The current and future uses of machine learning in ecosystem service research

Scowen, Matthew; Athanasiadis, Ioannis; Bullock, James; Eigenbrod, Felix; Willcock, Simon

Science of the Total Environment

DOI:
10.1016/j.scitotenv.2021.149263

Published: 10/12/2021

Peer reviewed version

The current and future uses of machine learning in ecosystem service research

Matthew Scowen[a,*], Ioannis N. Athanasiadis[b], James M. Bullock[c], Felix Eigenbrod[d], Simon Willcock[a,e]

[a] School of Natural Sciences, Bangor University, United Kingdom

[b] Wageningen University and Research, Wageningen University, The Netherlands

[c] UK Centre for Ecology and Hydrology, Wallingford, United Kingdom

[d] Geography and Environment, University of Southampton, United Kingdom

[e] Rothamsted Research, Harpenden, United Kingdom

∗ Corresponding author at: School of Natural Sciences, Bangor University, Thoday Building, Deiniol Road, Bangor LL57 2UW, United Kingdom.

E-mail addresses: mattscowen@bangor.ac.uk (M. Scowen), ioannis.athanasiadis@wur.nl (I. Athanasiadis), jmbul@ceh.ac.uk (J.M. Bullock). F.Eigenbrod@soton.ac.uk (F. Eigenbrod), s.willcock@bangor.ac.uk (S. Willcock).

Abstract

Machine learning (ML) expands traditional data analysis and presents a range of opportunities in ecosystem service (ES) research, offering rapid processing of 'big data' and enabling significant advances in data description and predictive modelling. Descriptive ML techniques group data with little or no prior domain specific assumptions; they can generate hypotheses and automatically sort data prior to other analyses. Predictive ML techniques allow for the predictive modelling of highly non-linear systems where casual mechanisms are poorly understood, as is often the case for ES. We conducted a review to explore how ML is used in ES research and to identify and quantify trends in the different ML approaches that are used. We reviewed 308 peer-reviewed publications and identified that ES studies implemented machine learning techniques in data description (63%; n= 308) and predictive modelling (44%), with some papers containing both categories. Classification and

Regression Trees were the most popular techniques (60%), but unsupervised learning techniques were also used for descriptive tasks such as clustering to group or split data without prior assumptions (19%). Whilst there are examples of ES publications that apply ML with rigour, many studies do not have robust or repeatable methods. Some studies fail to report model settings (43%) or software used (28%), and many studies do not report carrying out any form of model hyperparameter tuning (67%) or test model generalisability (59%). Whilst studies use ML to analyse very large and complex datasets, ES research is generally not taking full advantage of the capacity of ML to model big data (1138 medium number of data points; 13 median quantity of variables). There is great further opportunity to utilise ML in ES research, to make better use of big data and to develop detailed modelling of spatial-temporal dynamics that meet stakeholder demands.

Keywords: Machine learning; Ecosystem services; Big Data; Methodology; Validation; Data-driven modelling.

1. Introduction

Ecosystem service (ES) research involves the study of complex systems comprising interactions between biodiversity, human activity and the abiotic environment (MEA, 2005). The interactions underpinning ES are highly nonlinear and our mechanistic understanding of these processes is under-developed (Daw *et al.*, 2016; Spake *et al.*, 2017). This complexity makes implementing standard process-based modelling and statistical null hypothesis testing in ES problematic (Mouchet *et al.*, 2014; Villa *et al.*, 2014; Martínez-López *et al.*, 2019). Furthermore, data relevant to ES research, e.g. remotely sensed data, often has high-dimensionality, can be unstructured, and the volume of data is increasing at a rate beyond our ability to make sense of it using traditional approaches (Reichstein *et al.*, 2019).

Machine learning (ML) is an emerging and rapidly developing discipline and what constitutes ML, as opposed to other, more traditional statistical approaches, remains fuzzily defined (Bock *et al.*, 2019). Here we broadly define ML according to (Reichstein *et al.*, 2019) as '*a field of statistical research for* *training computational algorithms that split, sort and transform a set of data to maximize the ability to classify, predict, cluster or discover patterns in a target dataset*'. Using ML, data are empirically modelled with few or no prior assumptions about the system, using computer algorithms that can automatically learn from data. Since ML techniques can make data inferences without relying on causal theory, they can have useful application in highly non-linear, complex, and poorly characterised systems such as those producing ES. Furthermore, due to automation, ML approaches are particularly advantageous considering recent developments in social and environmental 'big data' relevant to ES research (Ghani *et al.*, 2019; Xia, Wang and Niu, 2020). ML approaches are therefore a valuable expansion to traditional data analyses and the diversity of ML techniques presents a range of opportunities as a data-driven approach to ES research (Willcock *et al.*, 2018). As such, ML is increasingly utilised within ecology and the environmental sciences and is enabling useful data inferences in domains in which traditional data analyses have had limited utility (Lucas, 2020). ML has enabled useful data inferences using data that has been collected automatically i.e. via remote sensing or other autonomous sensors (Lary *et al.*, 2016), or without experimental design (e.g. recording of species sightings by the public; (Torney *et al.*, 2019)), or open data that has been collected often for another purpose (Rammer and Seidl, 2019). ML is also used to analyse environmental data collected via social media platforms (Wäldchen and Mäder, 2018) or that has been generated synthetically via another modelling process (Chen, Roy and Hutton, 2018).

ML approaches can be divided to two main categories according to the type of task or research objective being pursued: *descriptive* (e.g. identifying unknown groups) and *predictive* (e.g. projections of future outcomes; Box 1) (Delen and Ram, 2018). Descriptive ML approaches group data with little or no prior domain specific assumptions, they can aid in hypothesis generation and can be used to automatically sort data prior to other data analyses. This allows for rapid processing of 'big data', where dataset size and high-dimensionality make organising or describing ES data by traditional methods not practically viable (Willcock *et al.*, 2018). ML clustering and ordination can be viewed as descriptive techniques, and in ES research they can identify ES bundles or hotspots in ES supply and demand, i.e. areas where two or more ES are consistently associated (Raudsepp-Hearne *et al.*, 2010; Mouchet *et al.*, 2014). ML classification of remotely sensed images involves describing large and complex datasets by grouping the data into meaningful classes, often for further analyses or to aid in hypothesis generation (Maxwell, Warner and Fang, 2018).

Predictive ML techniques are used to complete classification and regression tasks to use in models and make predictions about a system. This can allow for predictive modelling of highly non-linear systems where causal mechanisms are poorly understood (Huntingford *et al.*, 2019). The potential for the use of ML in a data-driven approach to predictive modelling of ES has already been highlighted and ML ES models have been shown to have comparable accuracy to conventional predictive modelling techniques (Willcock *et al.*, 2018). ML has a range of potential advantages over other modelling approaches in ES. Firstly, the inherent difficulty in making inferences with patterns in ‘noisy’ biological data results in high levels of uncertainty, and different models of the same system often diverge in their predictions (Knudby, Brenning and LeDrew, 2010; Willcock *et al.*, 2019). As such ES models may not meet the needs of stakeholders (Willcock *et al.*, 2016; Martínez-López *et al.*, 2019). ML models often have in-built measures of uncertainty that may be useful to stakeholders (Willcock *et al.*, 2018). Secondly, ML often allows the combination of continuous with categorical predictor variables (Cutler *et al.*, 2007), which is a particular advantage in modelling ES where data is often of disparate forms (Burkhard *et al.*, 2012). Thirdly, datasets relevant to ES research can have missing or unknown data that can be problematic to model construction (Willcock *et al.*, 2020). However, several ML algorithms (e.g. Classification and Regression Trees, some Support Vector Machines, and Neural Networks) can operate with gaps in the data without the need to impute missing data points (García-Laencina, Sancho-Gómez and Figueiras-Vidal, 2010). Finally, ML approaches can deal with many predictors, are robust to correlations in explanatory variables, and can allow for varying functional relationships between predictor and response variables (Hochachka *et al.*, 2007). These features make ML well suited to the analysis of complex systems with high dimensionality such as those producing ES.

Although automation in ML allows for rapid processing of large and complex datasets, which is clearly advantageous for both descriptive and predictive tasks considering the current challenges of 'big data', the lack of reliance on causal theory is also a potential pitfall of ML approaches. Essentially, by modelling correlations ML does not standardly incorporate any process-based theory, and this limits the generalisability of ML inferences outside of the input space of the data. It is therefore especially important that predictive ML models incorporate a process of validation whereby models are tested on independent data (Lucas, 2020). Likewise, any hypotheses or subsequent analyses based upon descriptive applications of ML should consider that the inference may not be transferable outside the parameter space (Spake *et al.*, 2017). ML approaches are also criticised as being 'black-box', in that it can be difficult to understand how or why they work (Zednik, 2019). Whilst, to some extent, opacity can be an inherent characteristic of some ML algorithms, it is nevertheless important that ML methodologies are as transparent as possible if research utilising ML is to be robust. As such, the input data used should be available to other researchers and any model settings, software used or relevant computer code necessary to run the model should be reported.

Considering these possible benefits but also pitfalls of using ML, here we conduct a review to quantify the use of ML in ES research. The aim is to explore how ML is used in ES research for descriptive and predictive tasks, to identify and quantify trends in ML approaches for ES, and to assess ML methodological repeatability. Specifically, we: 1) quantify the use of ML for descriptive and predictive modelling tasks in ES; 2) assess the extent to which applications of ML in ES research follow transparent and repeatable methodologies; 3) quantify the extent to which ES publications report model generalisability; and 4) review the size and complexity of datasets that have been used in ML approaches to ES.

Box 1. Machine Learning (ML) techniques

ML algorithms can broadly be divided into two kinds, from a learning perspective: supervised and unsupervised learning. In supervised learning a response variable is specified *a priori*. The user first labels and groups the system input variables and supplies the algorithm with the target output variable. The algorithm then finds a function that links the inputs with the outputs such that it can then make predictions of what the output will be from a given set of input variables. Classification and regression tasks are carried out using supervised learning approaches (Jordan and Mitchell, 2015). Types of supervised learning methods include Classification and Regression Trees (CARTs), Support Vector Machines (SVMs) and Maximum Likelihood approaches. In supervised ML the dataset is split into two subsets. One subset, the training data, is used to 'train' the algorithm how to carry out the task e.g., how to classify. This training data contains the target output and the user indicates what this is. The second subset, the test data, is reserved to 'test' the performance of the algorithm in carrying out its task. In this phase the target is not supplied to the algorithm so that the output produced by the algorithm can be compared to target output data (Breiman, 2001). When model tuning is involved, a part of the training set is held out from training and used for evaluating the training performance (during training) and to assist in selecting the optimal hyperparameter values. Model tuning can substantially increase the accuracy of the ML model, with only the optimal (i.e. most accurate) model being then used on the test set (Willcock *et al.*, 2018). However, we note that there is potential for confusion as both the tuning and testing processes are sometimes referred to as validation in the ML literature. Some studies also test the generalisability of the model to either arbitrary model decisions (e.g. how the datasets are subset into training and testing data) and/or to data outside the parameter space of the training and testing data subsets. Supervised learning approaches are especially useful in predictive modelling and in the analysis of variable importance.

In unsupervised learning prior knowledge of what the output should be is not given to the algorithm; no variables are labelled as outputs by the user. Unsupervised algorithms structure data by identifying groups that the user has not indicated *a priori*. Cluster analysis is an example of unsupervised learning. Some types of ML e.g., Artificial Neural Networks (ANNs), include supervised and unsupervised approaches. Unsupervised techniques are useful for data exploration and hypothesis generation because they allow insights into unstructured data (Solomatine, See and

Abrahart, 2009). As with other forms of data analyses, a variety of ML techniques can be used to carry out different tasks within a single study and ML can also be used in combination with tradition techniques. For example, a clustering algorithm might be used to group data prior to a regression either by ML or another statistical approach (Crisci, Ghattas and Perera, 2012). Generally, unsupervised approaches are used for descriptive/organisational tasks whilst predictive modelling tasks tend to be carried out using more supervised approaches.

2. Methods

We followed a quantitative review methodology that involved a two-step search strategy. We used the Web of Science database to find publications from which information was extracted according to categorisation criteria. The aim of step one was to generate a list of relevant machine learning (ML) terms that represent the use of ML in ecosystem service (ES) research. In step one we entered the search string: “machine learning” AND (“ecosystem services” OR “ecosystem service”). The Keywords and Keywords Plus were taken from all the resulting articles, and these were then classified as being terms either relevant to ML or not according to the mutual agreement of the review team. Thus, we generated a list of 33 relevant ML terms that represent the use of ML in ES research e.g., ‘data mining’, ’neural network’, ‘decision tree’, etc. (see SI1 for list of all Keyword and Keyword Plus terms and how they were classified). We then ran a new search by entering the search string: “*relevant-key-word*” AND (“ecosystem services” OR “ecosystem service”) for all the relevant ML terms identified in step one. All papers for each relevant term were assessed according to inclusion criteria: a) papers with no mention of ES in the title or the abstract were not included in the review; b) papers which did not use a machine learning algorithm as part of the data analyses were not included. Here an ML algorithm was defined as one which splits, sorts and transforms a set of data enabling it to classify, predict, cluster or discover patterns in a target dataset (Reichstein *et al.*, 2019). Those that did not meet the inclusion criteria were not included in this review. Papers that met the inclusion criteria were categorised and data extracted (below). If there were over 100 papers for each term, then random numbers were used to select 100 for inclusion in the review. For example, for the relevant-key-word ‘classification’ there were 1779 papers, so we selected a random sample of 100; while for relevant-key-word ‘support vector machine’ there were only 74, so all papers were reviewed. Note the search was not exhaustive because the Web of Science database is not totally comprehensive (Martín-Martín *et al.*, 2018) but provides a representative sample of important research in this area.

2.1. Data extraction and categorisation criteria

From our pool of articles, we categorised all applications of ML as either descriptive or predictive. Publications that had applications of both descriptive and predictive ML were included in both descriptive and predictive categories. Such articles included, for example, studies that carried out an ML cluster analysis prior to predictive modelling. All applications of unsupervised ML (i.e., clustering, PCA etc., see Box 1) were classed as descriptive methods. We also categorised ML applications used in the classification of remotely sensed data, and ML image recognition, as descriptive because the primary aim is to describe the data by sorting it into meaningful classes, with those descriptive papers not falling into this category termed ‘organisational’. All other applications of ML were classed as predictive. These predictive models either directly predicted specified ES (hereafter ‘direct ES prediction’), or the model did not directly predict a specified ES but was indirectly relevant to ES (hereafter ‘indirect ES prediction’). For example, if a study predictively modelled carbon sequestration this would be categorised as direct ES prediction but if it predictively modelled forest land cover then this could be used to indirectly predict ES. Thus, descriptive publications could be subdivided into either a) organisational or b) remote sensed / image recognition; and predictive publications could be subdivided into either a) direct ES prediction; b) indirect ES prediction. Note that membership of the subdivisions is mutually exclusive (i.e., ‘a’ or ‘b’) but a publication could be categorised as using both descriptive and predictive approaches.

The following information was also extracted from each manuscript:

- Dataset size and complexity – The number of data points (often referred to as the number of instances in a machine learning problem) and the number of variables (attributes) in the dataset used by the ML algorithm were recorded. If more than one application of ML was used in the analysis, then the largest of the sample sizes and number of variables is recorded.
- Data availability – The data used in the ML analysis were classed as being freely available if the data could be accessed for free.
- ML rationale given – Papers were considered as presenting a rationale for their use of ML if they provided an explanatory justification for its use in the analysis with reference to supporting literature.
- Generalisability – Papers were classified as having tested the generalisability of the model if: i) the impact of the training-testing subsets on the model were investigated (e.g. using cross validation to indicate how robust the model is to different subsets of the data), and/or ii) the transferability of the model outside the parameter set of the training and testing data were investigated (i.e. how well the model performs in a different geographic location or time frame; Box 1).
- Model tuning – A paper was classed as carrying out model tuning if adjustments were made to the standard parameters of the ML algorithm and either these adjustments were justified with reference to the literature or through testing of the effects on the ML output (Box 1).
- Software – A paper was classed as reporting the software if the software used to carry out the ML analyses was detailed.
- ML technique – The type of approach(es) used was recorded for each study. Approaches included: Classification and Regression Trees, Artificial Neural Networks, Bayesian, Maximum Likelihood, Support Vector Machines, Clustering algorithms.

Firstly, the percentage of reviewed publications using each ML approach was calculated per category of ML study (Organisational, Remote sensed and Image recognition, Direct ES prediction, and Indirect Prediction). Secondly, the percentage of publications meeting each of the other above criteria was calculated per category of ML study. Finally, the median, maximum, and minimum number of data points and variables for each category were also calculated. All analyses were carried out in R (version 4.0.4.)

3. Results

A pool of 1012 publications resulted from the search with a total of 308 publications applying machine learning (ML) in ecosystem service (ES) related research between 01/2008 and 07/2021 (Fig. 1; see SI2 for a comprehensive list). ML is increasingly being used in ES research and a wide variety of ML techniques are utilised for provisioning, regulating and cultural ES. In some ES studies (e.g. Funk *et al.*, 2019; Schirpke *et al.*, 2019; Havinga *et al.*, 2020), ML represents part of a methodology involving a range of other statistical and modelling techniques, sometimes involving application of more than one type of ML technique. In other studies (e.g. Richards and Tunçer, 2018; Nguyen, Nong and Paustian, 2019), ML represents the entire modelling process. In a further set of studies, different approaches are compared in terms of their ability to model similar data, either a range of ML techniques (e.g. Hirayama *et al.*, 2019; Sannigrahi *et al.*, 2019; Wu *et al.*, 2019) or ML in comparison to process based modelling (e.g. Willcock *et al.*, 2018). The median number of data points in each publication using ML was 1138 (maximum = 9,500,430; minimum = 17; n = 225; Table 1). The median number of variables was 13 (maximum = 2317; minimum = 3; n = 215).

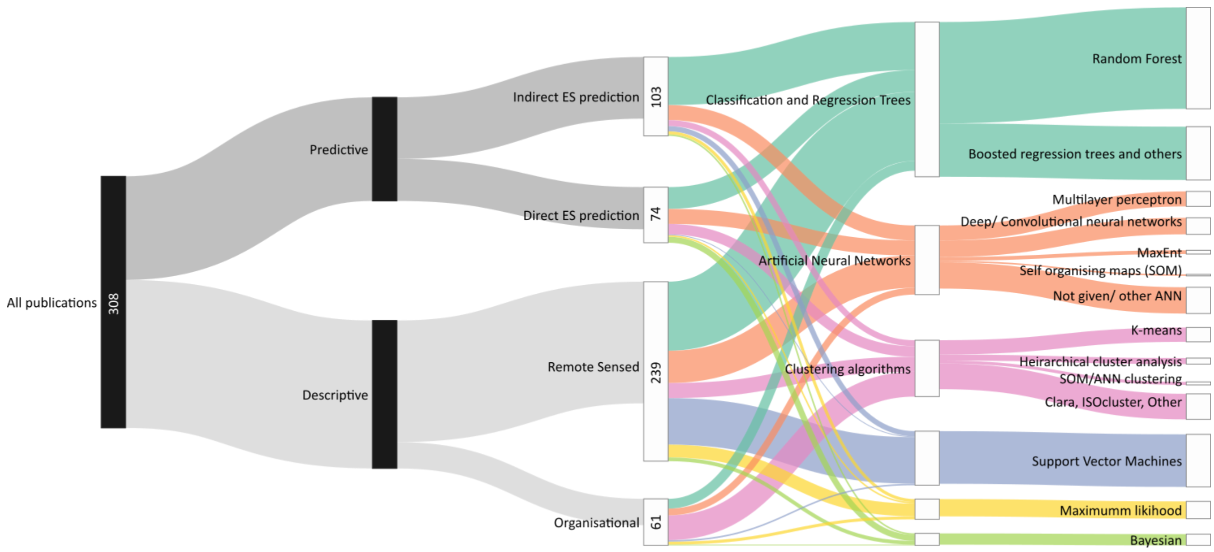

Fig. 1. Publications utilising Machine Learning (ML) for predictive or descriptive tasks and number of ML applications per ML technique. All publications = 308 papers. Height of black nodes are proportionate to number of publications. Height of white nodes proportionate to number of applications of ML (all applications = 477).

<u>3.1. ML for descriptive tasks</u>

ML was used for data description in 63% (n = 308) of studies., which can be divided into those using remotely sensed data or image recognition (52% of all studies; section 3.2.) and organisational studies (11%; Fig. 1). Clustering or ordination algorithms were commonly used to identify groupings, splits or other structure in data without theoretical assumptions (19%). Organisational studies used clustering algorithms to identify ES bundles or hotspots (7% of all studies). For example, *K-means* cluster analysis was used to describe bundles of supply, flow and demand of ES by identifying groups of ES according to spatial concurrence (Schirpke *et al.*, 2019), hierarchical cluster analysis was used to identify groups of ES according to social preferences (Martín-López *et al.*, 2012), and an Artificial Neural Network (ANN) with a clustering function was used to identify bundles of ES (Liu *et al.*, 2019).

In 16% of studies, ML clustering or dimensionality reduction was used in an additional methodological step for predictive modelling with a supervised learning technique. For example, Agglomerative Hierarchical Clustering was utilised to identify groups of structurally similar forest stands prior to the application of Random Forest to assess importance of structural variables on carbon storage (Thom and Keeton, 2019); and K-means cluster analysis was used to identify areas of homogeneous sets of species prior to the predictive modelling of floodplain biodiversity using a Bayesian Belief Network (BBN) (Funk *et al.*, 2019).

3.2. ML for remote-sensing and image recognition

ML was implemented in publications using remotely sensed data (53%; n= 308) for feature extraction or the classification of remotely sensed images to produce land cover maps (Zhang *et al.*, 2016; Traganos and Reinartz, 2018; Erker *et al.*, 2019; Pouliot *et al.*, 2019; Trinder and Liu, 2020) or landscape or vegetation feature extraction from remotely sensed images (Chen *et al.*, 2018; Jiang *et al.*, 2018; Dash *et al.*, 2019; Fujimoto *et al.*, 2019). In other studies (12%), remotely sensed data is used but as one of a range of spatially explicit predictor variables to model, e.g., carbon storage (Sanderman *et al.*, 2018; Silveira *et al.*, 2019; Havinga *et al.*, 2020), land use and ES change (Liu, 2014; Mahmoud and Gan, 2018; Hashimoto *et al.*, 2019), or for other ecological predictions such as Bark Beetle outbreaks (Rammer and Seidl, 2019). Ten studies utilised Deep Learning (an example of a Convolutional Neural Network which is a type of ANN) to model spatial-temporal dynamics from remote sensing images (Poggio, Lassauce and Gimona, 2019; Rammer and Seidl, 2019; Barbierato *et al.*, 2020; Du *et al.*, 2020; Samarin *et al.*, 2020; Timilsina, Aryal and Kirkpatrick, 2020; Arruda *et al.*, 2021; Bhargava, Sarkar and Friess, 2021; Caretti, Bohnenstiehl and Eggleston, 2021; Guo *et al.*, 2021).

ML was also utilised in descriptive image recognition tasks, such as cultural ES studies involving the analysis of large datasets from social media platforms using an ANN (3%). Online ANN image analysis models, specifically Deep Convolutional Neural Networks on cloud computing platforms Google Cloud Vision (*Google Cloud Vision*, 2021) and Clarifai (*Clarifai General Model*, 2021), were used to analyse the thematic content of user uploaded geo-tagged photographs on Flickr and clustering algorithms were used to group the photographs according to the themes. These themes were used as indicators of cultural ES, and were combined with spatial and temporal information associated with the photographs, enabling modelled cultural ES mapping (Richards and Tunçer, 2018; Bernetti, Chirici and Sacchelli, 2019; Gosal *et al.*, 2019; Chang *et al.*, 2020; Gosal and Ziv, 2020; Runge *et al.*, 2020) Similarly, an ANN image analysis model was used to classify geo-tagged photographs from Wikiloc – a sports photo-sharing platform – (*Wikiloc* , 2021) according to thematic content, and inferred cultural ES were mapped (Callau *et al.*, 2019).

3.3. ML for predictive modelling

ML was used in predictive modelling in 44% (n = 308) of publications. A wide range of ML techniques were used for predictive modelling of ES (Fig. 2). Classification and Regression Trees (CARTs) – a form of supervised learning (Box 1) – are the most widely used approach (60 %, n = 308; Fig. 2.), and Random Forest (RF) (44 %; Fig.2.) is an especially popular example of a CART. CARTs were used in supervised classification tasks to predict membership of a user-labelled class. For example, RF was used in the process of modelling timber production by predicting the age-class of forestry tree species from remotely sensed and historic forestry data (Gao *et al.*, 2016). CARTs were also used in supervised regression tasks. For example, RF was used in modelling carbon-diversity hotspots in agricultural soil from remote sensing, terrain and climate variables (Silveira *et al.*, 2019) and a regression tree model was used to predict soil carbon stocks under future land use and climate change from soil survey data (Adhikari *et al.*, 2019).

ES studies have used other supervised ML techniques in predictive modelling, 26% used an ANN, 4% used a BBN, 24 % used a Support Vector Machine (SVM). For example, an ANN was used in a regression task to predict rice crop yields from environmental and socio-economic variables (Dang *et al.*, 2019), and a BBN was used in a classification task to predict firewood use from environmental and socio-economic variables (Willcock *et al.*, 2018). ANNs were also used to predict future land use change (e.g. Akinyemi and Mashame, 2018; Beygi Heidarlou *et al.*, 2019; Hashimoto *et al.*, 2019). In addition to the prediction of target variables some techniques, most notably CARTs, were used to assess variable importance or for the selection of relevant predictor variables. For example, RF was used to identify the most important variables controlling organic carbon stocks in agricultural soils (Mayer *et al.*, 2019) and in forest stands (Thom and Keeton, 2019), and a CART was used to assess variable importance for the supply of a range of provisioning and regulating ES in an agroecosystem (Rositano *et al.*, 2018).

3.4. Repeatability, model tuning and generalisability.

Altering ML model settings can optimise model performance (Box 3). However, 67% (n =308; Fig. 3) of publications reviewed applied ML techniques 'off-the-shelf' without reporting any experimentation by altering model settings or model tuning. Indeed, 43 % of publications did not report the model settings used. 33% of all publications (n = 308) report model tuning (35% of organisational publications [n =34]; 35% of remote sensing [n = 162]; 23% of predictive ES direct [n = 56]; and 36% of predictive indirect [n =80]).

56% of all publications report model settings used (50%, organisational; 57%, remote sensing; 50% predictive direct; 63% predictive indirect). For those studies that do detail the model setting used, but do not experiment with model tuning, 51% (n = 102) give justification with reference to literature, but the rest of the studies provide no explanatory justification for the use of the particular model settings chosen. Some publications (4%; n = 308) do not report in their methods the kind of data used (e.g., categorical or nominal) as input or output in the ML model. Most publications (61%) give a rationale for the use of ML rather than an alternative modelling approach, but many studies do not. Publications tend to detail the software and the version used, but 28% do not report what software is used to carry out the ML technique. Model input data is sometimes freely available via supplementary material or an open data source but this is not the case in half of publications. Less than half of all publications reviewed report testing the generalisability of the ML model (Box 3) used within their study with an independent data set (41%, all publications).

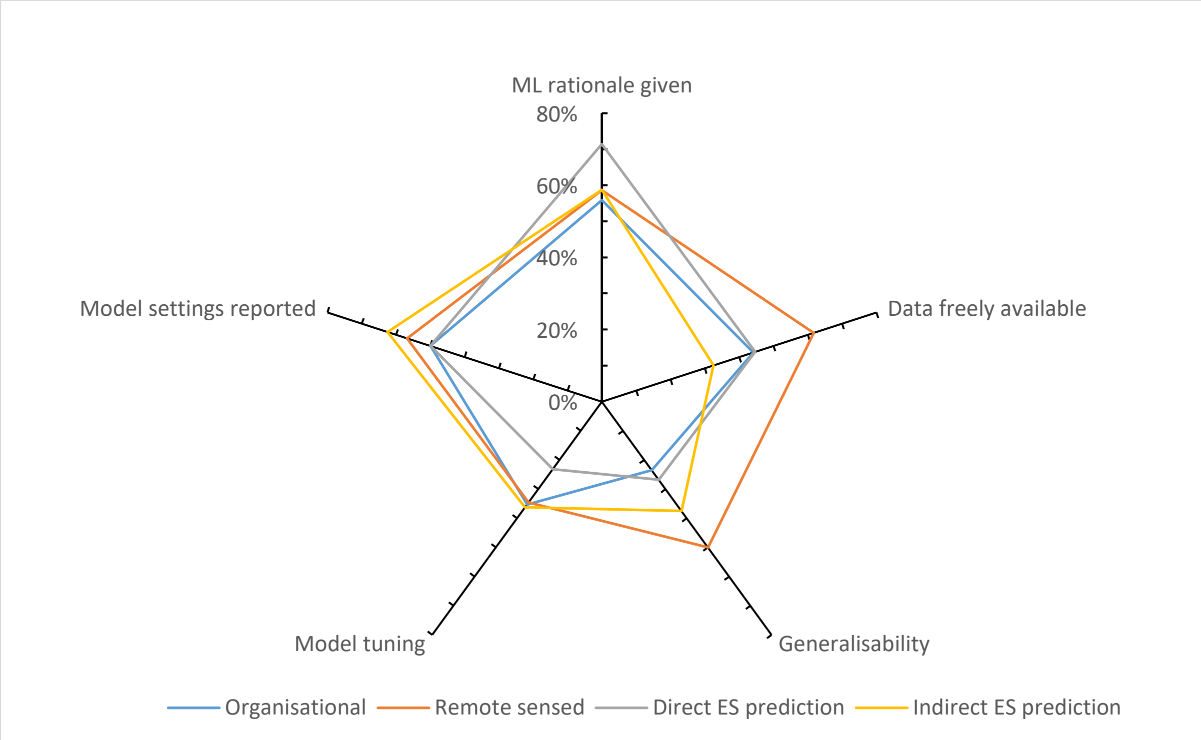

Fig.2. The percentage of publications which reported a rationale for using a machine learning technique, used data freely available data, reported model generalisability, model tuning and model settings used, categorised according to type of Machine Learning used: Organisational (n = 34); Remotely sensed (n = 162); Predictive direct (n = 56); Predictive indirect (n = 80).

Table 1. Median, maximum, and minimum number of datapoints (N), and variables (θ), used in each category of ML publications: Organisational; Remotely sensed; Predictive direct; Predictive indirect; All publications reviewed. NB. Number of samples (n) is lower than total number of publications reviewed for each category because not all publications reported number of datapoints, or variables used.

	Descriptive				Predictive				All publications reviewed	
	Organisational		Remotely sensed		Direct		Indirect			
	N	θ	N	θ	N	θ	N	θ	N	θ
Median	965	12	1509	12	1714	16	669	12	1138	13
Max	111884	363	2190763	2150	805684	2317	9500430	363	9500430	2317

Min	17	3	25	3	17	3	21	3	16	3
n	29	26	107	95	47	42	64	71	225	215

N = number of datapoints. Θ = number of variables

Box 2. Examples of ecosystem service (ES) studies using machine learning (ML) that demonstrate the benefits of ML approaches.

Many of the papers we reviewed highlight the benefits ES science can derive by adopting ML methods:

- **Big data** – ML allows for the rapid processing of data and one of its key strengths is that it can support analysis of larger datasets than many conventional methods (Reichstein *et al.*, 2019). Richards and Tunçer (2018) analyse over 20,000 images uploaded to photo sharing platform Flickr. They used Google Cloud Vision (an ML algorithm for image analysis) (*Google Cloud Vision*, 2021) to classify the thematic content of the images to map recreational beneficiaries. The time required to manually classify so many images would make this task impractical without the use of ML.

- **Clustering** – ML enables the grouping of data without the use of domain-specific theory. In ES science this can have useful application to identify bundles of ES provision or groups of ES beneficiaries. Schirpke *et al.* (2019) use K-means cluster analysis to identify areas where ES repeatedly occur together in the European Alps. Gosal *et al*. (2019) use the Ward-D clustering algorithm to identify six groups of recreational beneficiaries in the Camargue based on annotation of photos uploaded to Flickr.

- **Uncertainty measures** – Transparent estimates of model uncertainty are produced as an integral part of many ML predictive modelling algorithms. These measures can be useful to decision makers who can determine acceptable levels of uncertainty and use their own expertise for potentially contentious decisions. Willcock *et al. (*2018) model fuel use in South Africa and biodiversity in Sicily using ML Bayesian Belief Networks. They report associated estimates of uncertainty which were produced as part of the modelling process and highlight that the level of certainty might influence management decisions as well as the predicted level of ES.

- **Hypothesis generation and variable importance assessment** – In addition to the prediction of target variables, ML allows for the assessment of variable importance and the selection of relevant predictor variables. Mayer *et al* (2019) use the Random Forest algorithm (an example of a classification and regression tree) to identify the most important variables controlling organic carbon stocks in agricultural soils in Bavaria. They input 13 predictor variables and the algorithm identified the variables that explained the majority of variance in carbon stocks. This identification of important variables aids in the generation of hypotheses, e.g., theory about why these variables determine carbon stocks.

4. Discussion

Machine learning (ML) is used in ecosystem service (ES) research as both a descriptive tool, where aspects of automation enable speedy processing of high volumes of complex data, and in predictive modelling, in which accurate predictions can be made about ES. The variety of ways by which ML is incorporated in ES research methodologies highlights its value as an adaptable extension to traditional data analyses across all ES domains. Supervised ML approaches such as Classification and Regression Trees (CART) and Artificial Neural Networks (ANN) algorithms tend to be used for predictive model tasks, whilst descriptive tasks are often carried out using unsupervised ML, such as clustering algorithms to group data (Fig .1). While there are examples of studies that apply ML with a repeatable and rigorous methodology (Box 3), many studies fall short of methodological best practice; failing to report which software was used, model settings or tuning, or test of generalisability (Fig. 3). In some instances, these methodological shortcomings affect the repeatability of the study, such as not being able to identify the exact algorithm used, but in other instances they might mean that the findings of the study may be flawed. We suggest that future studies may use the findings of poorly reported models, but should do so with caution. Such models may well be valid, but the lack of repeatability means that that validity cannot be independently tested. For example, algorithm parameter optimisation has been shown to affect ML model accuracy (Daelemans *et al.*, 2003), so using default model settings might lead to reduced model performance. Thus, if a paper does not report model tuning then it is likely that the authors used the default parameters in the model settings in the relevant software. This may mean that, given the data the authors had at their disposal, the model presented may not be the best fit model to that data, and likely has higher uncertainty than could be achieved if tuning was performed. Similarly, without testing generalisability on an independent dataset, a ML model might be 'overfitted' to the data, this results in poor model accuracy when applied to new data from a parameter space that was not used to train the model, and so this should be done with caution (Hawkins, 2004; Kuhn and Johnson, 2013).

The potential impact of these methodological shortcomings varies with the type of ML approach used and the task for which the ML is being used. For example, the effect of altering algorithm hyperparameters away from defaults (Box 3) varies between ML techniques; e.g. increasing the of number of tree splits in a Random Forest above the default setting may have a marginal effect on model accuracy (Kulkarni and Sinha, 2012) compared to large effect on model performance that can result from altering the number of hidden layers in an ANN (Srivastava *et al.*, 2014). However, this largely depends on the problem at hand, therefore an investigation of hyperparameters is always recommended. Likewise, there is arguably less need to test for generalisability when, for example, using a CART to estimate variable importance, as compared to the need to a predictive classification model, because an estimation of variable importance does not explicitly generalise beyond the learnt parameter space (Kuhn and Johnson, 2013). Furthermore, for some descriptive tasks, testing generalisability may not be necessary; such as for some basic data sorting tasks or in applications to aid hypothesis generation (Lucas, 2020).

We found some examples of studies that use large and complex datasets (Box 2), but the capacity of ML to analyse available ‘big data’ has not yet been fully realised in ES research (Table 1). In remote sensing studies, large amounts of data are generated from satellites and manned and unmanned aerial vehicles. Automation in ML allows for rapid and accurate processing of these datasets (Lary *et al.*, 2016). Due to its capacity to process data of high dimensionality and to map classes with complex characteristics, ML is an effective and efficient geoscientific classification method, and now the standard approach for remote sensing image classification (Maxwell, Warner and Fang, 2018). In ES research, classification of remotely sensed images can provide estimates of the spatial distribution of ES supply via mapping of ES proxies, such as land use and land use change (Martnez-Harms and Balvanera, 2012) or factors that drive ES supply namely, ecosystem service providers, ecosystem processes and functional traits (Andrew *et al.*, 2015). That remote sensing ML methods tend to have a higher degree of repeatability and generalisability and utilise larger datasets compared to other methods (Fig. 5, Table 1) is likely testament to the maturity of the use of ML in the field of remote sensing. However, it suggests the under-utilisation of ML in other areas of ES research not associated with remote sensing, or that other areas of research have not amassed such high amounts of data.

In conducting our review, we noticed that the use of ML in ES research perhaps focuses on predictive modelling of the potential biophysical supply of ES, and often indirectly via ES proxies such as landcover or via hypothesised service providers. In these areas of ES research, ML can offer advantages over process-based models and standard statistical modelling in terms of improved predictive accuracy and ability to make use of disparate kinds of data. However, this is a relatively narrow subset of ES research and there is scope for further utilisation of ML in other areas, including ES demand and flows. For example, ES can be defined in terms of interactions between the service provider and service beneficiaries. In this sense they are co-produced, and to inform land management and policy decisions, ES research needs to quantify supply of ES relative to demand (Burkhard *et al.*, 2012).

Thus, ES modelling could better incorporate social science data (Daw *et al.*, 2016). This has been explored in part with the analysis of large datasets from social media platforms using deep convolutional neural networks (DCNNs; e.g. the automated content analysis tool, Google's Cloud Vision (*Google Cloud Vision*, 2021); Gosal *et al.*, 2019), which highlights the potential for ML to utilise very large social media datasets (Runge *et al.*, 2020b). However, to date, ES studies utilising social media have been largely limited to data from single social media platforms and there is further potential to use ML with a variety of social media platforms to analyse cultural ES (e.g. Ruiz-Frau *et al.*, 2020). More generally, social science datasets potentially relevant to ES research seem yet to be utilised. For example, it has been established there is a need to better understand the flows of ES beneficiaries (Bagstad *et al.*, 2013) and to better incorporate ES demand into predictive models (Martínez-López *et al.*, 2019). However, whilst big data from social science has recently been used effectively in some disciplines (e.g. in the development human mobility theory (Alessandretti, Aslak and Lehmann, 2020), such data has yet to be used by ES researchers. The availability of big data from social science together with the capacity of ML to both effectively utilise data from mixed sources and deal with a high number of variables, suggests that ML could be used in a more holistic system-scale modelling approach that captures the co-productive nature of ES.

The use of ML in ES research, whilst increasing, is still in its infancy. As such, ES scientists can benefit greatly from the experience of other disciplines. For example, recent developments in deep learning algorithms have enabled detailed modelling of spatial-temporal dynamics in the Earth Sciences (Reichstein *et al.*, 2019) and this is potentially applicable in a dynamic holistic ES modelling approach. In addition, hybrid ML models, which combine purely data-driven machine learning modelling with theory-bound, process-driven approaches, have been shown to have improved predictive power outside of the learnt parameter space in areas such as climate science (Huntingford *et al.*, 2019) and could be useful in the development of more transferable ES models.

In conclusion, this review found that a wide range of ML approaches have been used effectively in a variety of ES studies and that ML offers exciting potential in future ES research. However, for the full potential of ML in ES to be realised and confidently used by stakeholders, ML models should be transparently reported and readily repeatable (Martínez-López et al., 2019). Our review identifies 'gold standard' studies that exemplify methodological best practice and could be used as a benchmark for ML reporting in ES research.

Box 3. 'Gold-standard' ecosystem service (ES) studies using machine learning (ML), demonstrating best practice.

Our review of 200 ES papers using ML revealed a wide range in their ML protocols. Here, we highlight a sample of papers that we consider provide 'gold-standard' or best practice for key aspects of ML reporting.

• **Methodological transparency** – Each application of ML needs to be fully repeatable. As such, the input data used should be available to other researchers. Ideally the data would be open access and links to data sources provided in the publication. Furthermore, any model settings, the software used and relevant computer code necessary to run the model should be reported. Funk *et al.* (2019) is a good example of transparent reporting of ML methods. The authors develop a data-driven Bayesian Network to prioritise areas of floodplain for management interventions in the Danube River based on ES multifunctionality. They use open access data and provide links to all data sources. In addition, they detail data-discretisation (i.e., the method used to group data into discrete categorises as input to the model) and model parameterisation and the software used (i.e., they fully describe their approach to model development and validation).

• **Model tuning** – Hyperparameters are aspects of model architecture that can be altered by the user to optimise model performance. Many ML techniques have hyperparameters that can (and often should) be varied by the user. For example, the number of tree splits in a decision tree, and the number of layers in a neural network, are hyperparameters that may affect model performance. Such changes may alter the model outputs so, at the very least, authors should report the hyperparameter values used and, where appropriate, justify these hyperparameter settings. Rammer and Seidl (2019) provide a good example of how to investigate the sensitivity of the ML hyperparameters and hone them to form the most accurate model. They develop a Deep Neural Network to predict bark beetle outbreaks and systematically evaluate different network architectures to optimise predictive power. They alter model structure such as network size and parameters of the training process including the loss function and optimizer used. They report iteratively the evaluation of model variations by calculating performance measures including model accuracy, precision, recall, F1 Score, Conditional Kappa and True Skill Statistic for each model run. The source code they use to build the model can be found here: https://github.com/werner-rammer/BBPredNet.

- **Generalisability** – Model testing, where model performance is tested using a random subset of the data not used to train the algorithm, is an integral part of most supervised learning algorithms. However, without validation against an independent dataset outside the parameter space of the training-testing data, a ML model might be 'overfitted' and not generalisable to other spaces/times (Hawkins 2004). This can result in poor model accuracy when applied to new data which was not used to train the model (Alpaydin, 2020). This can be overcome by dividing the training dataset in two: with one set used for training and the other for testing generalisability, or by additionally testing the model on a dataset outside the learnt parameter space. For example, Hashimoto et al. (2019) use historical land use data to predict future land use change using an Artificial Neural Network. They model land use change using historic land use data for 1997 and 2006 and randomly split 50% of the data for training and 50% for testing the model (n = 1275), but also reserve an independent data set (data for 2014) for testing model generalisability (n=1275).

Acknowledgements

This work was supported by the Natural Environment Research Council (NERC) via the Envision DTP.

References

Adhikari, K. *et al.* (2019) 'Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change', *Science of The Total Environment*, 667, pp. 833–845. doi: 10.1016/J.SCITOTENV.2019.02.420.

Akinyemi, F. O. and Mashame, G. (2018) 'Analysis of land change in the dryland agricultural landscapes of eastern Botswana', *Land Use Policy*, 76, pp. 798–811. doi: 10.1016/J.LANDUSEPOL.2018.03.010.

Alessandretti, L., Aslak, U. and Lehmann, S. (2020) 'The scales of human mobility', *Nature*, 587(7834), pp. 402–407. doi: 10.1038/s41586-020-2909-1.

Andrew, M. E. *et al.* (2015) 'Spatial data, analysis approaches, and information needs for spatial ecosystem service assessments: A review', *GIScience and Remote Sensing*. Taylor and Francis Inc., pp. 344–373. doi: 10.1080/15481603.2015.1033809.

Arruda, V. L. S. *et al.* (2021) 'An alternative approach for mapping burn scars using Landsat imagery, Google Earth Engine, and Deep Learning in the Brazilian Savanna', *Remote Sensing Applications: Society and Environment*, 22, p. 100472. doi: 10.1016/J.RSASE.2021.100472.

Bagstad, K. J. *et al.* (2013) 'Spatial dynamics of ecosystem service flows: A comprehensive approach to quantifying actual services', *Ecosystem Services*, 4, pp. 117–125. doi: 10.1016/j.ecoser.2012.07.012.

Barbierato, E. *et al.* (2020) 'Integrating remote sensing and street view images to quantify urban forest ecosystem services', *Remote Sensing*, 12(2), p. 329. doi: 10.3390/rs12020329.

Bernetti, I., Chirici, G. and Sacchelli, S. (2019) ‘Big data and evaluation of cultural ecosystem services: an analysis based on geotagged photographs from social media in Tuscan forest (Italy)’, *http://iforest.sisef.org/*, 12(1), p. 98. doi: 10.3832/IFOR2821-011.

Beygi Heidarlou, H. *et al.* (2019) ‘Effects of preservation policy on land use changes in Iranian Northern Zagros forests’, *Land Use Policy*, 81, pp. 76–90. doi: 10.1016/J.LANDUSEPOL.2018.10.036.

Bhargava, R., Sarkar, D. and Friess, D. A. (2021) ‘A cloud computing-based approach to mapping mangrove erosion and progradation: Case studies from the Sundarbans and French Guiana’, *Estuarine, Coastal and Shelf Science*, 248, p. 106798. doi: 10.1016/j.ecss.2020.106798.

Bock, F. E. *et al.* (2019) ‘A review of the application of machine learning and data mining approaches in continuum materials mechanics’, *Frontiers in Materials*. Frontiers Media S.A. doi: 10.3389/fmats.2019.00110.

Breiman, L. (2001) *Random Forests*. Available at: https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf (Accessed: 4 December 2018).

Burkhard, B. *et al.* (2012) ‘Mapping ecosystem service supply, demand and budgets’, *Ecological Indicators*, 21, pp. 17–29. doi: 10.1016/j.ecolind.2011.06.019.

Callau, A. À. *et al.* (2019) ‘Landscape characterization using photographs from crowdsourced platforms: Content analysis of social media photographs’, *Open Geosciences*, 11(1), pp. 558–571. doi: 10.1515/geo-2019-0046.

Caretti, O. N., Bohnenstiehl, D. W. R. and Eggleston, D. B. (2021) ‘Spatiotemporal Variability in Sedimentation Drives Habitat Loss on Restored Subtidal Oyster Reefs’, *Estuaries and Coasts*, pp. 1–18. doi: 10.1007/s12237-021-00921-6.

Chang, C. chen *et al.* (2020) ‘Social media, nature, and life satisfaction: global evidence of the biophilia hypothesis’, *Scientific Reports*, 10(1), pp. 1–8. doi: 10.1038/s41598-020-60902-w.

Chen, L., Roy, S. B. and Hutton, P. H. (2018) ‘Emulation of a process-based estuarine hydrodynamic model’. doi: 10.1080/02626667.2018.1447112.

Chen, Y. *et al.* (2018) ‘Extraction of Urban Water Bodies from High-Resolution Remote-Sensing Imagery Using Deep Learning’, *Water*, 10(5), p. 585. doi: 10.3390/w10050585.

Clarifai General Model | Recognize Over 11,000 Concepts in Images (no date). Available at: https://www.clarifai.com/models/general-image-recognition (Accessed: 11 March 2021).

Crisci, C., Ghattas, B. and Perera, G. (2012) ‘A review of supervised machine learning algorithms and their applications to ecological data’, *Ecological Modelling*. Elsevier, pp. 113–122. doi: 10.1016/j.ecolmodel.2012.03.001.

Cutler, D. R. *et al.* (2007) ‘Random forests for classification in ecology’, *Ecology*, 88(11), pp. 2783–2792. doi: 10.1890/07-0539.1.

Daelemans, W. *et al.* (2003) ‘Combined optimization of feature selection and algorithm parameters in machine learning of language’, in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*. Springer Verlag, pp. 84–95. doi: 10.1007/978-3-540-39857-8_10.

Dang, K. B. *et al.* (2019) ‘Application of a hybrid neural-fuzzy inference system for mapping crop suitability areas and predicting rice yields’, *Environmental Modelling & Software*, 114, pp. 166–180. doi: 10.1016/J.ENVSOFT.2019.01.015.

Dash, J. P. *et al.* (2019) ‘Early Detection of Invasive Exotic Trees Using UAV and Manned Aircraft Multispectral and LiDAR Data’, *Remote Sensing*, 11(15), p. 1812. doi: 10.3390/rs11151812.

Daw, T. M. *et al.* (2016) ‘Elasticity in ecosystem services: exploring the variable relationship between ecosystems and human well-being’, 21(2). doi: 10.5751/ES-08173-210211.

Delen, D. and Ram, S. (2018) ‘Research challenges and opportunities in business analytics’, *Journal of Business Analytics*, 1(1), pp. 2–12. doi: 10.1080/2573234X.2018.1507324.

Du, L. *et al.* (2020) ‘Mapping forested wetland inundation in the delmarva peninsula, USA using deep convolutional neural networks’, *Remote Sensing*, 12(4), p. 644. doi: 10.3390/rs12040644.

Erker, T. *et al.* (2019) ‘A statewide urban tree canopy mapping method’, *Remote Sensing of Environment*, 229, pp. 148–158. doi: 10.1016/J.RSE.2019.03.037.

Fujimoto *et al.* (2019) ‘An End to End Process Development for UAV-SfM Based Forest Monitoring: Individual Tree Detection, Species Classification and Carbon Dynamics Simulation’, *Forests*, 10(8), p. 680. doi: 10.3390/f10080680.

Funk, A. *et al.* (2019) ‘Identification of conservation and restoration priority areas in the Danube River based on the multi-functionality of river-floodplain systems’, *Science of The Total Environment*, 654, pp. 763–777. doi: 10.1016/J.SCITOTENV.2018.10.322.

Gao, T. *et al.* (2016) ‘Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history’, *International Journal of Applied Earth Observation and Geoinformation*, 52, pp. 155–165. doi: 10.1016/J.JAG.2016.06.004.

García-Laencina, P. J., Sancho-Gómez, J. L. and Figueiras-Vidal, A. R. (2010) ‘Pattern classification with missing data: A review’, *Neural Computing and Applications*, 19(2), pp. 263–282. doi: 10.1007/s00521-009-0295-6.

Ghani, N. A. *et al.* (2019) ‘Social media big data analytics: A survey’, *Computers in Human Behavior*, 101, pp. 417–428. doi: 10.1016/j.chb.2018.08.039.

Google Cloud Vision (no date). Available at: https://cloud.google.com/vision/ (Accessed: 11 March 2021).

Gosal, A. S. *et al.* (2019) ‘Using social media, machine learning and natural language processing to map multiple recreational beneficiaries’, *Ecosystem Services*, 38, p. 100958. doi: 10.1016/j.ecoser.2019.100958.

Gosal, A. S. and Ziv, G. (2020) ‘Landscape aesthetics: Spatial modelling and mapping using social media images and machine learning’, *Ecological Indicators*, 117, p. 106638. doi: 10.1016/j.ecolind.2020.106638.

Guo, M. *et al.* (2021) ‘Me-net: A deep convolutional neural network for extracting mangrove using sentinel-2A data’, *Remote Sensing*, 13(7), p. 1292. doi: 10.3390/rs13071292.

Hashimoto, S. *et al.* (2019) ‘Scenario analysis of land-use and ecosystem services of social-ecological landscapes: implications of alternative development pathways under declining population in the Noto Peninsula, Japan’, *Sustainability Science*, 14(1), pp. 53–75. doi: 10.1007/s11625-018-0626-6.

Havinga, I. *et al.* (2020) ‘Spatial quantification to examine the effectiveness of payments for ecosystem services: A case study of Costa Rica’s Pago de Servicios Ambientales’, *Ecological Indicators*, 108, p. 105766. doi: 10.1016/J.ECOLIND.2019.105766.

Hawkins, D. M. (2004) ‘The Problem of Overfitting’. doi: 10.1021/ci0342472.

Hirayama, H. *et al.* (2019) ‘Evaluating multiple classifier system for the reduction of salt-and-pepper noise in the classification of very-high-resolution satellite images’, *International Journal of Remote Sensing*, 40(7), pp. 2542–2557. doi: 10.1080/01431161.2018.1528400.

HOCHACHKA, W. M. *et al.* (2007) ‘Data-Mining Discovery of Pattern and Process in Ecological Systems’, *Journal of Wildlife Management*, 71(7), p. 2427. doi: 10.2193/2006-503.

Huntingford, C. *et al.* (2019) ‘Machine learning and artificial intelligence to aid climate change research and preparedness’, *Environmental Research Letters*, 14(12), p. 124007. doi: 10.1088/1748-9326/ab4e55.

Introduction to Machine Learning - Ethem Alpaydin - Google Books (no date). Available at: https://books.google.co.uk/books?hl=en&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&dq=Alpaydin,+2020+machine+learning&ots=F3RV819swh&sig=FnLG7fDy0O-DXazcoCfLJHTECik&redir_esc=y#v=onepage&q=Alpaydin%2C 2020 machine learning&f=false

(Accessed: 26 November 2020).

Jiang, W. *et al.* (2018) ‘Multilayer Perceptron Neural Network for Surface Water Extraction in Landsat 8 OLI Satellite Images’, *Remote Sensing*, 10(5), p. 755. doi: 10.3390/rs10050755.

Jordan, M. I. and Mitchell, T. M. (2015) ‘Machine learning: Trends, perspectives, and prospects.’, *Science (New York, N.Y.)*, 349(6245), pp. 255–60. doi: 10.1126/science.aaa8415.

Kitchin, R. and McArdle, G. (2016) ‘What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets’, *Big Data & Society*, 3(1), p. 205395171663113. doi: 10.1177/2053951716631130.

Knudby, A., Brenning, A. and LeDrew, E. (2010) ‘New approaches to modelling fish–habitat relationships’, *Ecological Modelling*, 221(3), pp. 503–511. doi: 10.1016/J.ECOLMODEL.2009.11.008.

Kuhn, M. and Johnson, K. (2013) *Applied predictive modeling*, *Applied Predictive Modeling*. Springer New York. doi: 10.1007/978-1-4614-6849-3.

Kulkarni, V. Y. and Sinha, P. K. (2012) ‘Pruning of random forest classifiers: A survey and future directions’, in *Proceedings - 2012 International Conference on Data Science and Engineering, ICDSE 2012*, pp. 64–68. doi: 10.1109/ICDSE.2012.6282329.

Lary, D. J. *et al.* (2016) ‘Machine learning in geosciences and remote sensing’, *Geoscience Frontiers*, 7(1), pp. 3–10. doi: 10.1016/j.gsf.2015.07.003.

Lee, H. *et al.* (2019) ‘Mapping cultural ecosystem services 2.0 – Potential and shortcomings from unlabeled crowd sourced images’, *Ecological Indicators*, 96, pp. 505–515. doi: 10.1016/j.ecolind.2018.08.035.

Liu, Y. (2014) ‘Dynamic evaluation on ecosystem service values of urban rivers and lakes: A case study of Nanchang City, China’, *Aquatic Ecosystem Health & Management*, 17(2), pp. 161–170. doi: 10.1080/14634988.2014.907223.

Liu, Y. *et al.* (2019) 'Landscape functional zoning at a county level based on ecosystem services bundle: Methods comparison and management indication', *Journal of Environmental Management*, 249, p. 109315. doi: 10.1016/j.jenvman.2019.109315.

Lucas, T. C. D. (2020) 'A translucent box: interpretable machine learning in ecology', *Ecological Monographs*, 90(4). doi: 10.1002/ecm.1422.

Mahmoud, S. H. and Gan, T. Y. (2018) 'Impact of anthropogenic climate change and human activities on environment and ecosystem services in arid regions', *Science of The Total Environment*, 633, pp. 1329–1344. doi: 10.1016/J.SCITOTENV.2018.03.290.

Martín-López, B. *et al.* (2012) 'Uncovering Ecosystem Service Bundles through Social Preferences', *PLoS ONE*. Edited by K. Bawa, 7(6), p. e38970. doi: 10.1371/journal.pone.0038970.

Martín-Martín, A. *et al.* (2018) 'Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories', *Journal of Informetrics*, 12(4), pp. 1160–1177. doi: 10.1016/j.joi.2018.09.002.

Martínez-López, J. *et al.* (2019) 'Towards globally customizable ecosystem service models', *Science of The Total Environment*, 650, pp. 2325–2336. doi: 10.1016/J.SCITOTENV.2018.09.371.

Martnez-Harms, M. J. and Balvanera, P. (2012) 'Methods for mapping ecosystem service supply: A review', *International Journal of Biodiversity Science, Ecosystem Services and Management*. Taylor and Francis Ltd., pp. 17–25. doi: 10.1080/21513732.2012.663792.

Maxwell, A. E., Warner, T. A. and Fang, F. (2018) 'International Journal of Remote Sensing Implementation of machine-learning classification in remote sensing: an applied review Implementation of machine-learning classification in remote sensing: an applied review'. doi: 10.1080/01431161.2018.1433343.

Mayer, S. *et al.* (2019) 'Controlling factors of organic carbon stocks in agricultural topsoils and subsoils of Bavaria', *Soil and Tillage Research*, 192, pp. 22–32. doi: 10.1016/J.STILL.2019.04.021.

MEA (2005) *M I L L E N N I U M E C O S Y S T E M A S S E S S M E N T*. Available at: www.islandpress.org (Accessed: 24 May 2019).

Mouchet, M. A. *et al.* (2014) 'An interdisciplinary methodological guide for quantifying associations between ecosystem services', *Global Environmental Change*, 28, pp. 298–308. doi: 10.1016/J.GLOENVCHA.2014.07.012.

Nguyen, T. H., Nong, D. and Paustian, K. (2019) 'Surrogate-based multi-objective optimization of management options for agricultural landscapes using artificial neural networks', *Ecological Modelling*, 400, pp. 1–13. doi: 10.1016/j.ecolmodel.2019.02.018.

Poggio, L., Lassauce, A. and Gimona, A. (2019) 'Modelling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region', *Geoderma*, 346, pp. 63–74. doi: 10.1016/j.geoderma.2019.03.017.

Pouliot, D. *et al.* (2019) 'Assessment of convolution neural networks for wetland mapping with landsat in the central Canadian boreal forest region', *Remote Sensing*, 11(7), p. 772. doi: 10.3390/rs11070772.

Rammer, W. and Seidl, R. (2019) 'Harnessing Deep Learning in Ecology: An Example Predicting Bark Beetle Outbreaks', *Frontiers in Plant Science*, 10, p. 1327. doi: 10.3389/fpls.2019.01327.

Raudsepp-Hearne, C. *et al.* (2010) 'Ecosystem service bundles for analyzing tradeoffs in diverse landscapes', 107(11). doi: 10.1073/pnas.0907284107.

Reichstein, M. *et al.* (2019) 'Deep learning and process understanding for data-driven Earth system science', *Nature*, 566(7743), pp. 195–204. doi: 10.1038/s41586-019-0912-1.

Richards, D. R. and Tunçer, B. (2018) 'Using image recognition to automate assessment of cultural ecosystem services from social media photographs', *Ecosystem Services*, 31, pp. 318–325. doi: 10.1016/J.ECOSER.2017.09.004.

Rositano, F. *et al.* (2018) 'Identifying the factors that determine ecosystem services provision in Pampean agroecosystems (Argentina) using a data-mining approach', *Environmental Development*, 25, pp. 3–11. doi: 10.1016/J.ENVDEV.2017.11.003.

Ruiz-Frau, A. *et al.* (2020) 'Using graph theory and social media data to assess cultural ecosystem services in coastal areas: Method development and application', *Ecosystem Services*, 45, p. 101176. doi: 10.1016/j.ecoser.2020.101176.

Runge, C. A. *et al.* (2020) 'Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis', *Environmental Research Communications*, 2(7), p. 075001. doi: 10.1088/2515-7620/ab9c33.

Samarin, M. *et al.* (2020) 'Identifying soil erosion processes in alpine grasslands on aerial imagery with a u-net convolutional neural network', *Remote Sensing*, 12(24), pp. 1–21. doi: 10.3390/rs12244149.

Sanderman, J. *et al.* (2018) 'A global map of mangrove forest soil carbon at 30 m spatial resolution', *Environmental Research Letters*, 13(5), p. 055002. doi: 10.1088/1748-9326/aabe1c.

Sannigrahi, S. *et al.* (2019) 'Evaluating landscape capacity to provide spatially explicit valued ecosystem services for sustainable coastal resource management', *Ocean and Coastal Management*, 182, p. 104918. doi: 10.1016/j.ocecoaman.2019.104918.

Schirpke, U. *et al.* (2019) 'Integrating supply, flow and demand to enhance the understanding of interactions among multiple ecosystem services', *Science of The Total Environment*, 651, pp. 928–941. doi: 10.1016/J.SCITOTENV.2018.09.235.

Silveira, E. M. de O. *et al.* (2019) 'Carbon-diversity hotspots and their owners in Brazilian southeastern Savanna, Atlantic Forest and Semi-Arid Woodland domains', *Forest Ecology and Management*, 452, p. 117575. doi: 10.1016/J.FORECO.2019.117575.

Solomatine, D., See, L. M. and Abrahart, R. J. (2009) 'Data-Driven Modelling: Concepts, Approaches and Experiences', in *Practical Hydroinformatics*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 17–30. doi: 10.1007/978-3-540-79881-1_2.

Spake, R. *et al.* (2017) 'Unpacking ecosystem service bundles: Towards predictive mapping of synergies and trade-offs between ecosystem services', *Global Environmental Change*, 47, pp. 37–50. doi: 10.1016/j.gloenvcha.2017.08.004.

Srivastava, N. *et al.* (2014) *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, *Journal of Machine Learning Research*.

Thom, D. and Keeton, W. S. (2019) 'Stand structure drives disparities in carbon storage in northern hardwood-conifer forests', *Forest Ecology and Management*, 442, pp. 10–20. doi: 10.1016/J.FORECO.2019.03.053.

Timilsina, S., Aryal, J. and Kirkpatrick, J. B. (2020) 'Mapping urban tree cover changes using object-based convolution neural network (OB-CNN)', *Remote Sensing*, 12(18), p. 3017. doi: 10.3390/RS12183017.

Torney, C. J. *et al.* (2019) 'A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images', *Methods in Ecology and Evolution*. Edited by R. McCrea. doi: 10.1111/2041-210X.13165.

Traganos, D. and Reinartz, P. (2018) 'Mapping Mediterranean seagrasses with Sentinel-2 imagery', *Marine Pollution Bulletin*, 134, pp. 197–209. doi: 10.1016/J.MARPOLBUL.2017.06.075.

Trinder, J. and Liu, Q. (2020) 'Assessing environmental impacts of urban growth using remote sensing', *Geo-spatial Information Science*, pp. 1–20. doi: 10.1080/10095020.2019.1710438.

Wäldchen, J. and Mäder, P. (2018) 'Machine learning for image based species identification', *Methods in Ecology and Evolution*. Edited by N. Cooper, 9(11), pp. 2216–2225. doi: 10.1111/2041-210X.13075.

Wikiloc (2021). Available at: https://www.wikiloc.com/ (Accessed: 11 March 2021).

Willcock, S. *et al.* (2016) 'Do ecosystem service maps and models meet stakeholders' needs? A preliminary survey across sub-Saharan Africa', *Ecosystem Services*, 18, pp. 110–117. doi: 10.1016/j.ecoser.2016.02.038.

Willcock, S. *et al.* (2018) 'Machine learning for ecosystem services', *Ecosystem Services*, 33, pp. 165–174. doi: 10.1016/J.ECOSER.2018.04.004.

Willcock, S. *et al.* (2019) 'A Continental-Scale Validation of Ecosystem Service Models', *Ecosystems*, 22(8), pp. 1902–1917. doi: 10.1007/s10021-019-00380-y.

Willcock, S. *et al.* (2020) 'Ensembles of ecosystem service models can improve accuracy and indicate uncertainty', *Science of the Total Environment*, 747, p. 141006. doi: 10.1016/j.scitotenv.2020.141006.

Wu, C. K. *et al.* (2019) 'An IoT Tree Health Indexing Method Using Heterogeneous Neural Network', *IEEE Access*, 7, pp. 66176–66184. doi: 10.1109/ACCESS.2019.2918060.

Xia, J., Wang, J. and Niu, S. (2020) 'Research challenges and opportunities for using big data in global change biology', *Global Change Biology*. Blackwell Publishing Ltd, pp. 6040–6061. doi: 10.1111/gcb.15317.

Zednik, C. (2019) 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence', *Philosophy and Technology*, pp. 1–24. doi: 10.1007/s13347-019-00382-7.

Zhang, Z. *et al.* (2016) 'Object-Based Tree Species Classification in Urban Ecosystems Using LiDAR and Hyperspectral Data'. doi: 10.3390/f7060122.