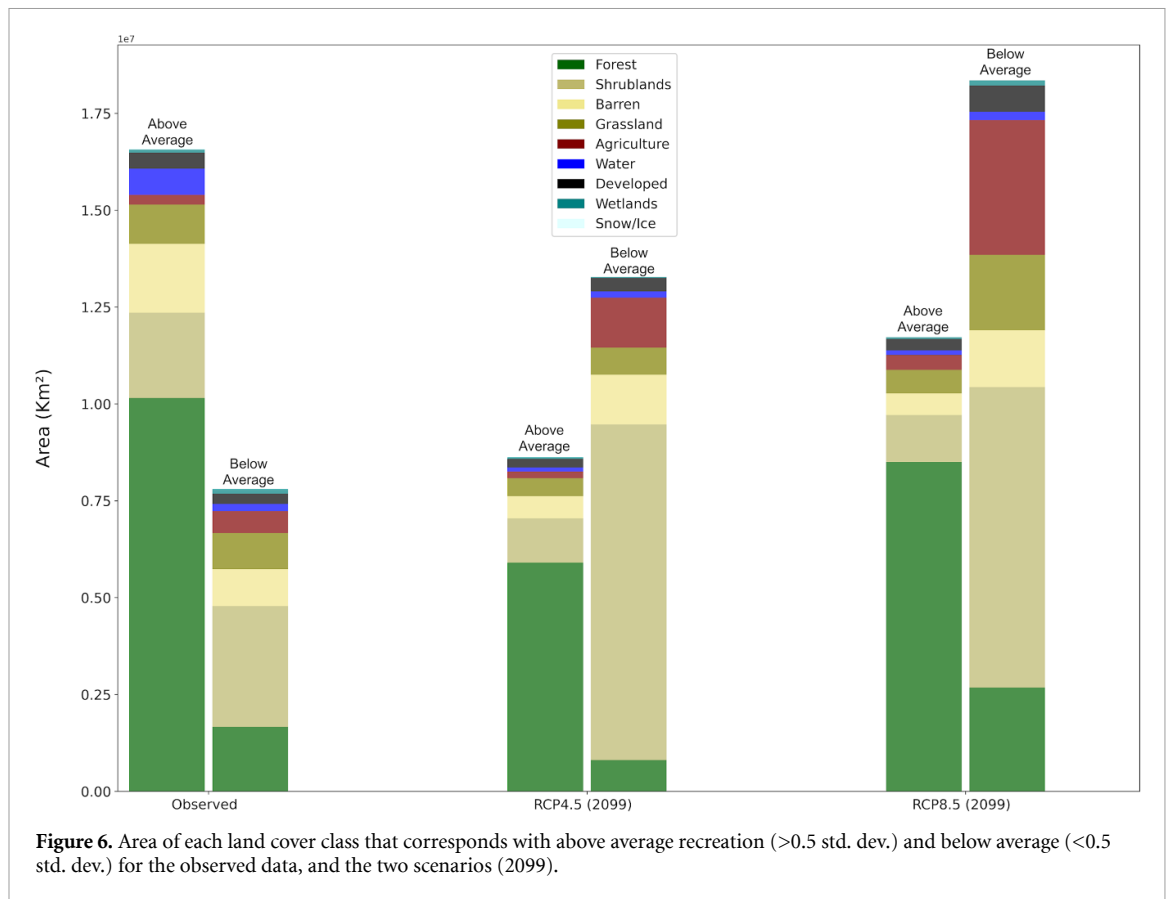

Figure 4. Comparison of the recreation visitation estimates from empirical Flickr data (left) and the model validation test data (right), at 3 km resolution.

Figure 5. Results (3 km resolution) from the recreation ecosystem service model with scenarios (RCP 4.5 and 8.5) separated by column and years separated by row, darker green represents greater suitability darker purple represents less suitability.

activities in the summer. Shrublands and agricultural lands see the greatest increases in unsuitable conditions for recreation, while forest areas seem to be relatively robust against these impacts. Under a higher end warming scenario, there are increases in both suitable and unsuitable recreation area. We note

that future projections should not be directly compared to observations of recreation as they are not predictions of recreation, rather projections of suitability. Patterns of the proportion of above average compared to below average area and proportion of land cover types, or direct comparisons

Figure 6. Area of each land cover class that corresponds with above average recreation (>0.5 std. dev.) and below average (<0.5 std. dev.) for the observed data, and the two scenarios (2099).

between the two RCP scenarios are more relevant to assess.

4. Discussion and conclusions

This study set out to address gaps in our understanding of climate change impacts on human well-being by connecting the two through ecosystem services. Specifically, there is a major lack of connecting the biophysical and social aspects of ecosystem services that results in an incomplete understanding of impacts to human well-being. Due to the recent availability and access to crowdsourced data, we can start bridging the socio-ecological modeling gap. We set out to test the use of this data for assessing climate change impacts on ecosystem services, specifically recreation. Along with crowdsourced data, we now also have machine learning methods that give us a tool to more effectively and efficiently use big data. For example, due to the data driven aspect of random forest, we were able to better integrate the nonlinear dynamics of socio-ecological interactions within our model rather than rely on the linear assumptions of traditional regression techniques (figure 3). Furthermore, the use of random forest allowed us to test the importance of relationships between multiple different environmental and climatic factors to recreational ecosystem services. Our study adds to the past literature demonstrating how effective a tool random forest is for better incorporating socio-ecological factors in

ecosystem service assessments (Richards and Tuncer 2018, Shiferaw *et al* 2019, Lorilla *et al* 2020).

Understanding how social, environmental, and climate variables influence recreational ecosystem services is essential for effectively predicting how they will be impacted by a changing climate. Our analysis showed that of the variables tested, trail density and the average maximum temperature were the two most influential factors for recreation (figure 3), clearly demonstrating the distinct link between social and biophysical systems. This is consistent with previous findings that show temperature is a major factor in outdoor physical activity (Tucker and Gilliland 2007, Obradovich and Fowler 2017). This also demonstrates the importance of recreational infrastructure (i.e. trails) that allow people to take advantage of recreational ecosystem services. Although socio-ecological factors are highly influential on the recreational behaviors of people, studies linking them have been surprisingly limited (de Freitas 2015, Chan and Wichman 2018). Many of these past studies have relied on only temperature and have focused on a small scale (e.g. within the boundaries of one park), a specific environment (e.g. Chaparral), or a specific recreational service (e.g. fishing) (Lane *et al* 2015, Tomczyk *et al* 2016, Pröbstl-Haider *et al* 2020, Jaung and Carrasco 2021). Also, within the ecosystem services community, studies have focused on the environmental aspect such as accessibility, forest intactness or naturalness, heritage or touristic

sites, and protected areas (Egoh *et al* 2012). This likely misrepresents the reality of the nonlinear socio-ecological relationships between recreation and climate (Dundas and von Haefen 2020) and highlights the need for larger scale studies beyond arbitrary political boundaries (e.g. a national park boundary) (Brice *et al* 2017). Further, this demonstrates the importance of taking into account coupled social and ecological systems. This is clearly illustrated in our partial dependence calculations showing how non-linear the relationships are between recreation and climate, environmental, and social factors (figure 3(b)). This demonstrates that a multitude of such factors need to be considered when trying to understand future impacts. Furthermore, the fluctuations shown in the box plots demonstrates how variable the influence of different factors can be on human behavior as different recreational services are provided across multiple diverse ecoregions and landscapes, highlighting the importance of regional studies that assess a multitude of plausibly influential factors.

Our model results generally show that current patterns of recreational ecosystem services are likely to be exacerbated by climate change, with greater exacerbation under higher warming scenarios (figure 5). For example, popular summer recreation regions identified with our Flickr data (i.e. Sierra Nevada, northern CA, and coastal regions) see spatial and magnitudinal increases in suitability. Unpopular summer regions (i.e. desert regions, southern CA, and the Central Valley) see decreases in suitability. This demonstrates that climate change impacts to recreational ecosystem services are not uniform across landscapes and cannot be extrapolated from small scale studies. This also demonstrates that similar regionally focused studies are necessary to properly inform of the spatial variations in impacts. Furthermore, we can see this heterogeneous impact within the various land cover classes across the region (figure 6). As the climate changes, more land area is unsuitable than suitable for summer recreation, but this effect is experienced differently between land cover classes. Forests see the greatest amount of recreation in the present and see the greatest positive influence on suitability in the future, whereas unsuitable summer conditions expand mostly into shrublands in the future. These results have major land management implications and give insight into which biomes' potential supply of recreational ecosystem services are more resilient and which are more vulnerable to future climate change within the region. There are already issues with overcrowding of popular recreational regions (mainly forested), like Yosemite and Muir Woods, in the summer. These already popular summer recreation sites are set to experience general increases in visitation due to population growth, but as our results show, should also expect increased suitability and subsequently even greater visitation in the summer as the climate changes. Beyond impacting the recreational ecosystem service itself, such heavy visitation can have impacts to the surrounding ecosystems, ecosystem services, and infrastructure. For example, high visitation rates to natural areas can impact soil loss (Barros *et al* 2013), habitat and species (Micheli *et al* 2016), vegetation composition and cover (Pickering and Hill 2007), and much more. Furthermore, high visitation rates can detrimentally impact the already under-maintained recreation infrastructure and overcrowd small towns in close proximity to popular areas. Our results illustrate that the already popular recreation regions that are most commonly within California forests are at risk of greater impacts from visitation due to the positive effect of climate change on recreation. This information can be used to include more comprehensive climate change impacts into land management plans and inform investments in areas that are predicted to see increased recreation. Furthermore, decision-makers can use these insights along with others from similar climate change impact studies to assess trade-offs and identify where land management might be most effective and efficient. For example, Coffield *et al* (2021) find that management efforts in California should be focused on stabilizing existing forest carbon stocks, rather than emphasizing just carbon gain, by reducing fire risks, thinning, and restoration. Together, such insights on ecosystem service vulnerability to climate change can help managers identify at-risk regions that provide significant and socially important ecosystem service bundles, like California forests, and most effectively and efficiently mitigate impacts to services and subsequently human well-being.

Uncertainties exist when assessing future climate change impacts to socio-ecological systems like recreational ecosystem services. The social media data inherently has uncertainties as it is just a proxy for visitation. Furthermore, social media data popularity can change from year to year (figure S3) and the demographics of users can be biased towards certain age groups, genders, and locations (Tenkanen *et al* 2017, Wood *et al* 2020). Yet, Flickr data has previously been validated as an effective proxy (Wood *et al* 2013, 2020) and we also conducted a simple validation on our data that demonstrated the efficacy of this proxy within the study region (figure S4). Our analysis did not assess all influential factors like air quality and wildfire. Both of these factors are bound to play a role in future recreation. For example, it is known that wildfire impacts recreation through closures and post-fire impacts (e.g. lower aesthetic appreciation) (Starbuck *et al* 2006), while higher air pollution levels have been shown to decrease recreation (Zajchowski *et al* 2021). This analysis focused specifically on direct climate changes (i.e. temperature, precipitation, and wind), but future analyses should also consider these compounding effects of a changing climate.

Furthermore, as with any model there are uncertainties within our model. To better quantify these uncertainties, we calculate the 90% confidence interval to show how much uncertainty surrounds our predictions (figure S6). We also calculate and map the model residuals, showing that the model tends to underpredict in a few areas within the mountain regions of CA (figure S7). Although our validation tests resulted in a relatively low R^2, machine learning is not necessarily designed to test significance, but is more so for isolating patterns within data and subsequently developing an algorithm that best describes and predicts those patterns (Huettmann *et al* 2018). Ultimately, our goal was to understand general recreational patterns and their relationship with social and biophysical factors, which our model was effective in doing (figure 4). Future ecosystem service research should take advantage of these big data and machine learning resources to reduce data limitations and better integrate ecosystem service beneficiaries into assessments. Our study demonstrates that the concurrent use of crowdsourced data and machine learning has great potential to close socio-ecological gaps in ecosystem service research and brings the possibility to more holistically assess climate change impacts to human well-being.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

ORCID iDs

Kyle Manley ⬤ https://orcid.org/0000-0002-6992-168X
Benis N Egoh ⬤ https://orcid.org/0000-0002-9126-7784

References

Abatzoglou J T, Dobrowski S Z, Parks S A and Hegewisch K C 2018 TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015 *Sci. Data* **5** 170191
Askew A and Bowker J M 2018 Impacts of climate change on outdoor recreation participation: outlook to 2060 *J. Park Recreat. Adm.* **36** 97–120
Barros A, Gonnet J and Pickering C 2013 Impacts of informal trails on vegetation and soils in the highest protected area in the Southern Hemisphere *J. Environ. Manage.* **127** 50–60
Beck H E, Zimmermann N E, McVicar T R, Vergopolan N, Berg A and Wood E F 2018 Present and future Köppen-Geiger climate classification maps at 1-km resolution *Sci. Data* **5** 180214
Bedsworth L, Cayan D, Franco G, Fisher L and Ziaja S 2018 California's fourth climate change assessment: statewide summary report (Sacramento, CA) (https://doi.org/10.1080/10962247.2018.1459325)
Berkes F and Folke C 1998 *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience* (Cambridge: Cambridge University Press)
Brauman K *et al* 2019 Status and trends – nature's contributions to people (NCP) *Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* ed E S Brondízio, J Settele, S Díaz, and H T Ngo (Bonn: IPBES Secretariat) ch 2.3
Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
Brice B, Fullerton C, Hawkes K, Mills-Novoa M, O'Neill B and Pawlowski W 2017 The impacts of climate change on natural areas recreation: a multi-region snapshot and agency comparison *Nat. Areas J.* **37** 86–97
Burge D O *et al* 2016 Plant diversity and endemism in the California Floristic Province *Madroño* **63** 3–206
Burkhard B and Maes J 2017 Mapping ecosystem services *Adv. Books* **1** e12837
Chan N W and Wichman C J 2018 *The Effects of Climate on Leisure Demand: Evidence from North America* (Resources for the Future)
Cheng X, van Damme S, Li L and Uyttenhove P 2019 Evaluation of cultural ecosystem services: a review of methods *Ecosyst. Serv.* **37** 100925
Coffield S R, Hemes K S, Koven C D, Goulden M L and Randerson J T 2021 Climate-driven limits to future carbon storage in California's wildland ecosystems *AGU Adv.* **2** e2021AV000384
Convention on Biological Diversity (CBD) 2021 First draft of the post-2020 global biodiversity framework
CPAD 2021 California protected areas database 2021b release GreenInfo Network
de Freitas C R 2015 Weather and place-based human behavior: recreational preferences and sensitivity *Int. J. Biometeorol.* **59** 55–63
Dundas S and von Haefen R 2020 The effects of weather on recreational fishing demand and adaptation: implications for a changing climate *J. Assoc. Environ. Resour. Econ.* **7** 209–42
Egoh B, Drakou E G, Dunbar M B, Maes J and Willemen L 2012 Indicators for mapping ecosystem services: a review (European Commission, Joint Research Centre (JRC))
European Commission *et al* 2020 Mapping and assessment of ecosystems and their services : an EU wide ecosystem assessment in support of the EU biodiversity strategy (Publications Office) (https://doi.org/10.2760/757183)
Fox E W, Ver Hoef J M and Olsen A R 2020 Comparing spatial regression to random forests for large environmental data sets *PLoS One* **15** e0229509
Haines-Young R and Potschin M 2012 Common International Classification of Ecosystem Services (CICES): Consultation on Version 4 (Accessed August–December 2012)
Halofsky J E, Peterson D L, Buluç L and Ko J 2021 Climate change vulnerability and adaptation for infrastructure and recreation in the Sierra Nevada *General Technical Report PSW-GTR-272* (Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station) p 275
Havinga I, Bogaart P W, Hein L and Tuia D 2020 Defining and spatially modelling cultural ecosystem services using crowdsourced data *Ecosyst. Serv.* **43** 101091
Hermes J, van Berkel D, Burkhard B, Plieninger T, Fagerholm N, von Haaren C and Albert C 2018 Assessment and valuation of recreational ecosystem services of landscapes *Ecosyst. Serv.* **31** 289–95

Huettmann F *et al* 2018 Use of machine learning (ML) for predicting and analyzing ecological and 'presence only' data: an overview of applications and a good outlook BT *Machine Learning for Ecology and Sustainable Natural Resource Management* ed G Humphries, D R Magness and F Huettmann (Cham: Springer International Publishing) pp 27–61

IPBES 2019 Summary for policymakers of the global assessment report on biodiversity and ecosystem services (https://doi.org/10.5281/ZENODO.3553579)

Jaung W and Carrasco L R 2021 Using mobile phone data to examine weather impacts on recreational ecosystem services in an urban protected area *Sci. Rep.* **11** 5544

Kim Y, Kim C-K, Lee D K, Lee H-W and Andrada R I T 2019 Quantifying nature-based tourism in protected areas in developing countries by using social big data *Tourism Manage.* **72** 249–56

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

Lane D, Jones R, Mills D, Wobus C, Ready R C, Buddemeier R W, English E, Martinich J, Shouse K and Hosterman H 2015 Climate change impacts on freshwater fish, coral reefs, and related ecosystem services in the United States *Clim. Change* **131** 143–57

Lautenbach S, Mupepele A-C, Dormann C F, Lee H, Schmidt S, Scholte S S K, Seppelt R, van Teeffelen A J A, Verhagen W and Volk M 2019 Blind spots in ecosystem services research and challenges for implementation *Reg. Environ. Change* **19** 2151–72

Lavorel S *et al* 2017 Pathways to bridge the biophysical realism gap in ecosystem services mapping approaches *Ecol. Indic.* **74** 241–60

Lorilla R S, Poirazidis K, Detsis V, Kalogirou S and Chalkias C 2020 Socio-ecological determinants of multiple ecosystem services on the Mediterranean landscapes of the Ionian Islands (Greece) *Ecol. Modell.* **422** 108994

Mandle L *et al* 2021 Increasing decision relevance of ecosystem service science *Nat. Sustain.* **4** 161–9

McDonough K, Hutchinson S, Moore T and Hutchinson J M S 2017 Analysis of publication trends in ecosystem services research *Ecosyst. Serv.* **25** 82–88

Micheli F, Heiman K W, Kappel C V, Martone R G, Sethi S A, Osio G C, Fraschetti S, Shelton A O and Tanner J M 2016 Combined impacts of natural and human disturbances on rocky shore communities *Ocean Coast. Manage.* **126** 42–50

National Academy of Sciences 1992 *Policy Implications of Greenhouse Warming: Mitigation, Adaptation, and the Science Base* (The National Academies Press) (https://doi.org/10.17226/1605)

Obradovich N and Fowler J H 2017 Climate change may alter human physical activity patterns *Nat. Hum. Behav.* **1** 97

Ochoa V and Urbina-Cardona N 2017 Tools for spatially modeling ecosystem services: publication trends, conceptual reflections and future challenges *Ecosyst. Serv.* **26** 155–69

Pachauri R K, Allen M R, Barros V R, Broome J, Cramer W, Christ R and Dubash N K 2014 *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change IPCC* (Geneva: IPCC)

Palomo I 2017 Climate change impacts on ecosystem services in high mountain areas: a literature review *Mt. Res. Dev.* **37** 179–87

Pickering C M and Hill W 2007 Impacts of recreation and tourism on plant biodiversity and vegetation in protected areas in Australia *J. Environ. Manage.* **85** 791–800

Pörtner H O *et al* 2021 IPBES-IPCC co-sponsored workshop report on biodiversity and climate change

Pröbstl-Haider U, Hödl C, Ginner K and Borgwardt F 2020 Climate change: impacts on outdoor activities in the summer and shoulder seasons *J. Outdoor Recreat. Tourism* **34** 100344

Rammer W and Seidl R 2019 Harnessing deep learning in ecology: an example predicting bark beetle outbreaks *Front. Plant Sci.* **10**

Reclamation 2013 *Release of Downscaled CMIP5 Climate Projections, Comparison with Preceding Information, and Summary of User Needs* (Denver, CO: U.S. Department of the Interior, Bureau of Reclamation, Technical Service Center)

Richards D R and Tunçer B 2018 Using image recognition to automate assessment of cultural ecosystem services from social media photographs *Ecosyst. Serv.* **31** 318–25

Runting R K, Bryan B A, Dee L E, Maseyk F J F, Mandle L, Hamel P, Wilson K A, Yetka K, Possingham H P and Rhodes J R 2017 Incorporating climate change into ecosystem service assessments and decisions: a review *Glob. Change Biol.* **23** 28–41

Scowen M, Athanasiadis I N, Bullock J M, Eigenbrod F and Willcock S 2021 The current and future uses of machine learning in ecosystem service research *Sci. Total Environ.* **799** 149263

Shiferaw H, Bewket W, Alamirew T, Zeleke G, Teketay D, Bekele K, Schaffner U and Eckert S 2019 Implications of land use/land cover dynamics and Prosopis invasion on ecosystem service values in Afar Region, Ethiopia *Sci. Total Environ.* **675** 354–66

Shin Y-J *et al* 2019 IPBES global assessment on biodiversity and ecosystem services Chapter 4. Plausible futures of nature, its contributions to people and their good quality of life

Sleeter B M, Wilson T S and Sherba J T 2017 Land use and land cover projections for California's 4th climate assessment *US Geological Survey Data Release*

Starbuck C M, Berrens R P and McKee M 2006 Simulating changes in forest recreation demand and associated economic impacts due to fire and fuels management activities *For. Policy Econ.* **8** 52–66

Tenkanen H, Di Minin E, Heikinheimo V, Hausmann A, Herbst M, Kajala L and Toivonen T 2017 Instagram, Flickr, or Twitter: assessing the usability of social media data for visitor monitoring in protected areas *Sci. Rep.* **7** 17615

The White House 2021 Executive order on protecting public health and the environment and restoring science to tackle the climate crisis

Thornton M M, Shrestha R, Wei Y, Thornton P E, Kao S and Wilson B E 2020 Daymet: monthly climate summaries on a 1-km grid for North America, version 4 (Oak Ridge, TN: ORNL DAAC) (https://doi.org/10.3334/ORNLDAAC/1855)

Tomczyk A M, White P C L and Ewertowski M W 2016 Effects of extreme natural events on the provision of ecosystem services in a mountain environment: the importance of trail design in delivering system resilience and ecosystem service co-benefits *J. Environ. Manage.* **166** 156–67

Tucker P and Gilliland J 2007 The effect of season and weather on physical activity: a systematic review *Public Health* **121** 909–22

U.S. Bureau of Economic Analysis 2019 2019 California Outdoor Recreation Satellite Account

UNEP 2016 *Loss and Damage: The Role of Ecosystem Services* (Nairobi: United Nations Environment Programme)

van der Geest K, de Sherbinin A, Kienberger S, Zommers Z, Sitati A, Roberts E and James R 2019 The impacts of climate change on ecosystem services and resulting losses and damages to people and society *Loss and Damage from Climate Change* (Cham: Springer) pp 221–36

Walls M, Darley S and Siikamäki J 2009 *The State of the Great Outdoors: America's Parks, Public Lands, and Recreational Resources* (Washington, DC: Resourc. Future)

Willcock S *et al* 2018 Machine learning for ecosystem services *Ecosyst. Serv.* **33** 165–74

Wilson T S, Sleeter B M and Davis A W 2015 Potential future land use threats to California's protected areas *Reg. Environ. Change* **15** 1051–64

Wood S A, Guerry A D, Silver J M and Lacayo M 2013 Using social media to quantify nature-based tourism and recreation *Sci. Rep.* **3** 2976

Wood S A, Winder S G, Lia E H, White E M, Crowley C S L and Milnor A A 2020 Next-generation visitation models using social media to estimate recreation on public lands *Sci. Rep.* **10** 15419

Xia J, Wang J and Niu S 2020 Research challenges and opportunities for using big data in global change biology *Glob. Change Biol.* **26** 6040–61

Zajchowski C A B, South F, Rose J and Crofford E 2021 The role of temperature and air quality in outdoor recreation behavior: a social-ecological systems approach *Geogr. Rev.* **1** 1–20

Zommers Z, van der Geest K, de Sherbinin A, Kienberger S, Roberts E, Harootunian G, Sitati A and James R 2014 Loss and damage: the role of ecosystem services (https://doi.org/10.1371/journal.pntd.0002780)