Ecosystem Services 57 (2022) 101478

Contents lists available at ScienceDirect

Ecosystem Services

journal homepage: www.elsevier.com/locate/ecoser

Review Paper

A review of machine learning and big data applications in addressing ecosystem service research gaps

Kyle Manley *, Charity Nyelele, Benis N. Egoh

Department of Earth System Science, University of California Irvine, Irvine, CA 92697, United States

ARTICLE INFO

Keywords:
Ecosystem services
Machine learning
Big data
Nature's contribution to people
Modeling
Socio-ecological systems
Uncertainty

ABSTRACT

Ecosystem services are essential for human well-being, but are currently facing many natural and anthropogenic threats. Modeling and mapping ecosystem services helps us mitigate, adapt to, and manage these pressures, but overall the field faces multiple major limitations. These include: 1) data availability, 2) understanding, estimation, and reporting of uncertainties, and 3) connecting socio-ecological aspects of ecosystem services. Recent technological advancements in machine learning coupled with rising availability of big data, offer an opportunity to overcome these challenges. We review studies utilizing machine learning and/or big data to overcome these limitations. We collect 56 papers that exemplify the current use of machine learning and big data to address the three identified gaps in the ecosystem service field. We find that although the use of these tools in ecosystem service research is relatively new, it is growing quickly. Big data can directly address data gaps, especially as new big data resources relevant to ecosystem service mapping become available (ex. social media data). Some properties of machine learning can also contribute to addressing data gaps in data sparse environments. Also, many machine learning algorithms can estimate and consider uncertainty, whereas big data can significantly increase sample size, reducing uncertainties in some situations. Some big data sources, like crowdsourced data, provide direct sources of social behaviors and preferences that relate to ecosystem service demand, thus allowing researchers to connect social and biophysical aspects of ecosystem services. Machine learning algorithms provide an effective and efficient tool for handling these large nonlinear socio-ecological datasets in tandem, giving researchers the ability to more realistically model and map ecosystem services without relying on oversimplified proxies or linear algorithms. Despite these opportunities, implementation is still lacking and limitations still hinder use.

1. Introduction

Ecosystems provide critically important ecosystem services, i.e., the benefits humans derive from ecosystems, such as raw materials, water purification, and recreation, which ultimately contribute to overall human well-being (Millennium Ecosystem Assessment, 2002; Díaz et al., 2018). To convey the importance of functioning ecosystems and justify the protection of forests, wetlands, and other ecosystems that provide these services, the concept of ecosystem services is increasingly being incorporated into global policies and assessments (European Commission, 2020; Convention on Biological Diversity, 2020; Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), 2018, Díaz et al, 2018). Additionally, the literature on ecosystem services is maturing as evidenced by the increasing number of studies across the world devoted to modeling and mapping ecosystem services from local to global scales (McDonough et al., 2017; Ochoa and Urbina-Cardona, 2017; Kosanic and Petzold, 2020; Xu et al., 2020). Despite this increased attention and recognition of the importance of ecosystem services, particularly after the Millennium Ecosystem Assessment (MA, 2005), several assessments are still reporting the continuous decline in ecosystem service provision. The IPBES global assessment of 2019 (IPBES, 2019), for example, showed that 14 of the 18 channels through which nature provides benefits to humans are being affected negatively. Considering this, the need for tools to model, map, and monitor ecosystem services in response to calls for the management, protection, and restoration of areas with high ecosystem service production could not be more important.

Following the demand to model and map ecosystem services, several models and tools have been developed and used around the world. The Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) suite

* Corresponding author at: 3200 Croul Hall St, Irvine, CA 92697, USA.
E-mail addresses: cnyelele@uci.edu (K. Manley), begoh@uci.edu (B.N. Egoh).

https://doi.org/10.1016/j.ecoser.2022.101478
Received 7 March 2022; Received in revised form 2 September 2022; Accepted 6 September 2022

of models (Tallis and Polasky, 2009) integrate biological, physical, chemical, and economic properties of landscapes to assess how changes in ecosystems impact a plethora of ecosystem services. The Soil and Water Assessment Tool (SWAT) is a conceptual and continuous time model developed to assess watershed resources (Gassman et al., 2007). These are two examples of some of the widely used tools for modeling and mapping ecosystem services (Burkhard and Maes, 2017; Ochoa and Urbina-Cardona, 2017; Kim et al., 2019; Chun et al., 2020). Other ecosystem service assessment models and tools that have seen increasing usage include ARtificial Intelligence for Ecosystem Services (ARIES), Social Values for Ecosystem Services (SolVES), and i-Tree (Hirabayashi et al., 2011; Villa et al., 2014; Sherrouse and Semmens, 2015). Despite the advances in developing ecosystem service modeling and mapping approaches, some of these resources are limited by data availability, particularly data that can properly account for the multidimensional traits of ecosystems and model the socio-ecological system in order to identify and quantify ecosystem service flows. This has prompted scientists to rely upon expert knowledge or use simple proxies, such as land use and land cover, to generate information that feeds into these models and tools in an attempt to overcome data availability barriers. However, this leads to greater uncertainty due to oversimplifications and generalizations of the nonlinear dynamics of ecosystem services (Spake et al., 2017; Lautenbach et al., 2019; Willcock et al., 2021). Furthermore, there remains a lack of information on the quality of model output and validation methods (Grêt-Regamey et al., 2017; Ochoa and Urbina-Cardona, 2017). As such, there is a need to reduce or better illustrate uncertainties, close data availability gaps, and improve on key indicators necessary for mapping both the supply and demand of ecosystem services at various scales and in different management contexts (Lavorel et al., 2017; Willcock et al., 2018; Mandle et al., 2021).

New tools such as machine learning (ML) algorithms and big data, which are increasingly being used in the ecological field, lend themselves to be used for modeling and mapping ecosystem services (Farley et al., 2018; Willcock et al., 2018; Willcock et al., 2021; Scowen et al., 2021). Although big data is relatively new within the ecological science community and has yet to be fully taken advantage of (Scowen et al., 2021), it is quickly gaining popularity as a way of filling in key gaps that exist in modeling and mapping ecosystem services. Multiple definitions of big data exist in the literature (Sagiroglu and Sinanc, 2013; Vitolo et al., 2015; Sun and Scanlon, 2019). Here, we adopt a commonly used definition by the National Institute of Standards and Technology (NIST, 2015) and refer to big data as datasets characterised by high volume, variety, velocity, and variability (4 V) which can not be handled efficiently by traditional data architectures. Volume refers to the size, variety describes the heterogeneity of the format and source, velocity refers to the rate of flow of the data, and variability refers to the changes in a dataset (i.e. flow rate, format/structure, and volume) (NIST, 2015). Currently there exists a diverse range of ecologically relevant datasets that would qualify as big data under this definition. Key examples include datasets in repositories such as AmeriFlux (https://ameriflux.lbl.gov), Nutrient Network (NutNet) (https://nutnet.org/home) and Neotoma (https://www.neotomadb.org) as well as datasets provided by ecological monitoring networks including the National Ecological Observatory Network (NEON), Long Term Ecological Research (LTER), and Critical Zone Observatory (CZO). Additionally, big data is collected across various social media sites, applications, Global Positioning System integrated devices, citizen science and other crowdsourced projects. These big data sources collect high volumes of data products (terabytes to exabytes every year), provide a wealth of information on a wide variety of environmental factors ranging from genomic to global scale, can be collected at high rates, and ultimately can give useful insights into ecosystems and the services they provide (Farley et al., 2018; Havinga et al., 2020a; Xia et al., 2020). Although big data has started to become relatively popular in ecology, its uptake for use of ecosystem service modeling and mapping has been slower and there remains a need to transition and apply this resource into the ecosystem service field (Havinga et al., 2020a; Scowen et al., 2021).

ML has emerged as a valuable tool for processing and analyzing big data as well as an effective and efficient way to address the key methodological concerns and challenges encountered in modeling and mapping ecosystem services. ML is a "subset of artificial intelligence, which builds a mathematical model based on sample data (not necessarily always big data), known as "training data", to make predictions or decisions without being explicitly programmed to perform the task" (Zhang et al., 2020). ML algorithms can process the huge collections of data that would otherwise be difficult or sometimes implausible to analyze using traditional techniques. Additionally, through validation of training data with a predefined set of testing data, ML can also automatically provide estimates of uncertainty enabling users to assign their own acceptable thresholds of uncertainty and ultimately increasing the decision relevance of such analyses (Grêt-Regamey et al., 2017; Hamel and Bryant, 2017). This is a powerful component considering that currently there remains a lack of information on the quality of model output and validation methods (Grêt-Regamey et al., 2017; Ochoa and Urbina-Cardona, 2017). These attributes and the data driven characteristics of ML can enhance current models in the modeling of nonlinear and highly dimensional data without relying on any assumptions about the data prior to analysis. This is particularly important in identifying indicators for mapping ecosystem services, given that many different indicators are being used for the same service (Egoh et al., 2012). ML also allows us to infer when data is missing or limited in availability, and oftentimes to more accurately represent reality, improving our overall understanding of the complex interactions and dynamics within ecosystems and the services they provide across various scales (Rammer and Seidl, 2019; Frey, 2020). ML can be divided into two sub-categories: supervised learning algorithms that require and make use of an input training dataset to learn a function that can most effectively approximate the relationship between the inputs and outputs from a dataset; and unsupervised learning algorithms that learn without predefined data by grouping similar attributes and identifying trends, patterns, and/or relationships within the data to infer natural structure within a dataset (McCue, 2015; Zhang, 2020).

Despite the availability of big data and ML algorithms, the ecosystem services community has not taken full advantage of these opportunities to model and map the supply and demand of ecosystem services. This is demonstrated by Scowen et al. (2021) who describe how ML usage within ecosystem service research has predominantly been for descriptive or predictive tasks. They highlight how ML can be used in ecosystem service assessments and identify how to further the repeatability of methodologies that utilize ML. Although Scowen et al. (2021) conduct a somewhat similar review, ours centres on how ML and big data are both being used to address the major gaps in the ecosystem service field. As big data becomes more readily available and temporally viable, and ML algorithms become increasingly routine and implementable for providing further insight into complex socio-ecological interactions and the provision of ecosystem services (Thessen, 2016; Huettmann et al., 2018; Rammer and Seidl, 2019; Willcock et al., 2018). As such, integrating big data and ML algorithms into ecosystem service research will help develop the field into a more robust, interdisciplinary, and extensive research field. Considering this, there is a need to take stock of how big data and ML have been used in the modelling and mapping of ecosystem services, particularly how they have addressed gaps within the field, and to explore what potential they hold in developing more robust approaches to respond to the ever-increasing policy and science demand for ecosystem service research. The aim of this review is to demonstrate the utility of ML and big data to ecosystem service research in addressing three key gaps that exist, namely: 1) data availability, 2) understanding and estimating uncertainty, and 3) modelling the social and biophysical (socio-ecological) aspects of ecosystems and the services they provide. While most ecosystems services research has focused on mapping the biophysical aspects and using ecological indicators such as land cover, few studies incorporate social data that captures the demand of these services (Reyers et al., 2013; Mandle et al., 2021), a major gap that can be filled by these new techniques.

2. Methods

In this study we used "Scopus" (https://www.scopus.com/), to compile literature published on the use of ML and big data in ecosystem service research. Scopus was chosen because of the vast, comprehensive, and multidisciplinary features of the abstract database that fit well for the purposes of this review. We conducted the search in July 2021 using search terms in the title, keywords, or abstract that included: "ecosystem service", OR "environmental service" OR "nature's contribution to people" AND "machine learning", OR "big data". We classified the paper as using big data if the authors specifically referred to their data as "big data". Although this method only captures studies that classify their data specifically as big data and does not comprehensively identify all studies that actually use big data, the method has the potential to capture the general patterns of big data use in ecosystem service research. To further confirm that this method captures the overall pattern of big data use, we conduct a Pearson's correlation analysis between big data use and ML use to assess whether big data usage is similar to ML usage over time. Furthermore, we assess the data used for all papers that label their data as "big data" to confirm whether it would qualify as big data under the previously stated definition used in this review. Our search terms resulted in a total of 256 papers. We then used the same search terms in Web of Science (https://www.webofscience.com/), a website with multiple databases that provide comprehensive citation data for many different academic disciplines, to create a more robust search. After removing duplicates, the Web of Science search resulted in 19 additional papers, bringing our total to 275 papers. We then read the titles and abstracts of all the 275 papers and removed papers that were conceptual, theoretical or reviews (95 in total) resulting in 180 potential papers to review. We conducted a second more in-depth screening by reading all 180 papers, resulting in 56 papers that were deemed appropriate for the objectives of the review, since the other 124 papers only mentioned ecosystem services and did not directly study them. Fig. 1 illustrates the steps undertaken to identify relevant papers for the review. In order to properly collect, organize, and analyze the data on the selected papers, we first had to standardize a few categories. For the scale of study, four categories were used: local (an area within a country), national (across an entire country), regional (within multiple countries), and global. For the ecosystem service classes, five categories were adopted: general (for unspecified ecosystem services), cultural, regulating, provisioning, and supporting following the Millennium Ecosystem Assessment classification (Millennium Ecosystem Assessment, 2002).

During analysis of the 56 papers included in the review, some basic initial data was gathered, including the year of publication, the region or country studied, the country of the first author's institute (to understand where expertise is), the publishing journal, as well as the scale of the study. The main analysis included collecting data on what ML algorithm was used, what big data was used, and which of the three ecosystem service gaps identified for this review were being filled using ML, big data or a combination of both. To identify whether and in what way the studies were addressing the three gaps, we read each paper and subjectively identified the knowledge gaps being filled (oftentimes, but not always, the paper would specifically refer to the gap being filled). This analysis allowed for an assessment of how ML and big data are currently being used to fill key knowledge and methodological gaps. In addition, it helps in the identification of emerging patterns in the use of these tools to fill the gaps and opportunities to further address these major gaps in

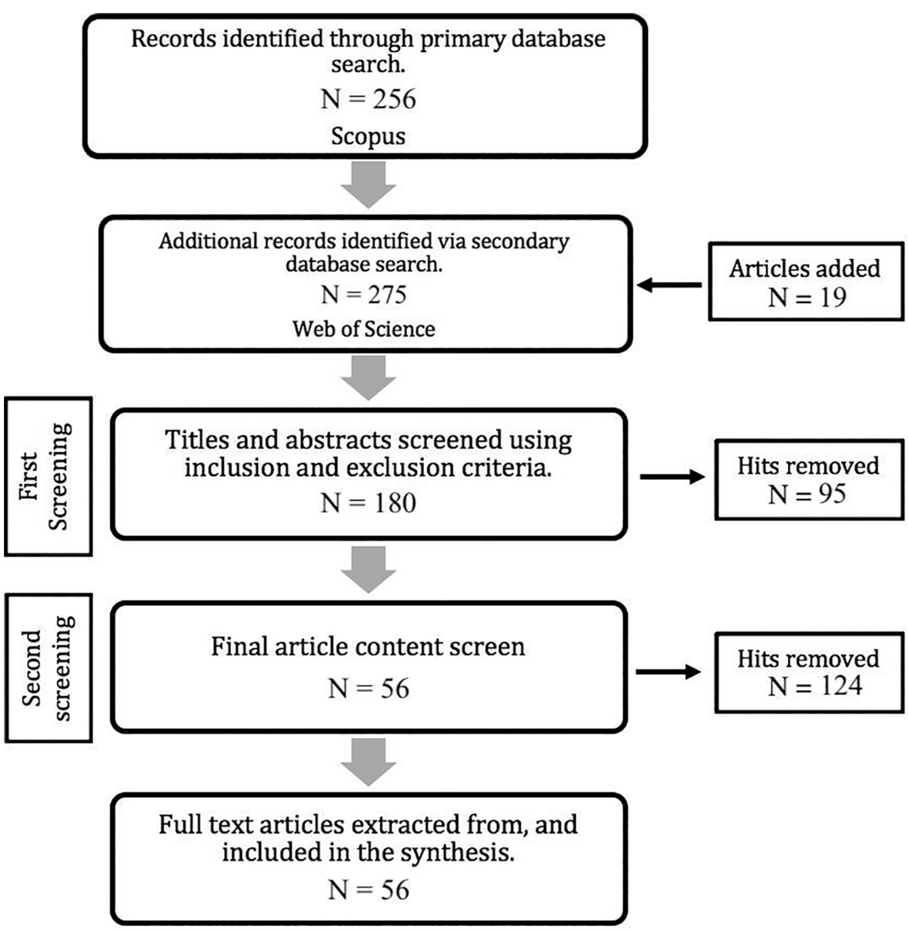

Fig. 1. Flow chart of peer-review procedure.

the future. We also collected supplementary information on the program or software used to implement ML or assess big data, what indicators were used for each service, the class of each indicator (biophysical, social, or economic), and whether the study was looking into ecosystem service supply, demand, or both.

We classified studies as filling the data availability gap if there is limited data in the study area or for the service being assessed that would cause an assessment to be difficult or impossible without the use of ML and/or big data, or if the authors explicitly stated they were filling the data availability gap. Additionally, we classified studies that filled the uncertainty gap if the authors had the aim to reduce uncertainty and stated this in the paper, if the authors tested multiple indicators or sets of covariates to identify the most suitable proxy for modeling and mapping, and/or if uncertainty would be increased or unknown without the use of ML/big data. Studies were classified as filling the third gap related to modeling the socio-ecological system if they considered the socio-ecological aspects together by using multiple indicators or covariates together. For example, the use of social media data to map recreation where both human behaviour (i.e. visitation) is combined with aspects of the ecosystem in which recreation occurs (i.e. environmental covariates), instead of just a reliance on a single biophysical proxy to linearly estimate human behavior.

3. Results

Our results show that the use of ML and big data in ecosystem service research started in 2012 and is increasing quickly. As illustrated in Fig. 2, results show that prior to 2017 only a few studies used these tools, followed by a steady increase peaking in 2019. The use of ML alone is seen more often (48 %) than the use of big data alone (32 %). Cultural services and regulating services saw a large increase in attention after 2018. Publications included in this study showed similar patterns in the use of big data and ML over time (Pearson correlation of 0.92). Analysis of the data used when authors claimed to be using big data showed that all such studies used data that would qualify as big data based on the definition adopted for this review. Additionally, most studies were focused on and had first authors from Asia (~46 % & ~45 % respectively) and Europe (~32 % & ~43 % respectively). For both the location of the first author and the location of the study, most countries with publications range from around 1 – 6 studies, of which studies in China clearly stand out from the rest of the group. China is the most studied country with 25 % of all studies carried out there and 29 % of all first authors being affiliated with institutions located in China. There is a clear lack of studies utilizing these tools within many regions of the world including the Middle East, Africa, Oceania, as well as South and Central America. Furthermore, when assessing the study scale, the majority of studies (~77 %) were conducted at a local scale. Studies focusing on the national scale were the second most popular (~12 %), while few studies were carried out at the global scale (~4%).

We find that a plethora of ML algorithms are used throughout the studies (Table 1). The algorithms used can be categorized into supervised and unsupervised algorithms, where the supervised algorithms are

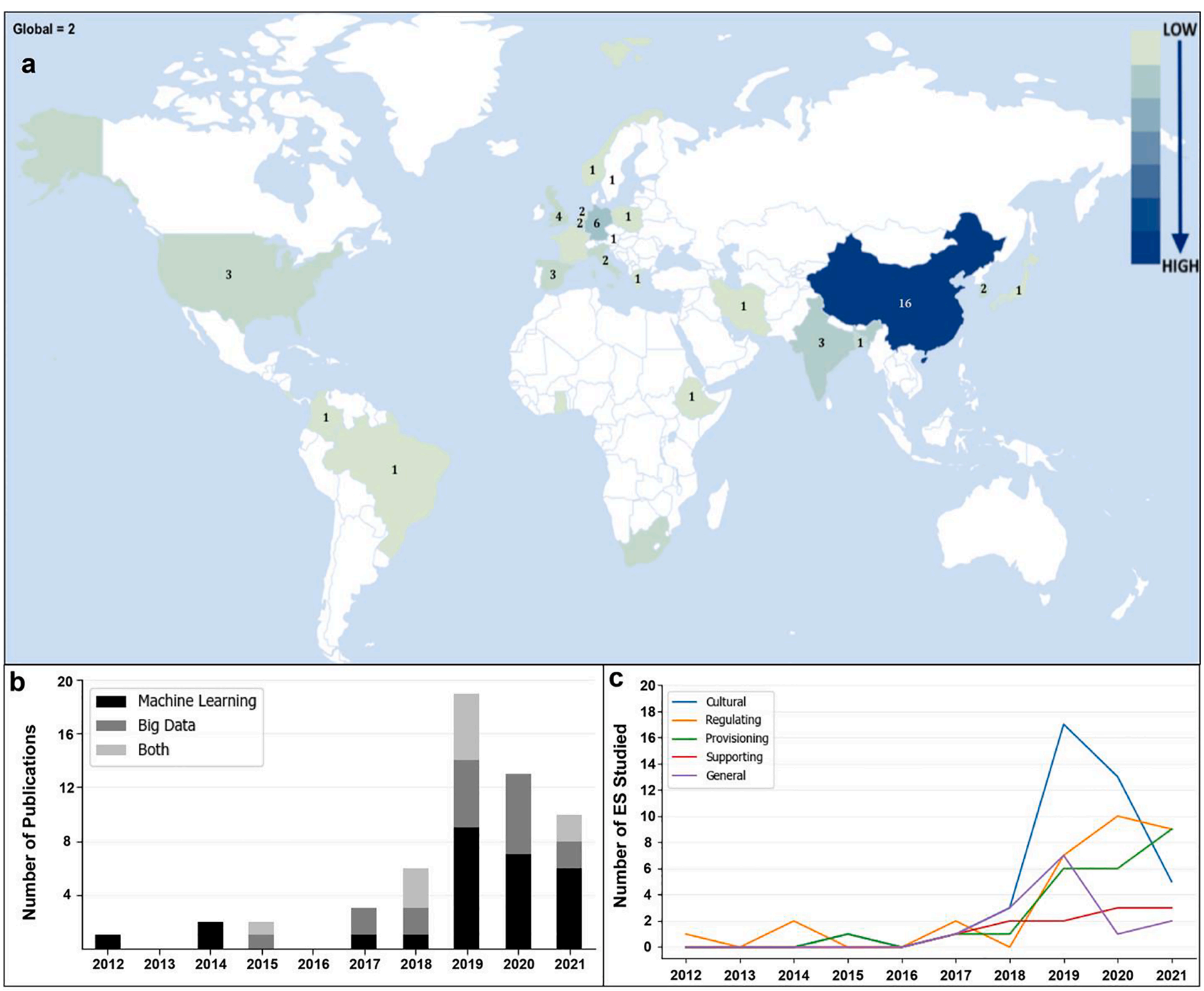

Fig. 2. Properties of the studies assessed in the review: a) map of the countries of study (shading) and countries of first author (numbering); b) distribution of studies over time, distinguished between studies using ML, big data, and both; and c) plot of the different ES classes assessed over time.

Table 1
Overview of the machine learning algorithms found within reviewed studies, what tasks they were used for, and references to specific example case studies.

Category	Subcategory	ML	Example Articles
Supervised			
	Classification	Convolutional Neural Net.	Callau et al. 2019
			Lee et al. 2019
		Decision Tree	Sannigrahi et al. 2019
		Random Forest	Tsai et al. 2019
			Strohbach and Haase 2012
		Rotation Forest	Shuangoa et al. 2021
		Support Vector Machine	Rathmann et al. 2020
			Sannigrahi et al. 2019
	Regression	Artificial Neural Net.	Morshed et al., 2022
			Debanshi and Pal 2020
		Bayesian	Balbi et al. 2019
			Ouyang et al. 2019
		Boosted Regression Tree	Ciesielski and Stereńczak, 2021
			Osborne and Alvares 2019
		CART	Rippy et al. 2021
			Escobedo et al. 2020
		Conditional Inference Tree	Liu et al. 2021
		Maxent	Bernetti et al. 2019
			He et al. 2019
		Random Forest	Havinga et al., 2020a; Havinga et al., 2020b
			Shiferaw et al. 2019
Unsupervised			
	Clustering	K-means	Kang et al. 2018
		Non-negative Matrix Fact.	Jaung 2021

used for classification and regression tasks and the unsupervised algorithms were used for clustering tasks. The use of supervised algorithms was much more common than the use of unsupervised algorithms. Table 1 provides a synopsis of the algorithms used and gives specific examples for readers to refer to. For more information on the studies assessed, what algorithms and software are used, as well as what big data sources are used we refer readers to Tables S2 and S4.

Our results indicate that ML and big data are used to overcome the three gaps examined in this review (i.e., data availability, model and output uncertainty as well as difficulties in mapping the socio-ecological system). Fig. 3 shows that as expected big data helped fill the data availability gap in the majority of the studies analysed (54 %). Within the reviewed studies, results also indicate that the high model and output uncertainty in most ecosystem service research was addressed mainly through the use of ML algorithms (58 %) such as random forest and Bayesian networks. The third gap related to difficulties of mapping the biophysical and social aspects of ecosystem services was addressed using both big data and ML (48 % and 40 % respectively). Big data sources were highly variable and ranged from social media data, to census data, to Google search engine data. The most commonly used big data came from social media sites (ex. Flickr and Weibo) with application data (ex. trail tracking application), mobile phone data, and citizen science being the second most commonly used big data sources. The majority of ML algorithms used were supervised (97 %) including random forest, support vector machines, and convolutional neural networks. The most commonly used ML algorithm was random forest (24 %) with Bayesian networks as the second most common (14 %). The most common software used to implement ML was R (36 %) most often using the random-forest package (Lorilla et al., 2020), while also using various other packages including "caret" (Degerickx et al., 2020), RoogleVision (Richards and Tuncer, 2018), and "clarifai" (Lee et al., 2019). It is also noteworthy that it was relatively common that the specific package used was not mentioned. Some other environments and programming languages used to carry out ML include WEKA (Debanshi and Pal, 2020), the Maxent software (Havinga et al., 2020b), Google Earth Engine (Escobedo et al., 2020), and Python (Landuyt et al., 2014).

There are also differences in the ecosystem service classes being studied and the gaps that are being filled within those studies as illustrated in Fig. 4. Papers assessing cultural ecosystem services were focused on filling the data availability gap often (~53 % of papers filling the data gap) and the socio-ecological gap (~61 %). Papers assessing multiple ecosystem service classes were commonly addressing the uncertainty gap (~32 %). From the studies analysed, supporting services were the least assessed class (~4% of studies). Within the studies on cultural ecosystem services, 38 % addressed the gap of connecting the biophysical and social aspects of ecosystems and modeling the ecosystem services realised from those systems, while 36 % of the studies sought to address the data availability gap. 44 % of the studies that assessed multiple ecosystem service classes addressed gaps

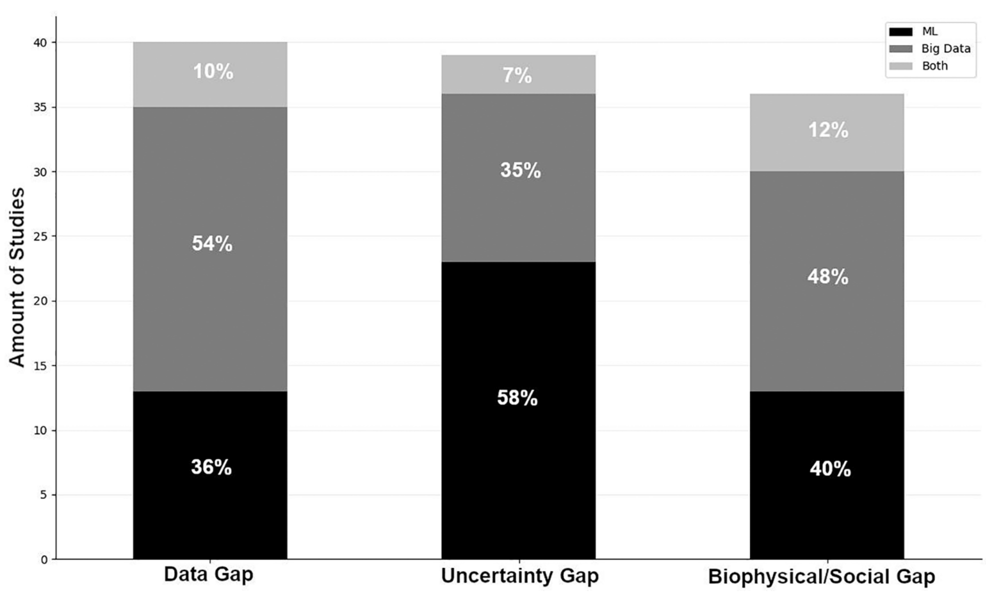

Fig. 3. Number of studies that use ML (black), big data (dark grey) and a combination of both (light grey) to address the three gaps identified in ecosystem service research.

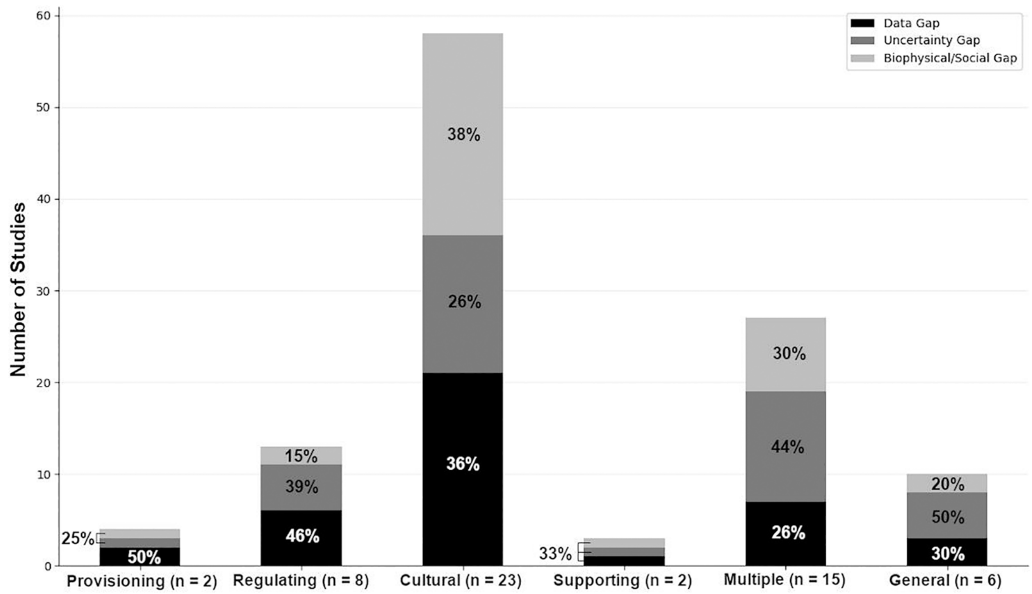

Fig. 4. Proportion of ecosystem service classes assessed and the specific gaps addressed in each of those classes.

associated with model and output uncertainty. Patterns within provisioning, regulating, and supporting services were difficult to assess due to the low number of studies analyzing just these service classes.

Studies that did not look at a specific ecosystem service were lumped into the general category (unspecified). We identified 16 different ecosystem services studied within the literature: recreation, aesthetic appreciation, tourism, sense of place, carbon sequestration and storage, local climate regulation, air quality regulation, erosion prevention, waste treatment, biological control, fresh water supply, raw materials, food provision, maintenance of genetic diversity, nutrient cycling, and habitat for species. Most studies assessed cultural ecosystem services (32 %), with recreation being the most studied cultural service. Regulating services were the second most studied service using big data and/or ML (26 %), with carbon sequestration and storage as the most studied regulating service. With the exception of sense of place, waste treatment, and nutrient cycling, most individual services addressed similar gaps as the overall service classes shown in Fig. 4. Fig. 5 summarizes the specific ecosystem services analysed while distinguishing between the gaps being addressed for each service.

4. Discussion

We set out to understand how big data and ML are increasingly being used to fill key knowledge and methodological gaps in ecosystem service research including data availability, model and output uncertainty, and linking socio-ecological systems. Our results show that the uptake of both big data and ML algorithms in ecosystem service research is increasing rapidly, especially since 2017, and that trend looks to

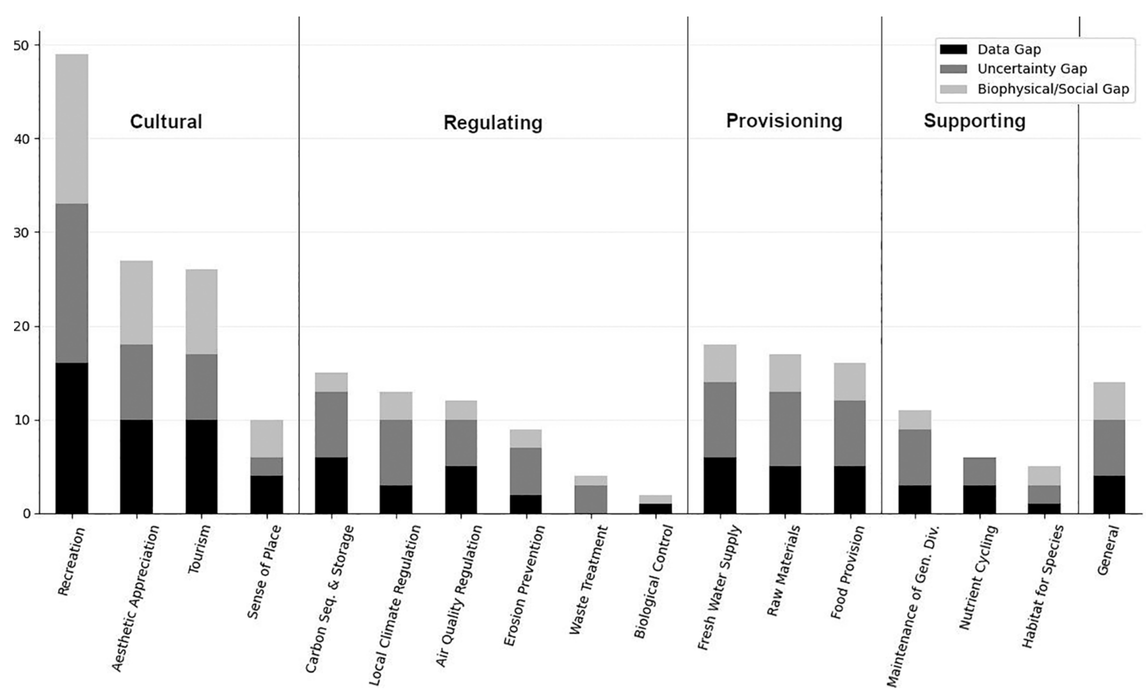

Fig. 5. Distribution of ecosystem services studied and the gaps being addressed.

continue beyond 2021 (Fig. 2). This is not surprising as we have seen similar increases in the use of ML in the ecology field for a variety of tasks such as species distribution modeling, species identification, and predicting the conservation status of species (Liu et al., 2018; Lucas, 2020). However, based on the number of studies identified during this review process (N = 56), it is evident that the field of ecosystem services has not fully utilized the opportunity to use these new tools and approaches, in particular the capacity to use ML to exploit big data (Willcock et al., 2018; Scowen et al., 2021). This is a missed opportunity because although traditional techniques, like linear regression or multinomial logistic regression, can integrate big data in ecosystem service research, the use of ML algorithms has the potential to discern more detailed patterns within the data, create more timely and precise predictions, and provide other benefits that can help fill major gaps in ecosystem service research (Zhou et al., 2017).

Using ML and big data together has even more potential to address the major gaps in ecosystem service research. For example, Callau et al. (2019) use 8-bit image recognition software to classify images, in order to analyze a crowdsourced social media dataset for the modeling and mapping of regional ecosystem services. This task would have otherwise been extremely difficult and time-consuming to complete manually, and without the big data, an alternative but similar dataset could likely not be found. Similarly, Bragagnolo et al. (2018) utilized a dataset of social media photographs to classify cultural ecosystem services within a region of Brazil, but acknowledged that using ML would have allowed for a more efficient larger scale analysis and less reliance on human interpretation.

4.1. Data availability and uncertainty gaps

In general, our findings show that most studies address the data availability as well as the model and output uncertainty gaps as illustrated in Fig. 3 (example case studies in appendix: Box 1). This is not surprising considering that big data and ML have inherent abilities to address data availability and uncertainty issues (Willcock et al., 2018). For example, big data introduces significant amounts of novel data that can directly and indirectly fill data availability gaps in areas of low data availability. Bragagnolo et al. (2018) demonstrate this when using crowdsourced data from Flickr to scale up their valuation of cultural services to cover their entire study region of the Caatinga in Brazil, which otherwise is too large an area to assess using traditional data. ML algorithms also give researchers the ability to directly quantify model and output uncertainty. Degerickx et al. (2020) illustrate this ability when using random forest to classify remotely sensed images and set a threshold for an acceptable amount of uncertainty to effectively highlight regions with higher uncertainty.

The data availability gap is most often filled using big data (54 %), but ML also plays a sizable role in filling this gap (36 %) (Fig. 3). This is due to the fact that multiple ML algorithms, especially Bayesian Networks, can be effective in modeling ecosystem services and their related dynamics in data sparse environments (Landuyt et al., 2014; Balbi et al, 2019). ML is used most often to address the uncertainty gap to produce robust outputs of ecosystem services and to understand the level of uncertainty associated with these outputs. Many ML algorithms, most notably random forest and Bayesian networks, automatically provide estimates of uncertainty (Cimburova and Barton, 2020; Degerickx et al., 2020), and can increase the accuracy of model predictions (Ciesielski and Sterenczak 2021; Morshed et al., 2021). Furthermore, ML increases the sample size that can be assessed, which can help mitigate uncertainty.

4.2. The socio-ecological gap

One of the biggest challenges in ecology, which has also been identified as one of the gaps hindering the attainment of global sustainability goals, is understanding and studying the socio-ecological components of ecosystems and finding indicators that can capture both the social and biophysical aspects of ecosystem services (Mastrángelo et al., 2019). Considering that the field of ecosystem services is relatively new, most scientists have used proxy data and/or expert knowledge in the modeling and mapping of services, relying on linear assumptions of the non-linear dynamics of socio-ecological connections. However, some studies have shown that simple proxies such as land use and land cover provide an oversimplified and often poor representation of ecosystem services in reality (Eigenbrod et al., 2010; Schulp et al., 2014; Lautenbach et al., 2019). In this study, we find the gap between the biophysical and social aspects of ecosystem services are addressed relatively equally by big data (48 %) and ML (40 %), although big data contributes slightly more. For example, new big data resources like social media and mobile phone data, have a huge potential and are being used to assess people's preferences for different ecosystems/ecosystem services and how they relate to human behavior (Lee et al., 2019; Jaung and Carrasco, 2020) (Also see appendix: Box 2). Such resources can address the third gap by providing a measure of the social behaviour and preferences of people and improving our understanding of the interaction between biophysical supply and social demand of ecosystem services within different landscapes. This gives ecosystem service researchers a tool to directly capture the connection between the social and biophysical aspects of ecosystems as well as the ability to directly assess the demand for ecosystem services.

ML can help bridge gaps in the modeling of socio-ecological systems through a variety of algorithms that have the unique abilities to both analyze and classify large crowdsourced datasets and handle the highly nonlinear and complicated relationships between the biophysical and social aspects of ecosystems (Debanshi and Pal., 2020; Rippy et al., 2021). ML deciphers relationships directly from the data with no prior assumptions or linear constraints, allowing for a more effective generalization of the complex structures and functions of ecological systems (Huettman et al., 2018; Humphries et al., 2018). This is especially the case when the modeled relationship between the dependent and independent variables is nonlinear and complex like those common when modeling cultural ecosystem services or the human demand or preference for the biophysical supply of ecosystem services (i.e., socio-ecological relationships), or when data is missing or low in availability (Thessen, 2016; Willcock et al., 2018; Debanshi and Pal., 2020; Frey, 2020; Lorilla et al., 2020). ML is not going to perform better than traditional techniques under every circumstance and should be thought of more as an additional toolset for ecosystem service research in which the method used, whether traditional or ML based, should be carefully chosen on a study-by-study basis (Thessen 2016; Willcock et al., 2018).

4.3. Patterns of addressing gaps within ecosystem services

The emerging patterns of which gaps are being addressed are also highly variable when looking at what ecosystem service class each paper is assessing (Fig. 4). Traditionally, cultural ecosystem services are the least mapped and most understudied compared to other service classes, due largely to limited data that fails to effectively interpret and capture these services (Brown et al., 2016; Cheng et al., 2019; Kosanic and Petzold, 2020). Cultural ecosystem services are difficult to map as they are subjective and intangible due to their dependence upon both social and biophysical factors (Hølleland et al., 2017; Cheng et al, 2019). However, in this review, 38 % percent of the studies addressing the data availability gap and 36 % of those modeling the socio-ecological system were focused on cultural ecosystem services. This is not surprising, considering that more recently, with the advent of crowdsourced data there has been an increase in the modeling of cultural services via geolocated crowdsourced data from a variety of sources. For example, traditionally, recreation is mapped using visitor numbers to protected areas that are often collected at an entrance or using stated preference methods, but this data is severely limited in both spatial and temporal resolution (Vaz et al., 2020, Wood et al., 2020). Big data offers thousands to millions of visitation data points with high spatial and temporal resolution. The combination of such data with ML allows researchers to predict where and when ecosystem services are provided, what specific ecosystems and services people are interested in, and what socio-ecological factors influence the demand for those services. Richards and Tuncer (2018) are a great example of this as they use Google Cloud Vision's image recognition software (https://cloud.google.com/vision) to classify 25,000 Flickr photographs in order to identify and map important cultural ecosystem services within Singapore. They proceeded to model the probability of other photograph occurrences as a proxy for each of the classified cultural services using Maxent and multiple socio-ecological covariates (Richards and Tuncer, 2018).

Studies that address multiple ecosystem services tend to focus on addressing uncertainty gaps and often utilize decision tree-based ML algorithms such as random forest, conditional inference tree, and Classification and Regression Trees (CART) to perform classification and regression tasks using the non-uniform and multifaceted variables required to assess multiple different service classes (Kang et al., 2018; Havinga et al., 2019; Degerickx et al., 2020). Since these studies do not focus on a single ecosystem service class, uncertainty has the potential to increase. For example, there will be an increase in types of data/indicators needed and generalizations will have to be made between multiple classes to assess them in a similar fashion. There are major concerns of the uncertainties in ecosystem service research relating to too many indicators used for the same service, a different level of understanding of various services, and different services being assessed using different methods (process based or knowledge based) (Schulp et al., 2014). Inherently, if you include multiple service classes in an assessment, such uncertainties increase and need to be addressed. ML algorithms assist with this issue by explicitly accounting for and quantifying uncertainty and improving the accuracy of models that handle the many variables that go into the assessment of multiple ecosystem services or service classes at various scales. For example, Lorilla et al. (2020) used random forest to identify the multitude of socio-ecological factors that contribute to the supply and demand of multiple ecosystem services across different service classes including provisioning, regulating, and cultural. Utilizing random forest allows the reduction of uncertainty at multiple levels as it has been proven to be robust to overfitting even with a large and variable sample of socio-ecological data.

4.4. Techniques used for addressing gaps

Different ML algorithms have different advantages and disadvantages and are thus used for a multitude of purposes within ecosystem service research (Table 1). Almost all studies that utilized ML algorithms chose to use supervised algorithms such as random forest, Bayesian networks, and Maxent particularly for regression and classification tasks. For example, Maxent has been used for modeling aesthetic appreciation services (Bernetti et al., 2019; He et al., 2019), Bayesian networks and random forest have been used to model carbon sequestration services (Landuyt et al., 2014; Havinga et al., 2020b; Cimburova and Barton, 2020), and convolutional neural networks have been used to classify various cultural services like recreation, aesthetic appreciation, and tourism (Richards et al., 2018; Callau et al., 2019; Gosal et al., 2019; Lee et al., 2019). The few times unsupervised learning was used within our data sample, it was utilized for data analysis and processing purposes (e.g. clustering) as in Jaung (2021) (See Box 2 in appendix). Unsupervised learning algorithms hold a potentially important key for future research in ecosystem services, especially relating to the rise in ecological big data, due to their efficacy in finding previously unidentified patterns in data with no a-priori assumptions (Rammer and Seidl, 2019).

Bayesian networks specifically are highly effective at modeling in low data availability situations in which many other alternative methods would have higher uncertainty (Landuyt et al., 2014; Cimburova and Barton, 2020). Random forest was used often for both classification and regression tasks. Many researchers used random forest due to the robustness against overfitting, high accuracy in past ecosystem service assessments, and the ability to quantify variable importance and estimate uncertainty, both helping address the uncertainty gap (Kang et al., 2018; Shiferaw et al., 2019; Lorilla et al., 2020). Random forest's efficacy in classification tasks along with the rise in big data relevant to ecosystem services helps address the data availability gap. This ability to effectively classify big data (social media, remotely sensed data, etc.) should make random forest a significant part of future ecosystem service research, especially in situations of low data availability. Maxent is the third most common ML algorithm and most often addressed the socio-ecological gap. Maxent often used for species distribution modeling, has seen an increase in use for modeling cultural ecosystem services (Bernetti et al., 2019; He et al., 2019). Maxent has the ability to explore the relationship between environmental covariates and people's actual use of cultural ecosystem services, providing researchers with tools to overcome the difficulties of connecting the socio-ecological system. Furthermore, some models require both presence and absence data, but Maxent only requires presence data, helping with data availability issues. There are many other supervised and unsupervised algorithms that have a multitude of relevant uses, but for a more in-depth discussion we refer readers to Mohri et al. (2018) or for more examples of ML use in ecology we suggest Humphries et al. (2018).

4.5. Limitations

While big data and ML present opportunities for future ecosystem service research, like any other data source or methodology there exist limitations. For example, one significant limitation for both tools is the high level of expertise needed to utilize ML algorithms and big data which can often lead to the impression that ML algorithms are "black boxes" (Zhou et al. 2017). Additionally, the use of ML as well as the collection and use of big data may require a significant amount of resources, whether they be training, internet, software, or other computational or technical resources (Thessen, 2016; Farley et al., 2018). For the most part, these resources are already in place (ex. established universities and research organizations) and have been collecting and analyzing relevant data for long periods of time (Martin et al., 2012), but only in certain parts of the world. For example, the majority of studies in this review were from China and western Europe, suggesting regions elsewhere need increased access and infrastructure for these resources. Some of the absence of countries in this study may be due to the fact that we do not fully capture all ecosystem service studies using ML or big data due to the keywords used, an English language focus, no grey literature being included, and other limitations of literature reviews.

5. Conclusions

As shown within this review, ML and big data are useful tools that can help address data availability, uncertainty, and socio-ecological gaps in ecosystem service research. Specifically targeting regions that are understudied due to these major gaps will benefit these regions and the ecosystem service field overall. Many regions, especially within the global South, are understudied within the ecosystem service field mainly due to low data availability. The application of newly relevant big databases (e.g. social media data), the increasing size and resolution of big data, and the ability of several ML algorithms to estimate uncertainty and effectively model the complex dynamics of ecosystem services in low-data environments all can assist in providing tools to these understudied and under-resourced regions. Also, improving relevant computing, data, and other resources as well as expertise within regions that lack studies of ecosystem services in general and specifically, using big data and/or ML, can help fill gaps in current knowledge. To this point, these tools have not been fully taken advantage of, especially within these regions, and thus, there remains significant opportunity for future studies to address major gaps and disparities in ecosystem service research.

Funding

We acknowledge funding from the UK Research and Innovation's Global Challenges.

Research Fund (UKRI GCRF) through the Trade, Development and the Environment Hub project (project number ES/S008160/1). This work was also supported by the Ridge to Reef NSF Research Traineeship, award DGE-1735040.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoser.2022.101478.

References

Balbi, S., Selomane, O., Sitas, N., Blanchard, R., Kotzee, I., O'Farrell, P., Villa, F., 2019. Human dependence on natural resources in rapidly urbanising South African regions. Environ. Res. Lett. 14 (4) https://doi.org/10.1088/1748-9326/aafe43.

Bernetti, I., Chirici, G., Sacchelli, S., 2019. Big data and evaluation of cultural ecosystem services: An analysis based on geotagged photographs from social media in tuscan forest (Italy). IForest 12 (1), 98–105. https://doi.org/10.3832/ifor2821-011.

Bragagnolo, C., Vieira, F.A.S., Correia, R.A., Malhado, A.C.M., Ladle, R.J., 2018. Cultural services in the Caatinga. In: Silva, J.M.C.d., Leal, I.R., Tabarelli, M. (Eds.), Caatinga. Springer International Publishing, Cham, pp. 335–355.

Brown, G., Pullar, D., Hausner, V.H., 2016. An empirical evaluation of spatial value transfer methods for identifying cultural ecosystem services. Ecol. Ind. 69, 1–11. https://doi.org/10.1016/j.ecolind.2016.03.053.

Burkhard, B., Maes, J., 2017. Mapping ecosystem services. *Advanced Books* 1, e12837.

Callau, A.À., Albert, M.Y.P., Rota, J.J., Giné, D.S., 2019. Landscape characterization using photographs from crowdsourced platforms: Content analysis of social media photographs. Open Geosciences 11 (1), 558–571. https://doi.org/10.1515/geo-2019-0046.

Cheng, X., Van Damme, S., Li, L., Uyttenhove, P., 2019. Evaluation of cultural ecosystem services: A review of methods. Ecosyst. Serv. 37, 100925 https://doi.org/10.1016/j.ecoser.2019.100925.

Chun, J., Kim, C.-K., Kim, G.S., Jeong, J., Lee, W.-K., 2020. Social big data informs spatially explicit management options for national parks with high tourism pressures. Tourism Manage. 81, 104136 https://doi.org/10.1016/j.tourman.2020.104136.

Ciesielski, M., Stereńczak, K., 2021. Using Flickr data and selected environmental characteristics to analyse the temporal and spatial distribution of activities in forest areas. Forest Policy Econ. 129, 102509 https://doi.org/10.1016/j.forpol.2021.102509.

Cimburova, Z., Barton, D.N., 2020. The potential of geospatial analysis and Bayesian networks to enable i-Tree Eco assessment of existing tree inventories. Urban For. Urban Green. 55, 126801 https://doi.org/10.1016/j.ufug.2020.126801.

Convention on Biological Diversity (2020): Update of the zero draft of the post-2020 global biodiversity framework. Preparations for the post-2020 biodiversity framework. 10 pages. https://www.cbd.int.

Debanshi, S., Pal, S., 2020. Modelling water richness and habitat suitability of the wetlands and measuring their spatial linkages in mature Ganges delta of India. J. Environ. Manage. 271 https://doi.org/10.1016/j.jenvman.2020.110956.

Degerickx, J., Hermy, M., Somers, B., 2020. Mapping functional urban green types using high resolution remote sensing data. Sustainability (Switzerland) 12 (5). https://doi.org/10.3390/su12052144.

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R.T., Molnár, Z., Hill, R., Chan, K.M.A., Baste, I.A., Brauman, K.A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P.W., van Oudenhoven, A.P.E., van der Plaat, F., Schröter, M., Lavorel, S., Aumeeruddy-Thomas, Y., Bukvareva, E., Davies, K., Demissew, S., Erpul, G., Failler, P., Guerra, C.A., Hewitt, C.L., Keune, H., Lindley, S., Shirayama, Y., 2018. Assessing nature's contributions to people. Science 359 (6373), 270–272.

Egoh, B., Drakou, E.G., Dunbar, M.B., Maes, J., Willemen, L., 2012. Indicators for Mapping Ecosystem Services: A Review. European Commission, Joint Research Centre (JRC).

Eigenbrod, F., Armsworth, P.R., Anderson, B.J., Heinemeyer, A., Gillings, S., Roy, D.B., Thomas, C.D., Gaston, K.J., 2010. The impact of proxy-based methods on mapping the distribution of ecosystem services. J. Appl. Ecol. 47 (2), 377–385. https://doi.org/10.1111/j.1365-2664.2010.01777.x.

Escobedo, F.J., Bottin, M., Cala, D., Sandoval Montoya, D.L., 2020. Spatial literacy influences stakeholder's recognition and mapping of peri-urban and urban ecosystem services. Urban Ecosyst. 23 (5), 1039–1049.

European Commission. (2020). EU Biodiversity Strategy for 2030. Bringing nature back into our lives.

Farley, S.S., Dawson, A., Goring, S.J., Williams, J.W., 2018. Situating ecology as a big-data science: Current advances, challenges, and solutions. Bioscience 68 (8), 563–576. https://doi.org/10.1093/biosci/biy068.

Frey, U.J., 2020. Putting machine learning to use in natural resource management—improving model performance. Ecol. Soc. 25 (4) https://doi.org/10.5751/ES-12124-250445.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Trans. ASABE 50 (4), 1211–1250.

Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B., Ziv, G., 2019. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosyst. Serv. 38 https://doi.org/10.1016/j.ecoser.2019.100958.

Grêt-Regamey, A., Sirén, E., Brunner, S.H., Weibel, B., 2017. Review of decision support tools to operationalize the ecosystem services concept. Ecosyst. Serv. 26, 306–315. https://doi.org/10.1016/j.ecoser.2016.10.012.

Hamel, P., Bryant, B.P., 2017. Uncertainty assessment in ecosystem services analyses: Seven challenges and practical responses. Ecosyst. Serv. 24, 1–15. https://doi.org/10.1016/j.ecoser.2016.12.008.

Havinga, I., Bogaart, P.W., Hein, L., Tuia, D., 2020a. Defining and spatially modelling cultural ecosystem services using crowdsourced data. Ecosyst. Serv. 43, 101091 https://doi.org/10.1016/j.ecoser.2020.101091.

Havinga, I., Hein, L., Vega-Araya, M., Languillaume, A., 2020b. Spatial quantification to examine the effectiveness of payments for ecosystem services: A case study of Costa Rica's Pago de Servicios Ambientales. Ecol. Ind. 108 https://doi.org/10.1016/j.ecolind.2019.105766.

He, S., Su, Y., Shahtahmassebi, A.R., Huang, L., Zhou, M., Gan, M., Deng, J., Zhao, G., Wang, K., 2019. Assessing and mapping cultural ecosystem services supply, demand and flow of farmlands in the Hangzhou metropolitan area, China. Sci. Total Environ. 692, 756–768. https://doi.org/10.1016/j.scitotenv.2019.07.160.

Hirabayashi, S., Nowak, D., Endreny, T., Kroll, C., Maco, S., 2011. i-Tree: Tools to assess and manage structure, function, and value of community forests. *AGU Fall MeetIng Abstracts*.

Hølleland, H., Skrede, J., & Holmgaard, S. B. (2017). Cultural heritage and ecosystem services: A literature review. *Conservation and Management of Archaeological Sites*, *19* (3), 210–237. doi: 10.1080/13505033.2017.1342069.

Huettmann, F., Craig, E. H., Herrick, K. A., Baltensperger, A. P., Humphries, G. R. W., Lieske, D. J., Miller, K., Mullet, T. C., Oppel, S., Resendiz, C., Rutzen, I., Schmid, M. S., Suwal, M. K., & Young, B. D. (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* (G. Humphries, D. R. Magness, & F. Huettmann (eds.); pp. 27–61). Springer International Publishing. doi: 10.1007/978-3-319-96978-7_2.

Humphries, G., Magness, D.R., Huettmann, F. (Eds.), 2018. Machine Learning for Ecology and Sustainable Natural Resource Management. Springer International Publishing, Cham.

IPBES (2019): Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele, E. S. Brondízio E.S., H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, and C. N. Zayas (eds.). IPBES secretariat, Bonn, Germany. 56 pages. doi: 10.5281/zenodo.3553579.

Jaung, W., 2021. Changes in human-nature relations during pandemic outbreaks: a big data analysis. Sci. Total Environ. 768, 144530 https://doi.org/10.1016/j.scitotenv.2020.144530.

Jaung, W., Carrasco, L.R., 2020. Travel cost analysis of an urban protected area and parks in Singapore: a mobile phone data application. J. Environ. Manage. 261, 110238 https://doi.org/10.1016/j.jenvman.2020.110238.

Kang, P., Chen, W., Hou, Y., Li, Y., 2018. Linking ecosystem services and ecosystem health to ecological risk assessment: A case study of the Beijing-Tianjin-Hebei urban agglomeration. Sci. Total Environ. 636, 1442–1454. https://doi.org/10.1016/j.scitotenv.2018.04.427.

Kim, I., Arnhold, S., Ahn, S., Le, Q.B., Kim, S.J., Park, S.J., Koellner, T., 2019. Land use change and ecosystem services in mountainous watersheds: Predicting the consequences of environmental policies with cellular automata and hydrological modeling. Environ. Modell. Software 122, 103982.

Kosanic, A., Petzold, J., 2020. A systematic review of cultural ecosystem services and human wellbeing. Ecosyst. Serv. 45, 101168 https://doi.org/10.1016/j.ecoser.2020.101168.

Landuyt, D., Broekx, S., Van Der Biest, K., & Goethals, P. (2014). Probabilistic mapping with Bayesian belief networks: An application on ecosystem service delivery in

Flanders, Belgium. *Proceedings - 7th International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling, IEMSs 2014*, *2*, 790–797.
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., Volk, M., 2019. Blind spots in ecosystem services research and challenges for implementation. Reg. Environ. Change 19 (8), 2151–2172. https://doi.org/10.1007/s10113-018-1457-9.
Lavorel, S., Bayer, A., Bondeau, A., Lautenbach, S., Ruiz-Frau, A., Schulp, N., Seppelt, R., Verburg, P., van Teeffelen, A., Vannier, C., Arneth, A., Cramer, W., Marba, N., 2017. Pathways to bridge the biophysical realism gap in ecosystem services mapping approaches. Ecol. Ind. 74, 241–260. https://doi.org/10.1016/j.ecolind.2016.11.015.
Lee, H., Seo, B., Koellner, T., Lautenbach, S., 2019. Mapping cultural ecosystem services 2.0 – Potential and shortcomings from unlabeled crowd sourced images. Ecol. Ind. 96, 505–515. https://doi.org/10.1016/j.ecolind.2018.08.035.
Liu, Z., Peng, C., Work, T., Candau, J.N., DesRochers, A., Kneeshaw, D., 2018. Application of machine-learning methods in forest ecology: recent progress and future challenges. Environ. Rev. 26 (4), 339–350. https://doi.org/10.1139/er-2018-0034.
Lorilla, R.S., Poirazidis, K., Detsis, V., Kalogirou, S., Chalkias, C., 2020. Socio-ecological determinants of multiple ecosystem services on the Mediterranean landscapes of the Ionian Islands (Greece). Ecol. Model. 422 https://doi.org/10.1016/j.ecolmodel.2020.108994.
Lucas, T.C.D., 2020. A translucent box: interpretable machine learning in ecology. Ecol. Monogr. 90 (4), e01422.
Mandle, L., Shields-Estrada, A., Chaplin-Kramer, R., Mitchell, M.G.E., Bremer, L.L., Gourevitch, J.D., Hawthorne, P., Johnson, J.A., Robinson, B.E., Smith, J.R., Sonter, L.J., Verutes, G.M., Vogl, A.L., Daily, G.C., Ricketts, T.H., 2021. Increasing decision relevance of ecosystem service science. Nat. Sustainability 4 (2), 161–169.
Martin, L., Blossey, B., Ellis, E., 2012. Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. Front. Ecol. Environ. 10, 195–201. https://doi.org/10.2307/41480052.
Mastrángelo, M.E., Pérez-Harguindeguy, N., Enrico, L., Bennett, E., Lavorel, S., Cumming, G.S., Abeygunawardane, D., Amarilla, L.D., Burkhard, B., Egoh, B.N., Frishkoff, L., Galetto, L., Huber, S., Karp, D.S., Ke, A., Kowaljow, E., Kronenburg-García, A., Locatelli, B., Martín-López, B., Meyfroidt, P., Mwampamba, T.H., Nel, J., Nicholas, K.A., Nicholson, C., Oteros-Rozas, E., Rahlao, S.J., Raudsepp-Hearne, C., Ricketts, T., Shrestha, U.B., Torres, C., Winkler, K.J., Zoeller, K., 2019. Key knowledge gaps to achieve global sustainability goals. Nat. Sustainability 2 (12), 1115–1121.
McCue, C., 2015. Chapter 7 - Identification, characterization, and modeling. In: McCue, C. (Ed.), *Data Mining and Predictive Analysis (Second Edition)*. Butterworth-Heinemann, pp. 137–155. https://doi.org/10.1016/B978-0-12-800229-2.00007-9.
McDonough, K., Hutchinson, S., Moore, T., Hutchinson, J.M.S., 2017. Analysis of publication trends in ecosystem services research. Ecosyst. Serv. 25, 82–88. https://doi.org/10.1016/j.ecoser.2017.03.022.
Millennium Ecosystem Assessment, 2002. Ecosystem and Human well-being: a Framework for Assessment. Island Press, NY.
Mohri, M., Rostamizadeh, A., Talwalkar, A., 2018. Foundations of Machine Learning. MIT Press.
Morshed, S.R., Fattah, M.A., Haque, M.N., Morshed, S.Y., 2022. Future ecosystem service value modeling with land cover dynamics by using machine learning based Artificial Neural Network model for Jashore city, Bangladesh. Phys. Chem. Earth Parts A/B/C 126. https://doi.org/10.1016/j.pce.2021.103021.
NIST. (2015). *Big data interoperability framework: vol 1, definitions*. doi: https://doi.org/10.6028/NIST.SP.1500-1.
Ochoa, V., Urbina-Cardona, N., 2017. Tools for spatially modeling ecosystem services: Publication trends, conceptual reflections and future challenges. Ecosyst. Serv. 26, 155–169. https://doi.org/10.1016/j.ecoser.2017.06.011.
Ouyang, X., Wang, Z., Zhu, X., 2019. Construction of the ecological security pattern of urban agglomeration under the framework of supply and demand of ecosystem services using bayesian network machine learning: Case study of the Changsha-Zhuzhou-Xiangtan Urban Agglomeration, China. *Sustainability (Switzerland)* 11 (22). https://doi.org/10.3390/su11226416.
Rammer, W., Seidl, R., 2019. Harnessing deep learning in ecology: an example predicting bark beetle outbreaks. Front. Plant Sci. 10 https://doi.org/10.3389/fpls.2019.01327.
Reyers, B., Biggs, R., Cumming, G., Elmqvist, T., Hejnowicz, A., Polasky, S., 2013. Getting the measure of ecosystem services: A social–ecological approach. Front. Ecol. Environ. 11, 268–273. https://doi.org/10.2307/23470507.
Richards, D.R., Tunçer, B., Tunçer, B., 2018. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. Ecosyst. Serv. 31, 318–325. https://doi.org/10.1016/j.ecoser.2017.09.004.
Rippy, M.A., Krauss, L., Pierce, G., Winfrey, B., 2021. Plant functional traits and viewer characteristics co-regulate cultural services provisioning by stormwater bioretention. Ecol. Eng. 168, 106284 https://doi.org/10.1016/j.ecoleng.2021.106284.
Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. 2013 International Conference on Collaboration Technologies and Systems (CTS), 42–47. doi: 10.1109/CTS.2013.6567202.
Schulp, C.J.E., Burkhard, B., Maes, J., Van Vliet, J., Verburg, P.H., Yue, G.H., 2014. Uncertainties in ecosystem service maps: a comparison on the European scale. PLoS ONE 9 (10), e109643.
Scowen, M., Athanasiadis, I.N., Bullock, J.M., Eigenbrod, F., Willcock, S., 2021. The current and future uses of machine learning in ecosystem service research. Sci. Total Environ. 799, 149263 https://doi.org/10.1016/j.scitotenv.2021.149263.
Sherrouse, B.C., & Semmens, D.J. (2015). Social values for ecosystem services, version 3.0 (SolVES 3.0)—Documentation and user manual: U.S. Geological Survey Open-File Report 2015–1008, 65 p.
Shiferaw, H., Bewket, W., Alamirew, T., Zeleke, G., Teketay, D., Bekele, K., Schaffner, U., 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, 354–366. https://doi.org/10.1016/j.scitotenv.2019.04.220.
Spake, R., Lasseur, R., Crouzat, E., Bullock, J.M., Lavorel, S., Parks, K.E., Schaafsma, M., Bennett, E.M., Maes, J., Mulligan, M., Mouchet, M., Peterson, G.D., Schulp, C.J.E., Thuiller, W., Turner, M.G., Verburg, P.H., Eigenbrod, F., 2017. Unpacking ecosystem service bundles: Towards predictive mapping of synergies and trade-offs between ecosystem services. Global Environ. Change 47, 37–50. https://doi.org/10.1016/j.gloenvcha.2017.08.004.
Strohbach, M.W., Haase, D., 2012. Above-ground carbon storage by urban trees in Leipzig, Germany: Analysis of patterns in a European city. Landscape Urban Plann. 104 (1), 95–104. https://doi.org/10.1016/j.landurbplan.2011.10.001.
Sun, A.Y., Scanlon, B.R., 2019. How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environ. Res. Lett. 14 (7), 73001.
Tallis, H., Polasky, S., 2009. Mapping and valuing ecosystem services as an approach for conservation and natural-resource management. Ann. N. Y. Acad. Sci. 1162 (1), 265–283. https://doi.org/10.1111/j.1749-6632.2009.04152.x.
Thessen, A., 2016. Adoption of machine learning techniques in ecology and earth science. One Ecosyst. 1, e8621.
Tsai, Y.H., Stow, D., An, L.i., Chen, H.L., Lewison, R., Shi, L., 2019. Monitoring land-cover and land-use dynamics in Fanjingshan National Nature Reserve. Appl. Geogr. 111 https://doi.org/10.1016/j.apgeog.2019.102077.
Vaz, A.S., Moreno-Llorca, R.A., Gonçalves, J.F., Vicente, J.R., Méndez, P.F., Revilla, E., Santamaria, L., Bonet-García, F.J., Honrado, J.P., Alcaraz-Segura, D., 2020. Digital conservation in biosphere reserves: Earth observations, social media, and nature's cultural contributions to people. Conserv. Lett. 13 (3) https://doi.org/10.1111/conl.12704.
Villa, F., Bagstad, K.J., Voigt, B., Johnson, G.W., Portela, R., Honzák, M., Batker, D., Margalida, A., 2014. A methodology for adaptable and robust ecosystem services assessment. PLoS ONE 9 (3), e91001.
Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C.J.A., Buytaert, W., 2015. Web technologies for environmental Big Data. Environ. Modell. Software 63, 185–198. https://doi.org/10.1016/j.envsoft.2014.10.007.
Willcock, S., Martínez-López, J., Hooftman, D.A.P., Bagstad, K.J., Balbi, S., Marzo, A., Prato, C., Sciandrello, S., Signorello, G., Voigt, B., Villa, F., Bullock, J.M., Athanasiadis, I.N., 2018. Machine learning for ecosystem services. Ecosyst. Serv. 33, 165–174.
Willcock, S., Martinez-Lopez, J., Dandy, N., Bullock, J.M., 2021. High Spatial-temporal resolution data across large scales are needed to transform our understanding of ecosystem services. Land 10 (7). https://doi.org/10.3390/land10070759.
Wood, S.A., Winder, S.G., Lia, E.H., White, E.M., Crowley, C.S.L., Milnor, A.A., 2020. Next-generation visitation models using social media to estimate recreation on public lands. Sci. Rep. 10 (1), 15419. https://doi.org/10.1038/s41598-020-70829-x.
Xia, J., Wang, J., Niu, S., 2020. Research challenges and opportunities for using big data in global change biology. Glob. Change Biol. 26 (11), 6040–6061. https://doi.org/10.1111/gcb.15317.
Xu, X., Chen, M., Yang, G., Jiang, B., Zhang, J., 2020. Wetland ecosystem services research: A critical review. Global Ecol. Conserv. 22, e01027.
Zhang, X.-D. (2020). Machine Learning BT - A Matrix Algebra Approach to Artificial Intelligence (X.-D. Zhang (ed.); pp. 223–440). *Springer Singapore*. doi: 10.1007/978-981-15-2770-8_6.
Zhou, L., Pan, S., Wang, J., Vasilakos, A.V., 2017. Machine learning on big data: Opportunities and challenges. Neurocomputing 237, 350–361. https://doi.org/10.1016/j.neucom.2017.01.026.