Applied Geography 165 (2024) 103249

Contents lists available at ScienceDirect

Applied Geography

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

Satellite-based Machine Learning modelling of Ecosystem Services indicators: A review and meta-analysis

Bruna Almeida [b,c,*], João David [d], Felipe S. Campos [b,e,f], Pedro Cabral [a,b,**]

[a] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[b] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312, Lisboa, Portugal
[c] Centre for Ecology, Evolution and Environmental Changes (cE3c), Azorean Biodiversity Group (GBA), University of the Azores (UAc), Faculty of Sciences and Technology (FCT-UAc), Rua Mãe de Deus, Campus Universitário de Ponta Delgada, Edifício do Complexo Científico, 9500-321, Ponta Delgada, Portugal
[d] Institute of Geography, Humboldt University of Berlin, Rudower Chaussee 16, 12489, Berlin, Germany
[e] Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès, Catalunya, Spain
[f] Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), 08193, Cerdanyola del Vallès, Catalunya, Spain

A R T I C L E I N F O

Handling editor: J Peng

Keywords:
Remote sensing
Natural capital
Biodiversity
Data fusion
MLOps
Environmental modelling

A B S T R A C T

Satellite-based Machine Learning (ML) modelling has emerged as a powerful tool to understand and quantify spatial relationships between landscape dynamics, biophysical variables and natural stocks. Ecosystem Services indicators (ESi) provide qualitative and quantitative information aiding the assessment of ecosystems' status. Through a systematic meta-analysis following the PRISMA guidelines, studies from one decade (2012–2022) were analyzed and synthesized. The results indicated that Random Forest emerged as the most frequently utilized ML algorithm, while Landsat missions stood out as the primary source of Satellite Earth Observation (SEO) data. Nonetheless, authors favoured Sentinel-2 due to its superior spatial, spectral, and temporal resolution. While 30% of the examined studies focused on modelling proxies of climate regulation services, assessments of natural stocks such as biomass, water, food production, and raw materials were also frequently applied. Meta-analysis illustrated the utilization of classification and regression tasks in estimating measurements of ecosystems' extent and conditions and findings underscored the connections between established methods and their replication. This study offers current perspectives on existing satellite-based approaches, contributing to the ongoing efforts to employ ML and artificial intelligence for unveiling the potential of SEO data and technologies in modelling ESi.

1. Introduction

Nature offers society a diverse array of services and resources through Ecosystem Services (ES) (Costanza et al., 1997). Natural ecosystems, recognized for their ability to deliver a multitude of ES, are appraised as assets through physical and monetary evaluations recorded as stock and flow accounts (Vallecillo et al., 2019). The physical accounts involve quantifying ecosystem extent, measuring the total occupied area, and assessing indicators that reflect the conditions of ecosystems and their services (Edens et al., 2022). ES indicators (ESi) are measurable parameters that provide insights into the status and trends of the diverse benefits ecosystems offer to humanity, spanning ecological, economic, and social dimensions, as well as for making policy and management decisions (Olander et al., 2018). The aboveground and belowground biomass in forests play a crucial role in supporting water filtration services by collecting and filtering rainfall before it reaches streams and rivers. ESi, such as tree cover density, water and wetness probability index, and soil depth, serve as proxies to describe the status of the water filtration service, contributing to providing cleaner water and reducing water treatment costs (United Nations et al., 2021).

Therefore, assessing ESi is still a challenge due to the intricate relationships within ecological dynamic systems, which involve numerous processes and components that interact nonlinearly (Haines-Young & Potschin, 2012). Besides, traditional approaches relying on simplistic spatial surrogates, often lead to incomplete representations of biophysical variables, and extensive ground surveys for spatially explicit

* Corresponding author. NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312, Lisboa, Portugal.
** Corresponding author. School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, 210044, China.
E-mail addresses: balmeida@novaims.unl.pt (B. Almeida), cabral@nuist.edu.cn (P. Cabral).

https://doi.org/10.1016/j.apgeog.2024.103249
Received 19 September 2023; Received in revised form 20 February 2024; Accepted 5 March 2024
Available online 10 March 2024

mapping encounter limitations such as high costs and scale restrictions (Zergaw Ayanu et al., 2012).

The literature proposes various methods to quantify ESi and natural stocks (Boyd & Banzhaf, 2007; De Groot et al., 2002; Fisher et al., 2009; Kienast et al., 2009; Syrbe & Walz, 2012). These studies have emphasized the importance of taking a comprehensive and integrated approach to developing and implementing ESi. A frequently employed approach involves extrapolating measurements of landscape ability provisioning through the utilization of land use and land cover (LULC) maps (Bolliger & Kienast, 2010; Costanza et al., 1997). Previous reviews (Andrew et al., 2014; de Araujo Barbosa et al., 2015; Feng et al., 2010; Malinga et al., 2015; Martínez-Harms & Balvanera, 2012; Zergaw Ayanu et al., 2012) have remarked that the main proxies for mapping ES, monitoring environmental status, and assessing ecosystems' extent and conditions often involve LULC variables.

However, limitations persist in accurately quantifying the interconnections among ecosystem functions, services, and the benefits to human well-being (Cord et al., 2017; Lausch et al., 2016). Ramirez-Reyes et al. (2019) emphasized the importance of shifting from categorical to continuous conceptualization in modelling ESi, as parameter coefficients are generally derived from field research or literature reviews. An alternative approach extending beyond mapping the presence/absence of ecosystems or relying on lookup tables for analysis is to use statistical regression to connect in-situ information with remotely sensed data to quantify ecosystem structure and functional traits (Lobert et al., 2021).

Remote Sensing (RS) offers a distinctive opportunity to systematically assess ecosystems' status across diverse spatial and temporal scales (Skidmore et al., 2021), holding promise for robust monitoring mechanisms within the framework of global sustainability policies and governance (Braun et al., 2018). Particularly Satellite Earth Observation (SEO) emerges as a cost-effective solution, providing data for quantifying and mapping ESi (Vyvlečka & Pechanec, 2023). ESi derived from SEO data models are spatial proxies aiding in the ES assessments through measurements of ecosystem extent and conditions (Czúcz et al., 2021).

Recently, Schirpke et al. (2023) conducted a review of studies exploring the use of emerging technologies in ES assessments, revealing the increased application of passive Earth observation sensors and ML algorithms among other technologies. Their study pointed out that in the last decade, ES modelling based on open-access satellite data predominantly involved high and middle spatial resolution sensors (ranging from 10 m to 1 km), such as those on board the Landsat, Sentinel, and Terra/Aqua MODIS satellites. In their review, Vyvlečka and Pechanec (2023) showcased a preference for Sentinel-2 over Landsat 8, especially in larger scales and vegetated areas, highlighting the relevance of spatial, temporal, and spectral resolutions in influencing modelling accuracy. They emphasized that additional factors like imagery processing, algorithm selection, and validation data also played crucial roles, and these should be further explored to enhance the potential of applications.

The integration of SEO data and ML stands as a powerful approach for ES assessment across multiple scales (Liang & Wang, 2019; Y. Wang, Zhang, & Peng, 2021; Willcock et al., 2018). ML has become crucial in environmental and landscape modelling, particularly in leveraging satellite data to precisely map diverse landscape properties (Ez-zahouani et al., 2023). They serve as a powerful tool for addressing challenges related to data-sparse locations, providing efficient estimations, and accounting for uncertainties in modelling ESi (Scowen et al., 2021). ML algorithms, such as Support Vector Machine (SVM) and Random Forest (RF), are widely used in RS applications (Maxwell et al., 2018), whereas Deep Learning (DL) has emerged as an advanced technique excelling at discriminating complex and nonlinear data structures (Goodfellow et al., 2016; LeCun et al., 2015) with great potential to fostering environmental modelling based on RS data (Ma et al., 2019; Pritt & Chern, 2018).

Various methods, including SVM, Decision Trees (DT), RF, Artificial Neural Networks (ANN), and k-Nearest Neighbours (k-NN), are widely utilized to predict complex LULC classes (Maxwell et al., 2018). Additionally, ensemble methods like Gradient Boosting, CatBoost, and Stochastic Gradient Boosting are adopted to enhance predictions (Friedman, 2001; Prokhorenkova et al., 2017). Thus, ML offers a feasible means to assess ecosystems' status, handling large and multidimensional datasets without relying only on oversimplified approaches or algorithms (Willcock et al., 2018). Despite the potential benefits, the implementation of SEO data and ML in ES research remains limited, posing some constraints (Czúcz et al., 2021). Challenges persist in refining methods, such as model settings, conceptualization, and knowledge of ecosystem influentional factors, underscoring the need for ongoing research development (Kubiszewski et al., 2022).

Recent literature reviews were focused on unveiling technologies for overcoming limitations encountered in ES assessments: Manley et al. (2022) showcased the current application of ML and Big Data to tackle gaps related to data availability, uncertainty understanding, and socio-environmental connections. Scowen et al. (2021) conducted a review to investigate the utilization of ML in ES research and identify trends in ML approaches. Their findings revealed that ML techniques are applied in data description and predictive modelling in a variety of datasets, including RS data from satellites and aerial vehicles. Other studies have contributed to the field of ES mapping and monitoring through RS, such as Feng et al. (2010), exploring the role of RS in ES assessments. Zergaw Ayanu et al. (2012) reviewed RS systems, sensor types, and methodologies relevant to quantifying provisioning and regulatory ESi. Andrew et al. (2014) explored the capabilities of RS in describing biodiversity, plant traits, and various ecological variables, emphasizing their contributions to ES assessments. Lausch et al. (2016) highlighted the affordability and repeatability of Earth Observation methods for measuring taxonomic, functional, and structural diversity. They called for a systematic approach to cross-case comparisons and methods development. Pettorelli et al. (2018) addressed the lack of consensus on defining and tracking ecosystem functions beyond the site level, proposing a framework for worldwide monitoring using satellite RS. C. Ramirez-Reyes et al. (2019) organized a workshop with researchers and decision-makers to outline stages in the assessment process where SEO could be applied. Their findings stressed that the widespread adoption of SEO data and technologies in ES assessments requires addressing conceptual barriers, including adapting existing products and models to accommodate continuous data.

Regardless of existing studies researching ES assessments through RS and ML, there is a gap in understanding the satellite-based ML modelling components, such as instruments, data integration, type of analysis and assessments, modelling conceptualization, development, and implementation. In this review article, we aim to offer a comprehensive overview of the current state of research in this domain, providing a synthesis of key methodologies, and answering the following research questions: i. Which SEO data are used for modelling ESi? ii. Which ESi are assessed using SEO and ML techniques? iii. Which ML models are implemented? iv. What is the relationship between ML tasks and ESi? v. Is there a rule for model development within ESi?

To address these questions, a systematic literature review and meta-analysis are conducted following the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Page et al., 2021) guidelines covering the literature from 2012 to 2022, and applying content analysis to formulate assumptions tested through statistical analysis. The review process utilizes a meta-analytical approach to understand the ML tasks applied to model ESi, providing a further understanding of the satellite-based ML modelling components such as SEO technologies, instruments, data integration, and type of analysis and assessments. Major outcomes incorporate insights into the application of SEO data and ML in ESi modelling, contributing to advancing the knowledge of the latest approaches developed by researchers worldwide.

2. Material and methods

Systematic reviews play a role for scholars and decision-makers, offering a structured approach to navigate through the vast volume of research and providing a solid foundation for informed decision-making (Greco et al., 2013). A systematic review process requires adhering to a set of baselines to identify, select, and synthesize data from the literature, enabling users to evaluate, reproduce, or update the findings (Higgins & Thompson, 2002). Our study follows the methodology schema presented in Fig. 1. The process includes conceptualization and research design, search strategy and data collection, data processing, content analysis, synthesis of results, and discussion.

2.1. Research design

A top-down strategy was used in the problem conceptualization to break down the study into keywords, associated terms, and synonyms, considering the different ways of spelling. This step is crucial for defining the research concepts and constructing the search query.

The research design followed the building blocks of theory development proposed by (Whetten, 1989), including the essential elements for a phenomena explanation: *what, how, why, when, who* and *where*. Individual components of the problem are specified in detail at this step, and composition takes place. *What* and *how* questions helped to bring out the keywords and associated terms related to ESi, SEO data (checked for Sentinel, and all missions of Landsat), and ML techniques. The *why* asks for the purpose of developing such models, *when* asks for the temporal scale of studies to be analyzed, *who* identifies the authors, and *where* flags the study's location. The building blocks for the research design and for building the search query are detailed in Table S1, in the Supplementary Materials.

2.2. Search strategy and data collection

An objective search strategy was implemented to identify the most relevant studies in the research scope. The Scopus scientific database (Elsevier B.V, 2023) was the source of the scientific literature and was last queried on 30/09/2022. The data collection process comprises four main phases.

Phase one, identification, starts with the application of the search query to identify studies. As the search terms must be logically connected, the search mechanism applied advanced techniques to the query-building process such as Boolean operators, quotes, stop words, wildcards, and nesting. The OR operator is used to aggregate similar concepts (synonyms or associated words), while the AND operator is applied to narrow the results. To search for specific phrases, the terms must be enclosed in double quotes ("") or, for an exact match inside braces ({}). In Scopus, the operator OR has precedence to AND, i.e., first, the Scopus searching algorithm processes the OR connector by looking for documents containing the specified words, and last processes the AND operator, by returning any documents it finds. Besides, the following assumptions were considered: 1) The stage of publication should be set to Final; 2) The types of documents are limited to articles published in a scientific peer-reviewed journal; 3) The publication must be written in English language; and 4) The search is within article title, abstract and keyword (TITLE-ABS-KEY). These assumptions are expressed by reserved words from the Scopus search, i.e., LIMIT-TO (limits the search to the assumptions), DOCTYPE (refers to the type of document that can be an article, a review, or a book), PUBSTAGE (publication stage that can be final or article in press), PUBYEAR (is the year of publication), and SRCTYPE (refers to the publishing source). The search strategy implemented in Scopus is detailed in Table 1. The query identified a total of 151 studies.

Phase 2, screening, consisted of analysing the abstracts of 151 records. In this phase, we assessed if the identified studies developed a model that is: 1. addressing one or more ESi; 2. using Sentinel-2 and/or Landsat imagery; and 3. applying ML techniques. In the end, 67 records were selected to be full-text analyzed. In phase 3, eligibility, the studies were excluded if their content did not fulfil requirements 1, 2, and 3, which led to the exclusion of 12 records. In the fourth phase, inclusion, a total of 55 articles were included in the systematic review process. A complete list of the selected studies is provided in the Supplementary Materials (Table S2).

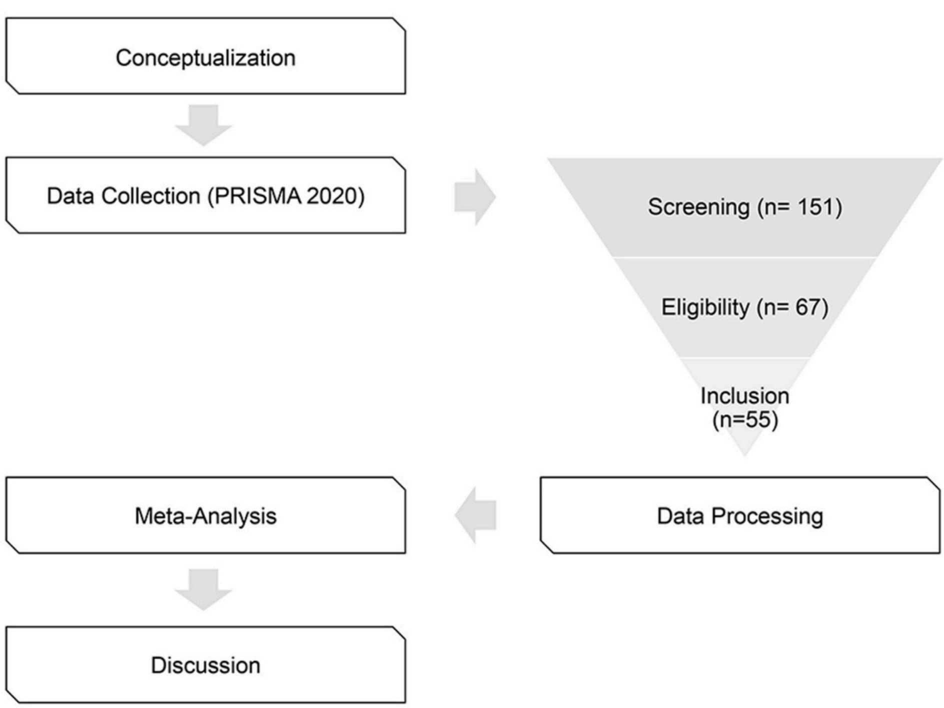

Fig. 1. Diagram flow of the meta-analytic reviewing process. "n" refers to the number of studies selected in each phase.

Table 1
Data Collection: Scopus searching query.

Scopus searching query
TITLE-ABS-KEY (({ecosystem service} OR {ecosystem services} OR (ecosystem-service*) OR (ecosystem* AND (demand* OR suppl* OR account* OR bundl* OR flow* OR value* OR function*)) OR {natural capital}) AND ((remote AND sens*) OR satellite* OR {earth observation} OR {big earth data} OR sentinel* OR sensor* OR {landsat} OR {imagery} OR {remotely sensed} OR {big-data}) AND (monitor* OR map* OR observ* OR detect* OR predic* OR classif*) AND (landscape* OR inland OR terrestr* OR natur* OR biophysi* OR socio-ecological) AND (metric* OR indic* OR index* OR prox* OR tradeoff* OR trade-off* OR footprint* OR parameter* OR variable*) AND ({spatiotemporal} OR {spatiotemporal} OR {spatiotemporal} OR {spacetime scales} OR {spatiotemporal scales} OR {spatiotemporal} OR spati* OR temp* OR time*) AND (*machine* OR (machine-learning) OR (artificial-intelligence) OR *learning*)) AND (PUBYEAR >2012) AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re")) AND (LIMIT-TO (LANGUAGE, "English"))

2.3. Content analysis and synthesis

The process of acquiring information began with content analysis and a summary of findings. Qualitative data were synthesized and processed for statistical analysis. The content examination included: 1) presenting where the selected studies were implemented and when they were published; 2) confirming the use of Sentinel-2 and Landsat missions within studies; 3) summarizing Biomes, ecosystems categories, and services; 4) identifying the ML algorithms (learners) employed, the tasks (classification or regression) carried out, and the evaluation metrics used to assess models uncertainties; 5) synthesizing modelling purposes, common limitations, and opportunities argued in the studies. To exemplify these key points and establish connections between them, we referred to specific examples from the selected literature.

Table 2 depicts current operational Landsat and Sentinel missions most applied to Earth monitoring services. NASA's Earth Observing System (NASA's EOS, 2022) comprises a coordinated constellation of satellites encompassing more than 20 missions. Notably, the Landsat satellites, operated jointly by NASA and the U.S. Geological Survey, represent the initial and most enduring terrestrial observation program focused on monitoring natural resources. Since 2009, all Landsat images have been made accessible to the public at no cost, contributing to open and widespread access to valuable SEO data (NASA, 2023).

Table 2
Most operational Landsat and Sentinel missions applied to Earth monitoring services.

Mission Name	Mission Agencies	Launch Year	End-Of-Life Year	Applications	Instruments
Landsat 7	USGS, NASA	1999	2024	Earth resources, land surface, environmental monitoring, agriculture and forestry, disaster monitoring and assessment, ice and snow cover.	ETM+
Landsat 8	USGS, NASA	2013	2028		OLI, TIRS
Landsat 9	USGS, NASA	2021	2031		OLI-2, TIRS-2
Sentinel-1 A; 2B	ESA	2014; 2016	2024; 2026	Monitoring of sea, ice zones, and the arctic environment, surveillance of marine environment, monitoring of land surface motion risks, and mapping in support of humanitarian aid in crises.	C-Band SAR
Sentinel-2 A; 2B	ESA	2015; 2017	2025; 2027	Land monitoring-related services: generation of land cover maps, risk mapping, and disaster relief, generation of leaf coverage, leaf chlorophyll content, and leaf water content.	MSI (Sentinel-2)

The Landsat 1 was launched in 1972 and was the first satellite of this program, which has been continuously improved over the last four decades, Landsat 9 was the last one launched (2021). Landsat 7 mission started in 1999 carrying the ETM+ (Enhanced Thematic Mapper Plus) sensor, which includes a thermal infrared band (TIR) with a spatial resolution of 60m. While Landsat 7 is still operational, it is in end-of-life activities. Landsat 8 was launched in 2013 and was equipped with two sensors: the OLI (Operational Land Imager) and the TIRS (Thermal Infrared Sensor). The OLI sensor produces images with 30m spatial resolution in the Visible (VIS), Near Infrared (NIR), and Short-Wave Infrared (SWIR) bands (in 8 distinct bands), and a panchromatic image with a resolution of 15m, the TIRS sensor records thermal images with a resolution of 100m. The enhancements introduced in Landsat 9 encompass higher radiometric resolution for OLI-2, enabling sensors to discern more nuanced differences, particularly in darker areas like water bodies or dense forests. Furthermore, TIRS-2 has substantially minimized stray light in comparison to its predecessor facilitating improved atmospheric correction and surface temperature measurements.

The European Space Agency (ESA, 2022) has successfully deployed over 70 Earth observation satellites into space. Among them, the Sentinel-1 and Sentinel-2 pairs are integral components of the Copernicus mission. This program is structured to encompass seven generations of Sentinel satellites equipped with specific technologies tailored for monitoring terrestrial, oceanic, and atmospheric processes. The Sentinel missions operate in both radar and multispectral wavebands, providing data for a range of Earth observation applications. Sentinel-1 and Sentinel-2A have been operational since 2014 and 2015 respectively, the first of which collects information through a Synthetic-aperture radar (SAR) sensor, and the second has a multi-spectral instrument (MSI) capturing optical information. Sentinel-2A/2B missions record information in 13 bands: VIS with 4 bands, NIR with 6 bands, and SWIR with 3 bands. The spatial resolution is 10, 20, and 60m, respectively, depending on the band, with a temporal resolution of 5 days. As many studies did not distinguish between Sentinel-2A and 2B data, this review grouped Sentinel-2A and 2B into Sentinel-2. The same happened for Sentinel-1.

In satellite-based applications, ML techniques are frequently employed for classification and regression problems (Ma et al., 2019). The targets are the response variables, also known as dependent variables; if the values are continuous, it is a regression problem that will fit the data as closely as possible to predict continuous values. Such as the prediction of tree cover density and leaf area index, as proxies for timber production included in the provision of raw materials service (Cilek et al., 2022; Mallinis et al., 2020), or the predictions of water and wetness probability index as a proxy for water provision and wetlands (Almeida & Cabral, 2023; Ludwig et al., 2019). If the values are discrete, it is a classification problem that will find the discriminant between discrete classes.

As many studies performed classification tasks without considering a recognized nomenclature to categorize ESi, we adopted the grouping system proposed by The Economics of Ecosystems and Biodiversity (TEEB) valuation database (McVittie & Hussain, 2013). TEEB provides a valuation framework to assist decision-makers in examining the impacts of international and national policies on biodiversity and ecosystems. Compared with other ES categorization systems, TEEB puts together habitat and supporting services to avoid the double-counting problem (Mustajoki et al., 2020).

ESi were categorized into four main functioning groups: (1) production functions, contributing to provisioning services; (2) regulation functions, involved in supplying regulating services; (3) habitat functions, essential for preserving ecological structures and processes, thereby offering supporting services; (4) cultural functions, delivering cultural and amenity services (Kienast et al., 2009).

ESi was employed, acknowledging their varied names under different categorization systems. Indicators were renamed and grouped as ESi based on the parameters or proxies mapped, not just the terms used in the original papers, as they varied considerably. We expand our typology to encompass biodiversity, recognizing its frequent treatment as an ES in the literature review. Since many records considered biodiversity in terms of genetic, functional, or habitat diversity, we follow this rationale, placing it in the habitat/supporting category. For instance, leaf coverage and leaf chlorophyll content act as indicators for climate regulating services (Egoh et al., 2012), the assessment of genetic resources serves as a measure of biodiversity (King et al., 2021), soil organic matter content functions as a proxy for soil formation and nutrient cycling (Vasenev et al., 2018), and water yield serves as an indicator for water provisioning service (Ludwig et al., 2019).

Meta-analysis uses statistical techniques to synthesize findings in a systematic review (Copas & Shi, 2000). Most meta-analysis methods are developed to measure variations across studies, improve research confidence, answer explicit questions, assess controversies in studies, and generate new hypotheses (Higgins et al., 2022). The log response ratio (logRR) (Hedges et al., 1999) was used to calculate the effect size of studies that developed ML models to assess ESi. Response ratio investigates weighted variance within groups, and the effect size estimates the change in the means in each ES (Durlak, 2009). The Q-test for heterogeneity (τ^2) (Cochran, 1954) and the I^2 statistic (Higgins & Thompson, 2002) were calculated to quantify variation across studies. Overall, this meta-analysis aimed to build evidence on the use of ML models by each group of ES. The main outcome is to evaluate the general applicability of employing hybrid models including classification and regression tasks instead of just one of them. The analysis was carried out using the R Studio software (R Core Team, 2022), through the "metafor" package version 4.0.0 (Viechtbauer, 2010).

3. Results

3.1. Location of studies

One of the reviewed studies had a global scope, while other studies covered 26 different countries within the selected literature. Studies locations were grouped into six categories according to the number of studies carried out in each country (Fig. 2). China and the United States of America (USA) gathered the highest number of studies with seven and ten studies, respectively. Bangladesh and India contributed with four and five studies, respectively. Brazil, Germany, Greece, and Spain had three selected studies, followed by other countries including, Italy, Kenya, Peru, Portugal, and South Africa, contributing two studies each. The remaining countries had one study each.

The global scale research of Sanderman et al. (2018) includes an ML model to map mangrove forest soil carbon at 30m spatial resolution. Six other studies were applied to more than one country as they were assessing transboundary ecosystems. All countries reported in the reviewed literature were taken into consideration when producing the above global map. Boutsoukis et al. (2019) developed an ML model to estimate the canopy in temperate forests of Germany and the Czech Republic; Guio Blanco et al. (2018) estimated soil water retention in Venezuela and Peru; Koskikala et al. (2020) mapped natural forest remnants in Tanzania and Kenya; Kundu et al. (2022) predicted induced wetland fragmentation and water richness in dams of India and Bangladesh; Mpakairi, Dube, Dondofema, and Dalu (2022a, 2022b) estimated vegetation heterogeneity within arid environments in Botswana and South Africa, and Sannigrahi, Joshi, et al. (2019) assessed coastal resources values in India and Bangladesh.

Among the selected studies, the earliest publishing dates go back to 2015. It relates to a work carried out in Alaska-USA proposing an ML model to estimate a spatial proxy for climate regulation services assessing carbon stored and sequestered in near-surface permafrost

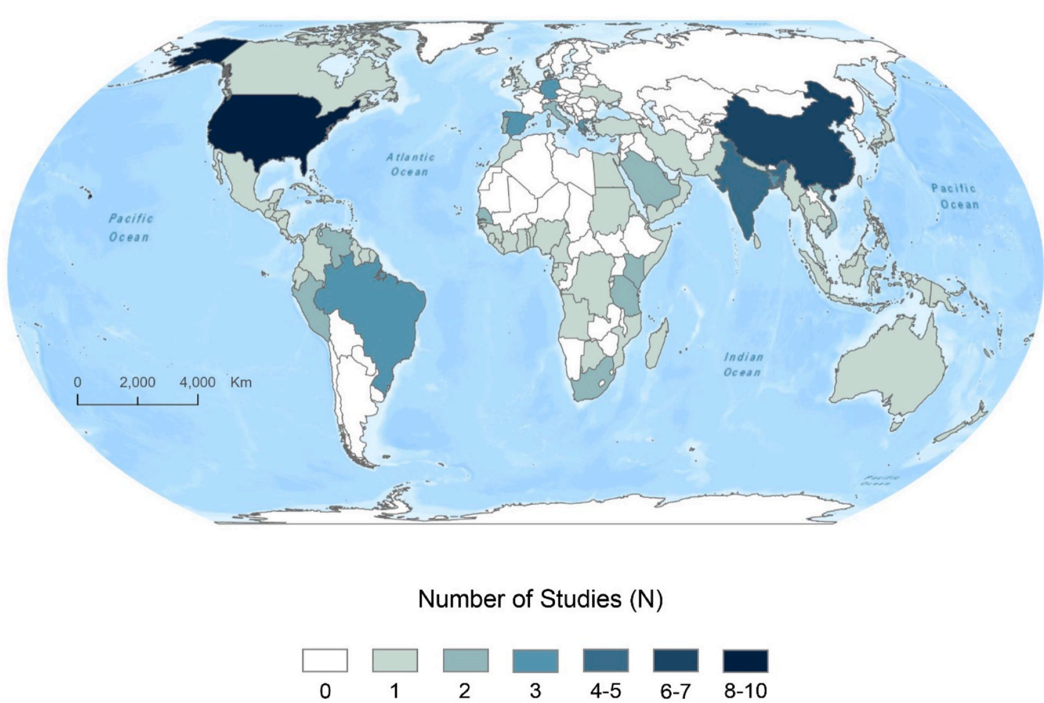

Fig. 2. Global representation of the studies included in the review.

(Pastick et al., 2015). The authors, used Landsat imagery as predictors in a decision tree model, advancing the application of multitemporal SEO data and benchmarking a transition between methods and technologies. For the years 2012, 2013, 2014, and 2016, none of the records fulfilled the proposed eligibility criteria to be included in this review.

3.2. Satellite Earth Observation data: Sentinel-2 and Landsat missions

The SEO data and technologies investigated in this research are publicly available and recognized as high spatial resolution sensors (5–100 m/pixel) (Liang & Wang, 2019) on board the Landsat (4, 5, 7 and 8) and Sentinel-2 (including 2A/2B) satellites. The following chart (Fig. 3) synthesizes the use of data from Landsat instruments and Sentinel-2, depicting also studies that used data from multiple sensors represented in the "Other" category.

Since 2017, at least one study has used data from Landsat 4 and 5, whereas in 2019 there were eight. Within studies, Landsat 8 was first used in 2017, and it was used 30 times until 2022. Sentinel-2 has been used 24 times since 2018. Implemented together, they are a powerful source of spatiotemporal information for ESi monitoring purposes (Vyvlečka & Pechanec, 2023). Mo et al. (2018) is the only study that used data from all Landsat sensors as well as the Sentinel-2. About 53% of reviewed studies used SEO data from more than one instrument, such as developing research combining data from MSI and OLI-TIRS sensors (Koskikala et al., 2020; Lobert et al., 2021; Mallinis et al., 2020; Pastick et al., 2018).

The "Other" category represents studies that included data from different satellites and instruments, such as combining Sentinel-2 and/or Landsat with radar imagery from Sentinel-1 (DeLancey et al., 2019; Hoffmann et al., 2022; Koskikala et al., 2020; Lobert et al., 2021) and other RS technologies such as manned and unmanned aerial vehicles (Hasan et al., 2021; Morell-Monzó et al., 2020; Sharma et al., 2018). DeLancey et al. (2019) highlighted that the combination of optical and radar data provides the highest potential for accurate and comprehensive landscape assessments through satellite-based modelling.

Many other studies combined drone imagery with data from other satellites such as Aqua/Terra MODIS (Moderate Resolution Imaging Spectroradiometer) (Arruda et al., 2021; Mo et al., 2018; Osborne & Alvares-Sanches, 2019; Pastick et al., 2018; Wang et al., 2017; J. Zhang, Du, et al., 2019), TripleSat (Cilek et al., 2022), IceSat-2 (Ice, Cloud and land Elevation Satellite) (Narine et al., 2019), SPOT-6 (Satellite pour l'Observation de la Terre) (Sharma et al., 2018), PlanetScope (Leroux et al., 2020), Rapid Eye (Hasan et al., 2021; Leroux et al., 2020; Mallinis et al., 2020), and WorldView-2 (Hasan et al., 2021; Mallinis et al., 2020). Additionally, spectral bands, derived indices, and a wide array of socio-biophysical variables (i.e., climatic, topographic, population density, road network information, census data, soil type, and geology) were considered as features by a few studies (Agrillo et al., 2021; Fitts et al., 2021; Hauser et al., 2021; Hudak et al., 2020; Kudzai S. Mpakairi et al., 2022a, 2022b; Mallinis et al., 2014; Mouta et al., 2021; Nzuza et al., 2021; Pipia et al., 2021; Pizarro et al., 2022; Vidal-Macua et al., 2020).

3.3. Which ESi are assessed using ML techniques?

A considerable number of the studies that performed classification tasks did not use any recognized system to classify ecosystems. Hence before the statistical analysis, we standardized all the classes following the TEEB database framework which recognizes 12 biomes, Marine/Open Ocean, Coastal systems, Wetlands, Rivers and Lakes, Forests, Woodland and shrubland, Grassland, Cultivated, Desert, Tundra, Ice, Rock and Polar, and Urban areas; and 22 ES grouped into 4 main categories: provisioning, regulating, habitat/support and cultural services. The most classified ecosystems were Forests (22), followed by Cultivated lands (13), Urban areas (12), Woodland and shrubland (10), Grass/Rangeland (9), Rivers and lakes, and Wetlands (8) (Fig. 4a).

Rivers, lakes, and Wetlands were mostly assessed regarding anthropogenic pressures affecting aquatic ecosystems. Kundu et al. (2022) estimated hydro-period, water depth, and water presence consistency, after classifying lakes, rivers and wetlands. Sannigrahi, Chakraborti, et al. (2019) developed a model to assess mangrove ecosystems, based on the LULC dynamics. Alqadhi et al. (2022) assessed the impact on ES valuation of future landslide events, from detecting changes in Grass/-Rangeland, Urban, Cultivated, Ice/Rock/Polar, Desert, and Rivers and Lakes; Mpakairi et al. (2022a, 2022b) implemented a model to classify the presence of woodland as a proxy for timber production. Gwal et al. (2020) classified natural forests to estimate biomass density; Han et al. (2022) estimated carbon storage through the spatiotemporal dynamics of vegetation, classifying land use into six types: Cultivated, Woodland and shrubland, Grass/Rangeland, Urban, Rivers and Lakes and Wetlands. Morell-Monzó et al. (2020) assessed crop yield for monitoring cultivated lands and maximizing food production and water use. The ES categories with more ML models developed were regulation (35%), followed by provision (29%), habitat/support (25%), and Cultural (12%) (Fig. 4b).

A total of 16 ESi were identified in the review (biological control, climate regulation, cultural heritage, flood regulation, food production, genetic resources, habitat, inland wetlands, natural hazard regulation, nutrient cycling, public health, raw materials, recreation, soil formation, water provision, water regulation). Within the regulation category, climate regulation is the most representative (32) service, followed by genetic resources (28) from the provisioning category, and from the

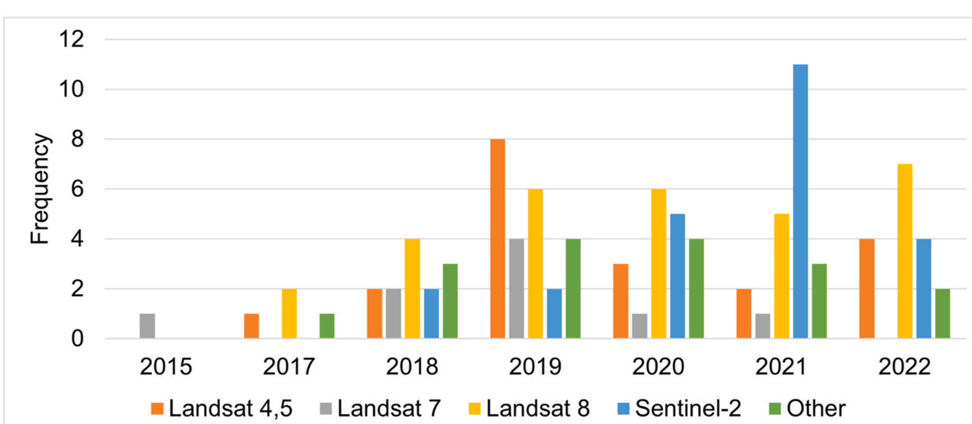

Fig. 3. Frequency of SEO data source identified in the literature review. The "Other" category refers to other satellites and instruments combined with any Landsat missions or Sentinel-2.

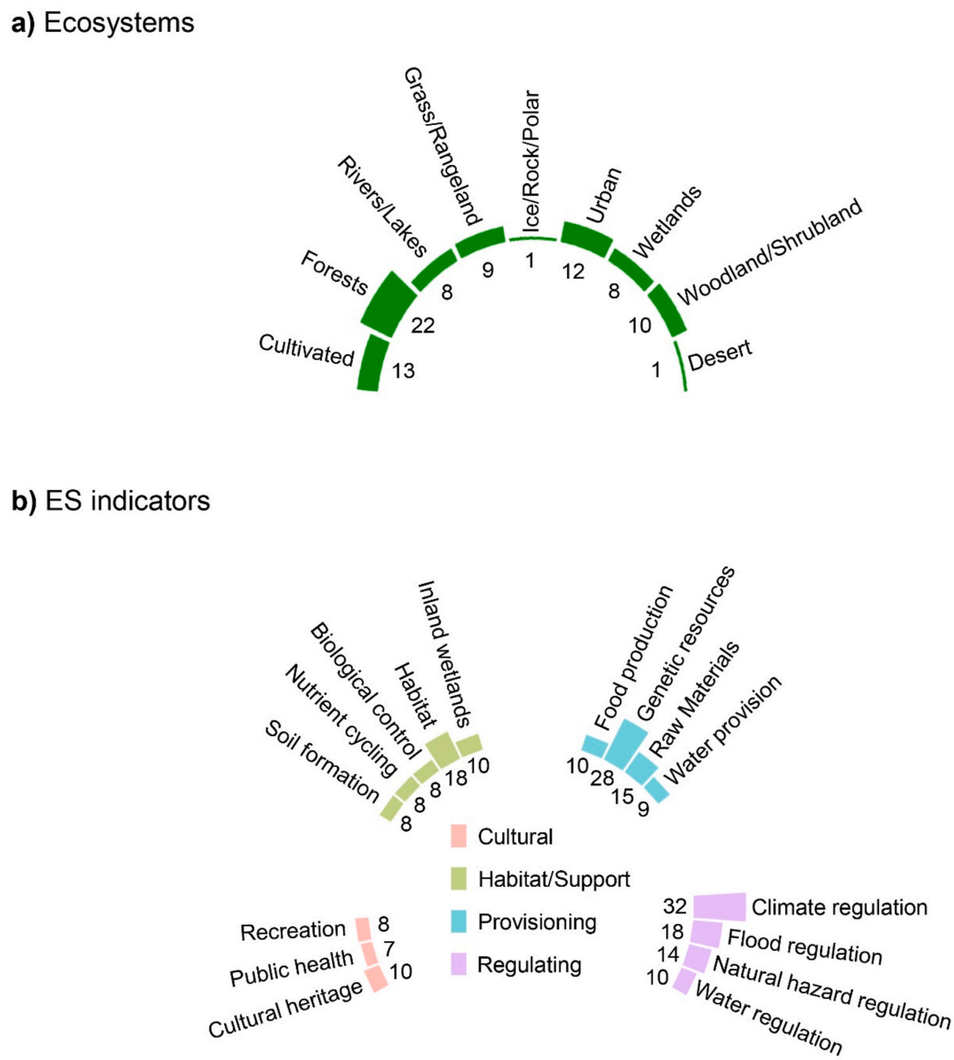

Fig. 4. a) Ecosystems and b) Ecosystems services indicators (ESi) modelled by the reviewed studies.

habitat/support category, the habitat service is the most modelled indicator (18 studies). These top three ESi were mostly identified in research that developed models to estimate carbon storage and sequestration, biomass density, water yield, and raw materials supply. The habitat service is responsible for the maintenance of genetic diversity, and its condition directly affects productivity within the service (McVittie & Hussain, 2013). Cultural ecosystem services (CES) are represented by social-environmental services that support human well-being (Karasov et al., 2020). Within all the studies, the public health service from the cultural category was the least assessed (7). The reviewed studies were organized into ES categories and are presented in Table S3 in the Supplementary Materials.

3.4. Which ML models are implemented?

Among the reviewed studies, 37 implemented classification models to discriminate or delineate ecosystems, 31 modelled ESi through regression, and 12 built both models. All ML models were implemented based on supervised learning (SL), in which the training data are labelled, allowing the improvement of the learner's performance directly. In unsupervised learning (UL) all data are unlabelled, and the machine learns to find structure and new features in the data. UL has the advantage that does not require training samples, but its accuracy can be lower when compared to SL, which relies on training samples to establish the relationships between inputs and results (Shao & Lunetta, 2012). Fig. 5a) presents ML algorithms by task (classification and regression) employed more than once in each study.

Six studies built two different classification models (Alqadhi et al., 2022b; DeLancey et al., 2019; Leroux et al., 2020; Matsala et al., 2020; Tarantino et al., 2021; Yang et al., 2021). The most employed algorithms were RF, applied in approximately 50% of the studies, followed by SVM (25%), Gradient Boosting (GB) (20%), Classification and Regression Trees (CART), and Maximum Likelihood (MLC) (more than 10%). Seven studies compared the performance between two to five algorithms (Alqadhi et al., 2022b; Boutsoukis et al., 2019; Fitts et al., 2021; Hunter et al., 2020; Mo et al., 2018; X. Wang, Zhang, & Peng, 2021; Wang et al., 2017). Five other studies tested more than five learners to compare performance between models (Mouta et al., 2021; Pizarro et al., 2022; S. Sannigrahi, Chakraborti, et al., 2019; L. , 2019b, Zhang, Sharma, et al., 2022).

Sannigrahi, Chakraborti, et al. (2019) and Sannigrahi, Joshi, et al. (2019) tested 10 ML algorithms to assess ES values and provisioning status based on the classification of ecosystem units. The authors concluded that SVM and RF produced the best classification predictions to assess 9 ESi including spatial proxies of climate regulation, water regulation, soil formation, nutrient cycling, biological control, food production, raw materials, recreation, and cultural services.

Hoffmann et al. (2022) demonstrated in their research the capability of DL to assess the status and functions of habitat and raw materials provisioning services of forest ecosystems. Tarantino et al. (2021)

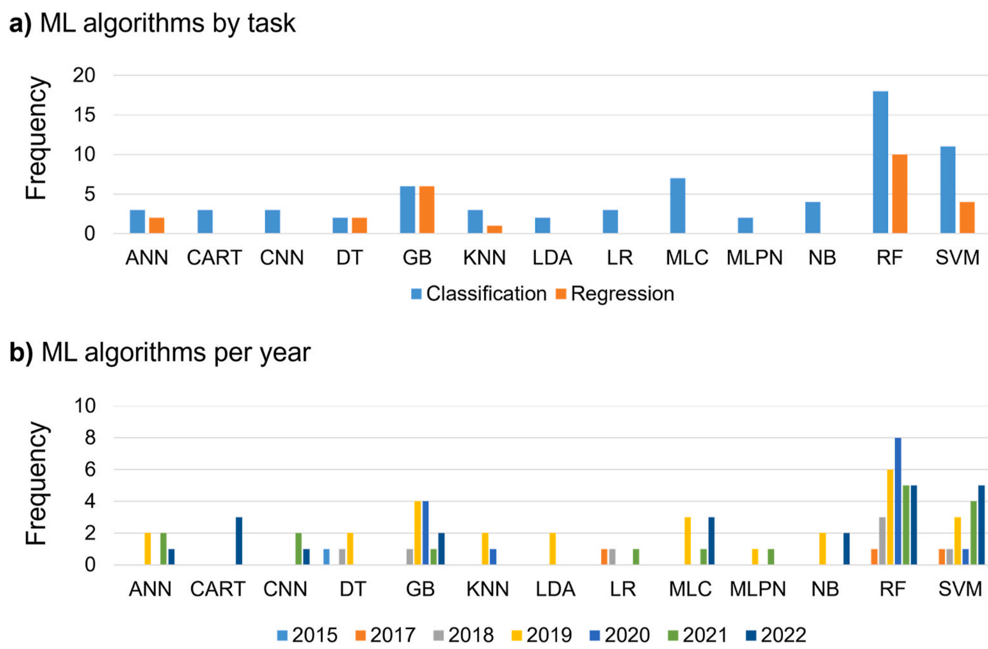

Fig. 5. a) ML algorithms used for modelling ESi. b) Evolution in the use of ML algorithms through publication years. Legend: ANN - Artificial Neural Networks; CART - Classification and Regression Trees; CNN - Convolutional Neural Networks; DT - Decision Trees; GB - Gradient Boosting; KNN - k-Nearest Neighbours; LDA - Linear Discriminant Analysis; LR - Logistic Regression; MLC - Maximum Likelihood Classifier; MLPN - Multi-Layer Perceptrons; NB - Naive Bayes; RF - Random Forest; SVM - Support Vector Machines.

argued that compared to DL algorithms, the SVM classifier was preferred since DL requires larger datasets and more computational resources. Even if it demands higher computational resources, these costs are being overcome with computing power improvements, efficiently allocating computing resources, and cloud computing (Mohanty et al., 2020). The authors used proprietary and/or open-source software to build their models and to compute performance comparisons between algorithms and models. Within records, two commercial software were used in ML modelling: See5/C5.0 was used in the classification task and the Cubist (RuleQuest Research, 2022) was used in the regression model. Cilek et al. (2022) classified forest ecosystems and predicted tree cover density as a proxy to estimate carbon storage in coniferous and broadleaf trees, using both software. Google Earth Engine (GEE) (Google, 2023), Weka (Waikato Environment for Knowledge Analysis) (Hall et al., 2008) and biomod2 (Ensemble Platform for Species Distribution Modeling) (Thuiller et al., 2022) were the open-source programs, used to support models' implementation and performance evaluation.

Mouta et al. (2021) employed biomod2 to perform a classification task fusion through eight ML algorithms mapping species' abundance as a proxy of genetic resources service. Boutsoukis et al. (2019) used Weka for modelling habitat service based on estimations of temperate forest canopy height. Around 60% of the studies published after 2019, used the GEE platform as a source of satellite data and/or for imagery and geospatial processing. Open-source platforms and libraries including DL algorithms were also employed by a few studies, such as TensorFlow (Martín Abadi et al., 2015), Keras (Chollet et al., 2015) and scikit-learn (Pedregosa et al., 2011).

Fig. 5b) depicts the use of ML algorithms through the publication years. Regarding the evolution in ML modelling applications, DT was the first algorithm identified in the literature (2015). In 2017, LR, RF and SVM were the algorithms identified; in 2018 GB was first observed, and RF increased their occurrence; in 2019 were identified modelling applications for all referred algorithms, except for CART, CNN, and LR. This year is coincident with the increased use of GEE for environmental modelling (Arruda et al., 2021; DeLancey et al., 2019; Fitts et al., 2021; Hoffmann et al., 2022; Hunter et al., 2020; Koskikala et al., 2020; Mallinis et al., 2020; Kudzai S. Mpakairi et al., 2022a, 2022b; Nawrocki et al., 2020; Pipia et al., 2021; Pizarro et al., 2022; Wall et al., 2021; J. Zhang, Du, et al., 2019; L. Zhang, Sharma, et al., 2022). Since 2019, there has been an increasing number of publications that performed algorithm comparisons, and hyperparameter tuning as computation power developed, and the use of cloud computing and other platforms and software became more attractive and spread (de Brito et al., 2021). Table S4 in the Supplementary Materials lists all ML algorithms employed by the examined papers.

Hyperparameter tuning was neglected or not reported in 83% of the studies. Selecting appropriate parameters is crucial as improper choices may lead to overfitting or underfitting issues. Table 3 summarizes the literature within reviewed studies that have carried out and reported hyperparameters tuning. L. Zhang, Hu, and Tang (2022) conducted hyperparameter tuning to identify the most effective model by comparing validation results across different models with varied hyperparameters. In RF models, the optimal model was determined by testing different numbers of decision trees (n_estimators). SVM models underwent tuning with various kernel settings and other hyperparameters to find the optimal SVM model for classification. NB models were compared based on different lambda values. GB models varied in the number of trees for hyperparameter optimization.

Ha et al. (2021) developed an ML model utilizing the CatBoost algorithm, which requires setting up a few hyperparameters, such as depth

Table 3
ML algorithm hyperparameters tuning.

ML task	ML algorithm and hyperparameter tuning method	Literature
Classification	RF (n_estimators), SVM (kernel), NB (lambda), GB (n_estimators)	L. Zhang, Hu, and Tang (2022)
Classification	CGB (learning_rate, depth, l2_leaf_reg, iterations)	Ha et al. (2021)
Regression	KNN (k), GB (n.trees, interaction.depth, shrinkage, bag.fraction, n.minobsinnode)	Matsala et al. (2020)
Regression	GB (n.trees, interaction.depth, shrinkage, bag.fraction, n.minobsinnode)	Leroux et al. (2020)
Regression	SVM (c, kernel, gamma)	X. Wang, Zhang, and Peng (2021)

of trees (depth), number of trees (iterations), learning rate (learning_rate), and L2 leaf regression (l2_leaf_reg) identified as the four most crucial ones. The optimization process employed the GridSearchCV library with five-fold cross-validation. Matsala et al. (2020) and Leroux et al. (2020) tuned GB hyperparameters through a grid search, exploring potential combinations of parameters: shrinkage, interaction depth, number of trees (n.trees), number of observations in the terminal nodes, and bagging fraction. X. Wang, Zhang, and Peng (2021) performed a regression task utilizing SVM. The process involved selecting an appropriate kernel function and determining the (c) cost and gamma (g) parameters.

Spatial analysis and visualization were implemented in Quantum GIS (QGIS) (QGIS Association, 2022), System for Automated Geoscientific Analyses (SAGA), ArcGIS (ESRI, 2023), GEODA1.6.7 (Anselin & et al., 2016), FRAGSTATS (McGarigal & Marks, 1995), and an R package, named "spdep: Spatial Dependence: Weighting Schemes, Statistics", (Bivand & Wong, 2018). Ten studies employed spatial techniques to evaluate spatial autocorrelation and neighbouring dependency (Han et al., 2022; Hauser et al., 2021; Hoffmann et al., 2022; Mallinis et al., 2020; Mpakairi et al., 2022a, 2022b; Nzuza et al., 2021; Osborne & Alvares-Sanches, 2019; Pizarro et al., 2022; S. Sannigrahi, Joshi, et al., 2019; Y. Zhang, Sharma, et al., 2022). Commonly utilized statistics methods were Spearman and Pearson's Correlations, Lee's L statistic, the Global Moran's I, and the Local Moran's I indices. These methods were applied to identify spatial autocorrelations, and correlated variables, reduce instability during modelling, mitigate the impact of noisy predictors and redundant data (Nzuza et al., 2021), enhance the efficiency of predictions, and improve the interpretability of final models (Hoffmann et al., 2022).

Applying metrics that do not align with the goals or characteristics of the problem may result in inaccurate evaluations of model performance (Astola et al., 2019). To validate the predictions' power of the classification tasks, the following metrics were commonly employed: Producer Accuracy (PA) (a total of 14 studies), User Accuracy (UA) (14), Overall Accuracy (OA) (13), and Kappa coefficient (K) (12), as shown in Fig. 6a). The ROC-AUC (Receiver Operating Characteristic - Area Under the Curve), Precision, and F1-score were used to assess prediction performance also between models in five studies. For imbalanced classification problems, the F1-score was reported to be the best option to measure model quality. Olofsson et al. (2013) do not recommend using just K to assess the accuracy of land use change maps, and Koskikala et al. (2020) evaluated their results with the True Skill Statistic (TSS) arguing that TSS is independent of prevalence, and superior reliability compared to K. The top four classification metrics were UA, PA, OA, and K.

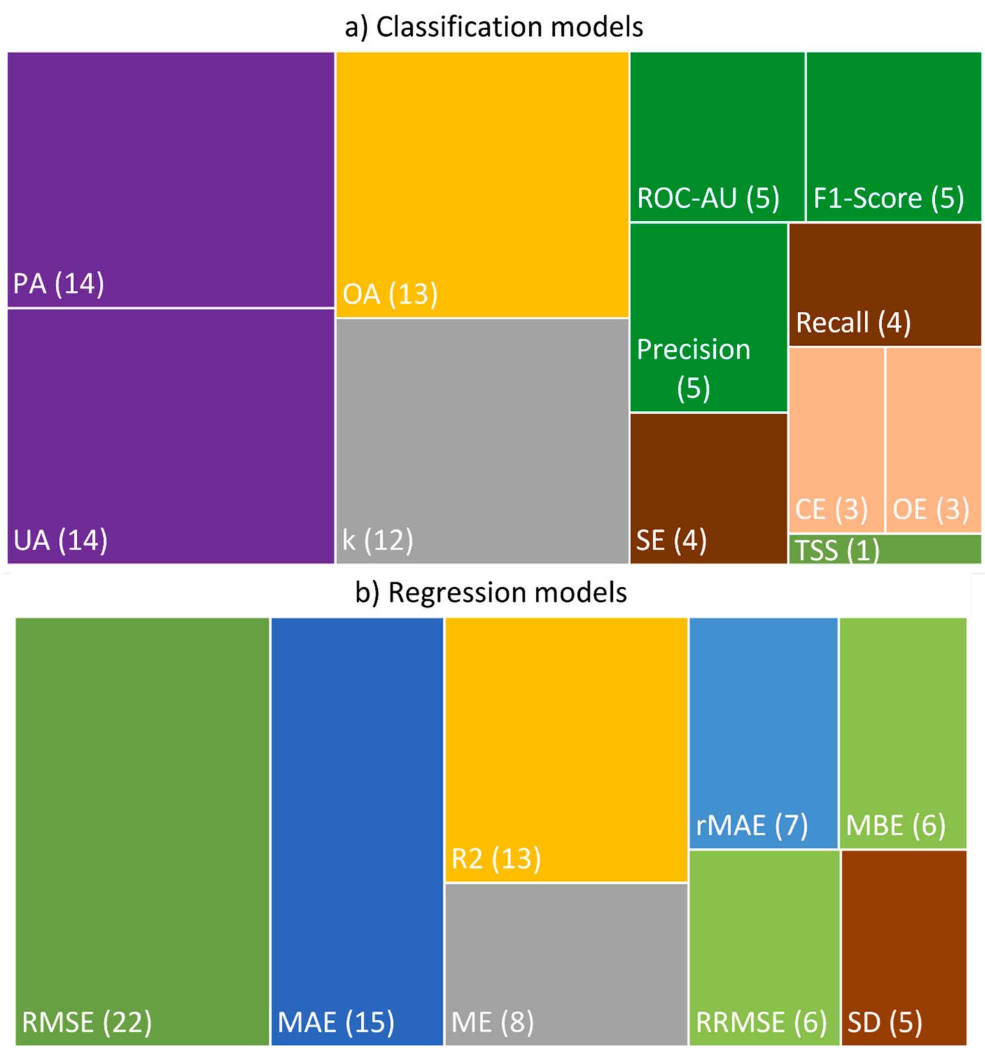

Fig. 6. Frequency of evaluation metrics used to assess models' quality. a) Legend (classification models): CE-Commission Error, OE-Omission Error, K-Kappa coefficient, OA-Overall Accuracy, PA-Producer Accuracy, TSS-True Skill Statistics, ROC-AUC- Receiver Operating Characteristic - Area Under the Curve, UA-User accuracy, SE-Sensitivity. b) Legend (regression models): MAE-Mean Absolute Error, MBE-Mean Bias Error, ME-Mean Error, R2 -Coefficient of determination, rMAE-relative Mean Absolute Error, RMSE-Root Mean Square Error, RRMSE-Relative RMSE, SD-Standard Deviation. The number in parenthesis refers to the number of studies.

Around 75% of the studies that performed regression tasks used Root Mean Square Error (RMSE) as the primary evaluation metric for assessing the accuracy of predictions (Fig. 6b). The assessment of the agreement between actual and predicted values frequently employed RMSE (24), Mean Absolute Error (MAE) (10), Coefficient of determination (R2) (6), and Relative RMSE (RRMSE) (5). Wang et al. (2017), Mouta et al. (2021), and Hauser et al. (2021) combined RMSE and R2 metrics with Pearson's R and Spearman Correlation to measure correlations between predicted values and observed values. Hauser et al. (2021), and Hoffmann et al. (2022) added Moran's I index to measure the spatial dependence of the residuals' neighbouring points, and Zhu et al. (2022), Pastick et al. (2015), Leroux et al. (2020), and Hudak et al. (2020) measured average model bias through MBE and RMBE.

3.5. What is the relationship between ML tasks and ESi?

The plot in Fig. 7 relates ESi with ML tasks, highlighting their correspondence. The classification was the most performed task within ES categories, habitat/supporting (21), regulating (19), provisioning (13), and cultural (4), against seven, nine, five, and one, respectively, for regression tasks.

Among classification tasks, 44% of studies were multiclass problems and 28% were binary classifications employed to create presence/absence maps. Binary classifications were mostly implemented in studies assessing habitat (10) and climate regulation (4). In a multiclass scenario where many values are discriminated into many classes, climate regulation (7), genetic resources (5), and habitat (4) were the most modelled indicators within studies. The classification tasks performed for taxonomic mapping or habitat discrimination were done in two steps. First, a map delineating the ecosystem was created, and then its output was used as a mask to discriminate species and habitats. Tarantino et al. (2021) first extracted the grassland layer by performing binary classification and then used this as an input for the second step, which was a multiclass problem to classify four grassland habitats. Matsala et al. (2020) first predicted land cover classes to produce a forest cover mask and then used the predicted map to classify dominant tree species.

Predicting continuous values through regression tasks was mostly performed by climate regulation (7), genetic resources (5), and habitat (4) indicators. Nearly 55% of studies that performed a regression task predicted carbon storage in forests, woodlands, grasslands, croplands, and tundra ecosystems, as a proxy of climate regulation services. Other representative dependent variables were tree cover density, leaf area index, and leaf mass area, estimated as spatial proxies for timber production within raw materials provisioning services, as well as modelling crop production as a proxy for food provisioning, and the prediction of dissolved inorganic nitrogen (DIN) and reactive phosphate (PO_4-P) levels, to assess nutrient cycling, water quality regulation, and water provision services.

3.6. Is there a rule for model development within ESi?

Approximately 24% of the studies built their models by applying both, classification and regression tasks (Table S5 – Supplementary Materials). Classification tasks were mostly employed to create land cover maps or discriminate ecosystems as the first step of nearly a quarter of the developed ML models. These maps sometimes combined with other variables were used as input for further regression models.

Meta-analysis was carried out by grouping reviewed studies within the 16 ESi identified. We found that 69% of estimates had negative log response ratios (log [RR]) varying from −0.92 to 0.41. With a 95% confidence interval (CI) (−0.80 to −0.05), the estimated average log [RR] based on the random-effects model was $\mu = -0.42$. As a result, the average outcome differed significantly from zero ($z = -2.2194$, $p = 0.0265$), and the Q-test supports the hypothesis that the outcomes were not statistically heterogeneous ($Q = 2.0623$, $p = 1.0000$, $\tau^2 = 0.0000$, $I^2 = 0.0000$ %). A forest plot with the estimated results grouped by indicators and based on the random-effects model is depicted in Fig. 8.

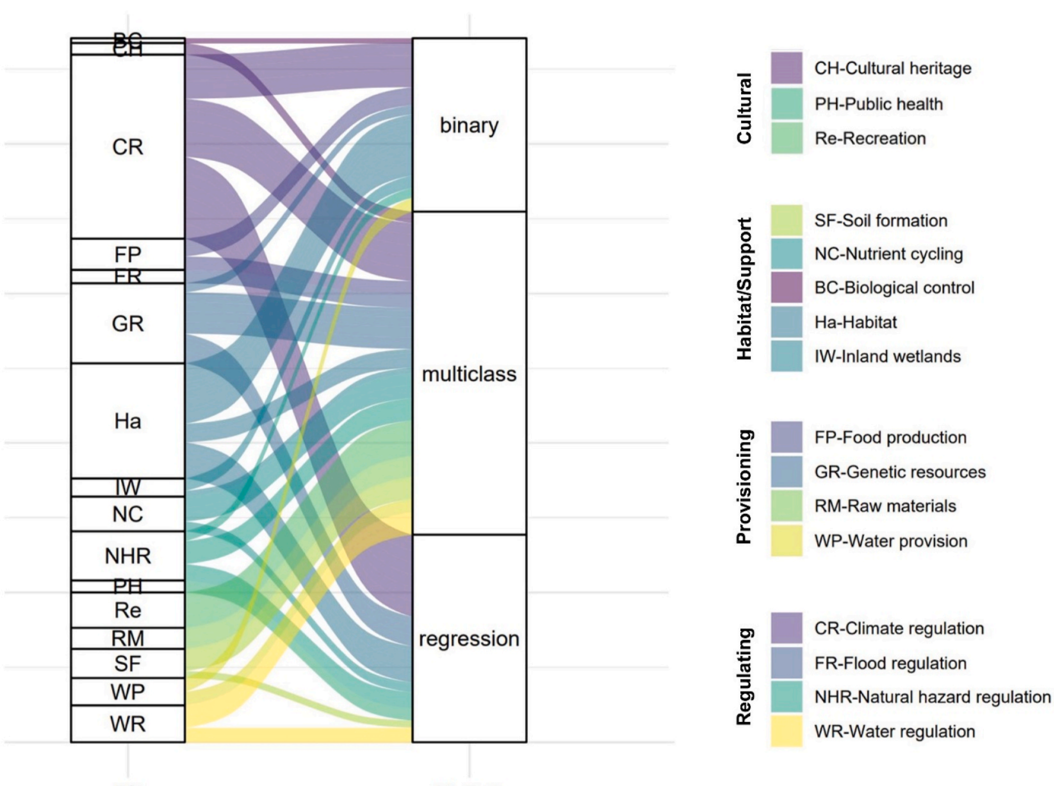

Fig. 7. Correspondence plot relating ML tasks within ESi.

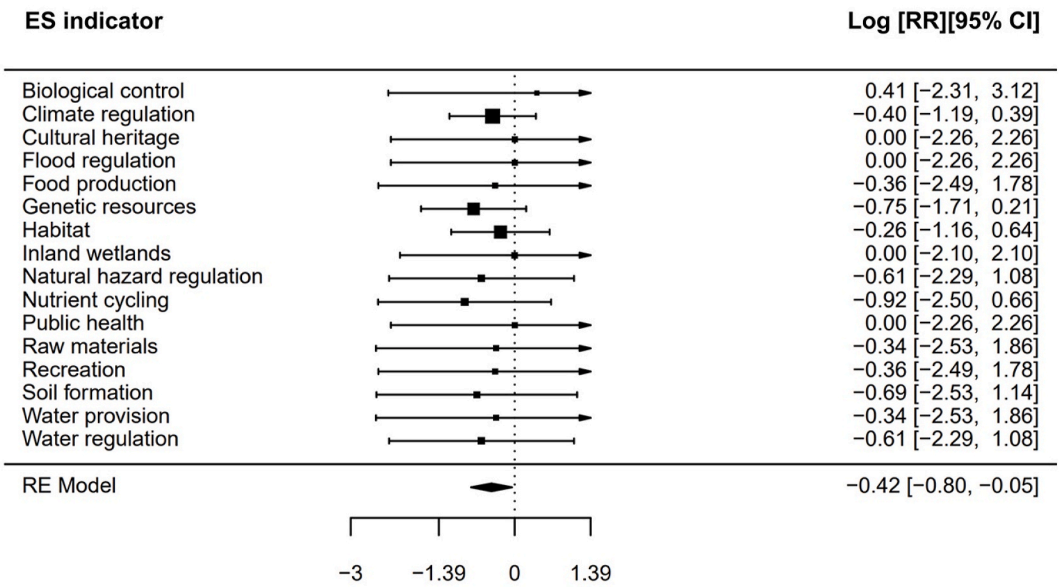

Fig. 8. Forest plot with the log [RR] estimates using 95% of CI.

Results indicate that, on average, there is a 65% likelihood of utilizing classification and regression for assessing ESi. Biological control, cultural heritage, flood regulation, food production, inland wetlands, public health, raw materials, recreation, and water provision were the indicators likely not to employ a classification as a first step in the models. On the other hand, studies assessing climate regulation, genetic resources, habitat, natural hazard regulation, nutrient cycling, soil formation, and water regulation were all investigated employing both tasks.

Analyzing the percentage of weights given to each indicator during the model fitting evidence indicators influence (Fig. 9a). Climate regulation (1), genetic resources (6), and habitat (7) had a relatively high weight compared to the other indicators, being influential indicators, but not outliers. The radial graph evaluates the consistency of the results considering standard errors and variances of samples (Fig. 9b). The value of the log [RR] is represented by the line projected from (0,0) through the corresponding value in the arc (−0.42). The graph indicates good consistency of results clustering the three influential indicators (climate regulation, genetic resources, and habitat).

4. Discussion

In this comprehensive literature review, we investigated the evolution of ML models for evaluating ESi using satellite data from Sentinel-2 and Landsat missions. The review synthesized the prevalent algorithms, tasks, and evaluation metrics employed in satellite-based ML modelling of ESi. It provided a benchmark for the conceptualization, implementation, and development of models across diverse indicators, while also shedding light on potential directions for future research.

4.1. Summary of results

Modelling spatial proxies of ES based on high spatial and spectral resolution satellites enables a more consistent characterization of ESi components that are essential to sustainably managing natural resources (Guirado et al., 2019). It facilitates rapid, frequent, and continuous observations, thereby enhancing monitoring capabilities (del Río-Mena et al., 2023).

High spatial resolution sensors capture more detailed and accurate information about an object when compared to medium spatial resolution satellites (Liang & Wang, 2019). Multispectral sensors can detect small differences in spectral signatures, as they have contiguous bands with small bandwidths (<20 nm), consequently, producing more accurate classifications (Hunter et al., 2020). Nevertheless, the utility of optical sensors is constrained by limitations such as daytime image acquisition and vulnerability to cloud cover, atmospheric haze, or dense vegetation canopies (X. Wang, Zhang, & Peng, 2021).

Data fusion integrating observations from different instruments

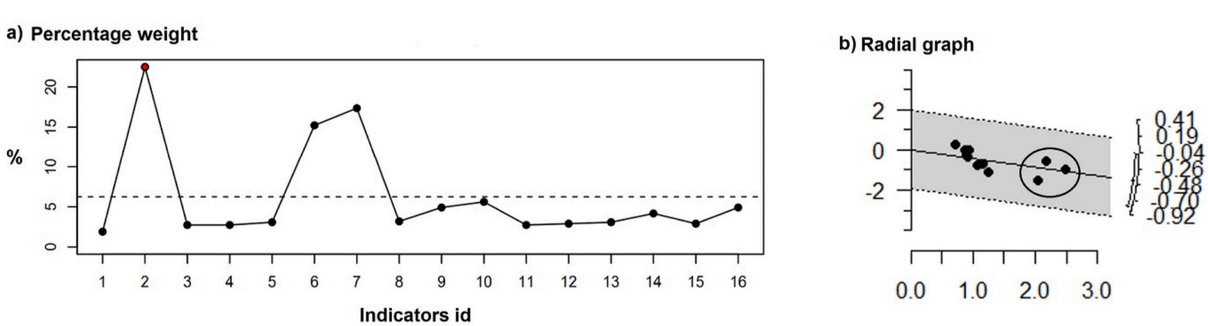

Fig. 9. a) Percentage weights plot for the 16 ESi examining the use of classification tasks in regression models. b) Radial graph. Legend Indicators id: 1-Biological control, 2-Climate regulation, 3-Cultural heritage, 4-Flood regulation, 5-Food production, 6-Genetic resources, 7-Habitat, 8-Inland wetlands, 9- Natural hazard regulation, 10-Nutrient cycling, 11-Public health, 12-Raw materials, 13- Recreation, 14-Soil formation, 15-Water provision, 16-Water regulation. Radial graph of the model fitted.

provides advanced knowledge of the relationship between landscape dynamics, ESi, and human activities (C. Ramirez-Reyes et al., 2019). The approach is becoming more frequent in ES-related research and was employed in 53% of the reviewed studies, including integration of Sentinel-2 and/or any Landsat missions with Sentinel-1, Rapid Eye, MODIS, World View-2, IceSat-2, Planet Scope, SPOT-6, and manned and unmanned aerial vehicles. This observation agreed with Schirpke et al. (2023), emphasizing the increased adoption of the data fusion approach, as well as the inclusion of a wide variety of ancillary data, such as climatic and topographic variables, population density, census data, road networks, number of trees, soil type, and geology. Vyvlečka and Pechanec (2023) demonstrated in their review a preference for Sentinel-2 over Landsat 8, with spatial and spectral resolutions being a crucial factor affecting accuracy. Our results are also aligned with these statements, and some reviewed papers clearly state that (Koskikala et al., 2020; Lobert et al., 2021; Pastick et al., 2018).

A total of 16 ESi were identified (i.e., biological control, climate regulation, cultural heritage, flood regulation, food production, genetic resources, habitat, inland wetlands, natural hazard regulation, nutrient cycling, public health, raw materials, recreation, soil formation, water provision, and water regulation). These indicators derived from satellite-based ML modelling, enable the assessment of ecosystem structure, configuration, and patterns of changes that influence nature's ability to provide goods and services (Olander et al., 2018).

The identified ESi aid in informed decision-making through extrapolating measurements of ecosystems' extent and conditions. These outcomes provide valuable support for landscape planning and management through physical measurements, enabling tracking changes in ecosystem assets, species distribution, and ecosystem health (King et al., 2024). When integrated into socio-economic data, ESi can contribute to the design of ecosystem accounting, providing information on natural stocks and aiding in the development of national accounts and public policies, enabling better-informed and sustainable management practices (Fleming et al., 2022). Nevertheless, ecosystem monetary accounting and monitoring still require direct measurements, and ESi would not be enough.

Within the reviewed studies, only 10% evaluated monetary accounts of ES calculating the benefits generated by each service or category of services (Alqadhi et al., 2022b; Han et al., 2022; Kundu et al., 2022; Mugiraneza et al., 2019; S. Sannigrahi, Chakraborti, et al., 2019; Sannigrahi, Joshi, et al., 2019).

Approximately, 80% of studies assessed 20% of the ESi identified. The ESi categories with the most frequently developed ML models and applications were regulating (with climate regulation), provisioning (with the provision of genetic resources and raw materials), and habitat (with habitat service providing maintenance of genetic diversity). The condition of the habitat providing the service directly affects the availability of many other ecosystem functions and services (McVittie & Hussain, 2013). Natural habitats are crucial to maintaining genetic resources, and biodiversity (Reddy, 2021). Mapping tree-species distribution, canopy height, and tree cover density allowed for measuring the ability functions of climate regulating services such as carbon storage and sequestration, and raw materials provisioning, such as timber, biomass, and other tree-derived products (Hudak et al., 2020).

ML serves as a powerful complement to current methods rather than a substitute (Scowen et al., 2021). Assessing ESi using ML models is a good decision when one follows a set of good practices and has well conceptualized the study's purposes (Willcock et al., 2018). When one chooses an algorithm just because it is trending or easier to implement, it is quite dangerous, because it can be effective in solving certain problems, but not for a specific problem and/or dataset (Craven et al., 2023). A good practice is to test performances between models and learners and monitor the results through a wide range of evaluation metrics (Domingos, 2012).

ML tasks are demanding in terms of computing, particularly when modelling big earth data (Ma et al., 2019). Nowadays, there is a wide variety of algorithms, and most of them do not have many requirements to implement, unlikely most of them are highly sensitive to the quality of the training data, as well as to hyperparameter configurations (Maxwell et al., 2018). Inevitably, before building ML models, it is essential to understand the algorithms' drawbacks and limitations, such as lack of transparency, hyperparameter sensitivity, generalizability through overfitting and underfitting, power computing needed and time-consuming (Scowen et al., 2021).

Biases, inaccuracies, or missing data can significantly impact model performance, as ML models require sufficiently large, representative, and diverse data to capture the underlying patterns, and generalize well to new, unseen instances (Bishop, 2006). The low accuracy of models often stems from inadequate training data, regarding insufficient quantity and quality, and imbalances in the distribution of classes or features (Hastie et al., 2009). To improve the robustness and reliability of ML models, data preprocessing, cleaning, augmentation, and oversampling techniques lead to better accuracy and more trustworthy results (Craven et al., 2023).

A lack of transparency can be a significant barrier to understanding how the model arrives at its predictions and contributes to the *black box* problem, making it challenging to decipher the reasoning behind the results (Schirpke et al., 2023). Nonetheless, learners can be transparent and repeatable, and algorithm development is calling for advances that better link inputs, outputs, and intermediary steps (Nikparvar & Thill, 2021).

RF was the most implemented algorithm, followed by SVM. The advantages of using RF instead of CART are the robustness to overfitting and the predictors' value range aggregation (Breiman, 1984). The number of hyperparameters required by the RF classifier is less than the number required for SVM and other algorithms (Boser et al., 1992). In contrast to many algorithms, including SVM, RF might face challenges when dealing with imbalanced data. On the other hand, SVM tends to perform better under such circumstances, thanks to its generalization capability, overall robustness, high accuracy, and suitability for situations involving limited training data (Pizarro et al., 2022). Choosing an algorithm that fulfils the requirements of each model is quite challenging. Making use of tools that compare algorithms and models is a better option, as they also commonly offer various elements of data handling, processing, and performance evaluation.

The primary result of the meta-analysis was to assess the overall feasibility of using hybrid models, which involve completing both classification and regression tasks. Measuring the effects of implementing hybrid models instead of a single application provides a better understanding of researchers' decisions and the applicability of combining tasks in the assessment of ESi.

Classification tasks were mostly employed to identify and delineate ecosystem assets, such as forests, and estimate their extent, while regression tasks were modelled to evaluate ecosystem conditions by predicting specific characteristics (e.g. tree cover density). Besides, some studies implemented a two-step modelling strategy, integrating both classification and regression tasks. This synergistic method allows for a comprehensive and optimized use and analysis of satellite-based ML techniques to assess ESi, as well as contributes to more robust and accurate modelling outcomes in ES assessments.

4.2. Common challenges and opportunities

Many studies highlighted opportunities for leveraging SEO in ESi assessments and stressed the need for addressing challenges through innovative funding and partnerships to enhance global ES assessments (Lobert et al., 2021). Challenges related to variable and model accuracy, data accessibility, technology availability, and ethical considerations were pointed out as topics that require attention to ensure inclusive and meaningful technology use (Narine et al., 2019). Few records suggest that integrating different technologies, fostering transdisciplinary collaboration, and exploring advancements in related research fields could provide further insights into ES research (Alqadhi et al., 2022b; Kundu et al., 2022; Mpakairi et al., 2022a, 2022b; Vidal-Macua et al., 2020).

Limitations have been overcome through methodological and technical advancements in cloud computing, efficient resource allocation, algorithm improvements, and quantum information technologies (European Spatial Agency, 2023; Willcock et al., 2023). These developments have been fast-evolving and fostered by institutions and agencies, such as NASA, USGS, and ESA (Vyvlečka & Pechanec, 2023), as well as, satellite data streams, e.g. GEE and the Open Data Cube (ODC, 2018), facilitating access, management, and analysis of SEO data. Coupled with advances in open-source data science algorithms and packages, cloud-based technologies are facilitating the modelling workflow for a broad spectrum of ecosystem functions and services (DeLancey et al., 2019). Furthermore, fostering the applicability of such models and technologies in landscape planning, governance, and policymaking (King et al., 2024).

The rapid technological evolution in RS is calling for the development of approaches and techniques that can be used on SEO-based monitoring strategies of ESi. However, fieldwork remains crucial for some ES assessments, as RS is not a substitute for field surveys, ground-based sensors and sensor networks, and measurements derived from SEO may need to be validated through field observations (Jullian et al., 2021). The constraints imposed by the available ground-based datasets were one of the main limitations reported in the studies. Matsala et al. (2020) stated the need for a national statistical inventory in Ukraine, as a barrier met in implementing their research. M. Zhang, Okin, and Zhou (2019) pointed out that the temporal misalignment between SEO and field survey data caused by the absence of cloud-free Landsat scenes that precisely match ground observation times introduces errors in result estimations. Sharma et al. (2018) emphasized the crucial role of extensive ground data in calibrating and validating classification outputs and the need for a robust methodology capable of accurately quantifying crop production at the field scale. Therefore, integrating ground sensor data helps validate the quantification of ecosystem functions, complementing the central role of satellite data (Vyvlečka & Pechanec, 2023).

Achieving accurate predictions of ESi from remotely sensed data requires a profound understanding of the intricate relationship between spectral information and each proxy (Ørka et al., 2022). This task is particularly challenging due to the unobservability, or measurement difficulties associated with many ES (del Río-Mena et al., 2023). Furthermore, adapting locally derived ES models to new locations or different scales faces limitations due to the need for parameterization, calibration, and validation, often impeded by the absence of ground truth data (Cord et al., 2017).

Leveraging a diverse range of SEO instruments and measurements, suitable datasets can be chosen or developed for estimating specific ESi. This approach is expected to accelerate progress in spatially characterizing ESi, thereby contributing to ecological conservation, management, and integrated land-use planning (Andrew et al., 2014).

Despite the growing application of ML in ESi modelling, a substantial number of studies lack a dedicated focus on research commitments to ensure the replicability of methods and reproducibility of existent datasets. To address this gap, an easily understandable and transparent framework capable of reliably modelling unstructured data, such as images, video, time-series data, and text is required. As these diverse data types are increasingly integrated into ESi modelling throughout data fusion approaches, effectively handling and reproducing them may need domain expertise and the application of reverse engineering techniques.

The development of automated functions, akin to those found in ML Operations (MLOps) (Kreuzberger et al., 2023), is essential to streamline the process of taking ML models to production and subsequently maintaining and monitoring them in the context of satellite-based ESi modelling (Urbanowicz et al., 2022). These automated functions play a key role in optimizing modelling performance, capturing complex associations in data, enhancing interpretability and reproducibility, and preventing and detecting common sources of errors throughout the model's life cycle (Poleshchuk et al., 2022).

To ensure the robustness of the chosen model, resampling techniques must be employed for model tuning and predictor selection (Guio Blanco et al., 2018). Accordingly, there is a growing need for wider reporting of the specific tuning, predictor selection, and resampling methods utilized, given their substantial impact on prediction uncertainty and the identification of crucial predictor variables.

The ongoing advancements in cloud computing, open-access data, and ML technologies are expected to propel the development of large-scale ecosystem inventories (DeLancey et al., 2019). These advancements will not only expand the user base for geospatial data but also foster collaboration, addressing both existing and emerging challenges in large-area mapping and monitoring.

Developing models that can be scalable and replicable must consider uncertainties related to each phenomenon and spatial scale and a well-developed proof of concept of its applicability (Kubiszewski et al., 2022). Therefore, the scalability of models is achievable due to their rapid and cost-effective development (Scowen et al., 2021). The utilization of open-source tools and models plays a crucial role in fostering the widespread adoption of ML techniques, particularly for high-dimensional and multi-modal data. Additionally, the automation of data analysis is a key requirement underlying the efficiency and effectiveness of these models (Nikparvar & Thill, 2021).

4.3. Research limitations, contributions, and directions for future studies

Synthesizing the review of the included studies through meta-analysis helped in building knowledge of the SEO data-based ML applications. However, biases can lead to underestimations or overestimations in any review conclusions (Sterne & Egger, 2005). Many types of systematic errors might affect a review, but selection biases are the internal biases more frequently recognized as causes of reviews' conclusions deviations (Nakagawa et al., 2022). The selection bias occurs when the selected studies represent a small population of all research, and the absence of specific studies could increase biases (Greco et al., 2013). We are aware of the consequences of limiting the research to the English language, as this excludes worldwide outstanding studies reported for example in Chinese, Spanish or Portuguese. Some biases can also exist due to our decision to focus the review on studies that employed Landsat (4, 5, 7 and 8) and/or Sentinel-2A/2B. We also acknowledge that many other studies not included in this review would develop ML models to assess ESi using data from Sentinel-1A/2B, MODIS, Lidar, or other RS technologies; as well as other studies that model ESi, but did not choose to use "ecosystem services" in their title, keywords, and abstract. The methodology implemented allows users to assess, replicate, or update the review's findings. These results contribute to a better understanding of the use of SEO data and technologies, and the applicability of ML models to assess ESi.

As the assessment and monitoring of changes in ESi are helping decision-makers to achieve sustainable development and management at multiscale (Guirado et al., 2019), the outcomes of this review are relevant information for society and the scientific community, highlighting the latest approaches that significant researchers have produced worldwide.

Meta-analysis built the evidence that even when a problem is to predict continuous values, a classification task is commonly taken first, confirming ES-related researchers' adherence to ML techniques. Studies applied hybrid models when there were no available data identifying and delineating the target ecosystem, and further implemented a regression task to estimate ESi.

In future studies, ES researchers may adopt standardized methods to present the intricate steps related to data collection and processing, analysis, visualization, and interpretation of results, as suggested by

Scowen et al. (2021). But for that, a conceptual framework that guides model replicability needs to be developed and disseminated through the scientific community and stakeholders to enhance the applicability and understanding of ESi modelling. Additionally, future research may provide the required background to advance beyond what can be done and what can yet be enhanced when applying ML to estimate ES conditions. As this study did not analyze the predictors used in the various models, this is one of the topics that needs further knowledge and discussion, addressing this and other gaps including reviewing the use of spectral indices, and other landscape metrics as predictors, and handling the increasing volume of data developing tools that lower the costs associated with ES modelling.

5. Conclusions

The purpose of this systematic review was to summarize studies that developed satellite-based ML models to assess ESi and analyze their relationship, adopting a methodology that enables users to evaluate, repeat or revise the review's findings. Results show that there are no rules regarding the selection of ML tasks, algorithms, or evaluation metrics, but meta-analysis indicates the likelihood of applying classification and regression tasks, specifically when assessing climate regulation, habitat, and genetic resources. ML models based on SEO data are a promising method to assess ESi, with increased potential if combined with ground truth data of monitoring campaigns. Advances in automated methods for ES valuation are cost-effective and allow for better management of natural capital. However, further effort is needed to establish a robust and user-friendly framework for ML applications in ES modelling, ensuring not only the reliability of results but also facilitating the integration of diverse data sources and the deployment of models into operational settings. This includes the development of automated processes, inspired by MLOps, to efficiently manage the lifecycle of ML models within satellite-based ES applications. The main subjects of the reviewed studies related to environment protection, ES assessment and evaluation, nature conservation, natural capital management, policy-making, and land use and landscape planning. Modelling climate regulation services, such as carbon storage and sequestration, remained the overarching goal of 30% out from the reviewed studies. Additionally, measurements of nature stocks, such as biomass, freshwater, food production, and raw materials, have shown to be important at a global scale. The outcomes of this review add value to the present state-of-the-art on the applicability of ML models to assess ES through ESi. It further identifies the most relevant techniques and applications that support environmental problems.

CRediT authorship contribution statement

Bruna Almeida: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **João David:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Felipe S. Campos:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Pedro Cabral:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization, Project administration, Supervision.

Acknowledgements

This study was supported by the research project MaSOT – Mapping Ecosystem Services from Earth Observations, funded by the Portuguese Science Foundation - FCT [EXPL/CTA-AMB/0165/2021], and the European Union-NextGenerationEU. The authors gratefully acknowledge the financial support of the FCT, through the MagIC research center (Centro de Investigação em Gestão de Informação - UIDB/04152/2020). João David was financially supported by the Portuguese Foundation for Science and Technology (FCT) under Grant [2021.06482.BD]. We are grateful for the constructive remarks from anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2024.103249.

References

Abadi, M., Agarwal, A., Paul, B., Brevdo, E., Chen, Z., Craig, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Olah, C. (2015). *TensorFlow: Large-Scale machine learning on heterogeneous systems*.

Agrillo, E., Filipponi, F., Pezzarossa, A., Casella, L., Smiraglia, D., Orasi, A., Attorre, F., & Taramelli, A. (2021). Earth observation and biodiversity big data for forest habitat types classification and mapping. *Remote Sensing, 13*, 1231. https://doi.org/10.3390/rs13071231

Almeida, B., & Cabral, P. (2023). Data-Driven modelling of freshwater ecosystems: A multiscale framework based on global geospatial data. In *Proceedings of the 9th international conference on geographical information systems theory, applications and management* (pp. 104–111). SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0012037800003473.

Alqadhi, S., Mallick, J., Talukdar, S., Ahmed, M., Khan, R. A., Sarkar, S. K., & Rahman, A. (2022b). Assessing the effect of future landslide on ecosystem services in Aqabat Al-Sulbat region, Saudi Arabia. *Natural Hazards, 113*, 641–671. https://doi.org/10.1007/s11069-022-05318-7

Andrew, M. E., Wulder, M. A., & Nelson, T. A. (2014). Potential contributions of remote sensing to ecosystem service assessments. *Progress in Physical Geography: Earth and Environment, 38*, 328–353. https://doi.org/10.1177/0309133314528942

Anselin, et al. (2016). *GeoDa: Spatial data analysis*.

Arruda, V. L. S., Piontekowski, V. J., Alencar, A., Pereira, R. S., & Matricardi, E. A. T. (2021). An alternative approach for mapping burn scars using Landsat imagery, Google Earth Engine, and Deep Learning in the Brazilian Savanna. *Remote Sensing Applications, 22*, Article 100472. https://doi.org/10.1016/j.rsase.2021.100472

Astola, H., Häme, T., Sirro, L., Molinier, M., & Kilpi, J. (2019). Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment, 223*, 257–273. https://doi.org/10.1016/J.RSE.2019.01.019

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Bivand, R. S., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST, 27*, 716–748. https://doi.org/10.1007/s11749-018-0599-x.

Bolliger, J., & Kienast, F. (2010). Landscape functions in a changing environment. *Landscape Online, 21*, 1–5. https://doi.org/10.3097/LO.201021

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152).

Boutsoukis, C., Manakos, I., Heurich, M., & Delopoulos, A. (2019). Canopy height estimation from single multispectral 2D airborne imagery using texture analysis and machine learning in structurally rich temperate forests. *Remote Sensing, 11*, 2853. https://doi.org/10.3390/rs11232853

Boyd, J., & Banzhaf, S. (2007). What are ecosystem services? The need for standardized environmental accounting units. *Ecological Economics, 63*, 616–626. https://doi.org/10.1016/J.ECOLECON.2007.01.002

Braun, D., Damm, A., Hein, L., Petchey, O. L., & Schaepman, M. E. (2018). Spatio-temporal trends and trade-offs in ecosystem services: An Earth observation based assessment for Switzerland between 2004 and 2014. *Ecological Indicators, 89*, 828–839. https://doi.org/10.1016/j.ecolind.2017.10.016

Breiman, L. (1984). *Classification and regression trees* (1st ed.). New York: Routledge. https://doi.org/10.1201/9781315139470

Chollet, et al. (2015). *Keras*.

Cilek, A., Berberoglu, S., Donmez, C., & Sahingoz, M. (2022). The use of regression tree method for Sentinel-2 satellite data to mapping percent tree cover in different forest types. *Environmental Science and Pollution Research, 29*, 23665–23676. https://doi.org/10.1007/s11356-021-17333-5

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129. https://doi.org/10.2307/3001666

Copas, J., & Shi, J. Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics, 1*(3), 247–262. https://doi.org/10.1093/biostatistics/1.3.247

Cord, A. F., Brauman, K. A., Chaplin-Kramer, R., Huth, A., Ziv, G., & Seppelt, R. (2017). Priorities to advance monitoring of ecosystem services using earth observation. *Trends in Ecology & Evolution, 32*, 416–428. https://doi.org/10.1016/j.tree.2017.03.003

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P., & van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature, 387*, 253–260. https://doi.org/10.1038/387253a0

Craven, G. T., Chen, R., Cooper, M. W. D., Matthews, C., Rizk, J., Malone, W., Johnson, L., Gibson, T., & Andersson, D. A. (2023). Data-driven methods for diffusivity prediction in nuclear fuels. *Computational Materials Science, 230*, Article 112442. https://doi.org/10.1016/J.COMMATSCI.2023.112442

Czúcz, B., Keith, H., Driver, A., Jackson, B., Nicholson, E., & Maes, J. (2021). A common typology for ecosystem characteristics and ecosystem condition variables. *One Ecosystem, 6*, 1–16. https://doi.org/10.3897/oneeco.6.e58218

de Araujo Barbosa, C. C., Atkinson, P. M., & Dearing, J. A. (2015). Remote sensing of ecosystem services: A systematic review. *Ecological Indicators, 52*, 430–443. https://doi.org/10.1016/j.ecolind.2015.01.007

de Brito, H. C., Rufino, I. A. A., & Djordjević, S. (2021). Cellular automata predictive model for man-made environment growth in a Brazilian semi-arid watershed. *Environmental Monitoring and Assessment, 193*, 1–19. https://doi.org/10.1007/s10661-021-09108-9

De Groot, R. S., Wilson, M. A., & Boumans, R. M. J. (2002). A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological Economics, 41*, 393–408. https://doi.org/10.1016/S0921-8009(02)00089-7delRío-Mena

DeLancey, E. R., Kariyeva, J., Bried, J. T., & Hird, J. N. (2019). Large-scale probabilistic identification of boreal peatlands using Google Earth Engine, open-access satellite data, and machine learning. *PLoS One, 14*, 1–23. https://doi.org/10.1371/journal.pone.0218165

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*. https://doi.org/10.1145/2347736.2347755

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*, 917–928. https://doi.org/10.1093/JPEPSY/JSP004

Edens, B., Maes, J., Hein, L., Obst, C., Siikamaki, J., Schenau, S., Javorsek, M., Chow, J., Chan, J. Y., Steurer, A., & Alfieri, A. (2022). Establishing the SEEA ecosystem accounting as a global standard. *Ecosystem Services, 54*, Article 101413. https://doi.org/10.1016/J.ECOSER.2022.101413

Egoh, B., Drakou, E. G., Dunbar, M. B., Maes, J., & Willemen, L. (2012). Indicators for mapping ecosystem services: A review. *JRC scientific and policy reports*. https://doi.org/10.2788/41823

Elsevier, B. V. (2023). Scopus database [WWW Document]. URL https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030. (Accessed 30 June 2023).

ESRI. (2023). *ArcGIS pro - ESRI*. Environmental Systems Research Institute.

European Spatial Agency. (2023). ESA - Space Science [WWW Document]. URL https://www.esa.int/Science_Exploration/Space_Science. (Accessed 11 June 2023).

Ez-zahouani, B., Teodoro, A., El Kharki, O., Jianhua, L., Kotaridis, I., Yuan, X., & Ma, L. (2023). Remote sensing imagery segmentation in object-based analysis: A review of methods, optimization, and quality evaluation over the past 20 years. *Remote Sensing Applications, 32*, Article 101031. https://doi.org/10.1016/J.RSASE.2023.101031

Feng, X., Fu, B., Yang, X., & Lü, Y. (2010). Remote sensing of ecosystem services: An opportunity for spatially explicit assessment. *Chinese Geographical Science, 20*, 522–535. https://doi.org/10.1007/s11769-010-0428-y

Fisher, B., Turner, R. K., & Morling, P. (2009). Defining and classifying ecosystem services for decision making. *Ecological Economics, 68*, 643–653. https://doi.org/10.1016/j.ecolecon.2008.09.014

Fitts, L. A., Russell, M. B., Domke, G. M., & Knight, J. K. (2021). Modeling land use change and forest carbon stock changes in temperate forests in the United States. *Carbon Balance and Management, 16*, 20. https://doi.org/10.1186/s13021-021-00183-6

Fleming, A., O'Grady, A. P., Stitzlein, C., Ogilvy, S., Mendham, D., & Harrison, M. T. (2022). Improving acceptance of natural capital accounting in land use decision making: Barriers and opportunities. *Ecological Economics, 200*, Article 107510. https://doi.org/10.1016/J.ECOLECON.2022.107510

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*, 1189–1232.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Google. (2023). Google Earth Engine [WWW Document]. URL https://earthengine.google.com/. (Accessed 9 May 2023).

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart Lung Vessel, 5*, 219–225.

Guio Blanco, C. M., Brito Gomez, V. M., Crespo, P., & Ließ, M. (2018). Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. *Geoderma, 316*, 100–114. https://doi.org/10.1016/j.geoderma.2017.12.002

Guirado, E., Blanco-Sacristán, J., Rigol-Sánchez, J. P., Alcaraz-Segura, D., & Cabello, J. (2019). A multi-temporal object-based image analysis to detect long-lived shrub cover changes in drylands. *Remote Sensing, 11*. https://doi.org/10.3390/rs11222649

Gwal, S., Singh, S., Gupta, S., & Anand, S. (2020). Understanding forest biomass and net primary productivity in Himalayan ecosystem using geospatial approach. *Model Earth Syst Environ, 6*, 2517–2534. https://doi.org/10.1007/s40808-020-00844-4

Ha, N. T., Pham, T. D., & Tran, T. T. H. (2021). Zoning seagrass protection in lap an lagoon, vietnam using a novel integrated framework for sustainable coastal management. *Wetlands, 41*, 122. https://doi.org/10.1007/s13157-021-01504-8

Haines-Young, R., & Potschin, M. (2012). The links between biodiversity, ecosystem services and human well-being. In *Ecosystem ecology* (pp. 110–139). Cambridge University Press. https://doi.org/10.1017/cbo9780511750458.007.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2008). The WEKA data mining software: An update. *SIGKDD Explor. Newsl., 11*, 10–18.

Han, J., Hu, Z., Mao, Z., Li, G., Liu, S., Yuan, D., & Guo, J. (2022). How to account for changes in carbon storage from coal mining and reclamation in eastern China? Taking yanzhou coalfield as an example to simulate and estimate. *Remote Sensing, 14*. https://doi.org/10.3390/rs14092014, 2014.

Hasan, M. E., Zhang, L., Mahmood, R., Guo, H., & Li, G. (2021). Modeling of forest ecosystem degradation due to anthropogenic stress: The case of Rohingya influx into the cox's bazar–teknaf peninsula of Bangladesh. *Environments - MDPI, 8*, 121. https://doi.org/10.3390/environments8110121

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. In *Springer series in statistics*. New York, New York, NY: Springer. https://doi.org/10.1007/978-0-387-84858-7

Hauser, L. T., Féret, J. B., An Binh, N., van der Windt, N., Sil, Â. F., Timmermans, J., Soudzilovskaia, N. A., & van Bodegom, P. M. (2021). Towards scalable estimation of plant functional diversity from Sentinel-2: In-situ validation in a heterogeneous (semi-)natural landscape. *Remote Sensing of Environment, 262*, Article 112505. https://doi.org/10.1016/j.rse.2021.112505

Hedges, L. v, Gurevitch, J., & Curtis, P. S. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology, 80*, 1150–1156. https://doi.org/10.2307/177062

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2022). *Cochrane handbook for systematic reviews of interventions | cochrane training* [WWW Document]. URL https://training.cochrane.org/handbook#how-to-cite. (Accessed 10 March 2022).

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558. https://doi.org/10.1002/sim.1186

Hoffmann, J., Muro, J., & Dubovyk, O. (2022). Predicting species and structural diversity of temperate forests with satellite remote sensing and deep learning. *Remote Sensing, 14*, 1631. https://doi.org/10.3390/rs14071631

Hudak, A. T., Fekety, P. A., Kane, V. R., Kennedy, R. E., Filippelli, S. K., Falkowski, M. J., Tinkham, W. T., Smith, A. M. S., Crookston, N. L., Domke, G. M., Corrao, M. V., Bright, B. C., Churchill, D. J., Gould, P. J., McGaughey, R. J., Kane, J. T., & Dong, J. (2020). A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA. *Environmental Research Letters, 15*, Article 95003. https://doi.org/10.1088/1748-9326/ab93f9

Hunter, F. D. L., Mitchard, E. T. A., Tyrrell, P., & Russell, S. (2020). Inter-seasonal time series imagery enhances classification accuracy of grazing resource and land degradation maps in a savanna ecosystem. *Remote Sensing, 12*, 198. https://doi.org/10.3390/RS12010198

Jullian, C., Nahuelhual, L., & Laterra, P. (2021). The Ecosystem Service Provision Index as a generic indicator of ecosystem service supply for monitoring conservation targets. *Ecological Indicators, 129*, Article 107855. https://doi.org/10.1016/j.ecolind.2021.107855

Karasov, O., Heremans, S., Külvik, M., Domnich, A., & Chervanyov, I. (2020). On how crowdsourced data and landscape organisation metrics can facilitate the mapping of cultural ecosystem services: An Estonian case study. *Land, 9*, 158. https://doi.org/10.3390/LAND9050158

Kienast, F., Bolliger, J., Potschin, M., de Groot, R. S., Verburg, P. H., Heller, I., Wascher, D., & Haines-Young, R. (2009). Assessing landscape functions with broad-scale environmental data: Insights gained from a prototype development for europe. *Environmental Management, 44*, 1099–1120. https://doi.org/10.1007/s00267-009-9384-7

King, S., Agra, R., Zolyomi, A., Keith, H., Nicholson, E., de Lamo, X., Portela, R., Obst, C., Alam, M., Honzák, M., Valbuena, R., Nunes, P. A. L. D., Santos-Martin, F., Equihua, M., Pérez-Maqueo, O., Javorsek, M., Alfieri, A., & Brown, C. (2024). Using the system of environmental-economic accounting ecosystem accounting for policy: A case study on forest ecosystems. *Environmental Science & Policy, 152*, Article 103653. https://doi.org/10.1016/J.ENVSCI.2023.103653

King, S., Vardon, M., Grantham, H. S., Eigenraam, M., Ferrier, S., Juhn, D., Larsen, T., Brown, C., & Turner, K. (2021). Linking biodiversity into national economic accounting. *Environmental Science & Policy, 116*, 20–29. https://doi.org/10.1016/J.ENVSCI.2020.10.020

Koskikala, J., Kukkonen, M., & Käyhkö, N. (2020). Mapping natural forest remnants with multi-source and multi-temporal remote sensing data for more informed management of global biodiversity hotspots. *Remote Sensing, 12*, 1429. https://doi.org/10.3390/RS12091429

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning Operations (MLOps): Overview, definition, and architecture. *IEEE Access, 11*, 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

Kubiszewski, I., Muthee, K., Rifaee Rasheed, A., Costanza, R., Suzuki, M., Noel, S., & Schauer, M. (2022). The costs of increasing precision for ecosystem services valuation studies. *Ecological Indicators, 135*, Article 108551. https://doi.org/10.1016/J.ECOLIND.2022.108551

Kundu, S., Pal, S., Mandal, I., & Talukdar, S. (2022). How far damming induced wetland fragmentation and water richness change affect wetland ecosystem services? *Remote Sensing Applications, 27*, Article 100777. https://doi.org/10.1016/j.rsase.2022.100777

Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J. M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M. E., Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A. K., & Cord, A. F. (2016). Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators, 70*, 317–339. https://doi.org/10.1016/j.ecolind.2016.06.022

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Leroux, L., Falconnier, G. N., Diouf, A. A., Ndao, B., Gbodjo, J. E., Tall, L., Balde, A. A., Clermont-Dauphin, C., Bégué, A., Affholder, F., & Roupsard, O. (2020). Using remote sensing to assess the effect of trees on millet yield in complex parklands of Central Senegal. *Agricultural Systems, 184*, Article 102918. https://doi.org/10.1016/j.agsy.2020.102918

Liang, S., & Wang, J. (2019). Advanced remote sensing: Terrestrial information extraction and applications. *Advanced remote sensing: Terrestrial information extraction and applications 1–986*. https://doi.org/10.1016/C2017-0-03489-4

Lobert, F., Holtgrave, A. K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., & Erasmi, S. (2021). Mowing event detection in permanent grasslands: Systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series. *Remote Sensing of Environment, 267*, Article 112751. https://doi.org/10.1016/j.rse.2021.112751

Ludwig, C., Walli, A., Schleicher, C., Weichselbaum, J., & Riffler, M. (2019). A highly automated algorithm for wetland detection using multi-temporal optical satellite data. *Remote Sensing of Environment, 224*, 333–351. https://doi.org/10.1016/J.RSE.2019.01.017

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing, 152*, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Malinga, R., Gordon, L. J., Jewitt, G., & Lindborg, R. (2015). Mapping ecosystem services across scales and continents - a review. *Ecosystem Services, 13*, 57–63. https://doi.org/10.1016/J.ECOSER.2015.01.006

Mallinis, G., Chrysafis, I., Korakis, G., Pana, E., & Kyriazopoulos, A. P. (2020). A random forest modelling procedure for a multi-sensor assessment of tree species diversity. *Remote Sensing, 12*, 1210. https://doi.org/10.3390/rs12071210

Mallinis, G., Galidaki, G., & Gitas, I. (2014). A comparative analysis of EO-1 hyperion, quickbird and landsat TM imagery for fuel type mapping of a typical mediterranean landscape. *Remote Sensing, 6*, 1684–1704. https://doi.org/10.3390/rs6021684

Manley, K., Nyelele, C., & Egoh, B. N. (2022). A review of machine learning and big data applications in addressing ecosystem service research gaps. *Ecosystem Services, 57*, Article 101478. https://doi.org/10.1016/J.ECOSER.2022.101478

Martínez-Harms, M. J., & Balvanera, P. (2012). Methods for mapping ecosystem service supply: A review. *Int J Biodivers Sci Ecosyst Serv Manag, 8*, 17–25. https://doi.org/10.1080/21513732.2012.663792

Matsala, M., Myroniuk, V., Bilous, A., Terentiev, A., Diachuk, P., & Zadorozhniuk, R. (2020). An indirect approach to predict deadwood biomass in forests of Ukrainian Polissya using Landsat images and terrestrial data. *Forestry Studies, 73*, 107–124. https://doi.org/10.2478/fsmu-2020-0018

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431161.2018.1433343

McGarigal, K., & Marks, B. J. (1995). *Fragstats: Spatial pattern analysis program for quantifying landscape structure. | pacific northwest research station*. PNW - US Forest Service.

McVittie, A., & Hussain, S. S. (2013). *The economics of ecosystems and biodiversity - valuation database manual*.

Mo, Y., Kearney, M. S., Riter, J. C. A., Zhao, F., & Tilley, D. R. (2018). Assessing biomass of diverse coastal marsh ecosystems using statistical and machine learning models. *International Journal of Applied Earth Observation and Geoinformation, 68*, 189–201. https://doi.org/10.1016/j.jag.2017.12.003

Mohanty, S. P., Czakon, J., Kaczmarek, K. A., Pyskir, A., Tarasiewicz, P., Kunwar, S., Rohrbach, J., Luo, D., Prasad, M., Fleer, S., Göpfert, J. P., Tandon, A., Mollard, G., Rayaprolu, N., Salathe, M., & Schilling, M. (2020). Deep learning for understanding satellite imagery: An experimental survey. *Front Artif Intell, 3*. https://doi.org/10.3389/frai.2020.534696

Morell-Monzó, S., Estornell, J., & Sebastiá-Frasquet, M. T. (2020). Article Comparison of Sentinel-2 and high-resolution imagery for mapping land abandonment in fragmented areas. *Remote Sensing, 12*, 2062. https://doi.org/10.3390/RS12122062

Mouta, N., Silva, R., Pais, S., Alonso, J. M., Gonçalves, J. F., Honrado, J., & Vicente, J. R. (2021). 'The best of two worlds'—combining classifier fusion and ecological models to map and explain landscape invasion by an alien shrub. *Remote Sensing, 13*, 3287. https://doi.org/10.3390/rs13163287

Mpakairi, K. S., Dube, T., Dondofema, F., & Dalu, T. (2022a). Spatio–temporal variation of vegetation heterogeneity in groundwater dependent ecosystems within arid environments. *Ecological Informatics, 69*. https://doi.org/10.1016/j.ecoinf.2022.101667

Mpakairi, K. S., Dube, T., Dondofema, F., & Dalu, T. (2022b). Spatial characterisation of vegetation diversity in groundwater-dependent ecosystems using in-situ and sentinel-2 MSI satellite data. *Remote Sensing, 14*, 2995. https://doi.org/10.3390/rs14132995

Mugiraneza, T., Ban, Y., & Haas, J. (2019). Urban land cover dynamics and their impact on ecosystem services in Kigali, Rwanda using multi-temporal Landsat data. *Remote Sensing Applications, 13*, 234–246. https://doi.org/10.1016/j.rsase.2018.11.001

Mustajoki, J., Saarikoski, H., Belton, V., Hjerppe, T., & Marttunen, M. (2020). Utilizing ecosystem service classifications in multi-criteria decision analysis – experiences of peat extraction case in Finland. *Ecosystem Services, 41*, Article 101049. https://doi.org/10.1016/J.ECOSER.2019.101049

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution, 13*, 4–21. https://doi.org/10.1111/2041-210X.13724

Narine, L. L., Popescu, S. C., & Malambo, L. (2019). Synergy of ICESat-2 and landsat for mapping forest aboveground biomass with deep learning. *Remote Sensing, 11*, 1503. https://doi.org/10.3390/rs11121503

NASA. (2023). Earthdata | Earthdata [WWW Document]. URL https://www.earthdata.nasa.gov/. (Accessed 3 February 2023).

NASA's. Earth observing system. WWW Document], n.d. URL https://eospso.nasa.gov/content/nasas-earth-observing-system-project-science-office. (Accessed 28 August 2023).

Nawrocki, T. W., Carlson, M. L., Osnas, J. L. D., Trammell, E. J., & Witmer, F. D. W. (2020). Regional mapping of species-level continuous foliar cover: Beyond categorical vegetation mapping. *Ecological Applications, 30*, Article e02081. https://doi.org/10.1002/eap.2081

Nikparvar, B., & Thill, J. C. (2021). Machine learning of spatial data. *ISPRS International Journal of Geo-Information, 10*(600 10), 600. https://doi.org/10.3390/IJGI10090600, 2021.

Nzuza, P., Ramoelo, A., Odindi, J., Kahinda, J. M., & Madonsela, S. (2021). Predicting land degradation using Sentinel-2 and environmental variables in the Lepellane catchment of the Greater Sekhukhune District, South Africa. *Physics and Chemistry of the Earth, 124*, Article 102931. https://doi.org/10.1016/j.pce.2020.102931

ODC. (2018). *Open data Cube project*.

Olander, L. P., Johnston, R. J., Tallis, H., Kagan, J., Maguire, L. A., Polasky, S., Urban, D., Boyd, J., Wainger, L., & Palmer, M. (2018). Benefit relevant indicators: Ecosystem services measures that link ecological and social outcomes. *Ecological Indicators, 85*, 1262–1272. https://doi.org/10.1016/J.ECOLIND.2017.12.001

Olofsson, P., Foody, G. M., Stehman, S.v., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment, 129*, 122–131. https://doi.org/10.1016/J.RSE.2012.10.031

Ørka, H. O., Jutras-Perreault, M.-C., Næsset, E., & Gobakken, T. (2022). A framework for a forest ecological base map – an example from Norway. *Ecological Indicators, 136*, Article 108636. https://doi.org/10.1016/j.ecolind.2022.108636

Osborne, P. E., & Alvares-Sanches, T. (2019). Quantifying how landscape composition and configuration affect urban land surface temperatures using machine learning and neutral landscapes. *Computers, Environment and Urban Systems, 76*, 80–90. https://doi.org/10.1016/j.compenvurbsys.2019.04.003

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., … Mckenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*. https://doi.org/10.1136/bmj.n160

Pastick, N. J., Jorgenson, M. T., Wylie, B. K., Nield, S. J., Johnson, K. D., & Finley, A. O. (2015). Distribution of near-surface permafrost in Alaska: Estimates of present and future conditions. *Remote Sensing of Environment, 168*, 301–315. https://doi.org/10.1016/j.rse.2015.07.019

Pastick, N., Wylie, B., & Wu, Z. (2018). Spatiotemporal analysis of landsat-8 and sentinel-2 data to support monitoring of dryland ecosystems. *Remote Sensing, 10*, 791. https://doi.org/10.3390/rs10050791

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Pettorelli, N., to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A. M., Keith, D. A., Wegmann, M., Schrodt, F., Stellmes, M., Sonnenschein, R., Geller, G. N., Roy, S., Somers, B., Murray, N., Bland, L., Geijzendorffer, I., Kerr, J. T., Broszeit, S., … Nicholson, E. (2018). Satellite remote sensing of ecosystem functions: Opportunities, challenges and way forward. *Remote Sens Ecol Conserv, 4*, 71–93. https://doi.org/10.1002/rse2.59

Pipia, L., Amin, E., Belda, S., Salinero-Delgado, M., & Verrelst, J. (2021). Green LAI mapping and cloud gap-filling using Gaussian process regression in google earth engine. *Remote Sensing, 13*, 403. https://doi.org/10.3390/rs13030403

Pizarro, S. E., Pricope, N. G., Vargas-Machuca, D., Huanca, O., & Ñaupari, J. (2022). Mapping land cover types for highland andean ecosystems in Peru using google earth engine. *Remote Sensing, 14*, 1562. https://doi.org/10.3390/rs14071562

Poleshchuk, E., Marshall, A., Tottrup, C., Ghafari, D., Coppens, L., & Crane, S. (2022). Use of big data for official environment statistics: The measurement of extent and quality of freshwater ecosystems. *Statistical Journal of the IAOS, 38*, 957–972. https://doi.org/10.3233/SJI-220041

Pritt, M., & Chern, G. (2018). Satellite image classification with deep learning. *Proceedings - applied imagery pattern recognition workshop 2017-octob*. https://doi.org/10.1109/AIPR.2017.8457969, 4–10.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2017). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 6638–6648, 2018-December.

QGIS Association. (2022). *QGIS geographic information system*.

R Core Team. (2022). *R: A language and environment for statistical computing*.

Ramirez-Reyes, C., Brauman, K. A., Chaplin-Kramer, R., Galford, G. L., Adamo, S. B., Anderson, C. B., Anderson, C., Allington, G. R. H., Bagstad, K. J., Coe, M. T., Watson, K., & Wright, T. M. (2019). Reimagining the potential of Earth observations for ecosystem service assessments. *Science of the Total Environment, 665*, 1053–1063. https://doi.org/10.1016/j.scitotenv.2019.02.150

Reddy, C. S. (2021). Remote sensing of biodiversity: What to measure and monitor from space to species? *Biodiversity & Conservation, 30*, 2617–2631. https://doi.org/10.1007/s10531-021-02216-5

RuleQuest Research. (2022). *Data mining tools available from RuleQuest research*.

Sanderman, J., Hengl, T., Fiske, G., Solvik, K., Adame, M. F., Benson, L., Bukoski, J. J., Carnell, P., Cifuentes-Jara, M., Donato, D., Duncan, C., Eid, E. M., Ermgassen, P. zu, Lewis, C. J. E., Macreadie, P. I., Glass, L., Gress, S., Jardine, S. L., Jones, T. G., … Landis, E. (2018). A global map of mangrove forest soil carbon at 30 m spatial resolution. *Environmental Research Letters, 13*, Article 055002. https://doi.org/10.1088/1748-9326/AABE1C

Sannigrahi, S., Chakraborti, S., Joshi, P. K., Keesstra, S., Sen, S., Paul, S. K., Kreuter, U., Sutton, P. C., Jha, S., & Dang, K. B. (2019). Ecosystem service value assessment of a natural reserve region for strengthening protection and conservation. *Journal of Environmental Management, 244*, 208–227. https://doi.org/10.1016/j.jenvman.2019.04.095

Sannigrahi, S., Joshi, P. K., Keesstra, S., Paul, S. K., Sen, S., Roy, P. S., Chakraborti, S., & Bhatt, S. (2019). Evaluating landscape capacity to provide spatially explicit valued ecosystem services for sustainable coastal resource management. *Ocean & Coastal Management, 182*, Article 104918. https://doi.org/10.1016/j.ocecoaman.2019.104918

Schirpke, U., Ghermandi, A., Sinclair, M., Van Berkel, D., Fox, N., Vargas, L., & Willemen, L. (2023). Emerging technologies for assessing ecosystem services: A synthesis of opportunities and challenges. *Ecosystem Services, 63*, Article 101558. https://doi.org/10.1016/J.ECOSER.2023.101558

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. *Science of the Total Environment, 799*, Article 149263. https://doi.org/10.1016/J.SCITOTENV.2021.149263

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing, 70*, 78–87. https://doi.org/10.1016/J.ISPRSJPRS.2012.04.001

Sharma, A., Hubert-Moy, L., Buvaneshwari, S., Sekhar, M., Ruiz, L., Bandyopadhyay, S., & Corgne, S. (2018). Irrigation history estimation using multitemporal landsat satellite images: Application to an intensive groundwater irrigated agricultural watershed in India. *Remote Sensing, 10*, 893. https://doi.org/10.3390/rs10060893

Skidmore, A. K., Coops, N. C., Neinavaz, E., Ali, A., Schaepman, M. E., Paganini, M., Kissling, W. D., Vihervaara, P., Darvishzadeh, R., Feilhauer, H., Wegmann, M., & Wingate, V. (2021). Priority list of biodiversity metrics to observe from space. *Nat Ecol Evol, 5*, 896–906. https://doi.org/10.1038/s41559-021-01451-x

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustment* (pp. 99–110). Chichester: Wiley.

Syrbe, R.-U., & Walz, U. (2012). Spatial indicators for the assessment of ecosystem services: Providing, benefiting and connecting areas and landscape metrics. *Ecological Indicators, 21*, 80–88.

Tarantino, C., Forte, L., Blonda, P., Vicario, S., Tomaselli, V., Beierkuhnlein, C., & Adamo, M. (2021). Intra-annual sentinel-2 time-series supporting grassland habitat discrimination. *Remote Sensing, 13*, 1–29. https://doi.org/10.3390/rs13020277

Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., & Patin, R. (2022). *biomod2: Ensemble platform for species distribution modeling*.

United Nations, et al. (2021). *System of environmental-economic accounting-ecosystem accounting*. SEEA-EA).

Urbanowicz, R. J., Zhang, R., Cui, Y., & Suri, P. (2022). Streamline: A simple, transparent, end-to-end automated machine learning pipeline facilitating data analysis and algorithm comparison. *Genetic Programming Theory and Practice XIX*, 201–231.

Vallecillo, S., La Notte, A., Ferrini, S., & Maes, J. (2019). How ecosystem services are changing: An accounting application at the EU level. *Ecosystem Services, 40*, Article 101044. https://doi.org/10.1016/J.ECOSER.2019.101044

Vasenev, V. I., Van Oudenhoven, A. P. E., Romzaykina, O. N., & Hajiaghaeva, R. A. (2018). The ecological functions and ecosystem services of urban and technogenic soils: From theory to practice (A review). *Eurasian Soil Science, 51*, 1119–1132. https://doi.org/10.1134/S1064229318100137/METRICS

Vidal-Macua, J. J., Nicolau, J. M., Vicente, E., & Moreno-de las Heras, M. (2020). Assessing vegetation recovery in reclaimed opencast mines of the Teruel coalfield (Spain) using Landsat time series and boosted regression trees. *Science of the Total Environment, 717*, Article 137250. https://doi.org/10.1016/j.scitotenv.2020.137250

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. https://doi.org/10.18637/jss.v036.i03

Vyvlečka, P., & Pechanec, V. (2023). Optical remote sensing in provisioning of ecosystem-functions analysis-review. *Sensors, 23*. https://doi.org/10.3390/s23104937

Wall, W. A., Hohmann, M. G., Just, M. G., & Hoffmann, W. A. (2021). Characterizing past fire occurrence in longleaf pine ecosystems with the Mid-Infrared Burn Index and a Random Forest classifier. *Ecological Management, 500*, Article 119635. https://doi.org/10.1016/j.foreco.2021.119635

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., & Shangguan, Z. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific Reports, 7*, 6940. https://doi.org/10.1038/s41598-017-07197-6

Wang, Y., Zhang, X., & Peng, P. (2021). Spatio-temporal changes of land-use/land cover change and the effects on ecosystem service values in Derong county, China, from 1992-2018. *Sustainability, 13*, 1–16. https://doi.org/10.3390/su13020827

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*, 490. https://doi.org/10.2307/258554

Willcock, S., Hooftman, D. A. P., Neugarten, R. A., Chaplin-Kramer, R., Barredo, J. I., Hickler, T., Kindermann, G., Lewis, A. R., Lindeskog, M., Martínez-López, J., & Bullock, J. M. (2023). Model ensembles of ecosystem services fill global certainty and capacity gaps. *Science Advances, 9*. https://doi.org/10.1126/SCIADV.ADF5492

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. *Ecosystem Services, 33*, 165–174. https://doi.org/10.1016/J.ECOSER.2018.04.004

Yang, X., Liu, S., Jia, C., Liu, Y., & Yu, C. (2021). Vulnerability assessment and management planning for the ecological environment in urban wetlands. *Journal of Environmental Management, 298*, Article 113540. https://doi.org/10.1016/j.jenvman.2021.113540

Zergaw Ayanu, Y., Conrad, C., Nauss, T., Wegmann, M., & Koellner, T. (2012). Quantifying and mapping ecosystem services supplies and demands: A review of remote sensing applications. *Environmental Science & Technology, 46*. https://doi.org/10.1021/es300157u

Zhang, M., Du, H., Zhou, G., Li, X., Mao, F., Dong, L., Zheng, J., Liu, H., Huang, Z., & He, S. (2019). Estimating forest aboveground carbon storage in Hang-Jia-Hu using landsat TM/OLI data and random forest model. *Forests, 10*, 1004. https://doi.org/10.3390/f10111004

Zhang, L., Hu, Q., & Tang, Z. (2022). Assessing the contemporary status of Nebraska's eastern saline wetlands by using a machine learning algorithm on the Google Earth Engine cloud computing platform. *Environmental Monitoring and Assessment, 194*, 193. https://doi.org/10.1007/s10661-022-09850-8

Zhang, J., Okin, G. S., & Zhou, B. (2019). Assimilating optical satellite remote sensing images and field data to predict surface indicators in the Western U.S.: Assessing error in satellite predictions based on large geographical datasets with the use of machine learning. *Remote Sensing of Environment, 233*, Article 111382. https://doi.org/10.1016/j.rse.2019.111382

Zhang, Y., Sharma, S., Bista, M., & Li, M. (2022). Characterizing changes in land cover and forest fragmentation from multitemporal Landsat observations (1993-2018) in the Dhorpatan Hunting Reserve, Nepal. *Journal of Research, 33*, 159–170. https://doi.org/10.1007/s11676-021-01325-9

Zhu, B., Bai, Y., Zhang, Z., He, X., Wang, Z., Zhang, S., & Dai, Q. (2022). Satellite remote sensing of water quality variation in a semi-enclosed bay (yueqing bay) under strong anthropogenic impact. *Remote Sensing, 14*, 550. https://doi.org/10.3390/rs14030550