



Characterising social-ecological drivers of landuse/cover change in a complex transboundary basin using singular or ensemble machine learning

Blessing Kavhu ^{a,b,c,*}, Zama Eric Mashimbye ^a, Linda Luvuno ^b

^a Department of Geography and Environmental Studies, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa

^b Centre for Sustainability Transitions, Stellenbosch University, Stellenbosch, 7600, South Africa

^c Scientific Services Unit, Zimbabwe Parks and Wildlife Management Authority, Headquarters, P. O. Box 140, Causeway, Harare, South Africa

ARTICLE INFO

Keywords:

Remote sensing
Landscape change
Water resources
Spatial modelling
Machine learning

ABSTRACT

Studies have focused on understanding land use/cover (LULC) change through regression techniques. However, machine learning (ML) techniques and their ensembles may provide more accurate results. Accurate determination of drivers of LULC can guide reliable land use planning and effective natural resource management. In this study, we tested the utility of ML techniques and ensemble modelling to explain the social-ecological drivers of LULC in the Okavango basin. The Deep Neural Network (DNN) coupled with climate-based regionalization of the study area were used for LULC classification of the years 2002, 2013 and 2020. Centroids of 22 LULC transitions of the same period were used to separately calibrate five (5) machine algorithms (namely Random Forests (RF), Gradient Boost Models (GBM) and Maximum Entropy (MaxEnt), Classification Tree Analysis (CTA) and their ensemble). Model performance was evaluated using the Receiver Operating Characteristic (ROC) and the True skill statistic (TSS). Variable importance was used to assess the contribution of social-ecological variables to each LULC transition. Variables that were determined to be driving LULC were then used to predict future LULC using the Artificial Neural Network and Cellular Automata (ANN-CA). Analysis results show that average LULC classification accuracy for the study period (2002, 2013 and 2020) was (Bsh Koppen zone; OA = 95.03, Kappa = 0.94), (Cwa Koppen zone; OA = 95.29, Kappa = 0.94), Cwb Koppen zone; OA = 96.04, Kappa = 0.95). The ML ensemble performed better (ROC > 95, TSS > 87) than singular ML models based on two separate model evaluation metrics. The Random Forest classifier outperformed other singular ML (ROC = 90.41, TSS = 84.2). Based on the top-performing ensemble model, distance from rivers, population density, annual average temperature, drought severity, fire frequency and distance from towns influence the conversion of natural to anthropogenic LULC classes (importance > 0.5). On the other hand, distance from rivers, soil organic carbon, precipitation, GDP, elevation, population density and annual average temperature importantly influenced conversion from one natural LULC class to another natural LULC class. The study revealed that natural classes (wetland, shrubland, water and woodlands) will gradually decrease at the expense of anthropogenic classes (built-up and cultivated) in future (2040). Despite proposing the necessity of a basin-wide land-use plan to minimise pressure on resources and ensure sustainable use, findings in this study illustrate the benefit of ensemble modelling in understanding LULC dynamics in trans-boundary basins.

* Corresponding author. Department of Geography and Environmental Studies, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa.
E-mail address: 24580538@sun.ac.za (B. Kavhu).

1. Introduction

Land use/cover (LULC) changes pose threats to social-ecological systems (SES) globally (Winkler et al., 2021; Opacka et al., 2021; Guo et al., 2021). According to Berkes (2011), SES refers to an integrated complex system that includes both social (human) and ecological (biophysical) subsystems in a two-way feedback relationship. Most of the changes that occur in SES are either due to anthropogenic activities or climate (Olson et al., 2008; Li et al., 2021; Kouassi et al., 2021). Not all climate-driven changes are easily detectable and manageable (Ellis et al., 2013). Changes due to climatic factors are mostly slower and subtle (Tasser et al., 2017). In some cases, climate-related changes may be beneficial to SES, for example when they facilitate the resilience of ecosystems (Schirpke et al., 2017). On the other hand, anthropogenic-driven changes are often rapid, direct, easy to detect and control measures to address them can be suggested (Sala et al., 2000; Ren et al., 2012; Li et al., 2021). Changes due to anthropogenic activities mostly replace the natural cover with artificial surfaces which can alter the state of ecosystem functions and services (Wang et al., 2021; Shiferaw et al., 2021; Yuan et al., 2021). Although identifying the drivers of LULC change can help to minimise the negative impacts of anthropogenic-driven LULC change, working on complex landscapes like Transboundary Drainage Basins (TDBs) can be challenging.

A common pattern observed in many studies is the continual loss of natural cover at the expense of increasing anthropogenic cover (Hishe et al., 2021; Doyle et al., 2021). For instance, the global urban extent is expanding at a rate of 9.687 km² per year (Liu et al., 2020). On the other hand, 2.3 million km² of forests were lost globally during the period 2000 to 2012 (Hansen et al., 2013). If these changes affect sensitive ecosystems which serve as habitats for species of high ecological value (Betts et al., 2017), the consequences can be enormous. TDBs are one of the most sensitive SES which support a large diversity of aquatic, floral and faunal species, and are also home to most of the endangered riparian and avian species (Sharma and Chettri, 2005). They constitute between 60 and 65% of the continent and are key livelihood support systems (Elhance, 1999; Domisch et al., 2019). LULC change has been one of the key environmental problems confronting the management of natural resources in TDBs. LULC change can alter the structure, diversity and function of SES by modifying nutrient cycling (Bickling et al., 2019), habitat fragmentation (Adhikari and Hansen, 2018), loss of native biodiversity (Shiferaw et al., 2021) and changing river flow regimes thus impacting on the quality and availability of water (Rimba et al., 2021; dos Santos et al., 2021). To develop and implement policies that can enhance monitoring and alter current LULC patterns in a transboundary context, it is vital to understand the drivers of LULC change.

One of the major obstacles that often hinder the monitoring of LULC change in TDBs is the intricacy associated with understanding drivers of change. Most TDBs are associated with a high degree of complexity in the political, social, economic, and biophysical factors that drive LULC change in countries that share the same drainage basin. Geographical Information Science (GIS) and Remote sensing (RS) offers the potential to identify changed sites and present social-economic and biophysical covariates of LULC change spatially (C. Singh et al., 2018; Chughtai et al., 2021). Spatial covariates coupled with spatial models are useful in explaining environmental changes across wide space. However, drivers of change have been largely understood using data acquired through social surveys (Chalmers and Fabricius, 2007; Kamwi et al., 2015; Simwanda et al., 2020; Gondwe, 2020). This could be mainly because manipulation of spatial data and calibration of spatial models require specialised skills. Unlike social surveys, spatial data coupled with spatial models provide extensive spatially explicit information which could accurately explain drivers of LULC change across extensive regions such as TDBs (Rai et al., 2018; Khoshnoodmotlagh et al., 2020). The use of robust spatial models in detecting drivers of LULC change is an important step towards creating effective mechanisms to change the negative effects of LULC in TDBs.

Different spatial models are often categorized according to the assumptions and the predictive nature of their algorithms. The common categories are machine learning (ML) and regression (RG) techniques. RG techniques mainly detect patterns in data by investigating the relationship between dependent and independent variables and an example of such is the logistic regression method (Uyanik and Güler, 2013). On the other hand, ML techniques build models based on known sample data and examples of these include the Random Forest (RF) and the Support Vector Machine (SVM) (Cracknell and Reading, 2014; Lary et al., 2016). Studies have mainly used spatial models based on RG techniques to understand drivers of LULC (Gibson et al., 2018; Kamwi et al., 2018; Marondedze and Schütt, 2019). Yet indications show that ML modelling techniques perform better in predictive analysis as compared to RG techniques (Camps-Valls and Bruzzone, 2009). Additionally, ML techniques have advanced in the past decades from analysing spectral data at the pixel to sub-pixel level particularly the Support vector regression (SVR) and Kernel Ridge Regression (KRR) (Zhang et al., 2014, 2018). Understanding the utility of ML techniques in evaluating the drivers of LULC, especially when operating in complex and extensive areas is crucial.

Studies have successfully used ML techniques such as Classification and Regression Tree (CART), Extreme Gradient Boost (XGboost), Random Forest (RF) in evaluating drivers of LULC in different environments (Phiri et al., 2019; Wang et al., 2020; Islam et al., 2021). However, other studies contend that single models are governed by a variety of assumptions that differ from place to place (Grenouillet et al., 2011). Hence, ensemble models have been used to overcome the shortcomings of singular models (Marmion et al., 2009; Buisson et al., 2010). Despite ML techniques' apparent potential, studies have not yet compared the utility of different ML techniques, let alone with their ensembles, in evaluating drivers of LULC change. Filling this gap is an important step towards accurately predicting future changes in the LULC and designing reliable land use plans for extensive areas such as TDBs.

This study used ML techniques and ensemble modelling to explain the social-ecological drivers of LULC in the Okavango basin. To do this, a set of singular ML techniques (namely, Random forest, Gradient boosting model, Maximum Entropy, Artificial Neural Network, Classification Tree Analysis) were used. Their performance was compared with that of their ensemble in teasing apart the drivers of LULC change. The specific objectives were to 1) compare the performance of singular and ensemble machine learning techniques in determining drivers of change, 2) identify the social-ecological drivers of LULC change in a complex transboundary basin, 3) to use the identified key drivers of LULC to predict future LULC. This study will reveal the drivers of land use/cover change within the Okavango Basin, and provide a future scenario on the impact of change on natural resources, for example, water availability

and distribution, and disruptions of crucial ecosystems. The findings of this study are envisaged to guide basin management and to inform transboundary policy formulation.

2. Materials and methods

2.1. Study site

This study was conducted in the Okavango drainage basin, a transboundary drainage basin that spans three countries, namely Angola, Namibia and Botswana (Fig. 1). The climate varies across the basin with an average annual rainfall of 430 mm (Eze et al.,

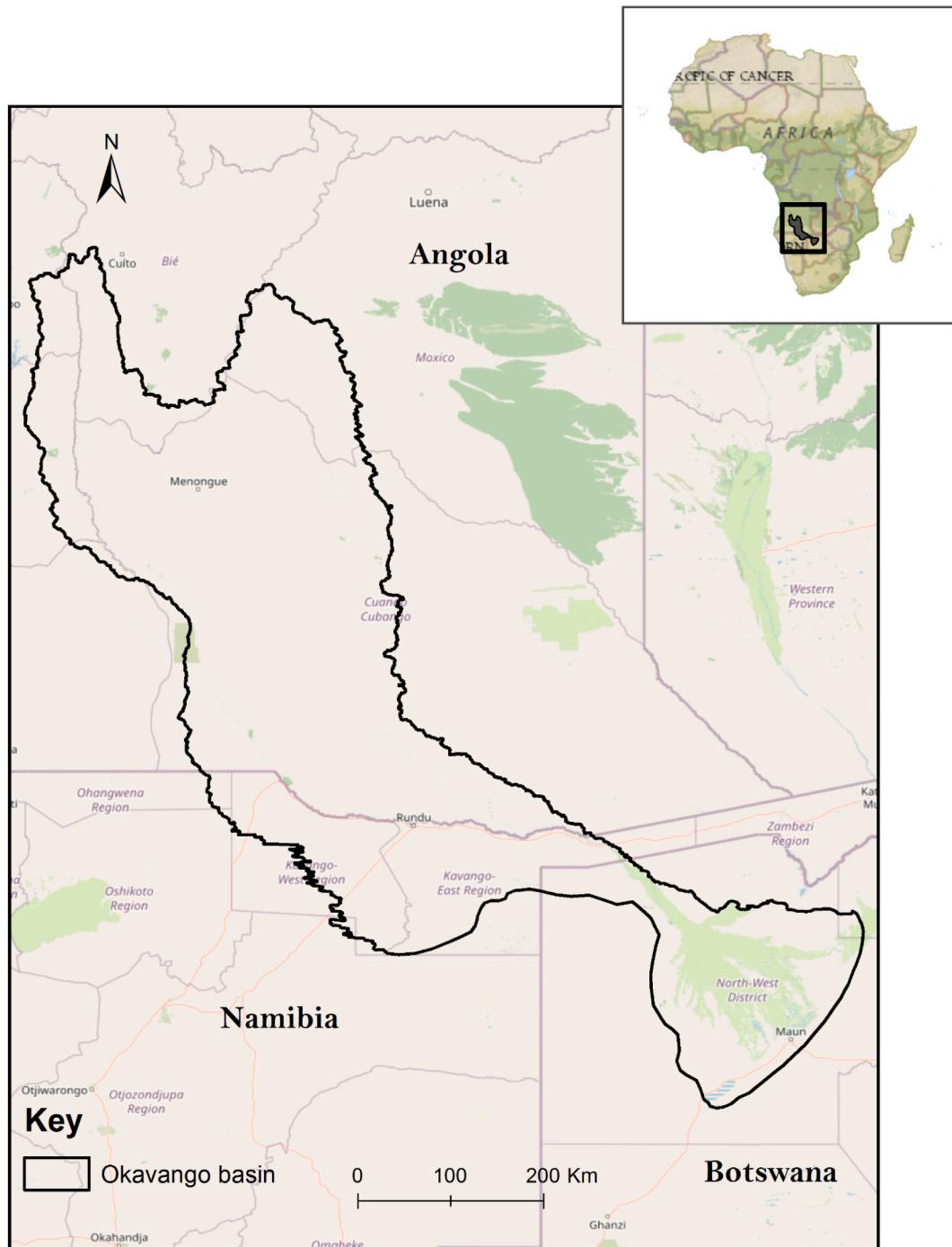


Fig. 1. Study area map.

2021). The temperature generally ranges from a minimum of 14.8 °C to a maximum of 33.7 °C (Ellery et al., 1989; Weber, 2013). The area's flora and fauna vary across the basin depending on climate and land use type (Revermann et al., 2016). One of the most important sources of water in this landscape is the Okavango River, which flows from the Angolan highlands through Namibia to Botswana and forms one of the world's largest inland deltas (Steudel et al., 2013).

This study site was chosen since it has been associated with many catastrophes such as war, drought, and floods in recent history. In addition, the end of the Angolan war in 2002 brought changes in land tenure that were associated with harmful land-use practices, such as illegal logging, overgrazing, and intensive tillage (Folwell et al., 2006; Porto and Clover, 2003; Mianabadi et al., 2020). Okavango basin is a biodiversity hotspot and a livelihood support system; therefore, it is empirically relevant to examine the social-ecological drivers of LULC change in this landscape.

2.2. Satellite image acquisition and processing

Analysis ready Landsat 5 and Landsat 8 OLI imagery that was sourced from the Google Earth Engine (GEE) platform were used for LULC classification. The images were acquired in June for the years 2004, 2013 and 2020. The temporal period is chosen based on the availability of cloud-free images that cover the entire drainage basin. The images were then exported to R (

(Team, 2010) for LULC mapping.

2.3. Spectral features

Spectral features used in the analysis comprises a combination of six bands (for Landsat 5 and Landsat 8), orthogonal spectral (OS) indices and Ratio-based Spectral indices. RBS indices were derived from pairs of spectral bands and OS indices are derived from transformation coefficients as indicated in Table 1 and Table 2 respectively.

2.4. Training and validation points

Sample data were sourced from the Okavango River Basin Water Commission (OKACOM) geodatabase and the National Geographic Okavango and Wilderness Project (NGOWP). OKACOM is responsible for jointly managing water resources in three countries found in the basin namely Angola, Botswana and Namibia. NGWOP conducted surveys to explore the least known and most accessible areas in the basin and the survey took geotagged images of riverine vegetation which helps to generate additional samples leaving the number of ground samples at 3420. Further samples were digitised from visual interpretation of high-resolution imagery available on Google Earth (de Sousa et al., 2020). Stratified random sampling based on the 2009 GLOBCOVER was used to minimise sample imbalance per class (Arino, 2010). The minimum fifty sample rule per class was adopted to ensure that a sufficient number of samples was used (Foody and Mathur, 2004). Overall, 5140 samples were generated for eight LULC classes, namely bare land, built-up land, bushland, forest/woodland, grassland, cultivated land, water and wetland. The spatial distribution of the training and validation data is depicted in Fig. 2.

2.5. Regionalization and inclusion of spectral indices for LULC classification

Previous studies have reported that spatial heterogeneity can reduce classification accuracy (Saha et al., 2005; Kassawmar et al., 2018), hence the study area is regionalised based on the Koppen-Geiger climate classifications (Bsh – Hot semi-arid, Cwa - Moonson and Cwb – sub-tropical highland) to reduce heterogeneity as recommended by Kavhu et al. (2021). Combinations of spectral bands (7), spectral indices (12) and elevation layer were implemented per year on each segment separately. Feature selection before LULC classification is crucial to deal with the so-called curse of dimensionality (Gilbertson et al., 2017). In this study, we used a feature selection technique with a wrapper called the Random Forest-based Recursive Feature Elimination (RF-RFE) (Ismail and Mutanga, 2011). The RF-RFE uses the provided input features and the Random Forest algorithm to select the best combination of features based on feature importance (Pal and Foody, 2010; Ma et al., 2017). To make the best combination of features, the RF-RFE iterates over various feature combinations through a repeated 10-cross validation and eliminates the least important features until the most

Table 1

Ratio based spectral indices used in this study.

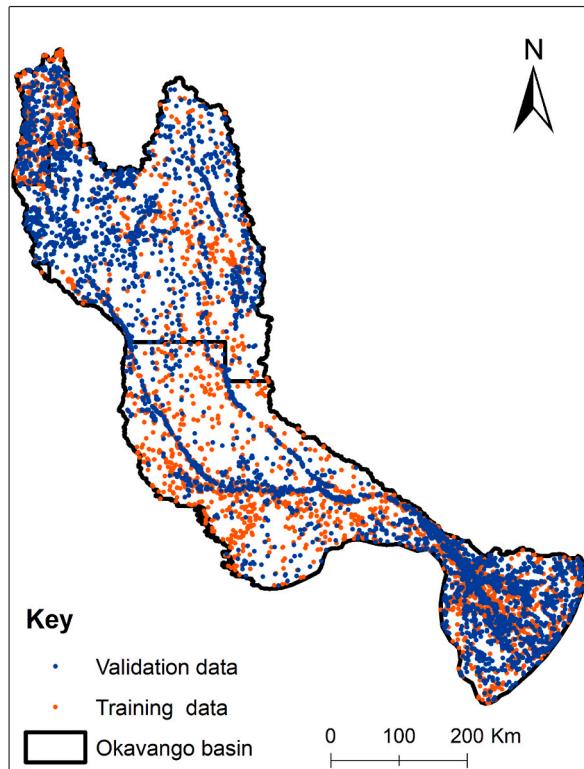
Name of spectral indices	Formulae	References
NDVI	$\frac{NIR - RED}{NIR + RED}$	Rouse et al. (1973)
NDBI	$\frac{SWIR1 - NIR}{SWIR1 + NIR}$	Zha et al. (2003)
NDWI	$\frac{GREEN - NIR}{GREEN + NIR}$	Chen et al. (2004)
MNDWI	$\frac{GREEN - SWIR1}{GREEN + SWIR1}$	Xu (2006)
NDTI	$\frac{SWIR1 - SWIR2}{SWIR1 + SWIR2}$	Van Deventer et al. (1997)
NDBal	$\frac{SWIR1 - TIRS1}{SWIR1 + TIRS1}$	Zhao and Chen (2005)
EVI	$2.5 \frac{NIR - RED}{NIR + 6 \times RED - 7.5 \times RED + 1}$	Huete et al. (2002)
SAVI	$1.5 \frac{NIR - RED}{NIR + RED}$	Huete (1988)

Table 2

Orthogonal spectral indices used to enhance LULC mapping in this study.

Landsat 5							
Name of spectral indices	Transformation coefficients						References
	(Blue) Band 1	(Green) Band 2	(Red) Band 3	(NIR) Band 4	(SWIR1) Band 5	(SWIR2) Band 7	Crist and Cicone, 1984
BTCAP	0.2043	0.4158	0.5524	0.5741	0.3124	0.2303	
GTCAP	-0.1603	0.2819	-0.4934	0.7940	0.0002	0.1446	
WTCAP	0.0315	0.2021	0.3102	0.1594	0.6806	0.6109	

Landsat 8							
Name of spectral indices	Transformation coefficients						References
	(Blue) Band 2	(Green) Band 3	(Red) Band 4	(NIR) Band 5	(SWIR1) Band 6	(SWIR2) Band 7	Ali Baig et al., 2014
BTCAP	0.3029	0.2786	0.4733	0.5599	0.5080	0.1872	
GTCAP	0.2941	0.2430	0.5424	0.7276	0.0713	0.1608	
WTCAP	0.1511	0.1973	0.3283	0.3407	-0.7117	0.4559	

**Fig. 2.** Training and validation points.

parsimonious model is identified (Kuhn, 2015). In this study, the RF-RFE is tuned to repeatedly iterate 30 times over 20 features (bands and spectral indices) based on the caret package (Kuhn, 2015) in R statistical software. The most parsimonious model is then used to map landcover change trajectory between 2013 and 2020 using the Deep neural network (DNN). The DNN classifier was chosen because of its great performance in previous land cover studies (Ma et al., 2017; da Silva et al., 2020). The basic structure of a DNN is a network of input layers that are connected to the output layer through hidden layers. This network of layers is responsible for transforming input data to output data with the help of activations and parameters. The weights on the nodes of each connection modify values at each layer to determine how the input values are translated to output values. This study used the DNN based on a multi-layered feedforward neural network that comprises more than 3 hidden layers (Abdi, 2020). Ideally, increasing the number of hidden layers and neurons increases the potential to make predictions in complex situations (S.K. Singh et al., 2018). The key parameters for DNN include the activation function (activation), number of hidden layers (hidden), size of each hidden layer (number of neurons per hidden layer) and the number of times to iterate (epoch). In this study, DNN is tuned (hyper-parameterisation) to select the optimal parameters using the H2O package (Cook, 2016) in R statistics. Table 3 shows the optimal parameters used in this study. Performance of the DNN is measured using the overall accuracy (OA) measure and the Kappa statistic (Kappa). The OA range from 0 to 100 where 0 represents the lowest accuracy and 100 represents the highest accuracy and the Kappa range from 0 to 1, where 0 represents the lowest accuracy and 1 represents the highest accuracy.

2.6. Change detection

To detect changed sites, the post-classification change analysis was performed through an overlay analysis of two-time steps ((2004–2013) and (2013–2020)) following Münch et al. (2019). A transition matrix for the intersection of each pair of land cover maps is generated. Centroids of changed sites for the period 2004 to 2020 were determined in ArcMap 10.6 and the number of centroid points for each transition was summarized in Table 4.

2.7. Changed sites data

In this study, only transitions with more than 30 centroid points are considered to model drivers of LULC change Griffith (2013). The transitions are categorized into 3 based on the initial class and final class; Category A-transitions from natural classes to anthropogenic classes, Category B- transitions from anthropogenic classes to natural classes, Category C- transitions from natural classes to another natural class. The list of transitions considered for modelling and their respective categories are provided in Table 5.

2.8. Experimental design

The centroid points for each transition are used to calibrate singular ML and ensemble ML models to evaluate drivers of LULC change. The workflow of the study design is illustrated in Fig. 3.

2.9. Social-ecological variables

Thirteen social-ecological variables that were collected using remote sensing are used to model drivers of change. The variables include Gross Domestic Product (GDP), population density, annual average temperature, drought severity, annual average precipitation, distance from rivers, distance from roads, distance from urban centres, and soil organic carbon, aspect, slope, fire frequency and elevation. The variables were chosen based on their availability and their performance in previous studies. Details on the variables and their use in previous studies are listed in Table 6.

Long term means monthly temperature and precipitation data available on GEE were averaged to obtain annual average temperature and annual average precipitation respectively. These were used to show how climate influences LULC change across space (Rutherford et al., 2007; Arficho and Thiel, 2020). According to Fick and Hijmans (2017), the long-term annual average temperature and precipitation data have an accuracy range of about 5–15% which can suffice for modelling at this study scale. Soil organic carbon is used as a proxy for soil fertility as it has been observed to influence selection sites for cropping fields (Gibson et al., 2018). The uncertainty of the soil organic data in the Okavango basin is $\pm 15 \text{ Mg C ha}^{-1}$ which is satisfactory to consider the variable. Population density and Gross domestic product (GDP) are used to assess the influence of demographic pressure and level of development on LULC change respectively (Wang et al., 2016). The influence of population density and GDP is crucial, for example, because, an increase in population may lead to overexploitation of natural resources, but may spur industrial development leading to innovative and cost-effective tools for mechanisation and ecologically friendly technology. GDP and population density data are a product of population censuses. For GDP, population census data were combined with national employment and output data to estimate output per capita by region (CIESIN, 2018).

Distance from rivers was used as a proxy for surface water availability. Surface water plays a critical role when selecting sites for residential settlements, areas for development, and people often settle close to surface water resources (Marondedze and Schütt, 2019). Vector layers for rivers were obtained from the OKACOM database and additional data are downloaded from the OpenStreetMap database (www.openstreetmap.org). Distance from roads is used to represent the influence of road networks in the distribution of change in LULC. In this study, distance from roads was used as a proxy for human access to places. Human access to places is known to be associated with increased anthropogenic activities that may pose a negative influence on natural ecosystems (Kleemann et al., 2017). To capture the effects of urban sprawl on natural classes because of in-migration during the post-war period, distance from urban centres was used following Phiri et al. (2019). Vector files for roads, rivers and urban centres are projected to Africa Albers Equal Area Conic and used to calculate Euclidean distances in ArcMap 10.7 (McCoy and Johnston, 2001).

Elevation, slope and aspect were used to represent the influence of variation in topography on LULC change following Wang et al. (2016) and, Kim et al. (2020). The Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) with a spatial resolution of 90 m was used as the elevation layer. Slope and aspect were then derived from the DEM in ArcGIS 10.7. The Palmer drought severity index was used as a proxy for drought conditions since frequent droughts have destroyed ecosystems, and displaced people and wildlife (Dudley et al., 2001; Ruan et al., 2016; McLeman et al., 2021). Monthly drought severity data sourced from Google Earth Engine (GEE) was averaged to produce long-term average drought severity for the study period. Fire frequency drives vegetation structure and pattern; thus, it was used to evaluate its effects on landcover change (Kamwi et al., 2015). Additionally, fire is a common feature in the Okavango basin (Ellery et al., 1989; Heinl, 2005). Fire frequency was derived from cumulating the fire scars using overlay analysis of monthly-burnt scars of the fire season (July to November). Monthly fire scars were validated using overlay analysis with

Table 3

A summary of optimal parameters for the DNN classifier as determined from hyper parameterisation.

Model	Parameters	Hyper-parameter values
DNN	activation	Rectifier
	hidden layers	5
	neurons per layer	200
	Epochs	300

Table 4

Shows a summary of the number of changed points derived from a change matrix. Transitions in red shows points of areas that did not change, green shows points that met the minimum required limit and yellow shows transitions in which the detected number of points were less than the minimum required limit (at least 30points) for further analysis.

2020								
	Bareland	Water	Builtup	Cultivated	Woodland	Grassland	Shrubland	Wetland
Bareland	67	20	3	5	13	10	25	24
Water	5	2167	18	299	438	70	12	8
Builtup	9	4	7793	3	7	4	10	0
Cultivated	18	58	1909	16856	236	4816	463	12
Woodland	26	13	508	23972	40452	66658	12784	3
Grassland	26	1042	142	7720	64359	82336	34238	21
Shrubland	24	27	20	453	10037	34654	23410	18
2004	2	23	13	56	72	56	46	18350

Table 5

Shows a summary of categories of LULC transitions bases on initial class and final class where Category A have transitions from natural classes to anthropogenic classes, Category B have transitions from anthropogenic classes to natural classes and Category C have transitions from natural classes to another natural class.

Category	Transition ID	Transition
A	A1	Water to Cultivated
	A2	Woodland to builtup
	A3	Woodland to cultivated
	A4	Grassland to builtup
	A5	Grassland to cultivated
	A6	Shrubland to cultivated
	A7	Wetland to cultivated
B	B1	Cultivated to builtup
	B2	Cultivated to woodland
	B3	Cultivated to grassland
	B4	Cultivated to shrubland
C	C1	Water to woodland
	C2	Water to grassland
	C3	Woodland to grassland
	C4	Woodland to shrubland
	C5	Grassland to water
	C6	Grassland to shrubland
	C7	Grassland to wetland
	C8	Shrubland to woodland
	C9	Shrubland to grassland
	C10	Shrubland to wetland
	C11	Wetland to woodland
	C12	Wetland to grassland
	C13	Wetland to shrubland

high-resolution imagery before generating the fire frequency dataset. Night-time lights was used as a proxy for anthropogenic activity following recommendations by Mpakairi and Muvengwi (2019)

Before modelling, all environmental variables were resampled from their original resolution to a 250 m spatial resolution using the nearest neighbour technique. Pairwise correlations between environmental variables were used to detect multicollinearity of predictor variables. A threshold of $r > 0.7$ was used to eliminate correlated variables and to create a parsimonious model (Dormann et al., 2013). Multicollinearity was tested in R (R Development CoreTeam, 2014). The elevation was correlated to annual average temperature ($r = 0.86$), however, both variables were retained as they are relevant to processes that drive LULC change. Night-time lights was correlated to distance from urban centres, hence it was excluded from model building. No other variables were excluded since they had no multicollinearity ($r < 0.7$).

2.10. Modelling approach

Five machine learning (ML) algorithms were used to separately model the drivers of LULC change. The ML algorithms include Random Forests (RF), Gradient Boost Models (GBM) and Maximum Entropy (MaxEnt), Classification Tree Analysis (CTA also referred to as classification and regression trees [CART]) and Artificial Neural Network (ANN). The selected models were used based on their extensive use in understanding social-ecological systems (Wang et al., 2020; Phiri et al., 2019).

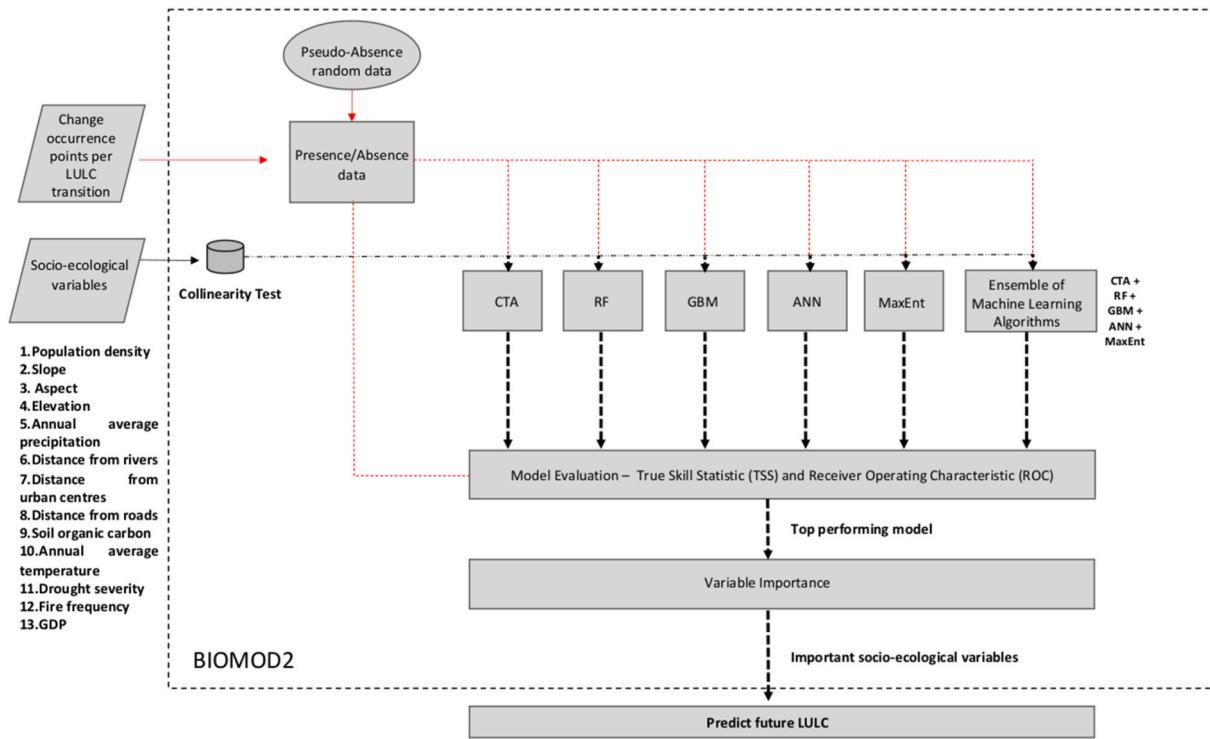


Fig. 3. Study workflow.

Table 6
Summary and details of social-ecological variables used in this study.

Variable	Data source	Resolution	Studies that used the same variable for LULC change modelling
Soil organic carbon	Isric soil database	250 m	Veldkamp and Fresco (1996)
Distance from roads	Openstreet + OKACOM	optional	(Hazen and Berry, 1997), (Kleemann et al., 2017)
Distance from urban centres	Openstreet + OKACOM	optional	Marondedze and Schütt (2019)
Distance from rivers	Openstreet + OKACOM	optional	(Kleemann et al., 2017; Gibson et al., 2018)
Elevation	SRTM DEM	90 m	Wang et al. (2016)
Slope	SRTM DEM	90 m	Gibson et al. (2018)
Aspect	SRTM DEM	90 m	Hazen and Berry (1997)
Population density	SEDAC	1°	Xu et al. (2013)
Annual average precipitation	Worldclim	30 s	Sibanda and Ahmed (2021)
Palmer drought severity	TerraCLIMATE	2.5 arc min	Phiri et al. (2019)
GDP	SEDAC	1°	Long et al. (2007)
Annual average temperature	Worldclim	30 s	Lin et al. (2014)
Fire frequency	MODIS(MCD64A1)	500 m	Kamwi et al. (2018)

All the models were run using the BIOMOD2 (Thuiller et al., 2013). BIOMOD2 was tuned to generate 500 pseudo-random points and split presence data using the 80%:20% ratio for training and validation (Kanagaraj et al., 2019). To address the shortfalls of using singular models at the same time strengthening our method of teasing apart drivers of LULC, an ensemble model was built from ML algorithms. The ensemble model was built based on a Receiver Operating Characteristic (ROC) threshold of 0.70. An ensemble model is used to account for the accuracy gaps which are associated with using singular models as illustrated by Buisson et al. (2010) and Grenouillet et al. (2011).

Variable importance scores from model outputs are determined, to identify the variables that were most influential in driving LULC change. Variable importance was calculated using correlation scores between predictions based on the model with all variables present against predictions where the variable of interest is removed (Thuiller et al., 2013).

2.11. Model evaluation and comparison

Model performance is evaluated using the True skill statistic (TSS) and Receiver Operating Characteristic (ROC) evaluation metrics. Model performance was considered poor if the ROC or TSS values were less than 0.5, good if the values are within the range 0.5–0.8 and excellent if the ROC or TSS values are greater than 0.8 (Phillips et al., 2006; Phillips and Dudík, 2008). ROC values are graphically presented against each model for each transition to visualise variations in model performances. Using ROC values per model from each

transition, an average ROC was determined and the One-way Analysis of variance (ANOVA) was used to test for statistical difference in model performance. A significant difference in overall model performance was observed ($P < 0.05$), hence a Tukey's posthoc was done for pairwise comparisons of model performance. To test the hypothesis that ensemble ML performs better than singular ML, the post-hoc results focus on pairs of models that involved an ensemble ML versus any of the singular ML.

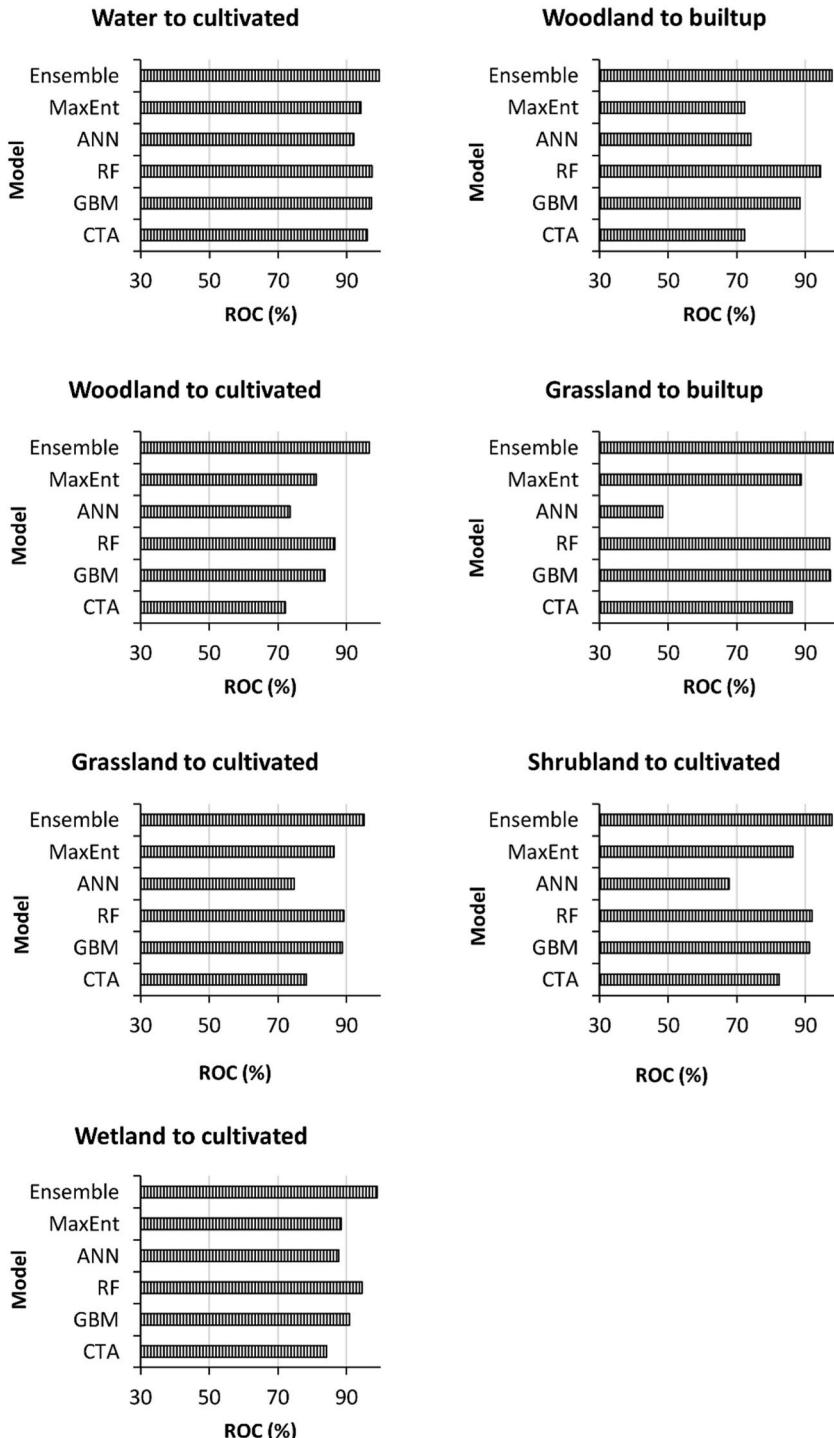


Fig. 4. Comparison of the performance of different machine learning models and ensemble machine learning model based on different transitions in category A.

2.11.1. Predict future LULC using the Cellular Automata and Artificial Neural Network model

To predict the potential distribution of future LULC in the study area, the Cellular Automata and Artificial Neural Network (CA-ANN) model was used based on the MOLUSCE (Modules for Land Use Change Simulation) plugin in QGIS Desktop V. 2.14 software (Rahman et al., 2017; Ashaolu et al., 2019). The CA is a discrete dynamic space that characterises every cell based on transition rules (the state of each cell as determined by its previous state and the state of its neighbourhood properties) (Santé et al., 2010). In this study, the prediction of future LULC was done in three steps: firstly, applying the ANN quantitative analysis to the classified LULC maps to compute transition probabilities; secondly, computation of transition potential maps, and thirdly, application of the CA spatial filter to the transition probabilities and the transition potential maps to simulate the future LULC. Classification results for the period between 2004 and 2013 and a set of top contributing socio-ecological variables were used to train the ANN to forecast 2020 LULC. A step size of 2 years with 6 iterations was adopted as recommended by (Alam et al., 2021). The simulated 2020 raster was then validated with the classified 2020 LULC. Hence, the CA is then tuned to predict the LULC for the year 2040.

3. Results

3.1. Accuracy of LULC maps

The average LULC classification accuracy results for the study period (2002, 2013 and 2020) in the three koppen zones (Bsh, Cwa and Cwb) were good, Bsh (OA = 95.03, Kappa = 0.94), Cwa (OA = 95.29, Kappa = 0.94), Cwb (OA = 96.04, Kappa = 0.95), hence change detection is done to identify changed sites for the same period to predict drivers of LULC change.

3.2. Performance of spatial models

Results for the model performance of singular ML and ensemble ML for category A, category B and category C transitions are provided in Figs. 4–6.

3.2.1. Category A transitions (natural classes to anthropogenic classes)

For category A transitions, the performance of models varies widely, with ROC values ranged between 52.1 and 99.3. The ensemble ML recorded the highest accuracy in all the transitions, with (A1:ROC = 99.4, TSS = 97.6), (A2:ROC = 97.7, TSS = 87.0), (A3:ROC = 96.7, TSS = 72.1), (A4:ROC = 99, TSS = 94.8), (A5:ROC = 95.2, TSS = 68.7), (A6:ROC = 97.8, TSS = 95.2) and (A7:ROC = 98.7, TSS = 98.9). In contrast, ANN consistently recorded the lowest accuracy for all the transitions, with (A1:ROC = 92.1, TSS = 90.3), (A2:ROC = 74.1, TSS = 55.5), (A3:ROC = 73.5, TSS = 34.3), (A4:ROC = 48.4, TSS = 40.2), (A5:ROC = 74.8, TSS = 46.7), (A6:ROC = 67.7, TSS = 27) and (A7:ROC = 84.2, TSS = 64). It can be seen that the RF performed better than most singular ML models (ROC < 86.7).

3.2.2. Category B transitions (anthropogenic classes to natural classes)

Like the category A, the ensemble ML recorded the highest accuracy in all category B transitions (Fig. 7). The accuracy values for ensemble ML are: (B1:ROC = 99.3, TSS = 91.3), (B2:ROC = 94.3, TSS = 81.0), (B3:ROC = 95.4, TSS = 72.1), (B4:ROC = 97, TSS =

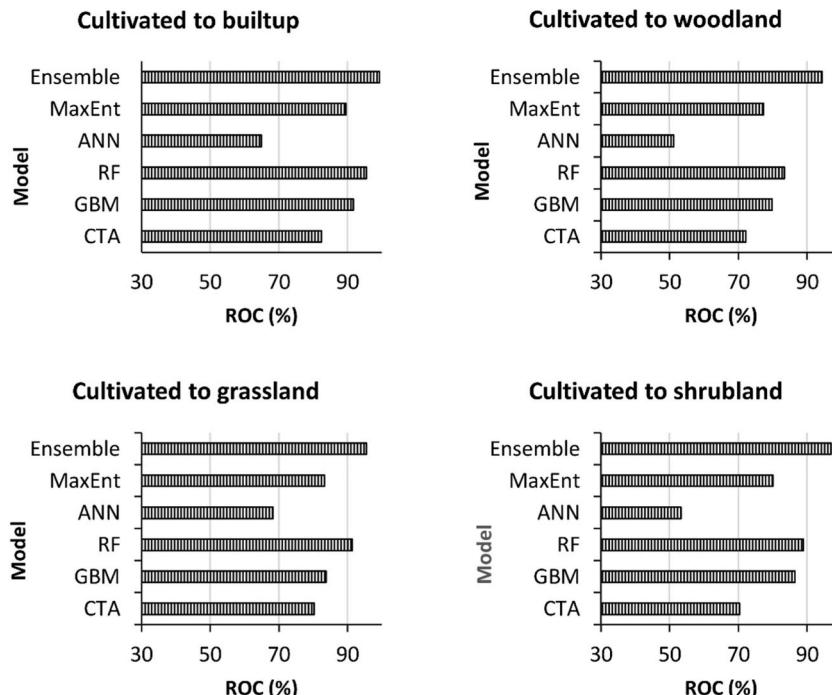


Fig. 5. Comparison of the performance of different machine learning models and ensemble machine learning model based on different transitions in category B.

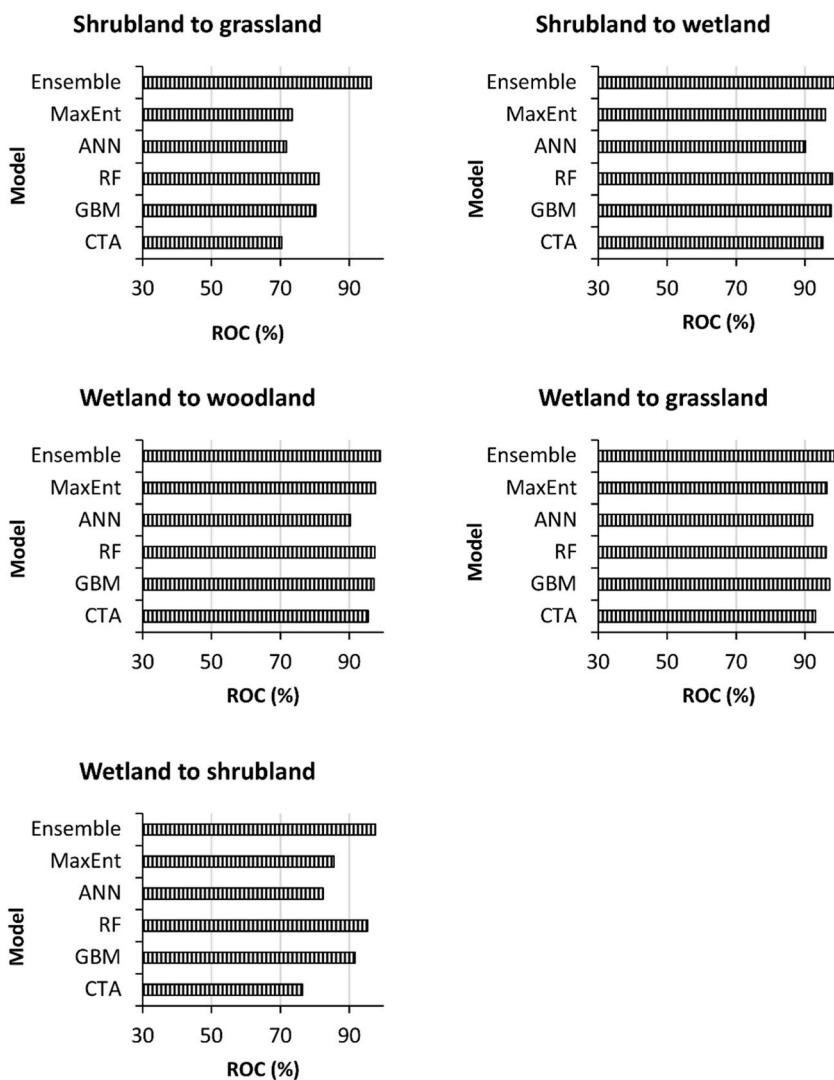


Fig. 6. Comparison of the performance of different machine learning models and ensemble machine learning model based on different transitions in category C.

82.0). On the other hand, ANN records the lowest accuracy for all transitions in this category, with accuracies of: (B1:ROC = 64.8, TSS = 60.1), (B2:ROC = 51.1, TSS = 30.9), (B3:ROC = 68.2, TSS = 34.5), (B4:ROC = 53.2, TSS = 24.8). With singular ML algorithm, it can be seen that there is a consistent pattern in performance of models in this transition, where RF (ROC >81.2) performs the best followed by GBM(ROC >80.1), MaxEnt (ROC >77.3, CTA(ROC >72.4 and lastly ANN(ROC >40.3).

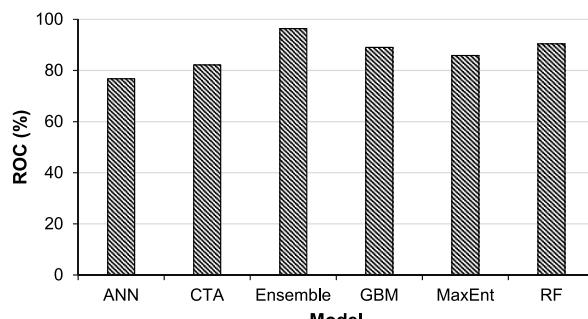


Fig. 7. Comparison of the average performance of different machine learning models and ensemble machine learning model when detecting social-ecological drivers of LULC in the Okavango basin.

3.2.3. Category C transitions (natural classes to natural classes)

As with the Category A and B transitions, ensemble ML consistently records the highest accuracy in all Category C transitions with ROC values ranging above 92.3 and TSS values above 71 (Fig. 8). Similar to category A and B transitions, ANN records the lowest accuracy with ROC and TSS values ranging below 92.1 and 90.3 respectively. A similar pattern is observed on the performance of singular models where RF is the best (ROC >98), followed by GBM (ROC >97.6), MaxEnt (ROC >97.3), CTA (ROC >95.4).

Overall, results reveal that on average ensemble ML consistently performs better than singular ML techniques when evaluating drivers of LULC change (ROC = 96.4) (Fig. 7). In addition, the RF consistently perform better than all singular ML used in this study (ROC = 90.41). In contrast, the ANN records the lowest accuracy (ROC = 76). The average performance of ensemble ML is statistically different from singular models except for the RF and GBM which also recorded considerably higher accuracies ($p < 0.05$).

3.3. Social-ecological drivers of LULC change

Variable importance results for all the transitions considered in this study and results of the top contributing variables ($VI > 0.5$) are illustrated in Table 7 and Table 8 respectively.

3.3.1. Category A

Of the variables used, distance from rivers, population density, annual average temperature, drought severity, fire frequency and distance from towns were important in influencing category A transitions (Tables 7 and 8). In contrast, slope, soil organic carbon and aspect were not influencing category A transitions. Although both natural and anthropogenic variables are importantly influencing transitions in this category, it can be seen that most of the natural variables contributing much potentially have an indirect influence on anthropogenic activities.

3.3.2. Category B

For category B transitions, elevation, population density and soil organic influence the most ($VI > 0.5$). On the other hand distance from urban centres, fire frequency, slope, distance from rivers, distance from roads, drought severity and aspect are not influencing much in category B transitions ($VI < 0.5$).

3.3.3. Category C

Distance from rivers, soil organic carbon, precipitation, GDP, elevation, population density and annual average temperature importantly influenced category C transitions (Table 7). On the other hand, slope, distance from urban centres, distance from roads and are not influencing much to any of the transitions in this category ($VI > 0.5$). Although the transitions in this category involve the conversion of one natural class to another natural class, it can be seen that anthropogenic variables still have relevant influence.

3.4. Current and future LULC

Results for current and future LULC extent, and LULC maps for the Okavango from 2020 to 2040 are shown in Table 9 and Fig. 8. The accuracy of the 2020 LULC map in the three Koppen zones (Bsh, Cwa and Cwb) were Bsh (OA = 92.85, Kappa = 0.918), Cwa (OA = 93.56, Kappa = 0.926), Cwb (OA = 91.61, Kappa = 0.9039) (see confusion matrices in appendix 1, 2 and 3). A decrease in the extent of wetlands (21.27%), woodlands (0.63%), water(3.5%) and shrubland (0.34%) was projected in the future (Table 8). Contrastingly, an

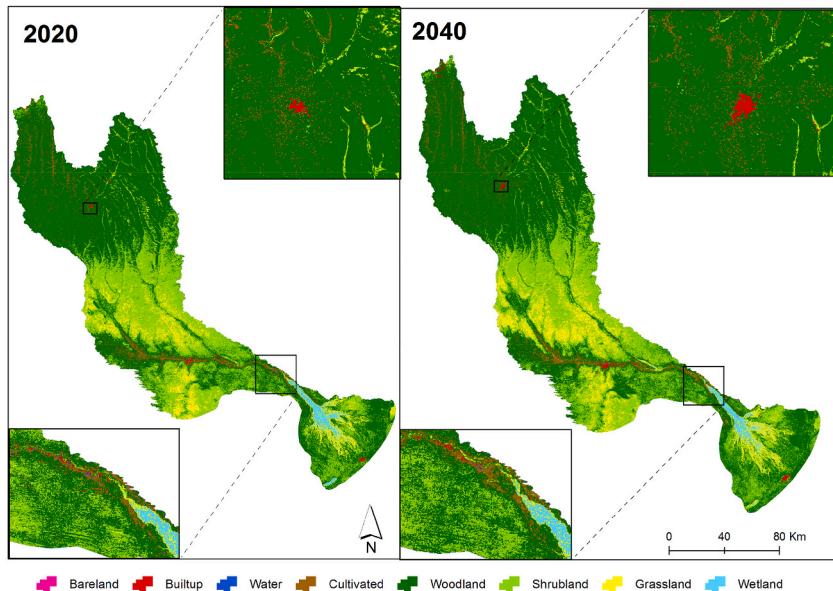


Fig. 8. Maps showing the current LULC (2020) and the projected LULC (2040). The map illustrates how builtup and cultivated areas are expanding at the expense of woodlands.

Table 7

Variable importance of social-ecological variables modelling drivers of LULC change in the Okavango basin. The colour of shading on each cell corresponds the level of importance with red shaded cells representing top contributing variables ($VI > 0.5$), yellow and orange representing moderate contributing and green representing less contributing variables ($VI < 0.1$) per transition.

Transition Variables	Category A							Category B							Category C									
	A1	A2	A3	A4	A5	A6	A7	B1	B2	B3	B4	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Aspect	0.001	0.001	0.029	0.015	0.058	0.051	0.003	0.038	0.069	0.142	0.028	0.033	0.024	0.053	0.042	0.021	0.031	0.013	0.021	0.031	0.061	0.044	0.021	0.063
Drought severity	0.013	0.081	0.103	0.491	0.121	0.149	0.000	0.001	0.028	0.328	0.141	0.141	0.001	0.338	0.224	0.361	0.284	0.009	0.223	0.317	0.004	0.000	0.012	0.044
Elevation	0.230	0.221	0.490	0.454	0.110	0.127	0.151	0.024	0.673	0.666	0.471	0.443	0.011	0.640	0.297	0.664	0.121	0.443	0.012	0.584	0.391	0.132	0.108	0.342
GDP	0.003	0.132	0.038	0.005	0.186	0.104	0.000	0.021	0.074	0.183	0.018	0.382	0.002	0.557	0.006	0.343	0.164	0.052	0.011	0.127	0.064	0.001	0.008	0.011
Soil organic carbon	0.032	0.121	0.089	0.003	0.007	0.002	0.021	0.009	0.021	0.004	0.671	0.087	0.582	0.183	0.031	0.593	0.279	0.148	0.006	0.362	0.001	0.395	0.117	0.011
Population density	0.218	0.202	0.671	0.746	0.311	0.175	0.321	0.181	0.578	0.042	0.004	0.008	0.272	0.412	0.123	0.211	0.613	0.002	0.457	0.489	0.009	0.681	0.310	0.211
Annual average precipitation	0.002	0.124	0.011	0.142	0.283	0.107	0.417	0.007	0.312	0.289	0.148	0.072	0.003	0.671	0.391	0.733	0.003	0.031	0.241	0.732	0.003	0.354	0.613	0.193
Distance from rivers	0.003	0.621	0.451	0.430	0.213	0.131	0.284	0.040	0.024	0.119	0.098	0.637	0.741	0.122	0.141	0.106	0.022	0.012	0.221	0.131	0.015	0.022	0.097	0.217
Distance from roads	0.011	0.234	0.492	0.082	0.442	0.142	0.026	0.229	0.094	0.113	0.008	0.037	0.014	0.181	0.037	0.118	0.187	0.022	0.023	0.178	0.061	0.094	0.118	0.038
Fire frequency	0.003	0.024	0.129	0.000	0.547	0.042	0.079	0.002	0.067	0.045	0.001	0.147	0.038	0.384	0.041	0.095	0.182	0.057	0.062	0.211	0.058	0.119	0.127	0.039
Slope	0.014	0.006	0.001	0.034	0.035	0.124	0.050	0.076	0.041	0.064	0.015	0.032	0.025	0.009	0.037	0.071	0.042	0.021	0.012	0.062	0.001	0.073	0.022	0.038
Annual average temperature	0.013	0.132	0.347	0.681	0.161	0.678	0.000	0.031	0.048	0.134	0.118	0.010	0.384	0.321	0.196	0.323	0.284	0.221	0.123	0.018	0.481	0.131	0.108	0.571
Distance from urban centres	0.001	0.032	0.298	0.002	0.091	0.061	0.676	0.021	0.018	0.041	0.002	0.127	0.039	0.194	0.025	0.103	0.004	0.021	0.143	0.073	0.024	0.034	0.072	0.022

Table 8

Overview of important social-ecological variables influencing LULC transitions in the Okavango basin.

Category	Transition ID	Transition	Important variables ($VI > 0.50$)
A	A1	Water to Cultivated	–
	A2	Woodland to builtup	● distance from rivers
	A3	Woodland to cultivated	● population density
	A4	Grassland to builtup	● population density, ● annual average temperature, ● drought severity
	A5	Grassland to cultivated	● fire frequency
	A6	Shrubland to cultivated	● annual average temperature
B	A7	Wetland to cultivated	● distance from urban centres
	B1	Cultivated to builtup	–
C	B2	Cultivated to woodland	● elevation
	B3	Cultivated to grassland	● population density
	B4	Cultivated to shrubland	● elevation
	C1	Water to woodland	● soil organic carbon
	C2	Water to grassland	● distance from rivers
	C3	Woodland to grassland	● soil organic carbon ● distance from rivers
	C4	Woodland to shrubland	● GDP
	C5	Grassland to water	● precipitation
	C6	Grassland to shrubland	–
C	C7	Grassland to wetland	–
	C8	Shrubland to woodland	–
	C9	Shrubland to grassland	● elevation ● precipitation
	C10	Shrubland to wetland	–
	C11	Wetland to woodland	● population density
	C12	Wetland to grassland	● precipitation
	C13	Wetland to shrubland	● annual average temperature

increase in the extent of grasslands (6.27%), croplands (10.61%), built-up areas (6.43%) and barelands (163.18%) was projected in the year 2040. It can be seen that most of the anthropogenic classes will increase by over 5% which reflects how anthropogenic activities will continue to transform LULC in the Okavango basin (Table 9). Also, results illustrate that anthropogenic classes will mostly expand in the northern and central parts of the basin (Fig. 8).

4. Discussion

LULC change driven by anthropogenic activities is currently threatening the availability and distribution of natural resources. These threats will likely intensify or increase in the future under global environmental change and increasing anthropogenic activities. Addressing these threats will allow sustainable management of resources which will go a long way in addressing National, Regional and International targets (e.g., Sustainable Development Goals). This study sought to tease apart the drivers of LULC change with the use of singular ML and ensemble ML models.

i) Ensemble vs Singular Machine Learning models

Results show that ensemble ML consistently performed better than singular models on all the LULC transitions that were considered in this study (Figs. 6–8). This corroborates with previous literature which claims that ensemble models are more robust and can effectively deduce patterns from data when tuned with both small or large datasets without overfitting (Marmion et al., 2009; Parvainen et al., 2009). Unlike singular models that rely on different assumptions which are more or less applicable to particular landscapes (Grenouillet et al., 2011), an ensemble model is built from multiple top-performing individual models which makes it more effective. Additionally, a weighted mean of probabilities algorithm used to build the ensemble models give more weight to models with higher predictive accuracy (Thuiller et al., 2013). The accuracy of ensemble ML was however not statistically significant from that of the RF and GBM ($p < 0.05$). This could be because RF and GBM are also robust and assumptions that govern them can withstand most prediction limitations that compromise model performance (Moisen et al., 2006; Belgiu and Drăguț, 2016). The GBM produces multiple trees using a random sample of selected data to perform boosting in the next step. The major strength of GBM lies in that it uses classification errors to refine the trees using a random sample and then combines multiple trees iteratively developed to make predictions (Lawrence et al., 2004). On the other hand, RF is an ensemble algorithm that produces multiple decision trees based on training samples via bagging using random split selection at each node (Breiman et al., 1984). Wang et al. (2016) used the RF to evaluate drivers of LULC change in China and recorded an average ROC value of 81. The average ROC for RF in this study is slightly higher (90.4) than the one obtained by Wang et al., (2016). This could be because their study was based on a smaller site (province) using slightly fewer (12) and less diverse variables (social (2), economic (8), topographic (1), climatic (1)) as compared to this study ((social (2), economic (1), topographic (3), climatic (3), proximity (3), edaphic (1)). RF is robust and performs better when properly tuned with diverse features (Belgiu and Drăguț, 2016; Saini and Ghosh, 2017). Notwithstanding that, ensemble ML are superior as they consistently outperformed all singular ML for all the transitions and with higher accuracies ($\text{ROC} > 91.3$). Accurately determining the drivers of LULC allow targeted mechanisms to alter the negative effects of LULC change, especially for anthropogenic drivers in developing countries where resources for monitoring are constrained. Previous studies have mainly used singular ML and RG algorithms to understand drivers of LULC change (Kamwi et al., 2015; Phiri et al., 2019; Gaur et al., 2020). However, these studies did not show the performance of ensemble ML in contrast to singular ML. Therefore, this study demonstrates for the first time the utility of ensemble ML in characterising the social-ecological drivers of LULC change.

ii) Drivers of Category A transitions – natural to anthropogenic classes

Results show that transitions in LULC classes from natural to anthropogenic classes (Category A) are mainly influenced by anthropogenic-related variables namely, population density, distance from towns and fire frequency. This is in line with previous studies that reported the effects of anthropogenic activities on the loss of natural LULC classes (Long et al., 2007; Simwanda et al., 2020). Areas with high population density are mostly associated with activities such as clearing of land (for example, woodlands and grasslands) to create space for settlements (builtup), practise agriculture (cultivated areas) and developmental projects (Wang et al., 2020; Kim et al., 2020; Song et al., 2021). Research has shown that most of the human activities (e.g. crop cultivation) mostly occur close to urban centres for easy access to agricultural inputs and the market (Schubert et al., 2018; Siddiqui et al., 2018). This study revealed that distance from urban centres influences the transition from wetlands to cultivated lands, which is supported by existing literature on the negative effects of peri-urban agriculture on wetland ecosystems (Ayambire et al., 2019; Das and Basu, 2020). According to Simorangkir (2007) fire is a common tool used to clear land for crop cultivation and prepare land for the coming agricultural season. This could explain findings from the current study that fire frequency influenced the transition from grassland to cultivated lands. Other factors such as annual average temperature, distance from rivers and drought severity have an indirect relationship with crop cultivation, human settlement and development. This study found them important to transitions from natural LULC classes to anthropogenic classes. This could be because most land use systems require specific conditions of annual average temperature and drought severity, for example, high temperature and extreme drought severity might be associated with drier conditions that are not suitable for some land uses (crop cultivation and human settlements) (Luo, 2011, (Valverde-Arias et al., 2018; Kim et al., 2020). In addition, distance from rivers, especially areas close to rivers are associated with the availability of water for irrigation to support crop cultivation in the Okavango basin (Kgathi et al., 2007).

iii) Drivers of Category B transitions –anthropogenic to natural classes

Elevation, population density and soil organic content influenced transitions from anthropogenic classes to natural classes

Table 9

Overview of change in the extent of LULC between the current LULC (2020) and projected LULC (2040).

LULC class	2020 LULC (ha)	2040 projected LULC (ha)	Projected percentage change
Bareland	129,38	340,51	163,18
Builtup	79175,89	84267,99	6,43
Water	37366,68	36056,48	-3,50
Cropland	677437,58	749359,29	10,61
Woodland	12689841,41	12609472,63	-0,63
Shrubland	7184016,50	7159585,67	-0,34
Grassland	1511686,02	1606488,32	6,27
Wetland	309810,76	243893,32	-21,27

(Category B). This could be due to the succession of cultivated fields that would have been abandoned over a long period (Osbornová et al., 2012; Jírová et al., 2012). High elevated areas are often difficult to access and research has shown that cultivated fields on such sites have high abandonment rates (Baldock, 1996; Ioffe et al., 2012). According to Verburg and Overmars (2009) and Müller et al. (2009), a decrease in population also influences the abandonment of cultivated fields. This is typical of rural communities in the Okavango basin during the post-war period when most people migrated to towns (Kgathi et al., 2006, 2007). Rural out-migration often results in decreased rural populations that are associated with low demand for agricultural space and limited workforce (Baumann et al., 2011). A finding from this study that soil organic carbon influence the transition of cultivated areas to shrubland is not new. Previous studies have reported that abandoned fields with fertile soils are associated with high succession rates (Moran et al., 2000).

iv) Drivers of Category C transitions – natural to natural classes

Distance from rivers and soil organic carbon influenced transitions from water to woodland and grasslands. This could be due to vegetation rejuvenation and succession following a drought event or a decrease in precipitation. This mostly affects areas close to rivers, especially floodplains that are vulnerable to water level fluctuations (Baptist et al., 2004; Breedveld et al., 2006). The past decades have been associated with decreasing trends in annual precipitation, floods and droughts in the Okavango basin (Motsholapheko et al., 2012). This often leads to water level fluctuations thereby facilitating rejuvenation of grasses and trees close to rivers (Odland and del Moral, 2002). The rate of vegetation rejuvenation depends on soil fertility (Van Cleve et al., 1996), which could explain the finding that soil organic carbon influences the transition from water to grassland. GDP and precipitation influence the transition from woodland to grassland. This suggests that the cutting down of trees by marginal communities which heavily depend on natural resources in the Okavango basin is exposing grass undergrowth resulting in the expansion of grasslands (Motsumi et al., 2012; Kgathi et al., 2007; Wilk et al., 2010). This is further compounded by negative trends in precipitation which leads to dying off of trees that exposes grass undergrowth (Breshears et al., 2009; Adams et al., 2012; Gaughan and Waylen, 2012). Kamwi et al. (2015) found that population density influences the transition from woodland to grassland in Sibinda communal area, Namibia. Their results differ from findings from this study possibly because they did not incorporate GDP that resamples the economic status of a locality. The advantage of understanding the economic status of an area is that it provides indications of how much community depend on natural resources (Kalimeris et al., 2020). Furthermore, variation in the scale and complexity of the social-ecological setting of a study by Kamwi et al. (2015) (communal area) to that of this study (transboundary basin) could explain how their results differ.

Population density, precipitation and average annual temperature influenced the transition from wetland to woodland, grassland and shrubland. Increased population density has often been associated with increased demand for food and housing that raise water demand and ultimately wetland loss (Mao et al., 2021; Athukorala et al., 2021). Receding wetlands may then leave more space for encroachment by herbaceous and tree species (Zedler and Kercher, 2005; Barbosa da Silva et al., 2016). The observation that population density influences wetland loss is consistent with previous literature that reported on the link between population increase and decrease in wetland extent (Hu et al., 2017; Donnelly et al., 2020). Luvuno et al. (2016) reported that the transition of wetlands to woodlands is mainly influenced by fire in Kwambonambi, South Africa which is different from findings from this study. The major limitation of their study is that it did not incorporate multiple social-ecological variables to evaluate this transition. Unlike their study, this study includes fire frequency amongst other social-ecological drivers that make the evaluation more robust. The use of various social-ecological variables provides an opportunity to accurately detect drivers from a large pool of factors influencing complex LULC transitions. Results also show that precipitation and average annual temperature influenced the transition from wetland to grassland and wetland to shrubland respectively. This could be due to decreasing annual precipitation accompanied by rising annual average temperature linked to climate change that often results in a recession of wetland and eventually encroachment by vegetation (Records et al., 2014; Mitsch and Gosselink, 2000).

v) Future LULC projections (2040)

Since the projected LULC trajectory reveal how social-ecological factors will continue to impact the availability and distribution of natural resources, key drivers of LULC change in the basin should guide management actions to closely monitor natural resources, especially the availability of water. The observation that the extent of water and wetland will likely decrease in future as a consequence of socio-economic (distance from urban centres, population density) and biophysical (distance from water, annual average precipitation, annual average temperature and soil organic carbon) factors leads to two recommendations. Firstly, a management approach to monitor water resource use in stages, involving targeting areas with high population densities and which are close to urban centres first before moving to areas with low population densities and away from urban centres, may be necessary. Secondly, an integrated land use

plan, guided by biophysical drivers which are mostly preferred for the establishment of human settlements and agriculture may be the best strategy to contain the continuous spread of anthropogenic classes across the basin. This may include the use of ensemble ML algorithms based on key social-ecological variables to map the most suitable areas for the establishment of anthropogenic activities, followed by estimating their maximum capacity which can be compared with the potential ecological implications of development on those areas. This will help in coming up with zones for a land-use plan, which is relevant to targeted monitoring and sustainable use of natural resources such as water at a transboundary scale.

vi) Study limitations and future recommendations

Although this study gave insights on drivers of LULC change in the Okavango basin, it could not factor in the effect of change in policies and environmental law during the post-war period. Mainly because laws and policies are changed at a country level and community buy-in to laws and policies cannot be easily presented spatially. Even though transboundary commissions such as OKACOM pass policies and laws, adoption by all parties involved may not be uniform and variation in the distribution of their adoption can be difficult to present across space. Future research should combine spatial models with social surveys to include the effects of change in environmental policies, variation in policy adoption and community perceptions on drivers of LULC change. Additionally, future studies should test the performance of other ML techniques reported to have enhanced performance in LULC classification in assessing drivers of LULC, particularly the SVR and KRR. KRR and SVR use kernel functions to combine ridge regression and classification (Vovk, 2013). The influence of carbon dioxide distribution on LULC change could not be tested due to the unavailability of valid spatial data to incorporate it in our analysis at such a scale.

5. Conclusion

This study aimed to evaluate the utility of ensemble ML in determining the social-ecological drivers of LULC change. The specific objectives were to 1) compare the performance of singular and ensemble machine learning techniques in determining drivers of change, 2) identify the social-ecological drivers of LULC. To do that, a set of singular ML techniques (Random forest, Gradient boosting model, Maximum Entropy, Artificial Neural Network, Classification Tree Analysis) was used and their performance was compared with that of their ensemble in teasing apart the drivers of LULC change transitions during the period between 2004 and 2020. Model performance was evaluated using the Receiver Operating Characteristic (ROC) and the True skill statistic (TSS). The variable importance was used to evaluate the contribution of social-ecological variables to each LULC transition.

The study demonstrated the superiority of ensemble models over singular models when evaluating drivers of LULC. The highest accuracy was obtained from the use of ML ensembles. A conclusion was drawn from this study that, evaluating social-ecological drivers of LULC change with ML ensembles results in the identification of drivers with greater statistical confidence. It is hoped that others, who wish to detect drivers of change with improved accuracy, may now consider combining several ML algorithms. However, in situations where computational power is a limitation to run ensemble, individual ML models such as GBM or RF could be used to detect drivers of change.

The study determined that a combination of both anthropogenic (distance from rivers, distance from urban centres, population density, fire frequency) and natural factors (drought severity, annual average temperature, precipitation) influence the expansion of anthropogenic classes. However, most of the top important natural variables seem to have an indirect influence on anthropogenic activities that suggest the dominance of anthropogenic conducts in LULC change in the Okavango basin. Furthermore, distance from urban centres, population density, distance from water, annual average precipitation, annual average temperature and soil organic carbon mainly influence the decrease in water-related classes. Altering the observed social-ecological drivers of loss of water and forest resources could help in sustainable conservation.

The study revealed that natural classes (wetland, shrubland, water and woodlands) have gradually been decreasing at the expense of anthropogenic classes (builtup and cultivated) which clearly shows how anthropogenic activities are affecting natural resources in the Okavango basin. A similar LULC trajectory is projected in the future (2040), which suggests the need for a basin-wide land-use plan to minimise pressure on resources and ensure sustainable use, particularly of water resources. Buy-in of the land-use plan by countries sharing the basin could be ensured through integrating existing structures spearheaded by transboundary commissions (OKACOM) with recommendations of the study.

Ethical statement

The authors declare that all ethical practices have been followed in relation to the development, writing, and publication of the article.

Author Agreement Statement

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

Declaration of competing interest

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

Acknowledgements

We thank USAID Resilient Waters for funding this project, implemented under prime contract number 72067418C00007. Our appreciation goes to the OKACOM and Okavango Wildland Trust for supplying us with some ground data which was used in this study. Our appreciation also goes to the editors and anonymous reviewers for their constructive comments on an earlier version of the manuscript.

Appendix

Appendix 1. Shows the confusion matrix of the Bsh climate zone

	Bareland	Builtup	Water	Cultivated	Woodland	Shrubland	Grassland	Wetland
Bareland	125	0	1	5	0	0	0	0
Builtup	0	130	0	0	0	2	0	1
Water	1	1	159	7	1	0	1	3
Cultivated	2	1	5	174	0	0	0	0
Woodland	0	2	3	2	108	0	4	5
Shrubland	0	9	2	1	2	142	1	5
Grassland	0	3	2	2	4	0	208	1
Wetland	0	0	4	0	1	2	6	150

Appendix 2. Shows the confusion matrix of the Cwa climate zone

	Bareland	Builtup	Water	Cultivated	Woodland	Shrubland	Grassland	Wetland
Bareland	43	0	0	0	0	0	1	0
Builtup	0	39	0	1	0	0	0	0
Water	1	0	36	0	0	0	0	0
Cultivated	3	0	0	38	1	1	0	0
Woodland	0	0	0	0	34	1	0	0
Shrubland	0	0	1	4	0	31	1	0
Grassland	0	0	0	1	0	1	33	0
Wetland	0	0	0	1	1	1	0	37

Appendix 3. Shows the confusion matrix of the Cwb climate zone

	Bareland	Builtup	Water	Cultivated	Woodland	Shrubland	Grassland	Wetland
Bareland	34	0	0	0	0	0	1	0
Builtup	0	41	0	0	0	0	0	0
Water	0	0	45	0	0	0	0	0
Cultivated	0	0	3	32	0	0	1	0
Woodland	0	0	0	0	44	2	1	0
Shrubland	0	0	4	1	3	30	1	1
Grassland	0	0	2	2	1	1	30	1
Wetland	0	0	0	0	2	0	0	39

References

- Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience Remote Sens.* 57, 1–20.
- Adams, H.D., Luce, C.H., Breshears, D.D., Allen, C.D., Weiler, M., Hale, V.C., Smith, A.M., Huxman, T.E., 2012. Ecohydrological consequences of drought-and infestation-triggered tree die-off: insights and hypotheses. *Ecohydrology* 5, 145–159.
- Adhikari, A., Hansen, A.J., 2018. Land use change and habitat fragmentation of wildland ecosystems of the North Central United States. *Landsc. Urban Plann.* 177, 196–216.
- Alam, N., Saha, S., Gupta, S., Chakraborty, S., 2021. Ann. In: Prediction Modelling of Riverine Landscape Dynamics in the Context of Sustainable Management of Floodplain: a Geospatial Approach. *GIS* 1–16.
- Arficho, M., Thiel, A., 2020. Does land-use policy moderate impacts of climate anomalies on LULC change in dry-lands? An empirical enquiry into drivers and moderators of LULC change in southern Ethiopia. *Sustainability* 12, 6261.
- Arino, O., 2010. GlobCover 2009.

- Ashaolu, E.D., Olorunfemi, J.F., Ifabiyi, I.P., 2019. Assessing the spatio-temporal pattern of land use and land cover changes in Osun drainage basin, Nigeria. *J. Environ. Geogr.* 12, 41–50.
- Athukorala, D., Estoqe, R.C., Murayama, Y., Matsushita, B., 2021. Impacts of urbanization on the Muthurajawela marsh and Negombo lagoon, Sri Lanka: implications for landscape planning towards a sustainable urban wetland ecosystem. *Rem. Sens.* 13, 316.
- Ayambire, R.A., Ampsonah, O., Peprah, C., Takyi, S.A., 2019. A review of practices for sustaining urban and peri-urban agriculture: implications for land use planning in rapidly urbanising Ghanaian cities. *Land Use Pol.* 84, 260–277. <https://doi.org/10.1016/j.landusepol.2019.03.004>.
- Baldock, D., 1996. Farming at the Margins. IIEP and LEI-DLO.
- Baptist, M.J., Penning, W.E., Duel, H., Smits, A.J., Geerling, G.W., Van der Lee, G.E., Van Alphen, J.S., 2004. Assessment of the effects of cyclic floodplain rejuvenation on flood levels and biodiversity along the Rhine River. *River Res. Appl.* 20, 285–297.
- Barbosa da Silva, F.H., Arieira, J., Parolin, P., Nunes da Cunha, C., Junk, W.J., 2016. Shrub encroachment influences herbaceous communities in flooded grasslands of a neotropical savanna wetland. *Appl. Veg. Sci.* 19, 391–400.
- Baumann, M., Kuemmerle, T., Elbakidze, M., Ozdogan, M., Radefol, V.C., Keuler, N.S., Prishchepov, A.V., Krughov, I., Hostert, P., 2011. Patterns and drivers of post-socialist farmland abandonment in Western Ukraine. *Land Use Pol.* 28, 552–562. <https://doi.org/10.1016/j.landusepol.2010.11.003>.
- Belgiu, M., Dragut, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogrammetry Remote Sens.* 114, 24–31.
- Berkes, F., 2011. Restoring Unity: The Concept of Marine Social-Ecological Systems. *World Fisheries. A Social-Ecological Analysis* 9–28.
- Betts, M.G., Wolf, C., Ripple, W.J., Phalan, B., Millers, K.A., Duarte, A., Butchart, S.H., Levi, T., 2017. Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* 547, 441–444.
- Bickling, S., Burkhardt, B., Kruse, M., Müller, F., 2019. Bayesian Belief Network-based assessment of nutrient regulating ecosystem services in Northern Germany. *PLoS One* 14, e0216053.
- Breedveld, M., Liefhebber, D., Geerling, G.W., Smits, A.J.M., Ragas, A.M.J., Leuven, R., van den Berg, J.H., 2006. Succession and rejuvenation in floodplains along the river Allier (France). In: *Living Rivers: Trends and Challenges in Science and Management*. Springer, pp. 71–86.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*. CRC press.
- Breshears, D.D., Myers, O.B., Meyer, C.W., Barnes, F.J., Zou, C.B., Allen, C.D., McDowell, N.G., Pockman, W.T., 2009. Tree die-off in response to global change-type drought: mortality insights from a decade of plant water potential measurements. *Front. Ecol. Environ.* 7, 185–189.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. *Global Change Biol.* 16, 1145–1157.
- Camps-Valls, G., Bruzzone, L., 2009. *Kernel Methods for Remote Sensing Data Analysis*. John Wiley & Sons.
- CIESIN. Center for International Earth Science Information Network, Columbia University, 2018. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals, Revision 11.
- Chalmers, N., Fabricius, C., 2007. Expert and generalist local knowledge about land-cover change on South Africa's Wild Coast: can local ecological knowledge add value to science? *Ecol. Soc.* 12.
- Chen, X., Vierling, L., Rowell, E., DeFelice, T., 2004. Using lidar and effective LAI data to evaluate IKONOS and Landsat 7 ETM+ vegetation cover estimates in a ponderosa pine forest. *Remote Sens. Environ.* 91, 14–26.
- Chughtai, A.H., Abbasi, H., Ismail, R.K., 2021. A review on change detection method and accuracy assessment for land use land cover. *Remote Sens. Appl. Soc. Environ.*, 100482.
- Cook, D., 2016. *Practical Machine Learning with H2O: Powerful, Scalable Techniques for Deep Learning and AI*. O'Reilly Media, Inc.
- CoreTeam, Rd, 2014, 2014. R foundation for statistical computing, Vienna.
- Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* 63, 22–33. <https://doi.org/10.1016/j.cageo.2013.10.008>.
- Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap. *IEEE Trans. Geosci. Rem. Sens.* (3), 256–263.
- da Silva, J.F., Cicerelli, R.E., Almeida, T., Neumann, M.R.B., de Souza, A.L.F., 2020. Land use/cover (LULC) mapping in Brazilian cerrado using neural network with sentinel-2 data. *Floresta* 50 (3), 1430–1438.
- Das, A., Basu, T., 2020. Assessment of peri-urban wetland ecological degradation through importance-performance analysis (IPA): a study on Chatra Wetland, India. *Ecol. Indicat.* 114, 106274.
- de Sousa, C., Fatoyinbo, L., Neigh, C., Boucka, F., Angoue, V., Larsen, T., 2020. Cloud-computing and machine learning in support of country-level land cover and ecosystem extent mapping in Liberia and Gabon. *PLoS One* 15, e0227438.
- Domisch, S., Kakouei, K., Martínez-López, J., Bagstad, K.J., Magrach, A., Balbi, S., Villa, F., Funk, A., Hein, T., Borgwardt, F., 2019. Social equity shapes zone-selection: balancing aquatic biodiversity conservation and ecosystem services delivery in the transboundary Danube River Basin. *Sci. Total Environ.* 656, 797–807.
- Donnelly, J.P., King, S.L., Silverman, N.L., Collins, D.P., Carrera-Gonzalez, E.M., Lafón-Terrazas, A., Moore, J.N., 2020. Climate and human water use diminish wetland networks supporting continental waterbird migration. *Global Change Biol.* 26, 2042–2059.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitao, P.J., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *EcoGraphy* 36, 27–46.
- dos Santos, J.Y.G., Montenegro, S.M.G.L., da Silva, R.M., Santos, C.A.G., Quinn, N.W., Dantas, A.P.X., Neto, A.R., 2021. Modeling the impacts of future LULC and climate change on runoff and sediment yield in a strategic basin in the Caatinga/Atlantic forest ecotone of Brazil. *Catena* 203, 105308.
- Doyle, C., Beach, T., Luzzadder-Beach, S., 2021. Tropical forest and wetland losses and the role of protected areas in Northwestern Belize, revealed from Landsat and machine learning. *Rem. Sens.* 13, 379.
- Dudley, J.P., Gibson, D.S.C., Haynes, G., Klimowicz, J., 2001. Drought mortality of bush elephants in Hwange national Park, Zimbabwe. *Afr. J. Ecol.* 39, 187–194.
- Elhance, A.P., 1999. *Hydropolitics in the Third World: Conflict and Cooperation in International River Basins*. US Institute of Peace Press.
- Ellery, W., Ellery, K., McCarthy, T.S., Cairncross, B., Oelofse, R., 1989. A peat fire in the Okavango Delta, Botswana, and its importance as an ecosystem process. *Afr. J. Ecol.* 27, 7–21.
- Ellis, E.C., Kaplan, J.O., Fuller, D.Q., Vavrus, S., Goldewijk, K.K., Verburg, P.H., 2013. Used planet: a global history. *Proc. Natl. Acad. Sci. Unit. States Am.* 110, 7978–7985.
- Eze, P.N., Molwalefhe, L.N., Kebonye, N.M., 2021. Geochemistry of soils of a deep pedon in the Okavango Delta, NW Botswana: implications for pedogenesis in semi-arid regions. *Geoderma Reg* 24, e00352. <https://doi.org/10.1016/j.geodrs.2020.e00352>.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315.
- Folwell, S., Farchuaron, F., Demuth, S., Gustard, A., Planos, E., Seatena, F., Servat, E., 2006. The impacts of climate change on water resources in the Okavango basin. *IAHS Publ.* 308, 382.
- Foody, G.M., Mathur, A., 2004. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sens. Environ.* 93, 107–117.
- Gaughan, A.E., Waylen, P.R., 2012. Spatial and temporal precipitation variability in the Okavango–Kwando–Zambezi catchment, southern Africa. *J. Arid Environ.* 82, 19–30.
- Gaur, S., Mittal, A., Bandyopadhyay, A., Holman, I., Singh, R., 2020. Spatio-temporal analysis of land use and land cover change: a systematic model inter-comparison driven by integrated modelling techniques. *Int. J. Rem. Sens.* 41, 9229–9255.
- Gibson, L., Münch, Z., Palmer, A., Mantel, S., 2018. Future land cover change scenarios in South African grasslands—implications of altered biophysical drivers on land management. *Heliyon* 4, e00693.
- Gilbertson, J.K., Kemp, J., Van Niekerk, A., 2017. Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. *Computers and Electronics in Agriculture* 134, 151–159.
- Gondwe, M.F.K., 2020. Influence of Miombo Woodlands Management, Drivers on Land Use/cover and Forest Change, Woody Composition/diversity, Population Structure in Malawi.

- Grenouillet, G., Buisson, L., Casajus, N., Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34, 9–17. <https://doi.org/10.1111/j.1600-0587.2010.06152.x>.
- Griffith, D.A., 2013. Establishing qualitative geographic sample size in the presence of spatial autocorrelation. *Ann. Assoc. Am. Geogr.* 103, 1107–1122. <https://doi.org/10.1080/00045608.2013.776884>.
- Guo, M., Ma, S., Wang, L.-J., Lin, C., 2021. Impacts of future climate change and different management scenarios on water-related ecosystem services: a case study in the Jianghuai ecological economic Zone, China. *Ecol. Indicat.* 127, 107732.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., 2013. High-resolution global maps of 21st-century forest cover change. *science* 342, 850–853.
- Hazen, B.C., Berry, M.W., 1997. The simulation of land-cover change using a distributed computing environment. *Simulat. Pract. Theor.* 5, 489–514.
- Heinl, M., 2005. Fire Regime and Vegetation Response in the Okavango Delta, Botswana. Technische Universität München.
- Hishe, H., Giday, K., Van Orshoven, J., Muys, B., Taheri, F., Azadi, H., Feng, L., Zamani, O., Mirzaei, M., Witlox, F., 2021. Analysis of land use land cover dynamics and driving factors in Desa's forest in Northern Ethiopia. *Land Use Pol.* 101, 105039.
- Hu, S., Niu, Z., Chen, Y., Li, L., Zhang, H., 2017. Global wetlands: potential distribution, wetland loss, and status. *Sci. Total Environ.* 586, 319–327. <https://doi.org/10.1016/j.scitotenv.2017.02.001>.
- Huete, A., 1988. Huete, AR A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*. *Remote Sens. Environ.* 25, 295–309.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Ioffe, G., Nefedova, T., Kirsten, D.B., 2012. Land abandonment in Russia. *Eurasian Geogr. Econ.* 53, 527–549.
- Islam, M.D., Islam, K.S., Ahasan, R., Mia, M.R., Haque, M.E., 2021. A data-driven machine learning-based approach for urban land cover change modeling: A case of Khulna City Corporation area. *Remote Sens. Appl. : Soc. Environ.* 24, 100634.
- Ismail, R., Mutanga, O., 2011. Discriminating the early stages of Sirex noctilio infestation using classification tree ensembles and shortwave infrared bands. *Int. J. Rem. Sens.* 32 (15), 4249–4266.
- Jírová, A., Kloudová, A., Prach, K., 2012. Spontaneous restoration of target vegetation in old-fields in a central European landscape: a repeated analysis after three decades. *Appl. Veg. Sci.* 15, 245–252.
- Kalimeris, P., Bithas, K., Richardson, C., Nijkamp, P., 2020. Hidden linkages between resources and economy: a "Beyond-GDP" approach using alternative welfare indicators. *Ecol. Econ.* 169, 106508.
- Kamwi, J.M., Chirwa, P.W., Manda, S.O., Graz, P.F., Kätsch, C., 2015. Livelihoods, land use and land cover change in the Zambezi Region, Namibia. *Popul. Environ.* 37, 207–230.
- Kamwi, J.M., Cho, M.A., Kaetsch, C., Manda, S.O., Graz, P.F., Chirwa, P.W., 2018. Assessing the spatial drivers of land use and land cover change in the protected and communal areas of the Zambezi Region, Namibia. *Land* 7, 131.
- Kanagaraj, R., Araujo, M.B., Barman, R., Davida, P., De, R., Digal, D.K., Gopi, G.V., Johnsingh, A.J.T., Kakati, K., Kramer-Schadt, S., 2019. Predicting range shifts of Asian elephants under global change. *Divers. Distrib.* 25, 822–838.
- Kassawar, T., Eckert, S., Hurni, K., Zeleke, G., Hurni, H., 2018. Reducing landscape heterogeneity for improved land use and land cover (LULC) classification across the large and complex Ethiopian highlands. *Geocarto Int.* 33, 53–69.
- Kgathi, D.L., Kniveton, D., Ringrose, S., Turton, A.R., Vanderpost, C.H.M., Lundqvist, J., Seely, M., 2006. The Okavango; a river supporting its people, environment and economic development. *J. Hydrol.*, Water Resources in Regional Development: The Okavango River 331, 3–17. <https://doi.org/10.1016/j.jhydrol.2006.04.048>.
- Kgathi, D.L., Ngwenya, B.N., Wilk, J., 2007. Shocks and rural livelihoods in the Okavango delta, Botswana. *Dev. South Afr.* 24, 289–308. <https://doi.org/10.1080/03768350701327186>.
- Khoshnoodmotlagh, S., Verrelst, J., Daneshi, A., Mirzaei, M., Azadi, H., Haghghi, M., Hatamimanesh, M., Marofi, S., 2020. Transboundary basins need more attention: anthropogenic impacts on land cover changes in aras River Basin, monitoring and prediction. *Rem. Sens.* 12, 3329. <https://doi.org/10.3390/rs12203329>.
- Kim, Y., Newman, G., Güneralp, B., 2020. A review of driving factors, scenarios, and topics in urban land change models. *Land* 9, 246.
- Kleemann, J., Baysal, G., Bulley, H.N.N., Fürst, C., 2017. Assessing driving forces of land use and land cover change by a mixed-method approach in north-eastern Ghana, West Africa. *J. Environ. Manag.* 196, 411–442. <https://doi.org/10.1016/j.jenvman.2017.01.053>.
- Kouassi, J.-L., Gyau, A., Diby, L., Bene, Y., Kouamé, C., 2021. Assessing land use and land cover change and farmers' perceptions of deforestation and land degradation in South-West Côte d'Ivoire, west Africa. *Land* 10, 429.
- Kuhn, M., 2015. Caret: classification and regression training. *1505.003 Astrophys. Source Code Libr. ascl: 1505.003*.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.*, Special Issue: Progress of Machine Learning in Geosciences 7, 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>.
- Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sens. Environ.* 90, 331–336. <https://doi.org/10.1016/j.rse.2004.01.007>.
- Li, Y., Sun, Y., Li, J., 2021. Heterogeneous effects of climate change and human activities on annual landscape change in coastal cities of mainland China. *Ecol. Indicat.* 125, 107561.
- Lin, Y., Deng, X., Li, X., Ma, E., 2014. Comparison of multinomial logistic regression and logistic regression: which is more efficient in allocating land use? *Front. Earth Sci.* 8, 512–523.
- Liú, X., Huang, Y., Xu, X., Li, Xieueco, Li, Xia, Ciais, P., Lin, P., Gong, K., Ziegler, A.D., Chen, A., 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nat. Sustain.* 1–7.
- Long, H., Tang, G., Li, X., Heilig, G.K., 2007. Social-economic driving forces of land-use change in Kunshan, the Yangtze River Delta economic area of China. *J. Environ. Manag.* 83, 351–364. <https://doi.org/10.1016/j.jenvman.2006.04.003>.
- Luo, Q., 2011. Temperature thresholds and crop production: a review. *Clim. Change* 109, 583–598.
- Luvuno, L.B., Kotze, D.C., Kirkman, K.P., 2016. Long-term landscape changes in vegetation structure: fire management in the wetlands of KwaMbonambi, South Africa. *Afr. J. Aquat. Sci.* 41, 279–288.
- Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X., Deliang, C., 2017. Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS Int. J. Geo-Inf.* 6 (2), 51.
- Mao, D., Tian, Y., Wang, Z., Jia, M., Du, J., Song, C., 2021. Wetland changes in the Amur River Basin: differing trends and proximate causes on the Chinese and Russian sides. *J. Environ. Manag.* 280, 111670.
- Marmion, M., Parviainen, M., Luoto, M., Heikkilä, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15 (1), 59–69.
- Marondedze, A.K., Schütt, B., 2019. Dynamics of land use and land cover changes in Harare, Zimbabwe: a case study on the linkage between drivers and the Axis of urban expansion. *Land* 8, 155.
- McCoy, J., Johnston, K., 2001. Using ArcGIS Spatial Analyst. Esri Redlands.
- McLeman, R., Fontanella, F., Greig, C., Heath, G., Robertson, C., 2021. Population Responses to the 1976 South Dakota Drought: Insights for Wider Drought Migration Research. *Popul. Space Place* e2465.
- Mianabadi, A., Davary, K., Mianabadi, H., Karimi, P., 2020. International environmental Conflict management in transboundary river basins. *Water resour. OR Manag.* 34, 3445–3464.
- Mitsch, W.J., Gosselink, J.G., 2000. The value of wetlands: importance of scale and landscape setting. *Ecol. Econ.* 35, 25–33.

- Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards, T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecol. Model.*, Predicting Species Distributions 199, 176–187. <https://doi.org/10.1016/j.ecolmodel.2006.05.021>.
- Moran, E.F., Brondizio, E.S., Tucker, J.M., da Silva-Forsberg, M.C., McCracken, S., Falesi, I., 2000. Effects of soil fertility and land-use on forest succession in Amazônia. *For. Ecol. Manag.* 139, 93–108. [https://doi.org/10.1016/S0378-1127\(99\)00337-0](https://doi.org/10.1016/S0378-1127(99)00337-0).
- Motsholapheko, M.R., Kgathi, D.L., Vanderpost, C., 2012. Rural livelihood diversification: a household adaptive strategy against flood variability in the Okavango Delta, Botswana. *Agrekon* 51, 41–62. <https://doi.org/10.1080/03031853.2012.741204>.
- Motsumi, S., Magole, L., Kgathi, D., 2012. Indigenous knowledge and land use policy: implications for livelihoods of flood recession farming communities in the Okavango Delta, Botswana. *Phys. Chem. Earth Parts ABC*, 12th WaterNet/WARFSA/GWP-SA Symposium: Harnessing the rivers of knowledge for social-economic development, 50–52 climate adaptation & environmental sustainability 185–195. <https://doi.org/10.1016/j.pce.2012.09.013>.
- Mpakairi, K.S., Muvengwi, J., 2019. Night-time lights and their influence on summer night land surface temperature in two urban cities of Zimbabwe: a geospatial perspective. *Urban Clim.* 29, 100468.
- Müller, D., Kuemmerle, T., Rusu, M., Griffiths, P., 2009. Lost in transition: determinants of post-socialist cropland abandonment in Romania. *J. Land Use Sci.* 4, 109–129.
- Münch, Z., Gibson, L., Palmer, A., 2019. Monitoring effects of land cover change on biophysical drivers in rangelands using albedo. *Land* 8, 33. <https://doi.org/10.3390/land8020033>.
- Odland, A., del Moral, R., 2002. Thirteen years of wetland vegetation succession following a permanent drawdown, Myrkdalen Lake, Norway. *Plant Ecol.* 162, 185–198. <https://doi.org/10.1023/A:1020388910724>.
- Olson, J.M., Alagarswamy, G., Andresen, J.A., Campbell, D.J., Davis, A.Y., Ge, J., Huebner, M., Lofgren, B.M., Lusch, D.P., Moore, N.J., 2008. Integrating diverse methods to understand climate-land interactions in East Africa. *Geoforum* 39, 898–911.
- Opacka, B., Müller, J.-F., Stavrakou, T., Bauwens, M., Sindelarova, K., Markova, J., Guenther, A.B., 2021. Global and regional impacts of land cover changes on isoprene emissions derived from spaceborne data and the MEGAN model. *Atmos. Chem. Phys. Discuss.* 1–42.
- Osbornová, J., Kovárová, M., Leps, J., Prach, K., 2012. Succession in Abandoned Fields: Studies in Central Bohemia, Czechoslovakia. Springer Science & Business Media.
- Pal, M., Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Rem. Sens.* 48 (5), 2297–2307.
- Parviainen, M., Marmion, M., Luoto, M., Thuiller, W., Heikkilä, R.K., 2009. Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biol. Conserv.* 142, 2501–2509.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175.
- Phiri, D., Morgenroth, J., Xu, C., 2019. Long-term land cover change in Zambia: an assessment of driving factors. *Sci. Total Environ.* 697, 134206.
- Porto, J.G., Clover, J., 2003. The Peace Dividend in Angola: Strategic Implications for Okavango Basin Cooperation. *Transbound. Rivers Sovereignty Dev. Hydropolitical Driv. Okavango River Basin Pretoria Geneva AWIRU Green Cross Int.*
- Rahman, M.T.U., Tabassum, F., Rasheduzzaman, Md, Saba, H., Sarkar, L., Ferdous, J., Uddin, S.Z., Zahedul Islam, A.Z.M., 2017. Temporal dynamics of land use/land cover change and its prediction using CA-ANN model for southwestern coastal Bangladesh. *Environ. Monit. Assess.* 189, 565. <https://doi.org/10.1007/s10661-017-6272-0>.
- Rai, R., Zhang, Y., Paudel, B., Acharya, B.K., Basnet, L., 2018. Land use and land cover dynamics and assessing the ecosystem service values in the trans-boundary gandaki River Basin, central Himalayas. *Sustainability* 10, 3052. <https://doi.org/10.3390/su10093052>.
- Records, R.M., Arabi, M., Fassnacht, S.R., Duffy, W.G., Ahmadi, M., Hegevisch, K.C., 2014. Climate change and wetland loss impacts on a western river's water quality. *Hydroclim. Earth Syst. Sci.* 18, 4509–4527.
- Ren, Y., Yan, J., Wei, X., Wang, Y., Yang, Y., Hua, L., Xiong, Y., Niu, X., Song, X., 2012. Effects of rapid urban sprawl on urban forest carbon stocks: integrating remotely sensed, GIS and forest inventory data. *J. Environ. Manag.* 113, 447–455.
- Revermann, R., Finckh, M., Stellmes, M., Strohbach, B.J., Frantz, D., Oldeland, J., 2016. Linking land surface phenology and vegetation-plot databases to model terrestrial plant diversity of the Okavango Basin. *Rem. Sens.* 8, 370.
- Rimba, A.B., Mohan, G., Chapagain, S.K., Arumansawang, A., Payus, C., Fukushima, K., Osawa, T., Avtar, R., 2021. Impact of population growth and land use and land cover (LULC) changes on water quality in tourism-dependent economies using a geographically weighted regression approach. *Environ. Sci. Pollut. Res.* 1–19.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. In: Proceedings of the Earth Resources Technology Satellite Symposium NASA SP-351, vol. 1, pp. 309–317 (Wash. DC NASA).
- Ruan, J., Kherbouche, F., Genty, D., Blamart, D., Cheng, H., Dewilde, F., Hachi, S., Edwards, R.L., Régner, E., Michelot, J.-L., 2016. Evidence of a prolonged drought ca. 4200 yr BP correlated with prehistoric settlement abandonment from the Gueldaman GLD1 Cave, Northern Algeria. *Clim. Past* 12, 1–14. <https://doi.org/10.5194/cp-12-1-2016>.
- Rutherford, G.N., Guisan, A., Zimmermann, N.E., 2007. Evaluating sampling strategies and logistic regression methods for modelling complex land cover changes. *J. Appl. Ecol.* 44, 414–424. <https://doi.org/10.1111/j.1365-2664.2007.01281.x>.
- Saha, A.K., Arora, M.K., Csaplócvics, E., Gupta, R.P., 2005. Land cover classification using IRS LISS III image and DEM in a rugged terrain: a case study in Himalayas. *Geocarto Int.* 20, 33–40. <https://doi.org/10.1080/10106040508542343>.
- Saini, R., Ghosh, S.K., 2017. Ensemble classifiers in remote sensing: a review. In: 2017 International Conference on Computing, Communication and Automation (ICCCA). IEEE, pp. 1148–1152.
- Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., 2000. Global biodiversity scenarios for the year 2100. *science* 287, 1770–1774.
- Santé, I., García, A.M., Miranda, D., Grecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape Urban Plann.* 96, 108–122.
- Schirpke, U., Kohler, M., Leitinger, G., Fontana, V., Tasser, E., Tappeiner, U., 2017. Future impacts of changing land-use and climate on ecosystem services of mountain grassland and their resilience. *Ecosyst. Serv.* 26, 79–94.
- Schubert, H., Caballero Calvo, A., Raucheker, M., Rojas-Zamora, O., Brokamp, G., Schütt, B., 2018. Assessment of land cover changes in the Hinterland of barranquilla (Colombia) using Landsat imagery and logistic regression. *Land* 7, 152.
- Sharma, E., Chettri, N., 2005. ICIMOD's transboundary biodiversity management initiative in the Hindu Kush-Himalayas. *Mt. Res. Dev.* 278–281.
- Shiferaw, H., Alamirew, T., Kassawmar, T., Zeleke, G., 2021. Evaluating ecosystems services values due to land use transformation in the Gojob watershed, Southwest Ethiopia. *Environ. Syst. Res.* 10, 1–12.
- Sibanda, S., Ahmed, F., 2021. Modelling historic and future land use/land cover changes and their impact on wetland area in Shashe sub-catchment, Zimbabwe. *Model. Earth Syst. Environ.* 7, 57–70.
- Siddiqui, Asfa, Siddiqui, Almas, Maithani, S., Jha, A.K., Kumar, P., Srivastav, S.K., 2018. Urban growth dynamics of an Indian metropolitan using CA Markov and Logistic Regression. *Egypt. J. Remote Sens. Space Sci.* 21, 229–236.
- Simorangkir, D., 2007. Fire use: is it really the cheaper land preparation method for large-scale plantations? *Mitig. Adapt. Strategies Glob. Change* 12, 147–164.
- Simwanda, M., Murayama, Y., Ranagalage, M., 2020. Modeling the drivers of urban land use changes in Lusaka, Zambia using multi-criteria evaluation: an analytic network process approach. *Land Use Pol.* 92, 104441.
- Singh, C., Murdoch, W.J., Yu, B., 2018. Hierarchical Interpretations for Neural Network Predictions. *ArXiv Prepr. ArXiv180605337*.
- Singh, S.K., Laari, P.B., Mustak, S.K., Srivastava, P.K., Szabó, S., 2018. Modelling of land use/land cover change using earth observation data-sets of Tons River Basin, Madhya Pradesh, India. *Geocarto Int.* 33, 1202–1222.
- Song, X., Feng, Q., Xia, F., Li, X., Scheffran, J., 2021. Impacts of changing urban land-use structure on sustainable city growth in China: a population-density dynamics perspective. *Habitat Int.* 107, 102296.

- Steudel, T., Göhmann, H., Flügel, W.A., Helmschrot, J., 2013. Assessment of hydrological dynamics in the upper Okavango River basins. *Biodivers. Ecol.* 5, 247–262.
- Tasser, E., Leitinger, G., Tappeiner, U., 2017. Climate change versus land-use change—what affects the mountain landscapes more? *Land Use Pol.* 60, 60–72.
- Thuiller, W., Georges, D., Engler, R., Breiner, F., 2013. biomod2: ensemble platform for species distribution modeling. *R Package Version 2*, r560.
- Uyank, G.K., Güler, N., 2013. A study on multiple linear regression analysis. *Procedia - Soc. Behav. Sci.*, 4th International Conference on New Horizons in Education 106, 234–240. <https://doi.org/10.1016/j.sbspro.2013.12.027>.
- Valverde-Arias, O., Garrido, A., Valencia, J.L., Tarquis, A.M., 2018. Using geographical information system to generate a drought risk map for rice cultivation: case study in Babahoyo canton (Ecuador). *Biosyst. Eng.* 168, 26–41.
- Van Cleve, K., Viereck, L.A., Dyrness, C.T., 1996. State factor control of soils and forest succession along the Tanana River in interior Alaska, USA. *Arct. Alp. Res.* 28, 388–400.
- Van Deventer, A.P., Ward, A.D., Gowda, P.H., Lyon, J.G., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogramm. Eng. Rem. Sens.* 63, 87–93.
- Veldkamp, A., Fresco, L.O., 1996. CLUE: a conceptual model to study the conversion of land use and its effects. *Ecol. Model.* 85, 253–270.
- Verburg, P.H., Overmars, K.P., 2009. Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecol.* 24, 1167–1181.
- Vovk, V., 2013. Kernel ridge regression. In: *Empirical Inference*. Springer, Berlin, Heidelberg. 105–116.
- Wang, L., Jia, Y., Li, X., Gong, P., 2020. Analysing the driving forces and environmental effects of urban expansion by mapping the speed and acceleration of built-up areas in China between 1978 and 2017. *Rem. Sens.* 12, 3929.
- Wang, Q., Ren, Q., Liu, J., 2016. Identification and apportionment of the drivers of land use change on a regional scale: unbiased recursive partitioning-based stochastic model application. *Agric. Ecosyst. Environ.* 217, 99–110. <https://doi.org/10.1016/j.agee.2015.11.004>.
- 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, 827.
- Weber, T., 2013. Okavango basin—Climate. *Biodivers Ecol* 5, 15–17.
- Wilk, J., Andersson, L., Wolski, P., Kgathi, D., Ringrose, S., Vanderpost, C., 2010. Changing flow in the Okavango basin: upstream developments and downstream effects. In: *Integrated Watershed Management*. Springer, pp. 99–112.
- Winkler, K., Fuchs, R., Rounsevell, M., Herold, M., 2021. *Spatiotemporal Patterns of Global Land Use Change: Understanding Processes and Drivers*. Copernicus Meetings.
- Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Rem. Sens.* 27, 3025–3033.
- Xu, Y., McNamara, P., Wu, Y., Dong, Y., 2013. An econometric analysis of changes in arable land utilization using multinomial logit model in Pinggu district, Beijing, China. *J. Environ. Manag.* 128, 324–334.
- Yuan, Z., Xu, J., Wang, Y., Yan, B., 2021. Analyzing the influence of land use/land cover change on landscape pattern and ecosystem services in the Poyang Lake Region, China. *Environ. Sci. Pollut. Res.* 1–14.
- Zedler, J.B., Kercher, S., 2005. Wetland resources: status, trends, ecosystem services, and restorability. *Annu. Rev. Environ. Resour.* 30, 39–74.
- Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Rem. Sens.* 24, 583–594.
- Zhang, Y., Kimber, D.Y., Coslett, H.B., Schwartz, M.F., Wang, Z., 2014. Multivariate lesion-symptom mapping using support vector regression. *Hum. Brain Mapp.* 35 (12), 5861–5876.
- Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., Zhuang, J., 2018. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* 565, 720–736.
- Zhao, H., Chen, X., 2005. Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+. In: *International Geoscience and Remote Sensing Symposium*, p. 1666.