



AALBORG UNIVERSITET
STUDENTERRAPPORT

Stimulating sustainable
development goals'
implementation and
conservation action:
PREDICTING FUTURE LAND USE
AND LAND COVER CHANGE IN THE
VIRUNGA NATIONAL PARK

MASTER THESIS CAND.TECH.
AALBORG UNIVERSITY COPENHAGEN
WORD COUNT: 21.370

07 JUNE 2019

BY:
MADS CHRISTENSEN

SUPERVISOR:
JAMAL JOKAR ARSANJANI

Stimulating sustainable development goals' implementation and conservation action

Predicting future land use and land cover change in the Virunga National Park

By: Mads Christensen

Abstract

The United Nations 2030 Agenda for Sustainable Development and the Sustainable Development Goals (SDG's) presents a roadmap and a concerted platform of action towards achieving sustainable and inclusive development, leaving no one behind, while preventing environmental degradation and loss of natural resources. However, population growth, increased urbanisation, deforestation and rapid economic development has decidedly modified the surface of the earth, resulting in dramatic land cover changes, which continue to cause significant degradation of environmental attributes and threaten planetary boundaries. In order to reshape policies and management frameworks, conforming to the objectives of the SDG's, it is paramount to understand the driving mechanisms of land use changes and determine future patterns of change.

The Virunga National Park is located in the surrounding area of the contentious North Kivu province in the north-eastern part of the Democratic Republic of the Congo and has been the scene of near-constant conflict, exploitation and extreme poverty. While contributing to the livelihoods of millions of people in one of the most densely populated regions in Africa, efforts to conserve this globally significant ecosystem and its catchment areas is threatened by uncontrolled agricultural expansion, natural resource extraction and deforestation. Thus, the Virunga National Park catchment has experienced significant land cover changes, which continues to undermine, not just the integrity of the national park, but the foundation of millions of livelihoods who depends on its ecosystem services.

This study aims to assess and quantify future land cover changes in the Virunga catchment by simulating a future landscape for the SDG target year of 2030, in order to provide evidence to support data-based decision-making processes conforming to the requirements of the SDG's. The study follows six sequential steps: (1) Creation of three land cover maps from 2010, 2015 and 2019 derived from satellite images; (2) Land change analysis by cross-tabulation of land cover maps; (3) Sub-model creation and identification of explanatory variables and dataset creation for each variable; (4) Calculation of transition potentials of major transitions within the case study area using machine learning algorithms; (5) Change quantification and prediction using Markov Chain analysis; (6) prediction of a 2030 land cover.

The model was successfully able to simulate future land cover and land use changes and dynamics and goes on to conclude that agricultural expansion and urban development is expected to significantly reduce Virunga's forest and open land areas in the next 11 years. Accessibility in terms of landscape topography and proximity to existing human activities are concluded to be primary drivers of forest cover change. Drawing on these conclusions, the discussion provides recommendations and reflections on how the predicted future land cover

changes can be used to support and underpin policy frameworks towards achieving the SDG's and the 2030 Agenda for Sustainable Development.

Keywords: Land cover modelling, Remote Sensing, Machine Learning, Sustainable Development Goals, Virunga National Park.

Stimulering og implementering af verdensmålene for bæredygtig udvikling

Forudsigtelse af den fremtidige arealanvendelse i Virunga National Parken

Af: Mads Christensen

Resumé

De Forenede Nationers 2030 dagsorden for bæredygtig udvikling og de 17 Verdensmål (SDG) fremlægger en klar køreplan og en samordnet handlingsplan for at opnå en mere bæredygtig og inkluderende udvikling, som samtidig forebygger miljøforringelse og tab af naturressourcer. Men befolkningstilvækst, øget urbanisering, skovrydning og hurtig økonomisk udvikling har ændret jordens overflade, hvilket har resulteret i dramatiske ændringer i arealanvendelse som fortsat medfører en betydelig forringelse af miljøet og dets naturressourcer og således truer planetens økologiske balance. For at omforme politikker og ledelsesrammer i overensstemmelse med SDG'erne er det afgørende at forstå drivmekanismerne bag de processer som forårsager negative ændringer i arealanvendelsen.

Virunga National Parken er beliggende i et omstridt område i den nordlige Kivu-provins i den nordøstlige del af Den Demokratiske Republik Congo, og den har været genstand for næsten konstant konflikt, udnyttelse og ekstrem fattigdom. På trods af at området bidrager til millioner af menneskers levebrød i en af de mest tætbefolkede regioner i Afrika, er bestræbelserne på at bevare dette globalt vigtige økosystem og dets nærområde truet af ukontrolleret landbrugsudvidelse, naturressourceudvinding og skovrydning. Således har Virunga National Parken oplevet betydelige ændringer i arealanvendelse og udnyttelse af naturressourcer, som fortsat underminerer, ikke kun nationalparkens integritet, men levegrundlaget for de millioner af mennesker som er afhængige af dets økosystemtjenester.

Dette studie forsøger at kvantificere omfanget af fremtidige ændringer i arealanvendelse i landskabet omkring Virunga ved at konstruere en model som kan simulere et fremtidigt landskab for år 2030. Eftersom 2030 også udgør målet for implementeringen af SDG'erne forsøger studiet samtidig at understøtte databaserede beslutningsprocesser i overensstemmelse med SDG'ernes målsætning. Studiet følger seks sekventielle komponenter: (1) Skabelse af tre landdækkekort fra 2010, 2015 og 2019 afledt af satellitbilleder; (2) Landændringsanalyse ved tværgående tabulering af landdækkekort; (3) Identificering og oprettelse af undermodeller og forklarende variabler og oprettelse af datasæt for hver variabel; (4) Identificering af de drivende transitioner i arealanvendelse indenfor studieområdet ved brug af maskinindlæringsalgoritmer; (5) Kvantificering og forudsigtelse ved brug af Markov Chain analyse; (6) Simulering af landdækket i 2030.

Modellen opnåede med succes at simulere fremtidige ændringer i arealanvendelse og konkluderer at landbrugsudvidelse og byudvikling forventes at reducere Virungas skovområder og åbne/græsarealer betydeligt i de næste 11 år. Tilgængelighed med hensyn til landskabstopografi og nærhed til eksisterende menneskelige aktiviteter konkluderes at være de primære drivkræfter bag ændringer i skovdækket. På

baggrund af disse konklusioner giver diskussionen anbefalinger og overvejelser om hvordan de simulerede fremtidige ændringer i arealanvendelse kan bruges til at understøtte udviklingen af de politiske rammer for at opnå SDG'erne og 2030-dagsordenen for bæredygtig udvikling.

Nøgleord: Landdækkemodellering, Jordobservationer, Maskinlæring, Verdensmålene, Virunga National Parken.

Table of content

1	Introduction	1
1.1	Problem statement and research questions	2
1.2	Study area	3
1.3	Background.....	5
1.3.1	MLP neural network.....	6
1.3.2	Markov Chains	7
1.4	Land cover classification	8
1.4.1	Random Forest.....	9
1.5	Toolbox and technology	10
1.5.1	Terrset.....	10
1.5.2	ArcGIS.....	12
1.5.3	Google Earth Engine	12
1.5.4	JavaScript	13
2	Methodology.....	14
2.1	Land cover classification	15
2.1.1	Satellite imagery	15
2.1.2	Collecting training/validation data	16
2.1.3	Land cover classification within the Google Earth Engine IDE.....	19
2.2	LULC modelling and prediction.....	27
2.2.1	Land change analysis.....	27
2.2.2	Modelling Transition Potentials – sub-models	29
2.2.3	Modelling Transition Potential – MLP calibration.....	40
2.2.4	Change prediction and model validation	46
3	Results	49
4	Discussion and conclusion.....	53
4.1	Policy response options, planning interventions and SDG implementation.....	55
4.2	Reproducibility of the study	58
4.3	Sensitivity analysis	59
4.3.1	Human factors and temporal variations.....	59
4.3.2	The Modifiable Area Unit Problem (MAUP).....	60
4.4	Conclusion.....	61
5	Bibliography	62
6	Appendices	66

List of acronyms & abbreviations

Application Programming Interface	API
Area Of Interest	AOI
Artificial Neural Networks	ANN
Cascading Style Sheets	CSS
Decision Trees	DT
Digital Elevation Model	DEM
Distance from artisanal mines	D _{am}
Distance from cities	D _{cities}
Distance from disturbance	D _{disturb}
Distance from forests	D _{forest}
Distance from mining concessions	D _{mining}
Distance from roads	D _{roads}
Distance from waterways	D _{water}
Earth Engine	EE
Earth Engine	EE
Enhanced Thematic Mappers	ETM+
Environmental Systems Research Institute	ESRI
Evidence Likelihood	EL
Geographic Information System	GIS
HyperText Markup Language	HTML
Interactive Development Environment	IDE
Interactive Development Environment	IDE
Kappa Index of Agreement	KIA
K-nearest neighbour	k-NN
Land Change Modeler	LCM
Land use/land cover	LULC
Modifiable Area Unit Problem	MAUP
Multilateral Environmental Agreement	MEA
Multi-Layer Perceptron	MLP
National Aeronautics and Space Administration	NASA
National Park	NP
Normalized Difference Vegetation Index	NDVI

Observation Land Images	OLI
Short-Wave InfraRed	SWIR
Support Vector Machines	SVM
Sustainable Development Goals	SDG
UN Educational, Scientific and Cultural Organization	UNESCO
United Nations	UN
Visible and Near-InfraRed	VNIR
World Wildlife Fund	WWF

List of figures

Figure 1 Study area around the Virunga NP in the Democratic Republic of the Congo (The data on the boundary lines of the Virunga NP has been downloaded from (UNEP-WCMC & IUCN, 2019))	4
Figure 2 Structure of an MLP neural network (from (Beysolow II, 2017)).....	7
Figure 3 Random Forest flowchart (adapted from Harris and Grunsky (2015))	10
Figure 4 Land Change Modeler working environment.....	11
Figure 5 Components of the Earth Engine Code Editor (Source:(GOOGLE, 2019))	13
Figure 6 LCM workflow to predict land cover change in Virunga in 2030.....	14
Figure 7 JavaScript code to add geometry collections of training data and import a cloud-free Landsat image composite.	20
Figure 8 JavaScript code used to subsample and randomise the training datasets	21
Figure 9 JavaScript code for the Random Forest classification of the Landsat composite using the subsampled training data	21
Figure 10 JavaScript code used to derive error matrixes, used for land cover validation.....	22
Figure 11 Land cover area per class in 2010, 2015 and 2019	23
Figure 12 Land cover map - 2010	24
Figure 13 Land cover map - 2015	25
Figure 14 Land cover map - 2019	26
Figure 15 Class transitions between 2010 and 2015.....	29
Figure 16 Histogram depicting the correlation between the distance (in meters) from disturbed areas in 2010 and the actual disturbance between 2010 and 2015.....	33
Figure 17 JavaScript code to acquire DEM data from Google Earth Engine.....	35
Figure 18 JavaScript code to process DEM data to acquire datasets for slopes and aspect	36
Figure 19 User interface of the Variable Transformation Utility in TERRSET LCM	36
Figure 20 Processed explanatory variable datasets used as input for the MLP modelling.....	39
Figure 21 Extract from the calibration report indicating accuracy scores and skill measure of the model when holding variables constant.....	42
Figure 22 Sub-model accuracy and skill measure from MLP	42
Figure 23 Transition potentials	45
Figure 24 Markov Chain transition probability matrix.....	46
Figure 25 Actual land cover map for 2019 versus the predicted 2019 land cover map	47
Figure 26 Predicted land cover maps from 2020 to 2030	50
Figure 27 Predicted 2030 land cover in Virunga	51
Figure 28 Predicted land cover change between 2020 and 2030, in % yearly (gain/loss) and total annual area coverage in km ² per class.....	52
Figure 29 Spatial location of forest loss/gain from 2019 to 2030	55

List of tables

<i>Table 1 Transition potential matrix example (Mishra et al., 2014)</i>	8
<i>Table 2 Key characteristics of Landsat 7-8</i>	16
<i>Table 3 Training data collection for each of the five land classes and extracts of final classification based on training data</i>	18
<i>Table 4 Accuracy scores for the 2010, 2015 and 2019 land cover maps</i>	23
<i>Table 5 LULC change matrix for the period from 2010 to 2015 (km2) Class</i>	28
<i>Table 6 Transition sub-models and descriptions</i>	30
<i>Table 7 Description of potential explanatory variables and associated Cramer's V scores</i>	34
<i>Table 8 Sub-models included in MLP with associated explanatory variables and selected performance indicators</i>	44
<i>Table 9 K scores for 2019</i>	48
<i>Table 10 Cross tabulation between actual 2019 land cover and the simulated land cover for 2030</i>	54
<i>Table 11 Confusion matrix and accuracy - 2010 land cover map</i>	69
<i>Table 12 Confusion matrix and accuracy - 2015 land cover map</i>	69
<i>Table 13 Confusion matrix and accuracy - 2019 land cover map</i>	69

1 Introduction

Established in 1925 as the first National Park (NP) in Africa, the Virunga NP is located in the Albertine Rift Valley in the eastern part of the Democratic Republic of the Congo (Andersen, 2018). Along with the Mgahinga Gorilla NP in Uganda and the Parc Nationale Des Volcans in Rwanda, Virunga is part of a triangle of NP's in central Africa, principally designated in order to enhance conservation efforts to protect the critically endangered mountain gorilla (*Gorilla Beringei Beringei*) (Kayijamahe, 2008). The park covers an area of 790,000 ha (UNESCO, 2019), and besides hosting majority fragments of the last remaining habitat suitable for the mountain gorilla, the multitude of variety in nature and climate variables, with large lakes, open land savannah, vast forest areas, snow-covered mountain tops and erupting volcanoes also provide critical habitats for a great variety of the other large species of mammals we associate with Africa (Andersen, 2018). For this reason, the park was inscribed as a United Nations (UN) Educational, Scientific and Cultural Organization (UNESCO) World Heritage site in 1979. However, the NP is located in one of the most densely populated regions in Sub-Saharan Africa, which has been the scene of prolonged political turmoil and social conflict (Rainer et al., 2001), causing severe pressure on the ecological integrity of the landscape and its biodiversity. Moreover, the rich volcanic soil and high rainfall within the Virunga NP catchment makes it highly suitable for agriculture, and thus an attractive opportunity to underpin subsistence and commercial farming operations (Kayijamahe, 2008).

The rapidly increasing population has significantly increased the demand for natural resources (land, water energy, food, etc.), causing rapid land clearing for agriculture and grazing, removal of plants for different purposes, including artisanal mining operations, and house building (Rainer et al., 2001). Besides the efforts of authorities to protect the integrity of the NP, and avoid land intrusion and habitat degradation within the park, it continues to be threatened by civil unrest, illegal activities, land conversion and encroachment, livestock farming / grazing of domesticated animals, widespread depletion of forests in the lowlands and a massive influx of 1 million refugees occupying adjacent areas of the park (UNESCO, 2018). Militia leaders and prospectors are threatening the borders of the park in search for the vast deposits of diamonds, gold, uranium and other coveted minerals, while the vast influx of destitute refugees resorts to poaching and charcoal production, resulting in further fragmentation and degradation of the forest landscape (Andersen, 2018). In fact, the majority of the total population of nearly 6 million people in the surrounding province of North Kivu rely entirely on charcoal for their cooking needs, and an estimated three-quarter of this charcoal is sourced from the Virunga catchment, most of it illicitly from within the NP (Yee, 2017).

Thus, the region is highly important, both ecologically and economically, and the conflicting demands for socio-economic development while maintaining the ecological integrity of the NP has underpinned the need to ensure continued conservation efforts and sustainable natural resource management in order to safeguard

critical biodiversity and habitat, while preserving the foundation of the livelihoods for millions of people. This agenda is fortified through the UN 2030 Agenda for Sustainable Development and the 17 Sustainable Development Goals (SDG)'s which were adopted at the UN general assembly in 2015. The SDG's calls on concerted action to pursue economic development while ensuring social inclusion and environmental sustainability, on the basis of good governance. The SDG framework provides a comprehensive agenda through which to mainstream policies and derive targeted actions for addressing core sustainability challenges. However, the ability to target policies and actions to address conservation issues, while pursuing economic development and prosperity, leaving no one behind, is hampered by lacking scientific evidence and data to direct and support informed decision making.

In order to derive targeted policies and actions to support effective land use planning, management and ecological restoration conforming to the requirements of the SDG's, it is imperative to understand the underlying processes of change (Liping et al., 2018). Up to date information on current land cover and land use provides critical information which can be used to underpin decision-making processes, while modelled predictions about plausible future land use/land cover (LULC) scenarios provide indications of potential trajectories and thus a platform for identifying interventions. Changes in land use and land cover can be described and projected through the use of land change models, which can be used to explain and assess the dynamics of land cover- and broader system change (National Research Council, 2014). Spatial land change models thus provide platforms for exploring potential future scenarios, which can be used to guide land use decision making and planning (National Research Council, 2014).

The purpose of the study is to assess and quantify past and plausible future land use and land cover changes and dynamics within the Virunga case study area. The primary analysis will be guided by a change analysis of classified satellite imagery to quantify past changes, and the development of a land change model, applying a coupled machine learning - Markov Chain approach, to derive a future land cover prediction for the year 2030. The aim is to assess the plausible future evolution of the landscape within the Virunga case study area and address an existing data gap in order to provide evidence to support data-based decision-making processes conforming to the requirements of the SDG's.

1.1 Problem statement and research questions

While several authors have already successfully applied predictive land change modelling to support land use management and decision-making processes (i.e. Gibson et al., 2018; Guerrero et al., n.d.; Shade & Kremer, 2019), a thorough literature review indicates that such an approach has been applied in just a few case studies and hence it is necessary to explore further cases in order to assure its applicability across different landscapes. Therefore, a remote area in Africa within the Virunga NP is targeted. This study aims to apply remotely sensed data, geospatial and modelling tools to detect, quantify, analyse, and predict future land change in the Virunga NP and its immediate vicinity.

Main hypothesis and research questions:

The study is framed around the hypothesis that there have been significant land cover changes within the study area, primarily caused by deforestation due to encroaching activities and cropland expansion. The study will test two main assumptions:

1. It is possible to remotely monitor and model a case study in Africa using a combination of remotely sensed data, Geographic Information System (GIS) tools and modelling techniques for studying the dynamics of the land cover within the study area.
2. There has been, and if unchecked and unregulated, will likely continue to be significant land use/land cover changes within the study area.

To assist the implementation of the main research framework and to guide the analysis, the following research questions were posed:

- Have there been major land cover changes within the study area in the last 10 years? And if so, what kind of land cover changes?
- What has the spatial extent of the land cover change been and which areas have experienced the highest rate of changes?
- What are the major driving forces behind these changes?
- What will the extent of land change be by 2030?
- How can the future land cover prediction for the Virunga study area be used to support and underpin policy frameworks towards achieving the SDG's?

1.2 Study area

The Virunga NP is located in Central Africa, in the Eastern part of the Democratic Republic of the Congo, on the border with Uganda and Rwanda. It is located in the equatorial zone, within the Albertine Rift, of the Great African Rift Valley (UNESCO, 2019). In this study, The Virunga NP and its immediate vicinity was included in order to fully assess of the NP the landscape dynamics of the entire Virunga catchment. This was considered critical in order to explore socio-economic changes, primarily in the form of urban development and cropland expansion, outside of the NP, and assess how these land cover dynamics could potentially impede conservation efforts and sustainable land management planning.

The study area as can be seen from Figure 1 below, covers a total of 14810 km² of which 7779 km² is within the Virunga NP.

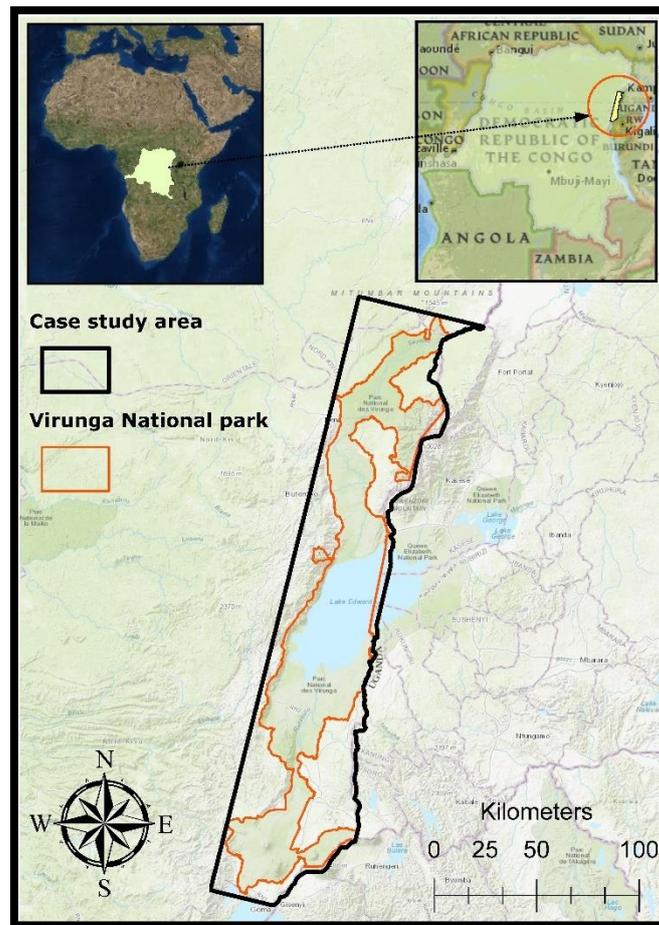


Figure 1 Study area around the Virunga NP in the Democratic Republic of the Congo (The data on the boundary lines of the Virunga NP has been downloaded from (UNEP-WCMC & IUCN, 2019))

As briefly outlined in the introduction, the area is characterized by an astonishing diversity of landscapes and biotopes and the varying topography lends itself to host more unique habitats than any other NP in Africa, ranging from swamps and steppes to the snowfields of Mount Stanley at an altitude of 5,109 m, and from the lava plains to savannah and the steppes of the low land plains at the feet of the many volcanoes (UNESCO, 2019).

The Northern part of the Virunga NP is characterized by high mountains, containing the third, fourth and fifth highest peaks on the continent of Africa (Crawford & Bernstein, 2008). The mountain massif is mainly covered by montane forests, however, cropland intrusion, particularly in the Western flank is also dominant as agriculture is the mainstay of the livelihoods in the region (Crawford et al., 2008). The central part of the park is dominated by Lake Edward which borders Uganda to the East. Smaller cities and villages are scattered along the coast of the lake, while open land and cropland characterize the hinterland. Until recently the lake was considered Africa's most productive for fisheries and hosted the largest concentration of hippo in the world, however, widespread poaching and overfishing has changed this, threatening the ecological balance of the ecosystem and the livelihoods of the people depending on it (Crawford et al., 2008). The Southern part of the

park is characterised by a series of active and extinct volcanoes, including Nyamulagira and Nyiragongo in the southwest, which are two of the most active volcanoes in the world (Crawford et al., 2008). In 2002, Nyiragongo erupted, resulting in the displacement of thousands of people living in the nearby city of Goma (Crawford et al., 2008). The volcanic landscape in the southern sector consists mainly of dense, humid montane forest within which the mountain gorilla reside (Crawford et al., 2008). However, the majority of the people in the region, including the main city of Goma, rely exclusively on charcoal for their energy, and this has translated into intense pressure on the nearby forests. According to Crawford et al. (2008), 24,000 hectares of forest is needed to satisfy this demand, and much of it comes from the park, as the old growth trees in the montane forests produce charcoal that burns longer and hotter. According to the World Resources Institute (2019), the total forest cover within the NP has been reduced by approximately 374 km² in the period from 2001 to 2018.

1.3 Background

Understanding the drivers and dynamics of LULC change is imperative in order to develop sustainable management strategies and policies and make informed planning decisions. Changes in LULC affect a wide range of environmental parameters, including soil erosion and accretion, hydrological balance, biodiversity, climate, all of which are factors that ultimately impact and drive societal wellbeing and influence the sustainability of local livelihoods (Zadbagher & Becek, 2018). The land cover changes are driven by an assembly of difference anthropogenic and natural processes operating at different spatiotemporal scales, each of which are driven by one or more variables (Zadbagher et al., 2018). The variables also referred to as explanatory variables, are drivers of the observed changes and typically consist of a range of biophysical and socioeconomic criteria.

The ability to determine the extent to which the drivers contribute to future LULC changes is fundamental in order to make accurate predictions about future LULC scenarios, which is vital in order to underpin and inform management decisions and interventions. LULC change models aim to predict or simulate the future behaviour of environmental and social systems in order to support the analysis of the causes and consequences of land use dynamics (Mishra et al., 2014). While LULC models are a simplified representation of complex, dynamic and nonlinear socioeconomic and natural structures, they are useful for determining plausible ways of how the future could potentially unfold (Noszczyk, 2018).

LULC change models consist of various methods aiming to aid the understanding of the spatial relationship between the historical change of land cover and their drivers (Meiyappan et al., 2014). The selection of the right method is a reflection upon the goals and aims of the research questions and is a critical component of the model construction process (Noszczyk, 2018). No model is able to comprehensively model all aspects of reality (Noszczyk, 2018), and thus the selection of an appropriate method is subject to compromise, capability and resources available.

According to literature, the following LULC change model types are identified (Noszczyk, 2018)

- Agent-based models
- Economics-based models
- Cellular automata
- Artificial neural networks (ANN)
- Markov chains
- Models based on statistical analysis

In this study, a Multi-Layer Perceptron (MLP) neural network is trained to analyse the empirical relationship between historical change and the explanatory variables, or drivers of change, in order to determine the transition potential of each pixel to change into another land cover class (Mas et al., 2014). A Markov Chain is used to derive future scenario predictions, based on the amount of historic change and a projection of the transition potential into a future state.

1.3.1 MLP neural network

A neural network is a type of computational framework for a collection of interconnected units or nodes (also called neurons or perceptrons) which aims to mimic the human brain (Yang, 2010). An MLP neural network consists of multiple layers of nodes, interconnected to the next node to form a feed-forward neural network (Beysolow II, 2017). The stronghold of neural networks is their ability to relate the representation of a training dataset to that of an output variable in order to make a prediction (Brownlee, 2016). As an MLP is a feed-forward neural network, data flows in one direction, from a set of input layers, through one or more hidden layers which are sets of computational nodes, to a set of computation/output layers (Gibson et al., 2018). The nodes are linked by a web of connections which are applied as weights, and a back-propagation algorithm is used to train the network iteratively by spreading errors from the output layer to the input layer by adjusting the value of the weights in order to minimise the error between the observed and predicted outcomes (Gibson et al., 2018). The back-propagation algorithm which is used to train the model is the key distinguishing feature of an MLP, compared to a single layer perceptron (SLP) model (Beysolow II, 2017). This algorithm is enabled by introducing hidden neurons and it allows the learning algorithm to alter the composition of the network based on a trial and error framework, by separating error by each node in the network (Yang, 2010).

The standard multilayer perceptron (MLP) is a cascade of single-layer perceptrons. There is a layer of input nodes, a layer of output nodes, and one or more intermediate layers. The interior layers are sometimes called "hidden layers" because they are not directly observable from the system inputs and outputs (Reed, Reed, and Marks 2014, page 31).

An MLP's capability to learn depends on the network architecture (number of hidden layers and nodes) and on the parameterisation of the model (i.e. learning rate, momentum factor, sigmoid value and number of

iterations). The performance of an MLP model is assessed by a precision value expressed in per cent, and networks that are too small tend to be unable to identify the internal structure of the data, resulting in lower accuracies, while networks that are too large tend to overfit the data (Gibson et al., 2018). Overfitting can occur when the algorithm produces a mathematical relationship between the observed changes and a set of explanatory variables, which fits the details of the calibration dataset but fails to represent the more general principles of changes that extend to other times and places (National Research Council, 2014).

Figure 2 below, shows an example of an MLP trained with a back-propagation algorithm where hidden neurons are introduced between the input layer (x_1 , x_2 and x_3) and the output layer (o_1 , o_2 , and o_3).

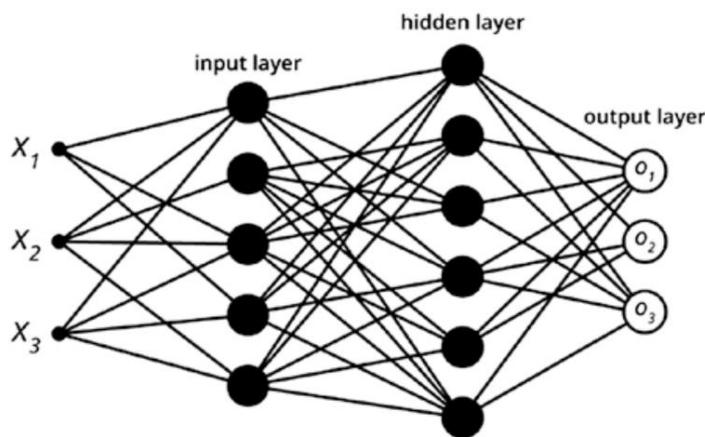


Figure 2 Structure of an MLP neural network (from (Beysolow II, 2017))

MLP's are suitable for classification prediction problems (Brownlee, 2016) and by using hidden neurons which affect the output of the model, they can be used for modelling complex nonlinear relationships allowing them to better handle Boolean XOR problems (Beysolow II, 2017).

1.3.2 Markov Chains

Named after Andrey Markov a Markovian process is “a stochastic process in which the conditional probability distribution of future states of the process, given the present state and all past states, depends only upon the present state” (Sammut & Webb, 2010). One of the most well-known Markovian processes is called Markov Chains, which are discrete time-series of different states with transition probabilities (Sammut et al., 2010). In a Markovian analysis of land class changes, a matrix is derived in order to represent changes between land cover categories (Noszczyk, 2018). Assuming that the pace of changes in time and the change itself is stationary, meaning that the rates of change observed during calibration (T1 to T2) will remain the same during simulation (T2 to T3), the matrix represents the likelihood of a land class to transform into another category, i.e. meaning that five land classes result in 25 possible changes (Noszczyk, 2018). This procedure determines the amount of land which is expected to transition from the later date to the prediction date, based on a

projection of the transition potentials into the future (Mishra et al., 2014). An example of a Markov Chain transition probability matrix is illustrated in Table 1 below.

Table 1 Transition potential matrix example (Mishra et al., 2014)

LU/LC Class	Agriculture land	Vegetation	Scrubs	Fallow land	Waste land	Built up area	Water bodies	River bed
Agriculture land	0.0000	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
Vegetation	0.3979	0.2812	0.0657	0.0992	0.0628	0.0669	0.0142	0.0122
Scrubs	0.3211	0.3091	0.0627	0.1027	0.0639	0.1005	0.0244	0.0156
Fallow land	0.3199	0.3032	0.0642	0.0932	0.0669	0.1073	0.0228	0.0226
Waste land	0.2475	0.2221	0.0542	0.0863	0.1066	0.1047	0.0950	0.0838
Built-up area	0.1612	0.1876	0.0849	0.1118	0.0552	0.3714	0.0151	0.0128
Water bodies	0.2690	0.2585	0.0697	0.1117	0.0797	0.1015	0.0738	0.0360
River bed	0.2469	0.2237	0.0559	0.0881	0.1055	0.1114	0.0896	0.0789

1.4 Land cover classification

The ability to provide a synoptic view over large areas and map land cover and land cover changes are one of the strongholds of satellite-based remote sensing (Rodriguez-Galiano et al., 2012).

Scientists and practitioners have made great advancements in improving existing and developing new advanced methods for multispectral image classification in order to improve accuracy and processing speed (Kulkarni & Lowe, 2016). There are many methods for land classification, spanning the range from unsupervised clustering algorithms to non-parametric machine learning algorithms. Prior knowledge about the area of interest is not needed when conducting an unsupervised classification, as these algorithms form clusters of pixels based on the statistical properties of each pixel. Supervised classifications, however, are dependent upon training data (ground-truth) which can be collected from existing maps, fieldwork observations or high-resolution satellite imagery (Al-Ahmadi & Hames, 2009). Supervised classifications determine the relationship between each pixel and the spectral signature defined in the training data set. In order to attain the best classification results, training data selection and segmentation need to be carefully considered (Phiri & Morgenroth, 2017). The selection of appropriate land cover classes (e.g. forest, cropland, urban areas) and careful consideration of stratification needs (i.e. subdividing urban into residential, industrial, fringe, etc.), reflecting the objective of the classification end use, mirrors the training data needs for the supervised classifier. Thus, the size and quality of the training data sample are key issues when planning for a supervised classification (Maxwell et al., 2018). For machine learning, the rule of thumb is that the minimum number of training samples should be 10 times the number of variables land classes (Maxwell et al., 2018). In order to assess the performance of the classifier, one option is to withhold a certain proportion of the training data, i.e. 30 %, and use this subset to test the classification accuracy (Maxwell et al., 2018). An alternative approach is to use the entire training dataset for training and derive a separate dataset to assess the performance of the classifier.

Machine learning has received much attention in recent years, largely due to its ability to improve efficiency and accuracy and handle data of high dimensionality and map classes with highly complex characteristics (Maxwell et al., 2018). Several machine learning methods are still relatively immature and experimental (e.g. extreme learning machines and deep convolution neural networks), however a range of methods have long been well-established and adopted by the scientific community as mature methods, including Support Vector Machines (SVM), single Decision Trees (DT), boosted DT's, Artificial Neural Networks (ANN), k-nearest neighbour (k-NN) and Random Forests (Maxwell et al., 2018).

Random Forests have received considerable attention in recent years, due to its robustness, high classification accuracy and easy parameterization, and it is well-established as one of the most efficient classification methods of satellite imagery (Gislason et al., 2006; Kulkarni et al., 2016; Maxwell et al., 2018; Ming et al., 2016; Pelletier et al., 2016).

1.4.1 Random Forest

Random Forests by Breiman (2001) is a supervised learning technique based on trees classifiers. It is an ensemble classifier, which means that it builds a “forest” of decision trees, each of which makes their own classification by using a random bootstrap sample selection (Ming et al., 2016). The idea behind combining multiple decision trees to produce ensembles is that a group of weak learners together form groups of strong learners, to increase predictive performance.

Bootstrapping is a randomization technique which helps to generate several subset datasets from a single set of data, by randomly choosing the same number of observations as the original data set, but with replacement (Suthaharan, 2016). This way, each subset dataset will have the same number of total observations as the original dataset, however, some of the observations are repeated due to multiple selections in the subsets.

Each tree in the ensemble is formed by the bootstrapped subset datasets, where approximately 2/3 of the data is used as training (in-the-bag), used for classification and the remaining 1/3 is used for validation (out-the-bag) (Kulkarni et al., 2016). The out-the-bag validation data is used to get a classification error score while trees are added to the ensemble. The training data is used to create multiple decision trees which ultimately makes independent classification votes. During the decision phase, each individual tree classifies the input data, and the final Random Forest output class label is based on the majority of votes within the ensemble (Suthaharan, 2016).

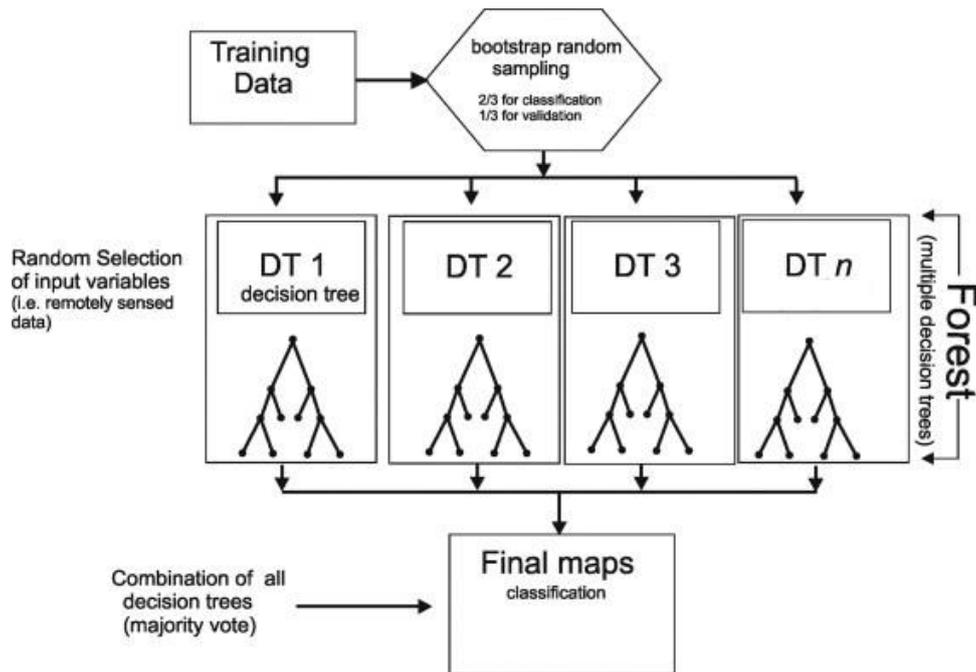


Figure 3 Random Forest flowchart (adapted from Harris and Grunsky (2015))

1.5 Toolbox and technology

A variety of different tools, software components and programming languages was utilized in order to frame the overall analysis and methodology. This section will provide a brief description of the software components, cloud platforms and programming languages used in the study.

1.5.1 Terrset

Developed in 2015, Terrset is an integrated successor of the IDRISI GIS and Image Processing software, first conceptualized by Clarks Labs in 1987 (<https://clarklabs.org/about/>). It constitutes a feature-rich software system which incorporates IDRISI GIS and Image Processing tools and offers a wide constellation of tools focused on monitoring and modelling of the Earth system for sustainable development (Eastman, 2016a).



The full constellation of the programme includes modules and processing tools for:

- GIS analysis
- Image processing
- Land Change modelling
- Habitat and biodiversity modelling
- Reducing Emissions from Deforestation and forest Degradation (REDD) planning
- Ecosystem services modelling
- Time series analysis
- Climate change adaptation modelling

Terrset supports both raster and vector-based analysis and processing, however, the mainstay of the modules focuses on raster-based analysis. This is largely due to Terrsets main analytical focus on remotely sensed data, which native structure is raster based.

Version 18.31 of the Terrset software package was utilised in this study.

1.5.1.1 Land Change Modeler

The primary tool applied in this study to analyze land-cover change and dynamics is the Land Change Modeler (LCM) module within the Terrset toolbox. By cross-tabulating change between two separate land cover maps, representing different time steps, LCM can model the empirical relationship between the land cover changes and a set of explanatory variables, to make a prediction of future land cover scenarios.

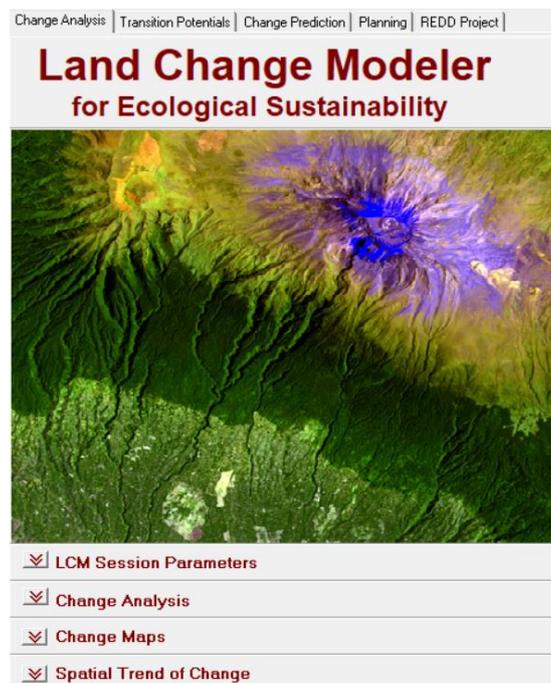


Figure 4 Land Change Modeler working environment

1.5.2 ArcGIS

The main software used in this project for data pre-processing and visualization is ArcGIS for Desktop, version 10.6.1. ArcGIS is a GIS system developed by the Environmental Systems Research Institute (Esri) and includes ArcMap which provides several powerful tools for displaying, analyzing, and creating GIS data (ESRI, 2019).

Furthermore, ArcGIS includes ArcCatalog which is particularly useful for browsing GIS data, viewing and managing metadata, creating and managing geodatabases, etc.

1.5.3 Google Earth Engine

Google Earth Engine, established in 2010, is a web-based cloud processing platform and satellite data repository that provides global-time series satellite imagery and vector data and access to software and algorithm for data processing (Kumar & Mutanga, 2018). The multi-petabyte analysis-ready data catalogue, including satellite imagery stored in the public data archive, includes historical earth images dating back more than forty years (Gorelick et al., 2017). Besides the collection of raw unprocessed satellite imagery, Google Earth Engine also provides access to various satellite-based products, including indices, composites, elevation models, land cover data, etc.

Data from the Earth Engine servers can be accessed using the JavaScript-based Google Earth Engine Internet-accessible application programming interface (API). The Earth Engine (EE) Code Editor, available from code.earthengine.google.com, is a web-based interactive development environment (IDE) for the Earth Engine JavaScript API, which allows users to create and run custom algorithms to retrieve and process data rapidly in the cloud.

As illustrated in Figure 5 below, the IDE includes the following components (GOOGLE, 2019);

- JavaScript code editor;
- Map display for visualizing geospatial datasets;
- API reference documentation (Docs tab);
- Git-based script manager (Scripts tab);
- Console output (Console tab);
- Task manager (Tasks tab) to handle long-running queries;
- Interactive map query (Inspector tab);
- Search of the data archive or saved scripts;
- Geometry drawing tools

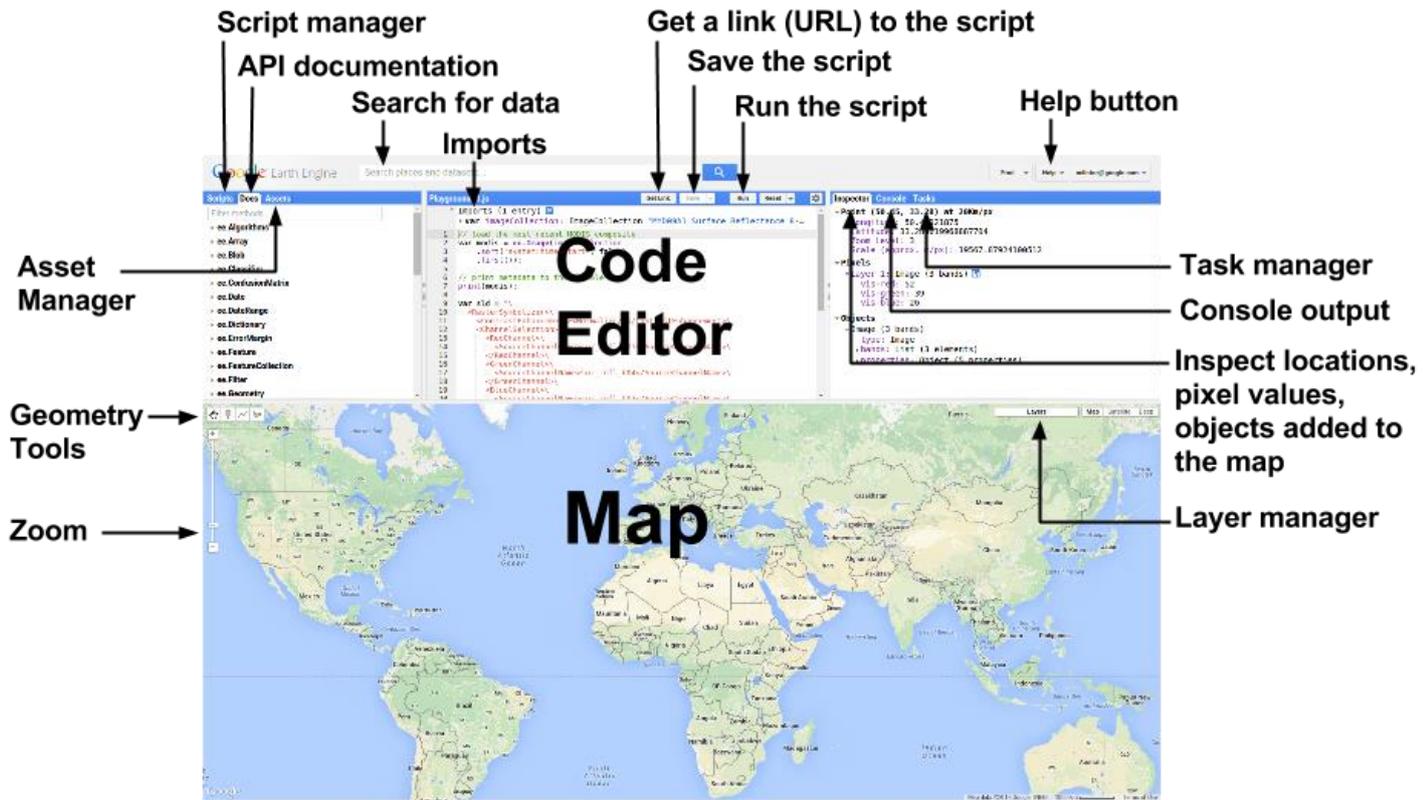


Figure 5 Components of the Earth Engine Code Editor (Source:(GOOGLE, 2019))

1.5.4 JavaScript

JavaScript is a lightweight, interpreted, object-oriented programming language, best known as one of three main pillars in web development along with HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) (MDN, 2019). It is a text-based and client-side programming language which is primarily used to make a webpage more interactive and responsive to the occurrence of a particular event (MDN, 2019).

In Google Earth Engine JavaScript commands can be used in the IDE to acquire, process and analyse geospatial data inputs.

2 Methodology

The methodological framework utilized in this study to predict the future landscape around the Virunga NP was developed using a variety of different tools and the theoretical framework outlined in section 1.3. The workflow is illustrated in Figure 6 below and the methodology follows six sequential components;

1. Creation of three land cover maps from 2010, 2015 and 2019 derived from satellite images;
2. Land change analysis by cross-tabulation of land cover maps;
3. Sub-model creation and identification of explanatory variables and dataset creation for each variable;
4. Calculation of transition potentials of major transitions within the case study area using an MLP neural network;
5. Change quantification and prediction using Markov Chain analysis, and accuracy assessment of the model performance by cross-comparing the predicted land cover map for 2019 with the actual 2019 land cover map;
6. Prediction of a 2030 land cover.

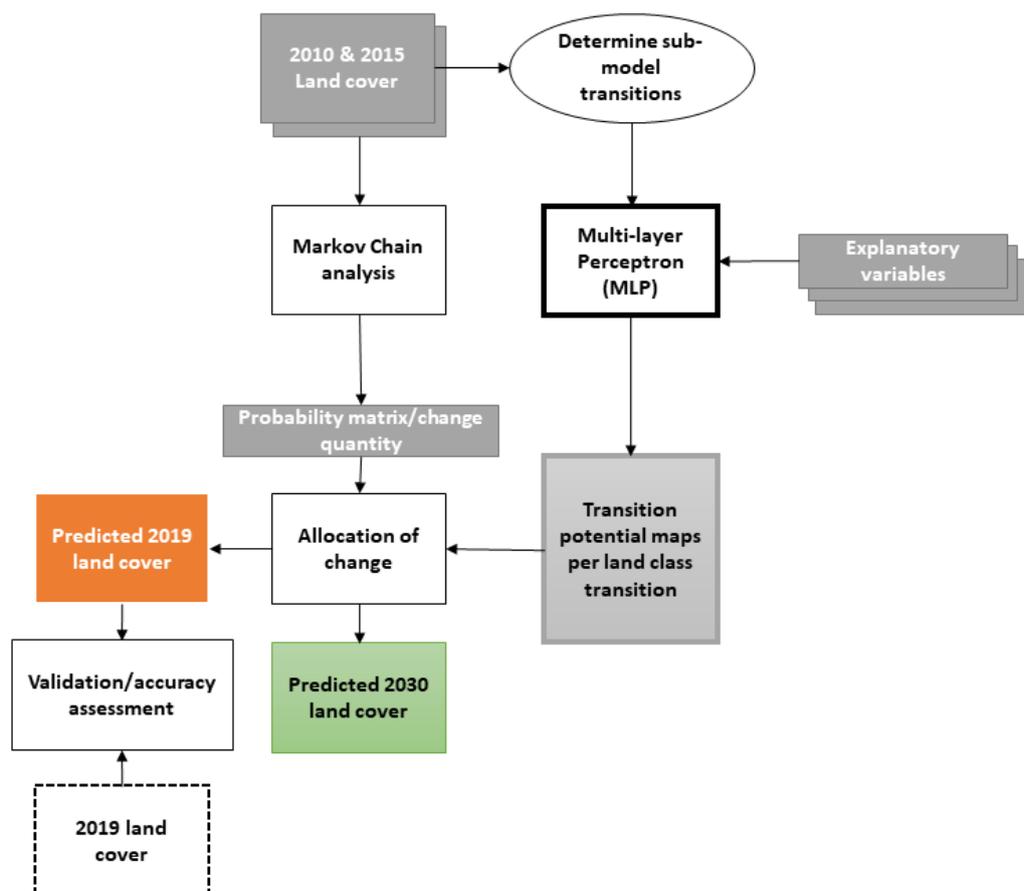


Figure 6 LCM workflow to predict land cover change in Virunga in 2030

In this section, the methodology applied in this study to derive land cover predictions for the year 2030, conforming to this sequential stepwise approach is described. All datasets were either created in, or reprojected to, a Réseau Geodesique de la RDC 2005 TM Zone 18 (EPSG:4051) projected coordinate system, suitable for use in the Democratic Republic of the Congo.

2.1 Land cover classification

Google Earth Engine provides a cloud-based platform for accessing and processing large amounts of both current and historical satellite imagery, including those acquired by the Landsat-7 and Landsat-8 satellites. The advantages of seamless integration of archived, and pre-processed satellite imagery, along with a powerful cloud processing platform made Google Earth Engine an ideal platform for conducting the land cover classification.

The land classification in Google Earth Engine is composed of several different steps;

- Choosing an appropriate satellite imagery dataset, fitting the objective of the study.
- Define land cover classes and collect training data to train the supervised classification algorithm.
- Developing a JavaScript code to acquire, process and classify the satellite imagery based on the choice of classification algorithm.

2.1.1 Satellite imagery

In this study, three land cover maps were needed, one for 2010, 2015 and 2019. As the National Aeronautics and Space Administration (NASA)'s Landsat satellites provides an archived and freely available dataset covering the entire study period with high resolution (30 m) multispectral imagery, these were selected for this study. Google Earth Engine provides integrated access to analysis-ready (already geometrically corrected and orthorectified), surface reflectance Landsat data from the Tier-1 collection.¹

For the 2010 land cover map, tier-1 data from the Landsat 7 Enhanced Thematic Mappers (ETM+) sensor was selected, while tier-1 data from the Landsat 8 Observation Land Images (OLI) was chosen for the 2015 and 2019 land cover maps. The Landsat 7 sensor has been in operation since 1999 and as seen from Table 2 below, the images contain 4 visible and near-infrared bands (VNIR), 2 short-wave infrared (SWIR) bands, 1 thermal infrared (TIR) band and a panchromatic band. The Landsat 8 sensor has been operative since 2013 and contains 5 VNIR bands, 2 SWIR bands, 2 TIR bands, a panchromatic band and a cirrus band.

¹ For further information on Landsat Collection 1 products: https://www.usgs.gov/land-resources/nli/landsat/landsat-collection-1?qt-science_support_page_related_con=1#qt-science_support_page_related_con

Table 2 Key characteristics of Landsat 7-8

Sensor	Spectral bands	Wavelength (μm)	Ground pixel size (m)
Landsat 7 - ETM+	Band 1 - Blue	0.45-0.52	30
	Band 2 - Green	0.52-0.60	30
	Band 3 - Red	0.63-0.69	30
	Band 4 - Near Infrared (NIR)	0.77-0.90	30
	Band 5 - Shortwave Infrared (SWIR)	1.55-1.75	30
	Band 6 - Thermal	10.40-12.50	60
	Band 7 - Shortwave Infrared (SWIR)	2.09-2.35	30
	Band 8 - Panchromatic	.52-.90	15
Landsat 8 - OLI	Band 1 - Ultra Blue	0.435 - 0.451	30
	Band 2 - Blue	0.452 - 0.512	30
	Band 3 - Green	0.533 - 0.590	30
	Band 4 - Red	0.636 - 0.673	30
	Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
	Band 6 - Shortwave Infrared (SWIR)	1.566 - 1.651	30
	Band 7 - Shortwave Infrared (SWIR)	2.107 - 2.294	30
	Band 8 - Panchromatic	0.503 - 0.676	15
	Band 9 - Cirrus	1.363 - 1.384	30
	Band 10 - Thermal Infrared (TIR) 1	10.60 - 11.19	100
	Band 11 - Thermal Infrared (TIR) 2	11.50 - 12.51	100

2.1.2 Collecting training/validation data

As a first step in preparing a training dataset for the land classification, the definition of a nomenclature of land cover classes fitting the objective of the study needed to be defined. For this study, the primary objective was to predict changes to the forest cover, and thus an elaborate definition of several land cover classes was not needed. Accordingly, the 5 mainland cover classes in the area of interest were enough to ensure a sufficient representation of the spatiotemporal variety of land cover changes and identify the primary drivers contributing to forest change dynamics.

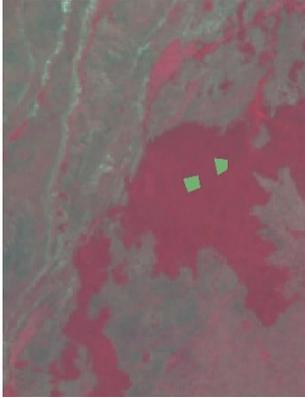
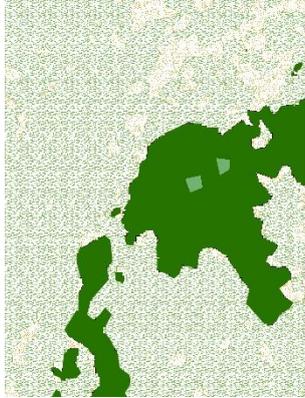
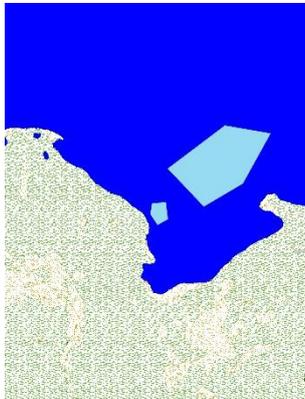
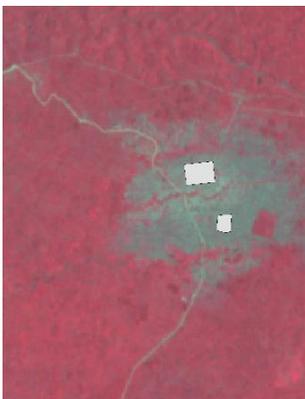
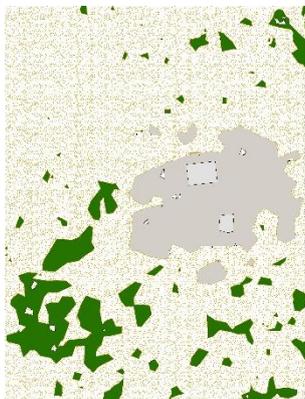
The five land cover classes were identified as;

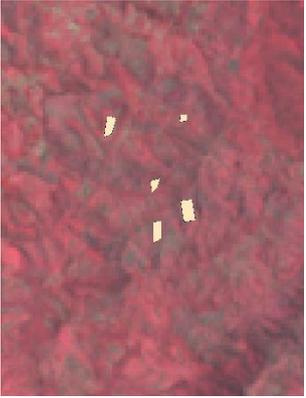
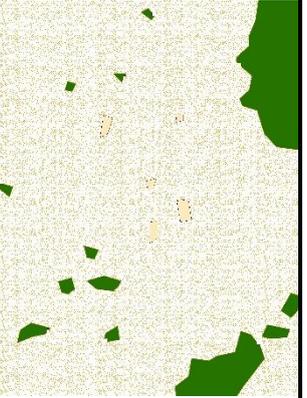
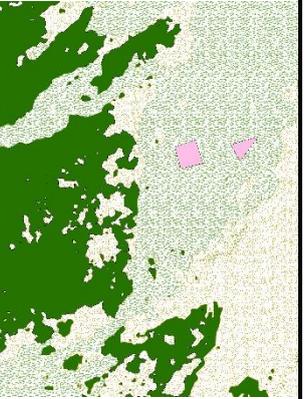
1. **Forest** – afforested and primary forest areas.
2. **Water** – lakes and rivers.
3. **Urban areas** – developed residential or industrial areas, roads and urban fringes.
4. **Cropland** – planted or bare crop fields.
5. **Open land/grassland** – areas with sparse vegetation, characterized by open grasslands, bare soil or volcanic ash.

To train and validate the land-cover classifications, a reference training dataset was collected within the study area. As reviewed in section 1.4, the minimum samples for machine learning based algorithms to perform optimally should be at least 10 times the number of land cover classes. Thus, in this case, the training data samples should be at least $10 \times 5 = 50$. These reference training datasets were collected by drawing polygons and clicking points within the Google Earth Engine map interface, on top of individual pixels or areas identified as belonging to one of the 5 classes by visual inspection. The reference maps for the sampling of the training datasets included time series of Landsat 7 images (2012 land cover map) and time series of Landsat 8 images (2015 and 2019 land cover maps) as well as Google Earth high-resolution images. The Landsat images were added to the map layer user interface as both a true colour (RGB) and false colour composites. The true colour composite represents the surface as the eye would see it, while the false colour composite enhances the spectral difference of vegetation patterns by replacing the red band with a near infrared band which has high reflectance values for vegetation. From the collection of training data polygons and points, a subsample of 500 points was used to train the model. An additional point dataset consisting of 50 individually sampled points was collected and used for validation. The validation dataset (i.e. ground truth) was sampled using the high-resolution images in Google Earth.

The two Landsat composites were used as the main reference for the sampling of the training data by visually inspecting each area, however, each sample unit was subsequently cross-referenced with the high-resolution dataset in order to substantiate whether the sampling unit feasibly belongs to that class. The cross-referencing for the 2019 training data was easily facilitated through the integration of current high-resolution satellite imagery from 2019, within the map interface of Google Earth Engine, allowing for seamless integration with the two composite images. However, as Google Earth engine does not integrate historic high-resolution imagery, cross-referencing for the 2010 and 2015 training datasets was resolved by comparing sample units to the high-resolution images archived in Google Earth Pro, for each of the two years. Table 3 provides an illustration of the training data collection method, as well as an extract of the final classification based on the training data sampling within that area.

Table 3 Training data collection for each of the five land classes and extracts of final classification based on training data

	True colour image	False colour image	High-resolution image	Classification
1: Forest				
2: Water				
3: Urban				

	True colour image	False colour image	High-resolution image	Classification
4: Crop and land				
5: Open land				

2.1.3 Land cover classification within the Google Earth Engine IDE

In order to create the three land cover maps, three individual scripts were prepared within the Google Earth Engine IDE, one for each of the three years. The JavaScript source code for the 2019 land cover classification is included in Appendix A.

The first component of the script was to import the area of interest (AOI) as table data, which is the Earth Engine equivalent to a dataset in a Shapefile format. Secondly, 5 empty containers for geometry collections for the training datasets were imported as variables. Subsequently, a cloud-free composite of satellite images was imported using the JavaScript code illustrated in Figure 7 below. The function *'maskClouds'* generates a cloud and a cloud shadow mask for the imported Landsat collection. Furthermore, within this function, the Normalized Difference Vegetation Index (NDVI) was calculated, using the formula $NDVI = \frac{(NIR-Red)}{NIR+Red}$, and added to the band collection of the satellite image composite. The NDVI was added to the band collection to enhance the contribution of vegetation in the spectral response for the classification. The variable *'L8collection'* imports the Landsat 8 (Landsat 7 for the 2010 land cover map) satellite image collection and filter the collection on the basis of image date, boundaries (within AOI) and cloud cover (< 35 %) and ultimately applies the cloud masking function. As the area of interest around the Virunga mountains generally

have a high percentage cloud cover, each of the three land cover images were filtered on the basis of satellite images acquired in the previous 2 years. The variable ‘testimage’ creates a cloud free composite from the satellite image collection, using a median reducer to derive the median values of each band in the stack, over time, and clips the composite to the AOI. The composited and clipped Landsat collection is ultimately added as map layers, as a false colour (line 21) and true colour (line 22) composite. The two composites, along with a high-resolution reference satellite image were subsequently used to create geometries for the training datasets (as detailed in section **Fejl! Henvisningskilde ikke fundet.**).

```

Imports (6 entries)
▶ var forest: Feature 0 (GeometryCollection, 1 property)
▶ var water: Feature 0 (GeometryCollection, 1 property)
▶ var city: Feature 0 (GeometryCollection, 1 property)
▶ var cropland: Feature 0 (GeometryCollection, 1 property)
▶ var OpenLand: Feature 0 (GeometryCollection, 1 property)
▶ var AOI: Table users/icemads/AOI

1 //For Landsat surface reflectance product cloud masking
2 function maskclouds(image) {
3   var cloudShadowBitMask = 1 << 3; // cloud shadow
4   var cloudsBitMask = 1 << 5; // cloud
5   var qa = image.select('pixel_qa');
6   var date = image.get('system:time_start');
7   var mask = qa.bitwiseAnd(cloudShadowBitMask).eq(0)
8     .and(qa.bitwiseAnd(cloudsBitMask).eq(0));
9   var ndvi = image.normalizedDifference(['B5', 'B4']).multiply(10000).rename('NDVI');
10  return image.addBands(ndvi).updateMask(mask).divide(10000).set('system:time_start',date);
11 }
12 //Landsat 8 image collection
13 var L8collection = ee.ImageCollection('LANDSAT/LC08/C01/T1_SR')
14   .filterDate('2017-01-01', '2019-03-23')
15   .filterBounds(AOI)
16   .filter(ee.Filter.lt('CLOUD_COVER', 35))
17   .map(maskclouds);
18 print (L8collection);
19 var testimage = L8collection.median().clip(AOI);
20
21 Map.addLayer(testimage.select(['B5', 'B4', 'B3']), {min:0, max:0.4}, 'false color');
22 Map.addLayer(testimage.select(['B4', 'B3', 'B2']), {min:0, max:0.4}, 'true color color');

```

Figure 7 JavaScript code to add geometry collections of training data and import a cloud-free Landsat image composite.

The training datasets were subsequently subsampled with random points to ensure that all classes within the training data have the same sample size (Figure 8). In this script, 500 sample points for each training layer is generated by looping over each training dataset and creating random points within the geometries of the training data layers. The classes for each point feature collection is added using the class properties of the training data geometries (Forest = 1, Water = 2, City = 3, Cropland = 4 and OpenLand = 5). The variable ‘training’ collects the points within each training class and samples the pixels within the Landsat composite to extract reflectance values for each point.

```

28 var trainingLayers = [forest, water, city, cropland, OpenLand];
29 var n = 500;
30 // loop over training layers
31 for (var i = 0; i < trainingLayers.length; i++) {
32   // sample points within training polygons
33   var pts = ee.FeatureCollection
34     .randomPoints(trainingLayers[i].geometry(), n);
35   // add class
36   var thisClass = trainingLayers[i].get('class');
37   pts = pts.map(function(f) {
38     return f.set({class: thisClass});
39   });
40   // extract raster cell values
41   var training = testimage.sampleRegions(pts, ['class'], 100);
42   // combine training regions together
43   if (i === 0) {
44     var trainingData = training;
45   } else {
46     trainingData = trainingData.merge(training);
47   }
48 }

```

Figure 8 JavaScript code used to subsample and randomise the training datasets

Following the finalisation of the subsampled training data and the acquisition of a cloud-free Landsat composite, the Random Forest algorithm was initiated in order to conduct the classification of the image. As seen from the JavaScript code extract in Figure 9, below, the individual bands used for the classification was first imported as a variable ‘bands’. The first 7 bands of the Landsat 8 composite + the NDVI added band, was used for the classification algorithm for the 2015 and 2019 land cover maps. Bands 1-5, band 7 and the added NDVI band was used for the 2010 land cover map, based on the Landsat 7 composite. Subsequently, the ‘classifier’ variable initiates the Random Forest algorithm and *train* it on the subsampled training dataset created in the previous step. Lastly the ‘classified’ variable use the ‘classifier’ variable to classify the rest of the Landsat composite, while the ‘p’ variable adds a palette of colours to display the classified image using the *Map.addlayer* function.

```

51 //// classify with random forests
52 // use bands 1-7+NDVI
53 var bands = ['B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'NDVI'];
54 // fit a random forests model
55 var classifier = ee.Classifier.randomForest(500)
56   .train(trainingData, 'class', bands);
57 // produce the land cover map
58 var classified = testimage.classify(classifier);
59 var p = ['00ff00', 'ff0000', '000000', '0000ff', 'orange',,];
60 // display
61 Map.addLayer(classified, {palette: p, min: 1, max: 5}, 'classification');

```

Figure 9 JavaScript code for the Random Forest classification of the Landsat composite using the subsampled training data

The final step in the land cover classification is the validation of the land cover maps. Using the code JavaScript code shown in Figure 10, a confusion matrix is created in order to assess the performance of the classification algorithm. In the first part of the code, the ground truth validation point FeatureCollections are compiled into one FeatureCollection using the function *featureCollection.merge()* and stored as a new variable “testingsep”.

Subsequently, a reducer function is applied in order to retrieve the land class attributes to compile a complete validation dataset with land class properties. The variable “validation_sep” compiles the validation data points within the classified land cover map in order to assess the land cover class of each validation point within the classified map. The variable errorMatrix_sep executes the errorMatrix function, using the validation_sep variable, in order to derive an error matrix. Lastly, the error matrix is exported as a table to Google Drive.

```

63 //Accuracy assessment
64 var testingsep = forestvali.merge(watervali).merge(cityvali).merge(croplandvali).merge(openlandvali);
65 // Add reducer output to the Features in the collection.
66 testingsep = testimage.sampleRegions(testingsep, ['class'], 30);
67 //print (testingsep)
68 var validation_sep = testingsep.classify(classifier);
69 //print (validation_sep)
70 var errorMatrix_sep = validation_sep.errorMatrix('class', 'classification');
71 //print('Error Matrix:', errorMatrix_sep);
72 var ft = ee.FeatureCollection([ee.Feature(null, {'Accuracy': errorMatrix_sep.accuracy()})]);
73 Export.table.toDrive({collection: ft, description: 'accu_2018',
74   fileNamePrefix: 'accu_2018', folder: 'Master thesis', selectors: ['Error Matrix']});

```

Figure 10 JavaScript code used to derive error matrixes, used for land cover validation

The confusion matrix for all three land cover maps can be seen in Table 11, Table 12 and Table 13 for 2010, 2015 and 2019, respectively, in Appendix B. The rows in the matrix correspond to instances of the actual class (validation data), while the columns are instances of the predicted class. The diagonal within the matrix indicates the correctly classified instances, while the off-diagonal instances are the number of incorrect classifications. The overall accuracy of the classification, as well as the users and producer’s accuracy, can be derived from the error matrix. The overall accuracy is calculated by adding the correctly classified diagonal values and dividing it with the total number of reference points (50 points*5 classes= 250). The producer’s accuracy is calculated by dividing the accurately classified instances in each class with the total number of reference sites within that class. The user’s accuracy is calculated by dividing the correct classifications for a particular class with the row total.

The overall accuracy, producers- and users’ accuracy for each of the three land cover maps (2010, 2015 and 2019) can be seen from Table 4 below. As can be seen, the overall accuracy, ranging from 92,8 to 94 % indicates a very high accuracy for all land cover maps. However, while the total producers and users’ accuracies are also high, individual classes such as cropland, open land and urban have lower accuracy scores in a few cases. This likely has to do with the similarity in spectral response between these three classes in the case study area.

Table 4 Accuracy scores for the 2010, 2015 and 2019 land cover maps

Land cover class	Overall accuracy			Producers accuracy			Users accuracy		
	2010	2015	2019	2010	2015	2019	2010	2015	2019
Forest				96	90,9	100	96	100	86
Water				100	100	100	100	100	100
Urban				95,5	97,8	97,9	84	90	92
Cropland				91,8	100	80	90	82	88
Open Land				82,5	84,5	85,5	94	98	94
Total (in %)	92,8	94	92	93,2	94,6	92,7	92,8	94	92

The quantified land cover area under each land class, and for each year, can be seen from the graph presented in Figure 11 below. The final classified land cover maps for 2010, 2015 and 2019 can be seen from Figure 12, Figure 13 and Figure 14, respectively, on the next three pages.

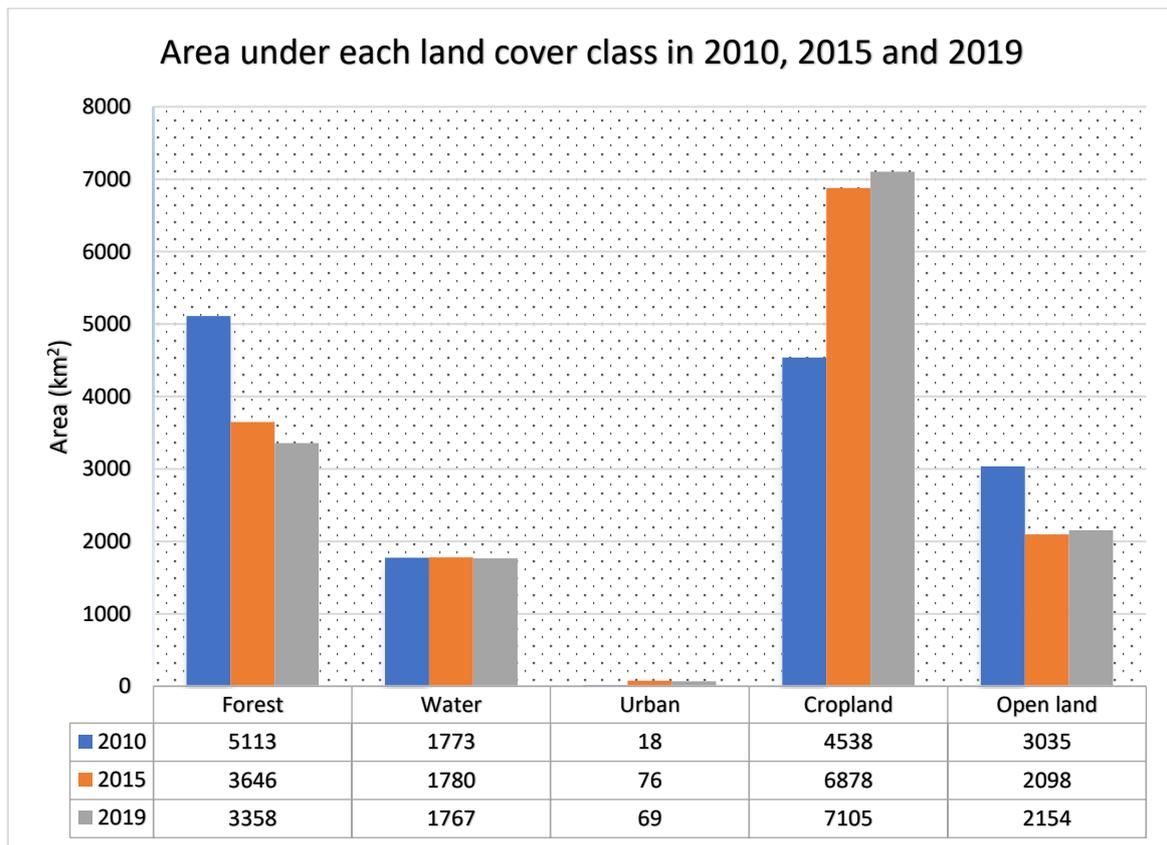


Figure 11 Land cover area per class in 2010, 2015 and 2019

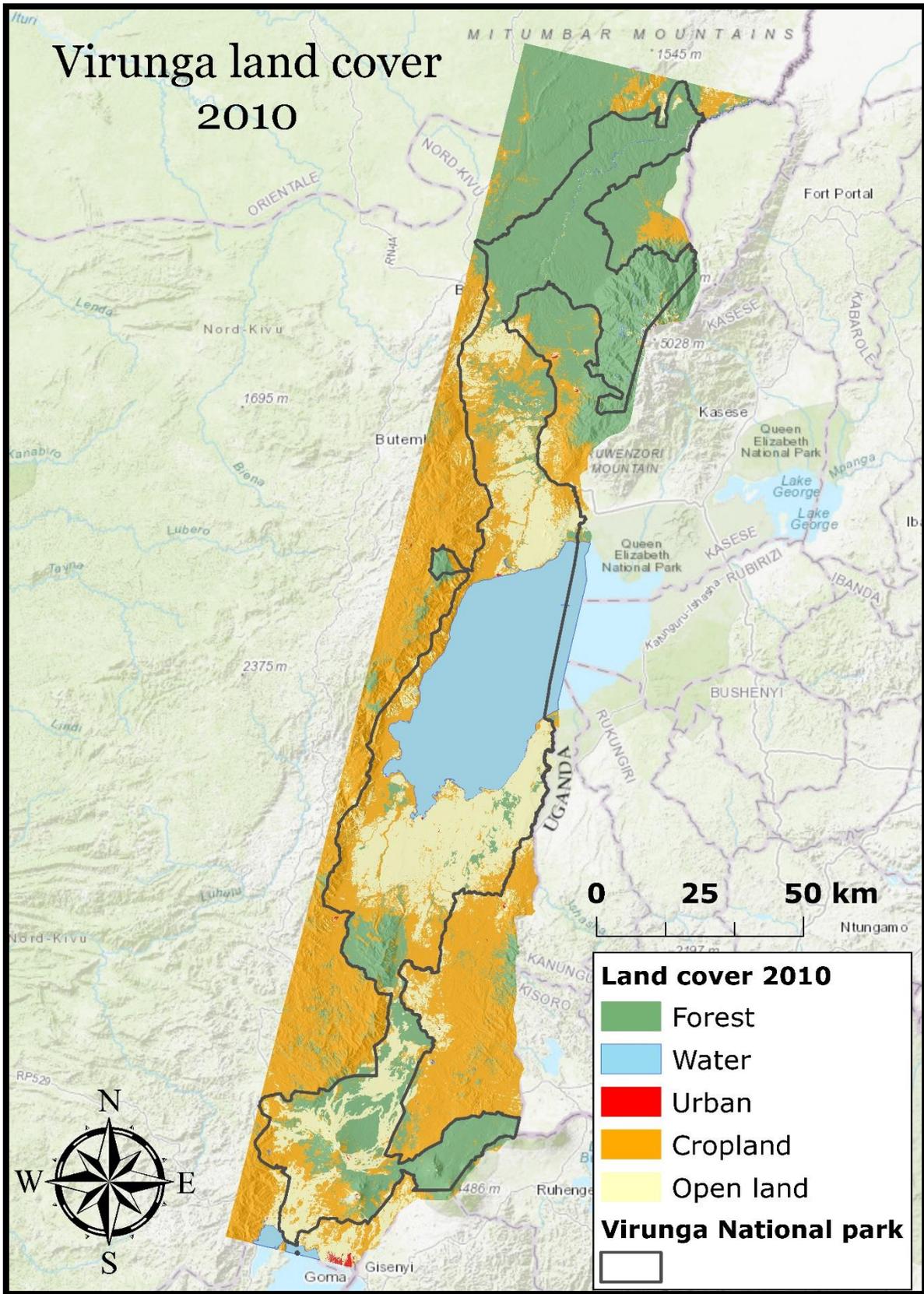


Figure 12 Land cover map - 2010

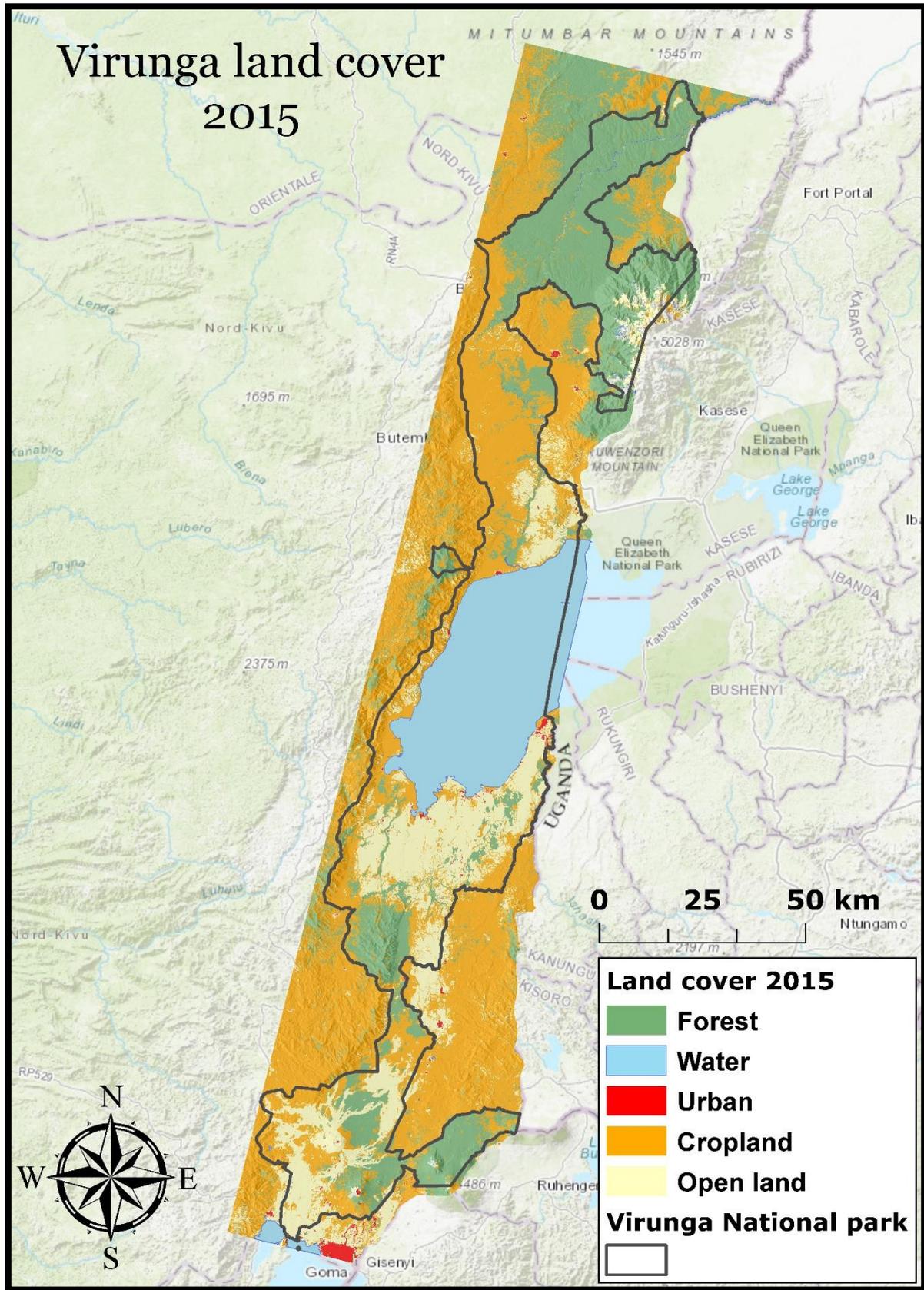


Figure 13 Land cover map - 2015

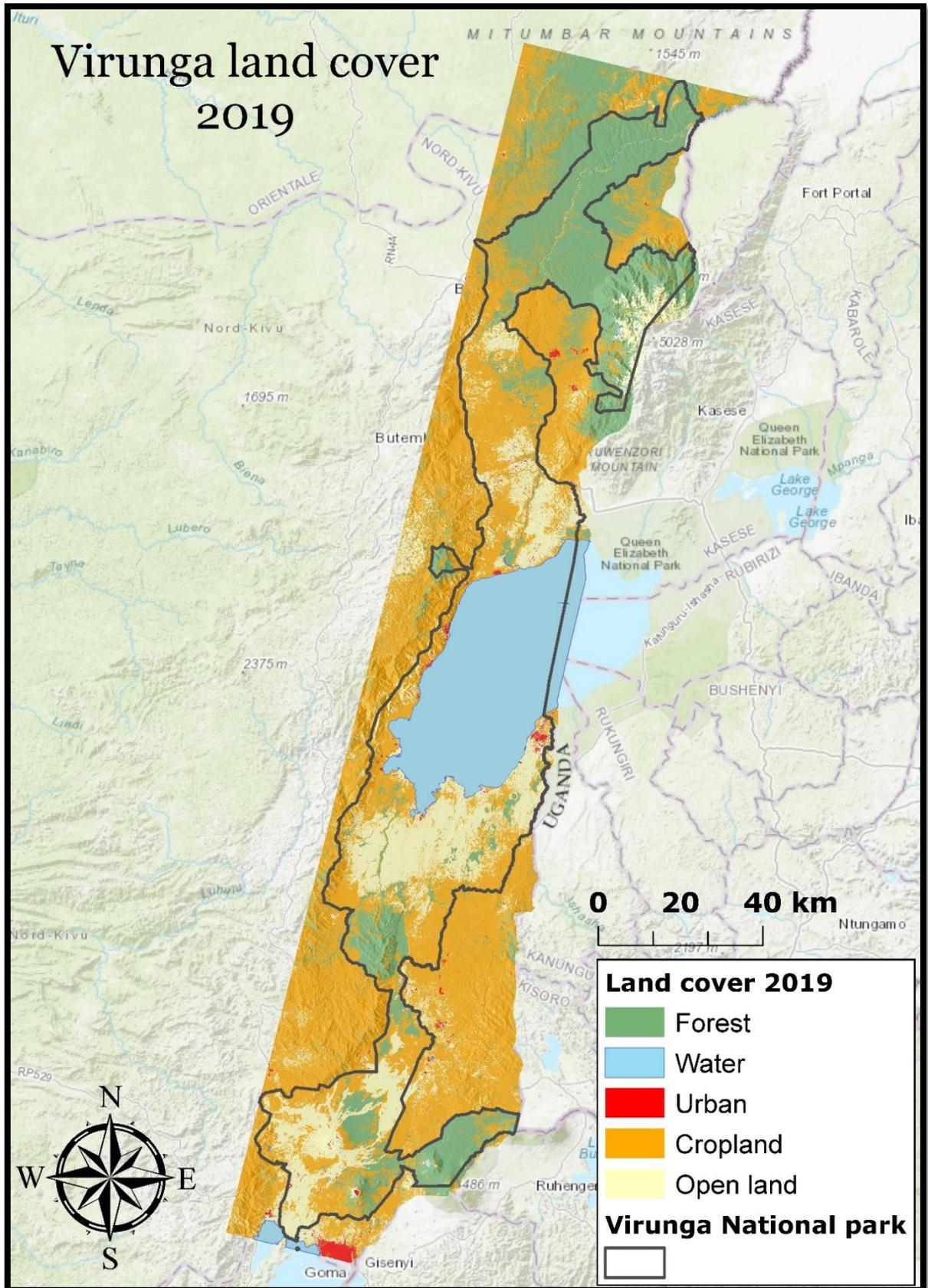


Figure 14 Land cover map - 2019

2.2 LULC modelling and prediction

The LCM module within TERRSET was used to conduct the sequential steps conforming to the requirements of LULC modelling using an MLP-Markov Chain approach. In this section, each step of the LULC modelling process is described.

2.2.1 Land change analysis

In order to assess the spatiotemporal changes between 2010 and 2015, the earlier and latter land cover maps were cross-tabulated using the CrossTab tool in Terrset. Cross-tabulation is a means to determine the amounts of changes between two different land cover maps and determine conversions between land classes (Alphan et al., 2008). The cross-tabulation table shown in Table 5 below, shows the frequencies with which the land classes remained the same (Diagonal) or changed into other categories (off-diagonal frequencies). The table represents quantities of conversion from the earlier to the later land cover data, and it clearly depicts significant changes, primarily between forest and cropland.

The following information was obtained about the changes in each class from the table:

1. Between 2010 and 2015 the forest cover was reduced by 28,7 % from 5113,4 km² in 2010 to 3646,4 km² in 2015. Even though there was a forest gain of 318,9 km² largely caused by afforestation from cropland and open land, the net loss of 1467 km² is almost exclusively attributed to forest conversion into cropland.
2. The water bodies have remained largely unchanged, which is to be expected as there have been no waterworks (e.g. dam construction) in the study period. Thus, the water bodies, largely consisting from the two major lakes in the study area, lake Édouard and lake Kivu, has remained relatively consistent. The rivers and waterways in the area study area are relatively small, and mainly runs through dense forest areas, and thus the main differences and interchanges in water bodies are likely due to classification anomalies in separating river bodies from other classes, mainly forest and open land.
3. Accounting for the least prevalent land class in the case study area, urban areas have experienced a large increase between 2010 and 2015, from 17,8 km² to 75,6 km², resulting in a 57,9 km² net gain. This is largely attributed with rapid urbanisation processes in the Democratic Republic of Congo in general, which has an estimated average annual urban population growth rate of 4.3% (United Nations, 2013). The population of the capital city in the North Kivu province, Goma, located in the south-eastern corner of the case study area, increased from 150,000 people in 1990 to more than one million in 2017 (Yee, 2017). Thus, the majority of the urban class increase is caused by the expansion of Goma.
4. Cropland is the most dynamic land class in the case study area and represents the most dominant land cover type. The total area under cultivation increased by 51,5 %, from 4538,4 km² in 2010 to 6877,6 km² in 2015. As mentioned previously, cropland is the main driver of deforestation and thus the

majority of the agricultural expansion is caused by forest conversion. However, another 47 % of cropland expansion is attributed with the cultivation of previously open land/grassland areas.

- The open land cover class was reduced by 30,8 %, from 3035,0 km² in 2010 to 2098,3 km² in 2015. Even though the open land class received a total net loss, 236,2 km² was gained, caused by agricultural abandonment. Another 66 km² gain of open land is attributed to deforestation. The majority of the net loss of the open land class (1109,2 km²) is associated with agricultural expansion, while another 59,5 km² is attributed to urbanisation processes. A net loss of 69,8 km² is associated with afforestation processes.

Table 5 LULC change matrix for the period from 2010 to 2015 (km²) Class

		LC_2010					
		Forest	Water	Urban	Cropland	Open land	Total (km²)
LC_2015	Forest	3327,5	1,6	0,0	247,5	69,8	3646,4
	Water	3,8	1770,2	0,0	0,3	5,2	1779,5
	Urban	0,6	0,0	13,0	2,5	59,5	75,6
	Cropland	1715,4	0,3	0,9	4051,8	1109,2	6877,6
	Open land	66,1	0,9	3,8	236,2	1791,2	2098,3
	Total (km²)	5113,4	1773,0	17,8	4538,4	3035,0	14477,5

While the change matrix provides a quantitative means of assessing the land cover changes between 2010 and 2015, the map in Figure 15 below illustrates the spatial trends of change. The map includes all major land cover changes which occurred over areas larger than 1 km² in total, between 2010 and 2015. It clearly illustrates that the major transitions are associated with agricultural expansion, primarily at the expense of forest areas. Spatially, the majority of the conversion of forest to cropland is concentrated in the north-northwest and southern parts of the case study area.

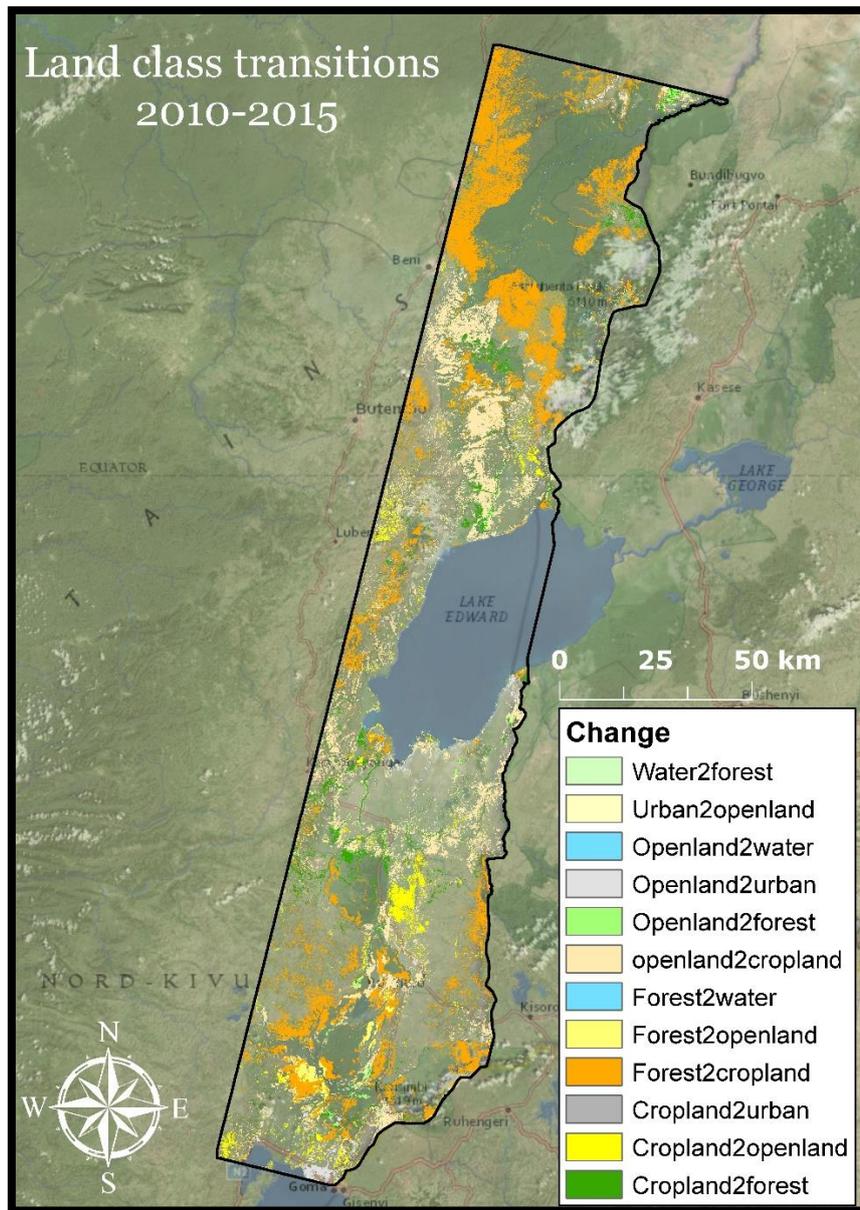


Figure 15 Class transitions between 2010 and 2015

2.2.2 Modelling Transition Potentials – sub-models

The second step in the LULC change prediction process is to model the transition potentials, which are in essence, maps of suitability/likelihood of one land cover changing into another (Eastman, 2016a). Following Pérez-Vega, Mas, and Ligmann-Zielinska (2012) the land cover transitions can be grouped together into empirically evaluated transition sub-models when the common underlying drivers are assumed to be the same. The sub-models can consist from a single land cover transition (e.g. from open land to cropland) or from multiple transitions, grouped together based on the assumption that transitions are caused by the same underlying drivers of change. The explanatory variables are used to model the historical change process based on the empirical relationship between the measured change and the explanatory variable.

Based on the major land class transitions illustrated in Figure 15, the 12 predominant transitions were grouped together based on transition type, to form 6 individual sub-models. The composition of transition groups and a description of the types of changes under each sub-model can be seen from Table 6 below. Although persistence, i.e. areas that did not change, can be considered a trajectory, it cannot be considered as a transition class, and thus areas of persistence are ignored in LCM (Gibson et al., 2018).

Table 6 Transition sub-models and descriptions

Transition sub-model	Description	Land cover transitions
Abandonment/reclamation	<i>Urban and agricultural areas converted to grassland and open land</i>	<ul style="list-style-type: none"> • Urban to open land • Cropland to open land
Afforestation	<i>Land cover classes converted to tree plantation</i>	<ul style="list-style-type: none"> • Cropland to forest • Open land to forest
Agricultural intensification	<i>Agricultural areas substituting grasslands and open land areas</i>	<ul style="list-style-type: none"> • Open land to cropland
Deforestation	<i>Forested areas converted into other land class types</i>	<ul style="list-style-type: none"> • Forest to cropland • Forest to open land
Natural dynamics	<i>Areas where natural changes cause land conversion</i>	<ul style="list-style-type: none"> • Forest to water • Water to forest • Open land to water
Urban intensification	<i>Urban areas substitute other land classes</i>	<ul style="list-style-type: none"> • Cropland to urban • Open land to urban

2.2.2.1 Explanatory variables

As reviewed in section 1.3, LULC change processes are dynamic and result from the interaction between a range of different, primarily biophysical and socioeconomic criteria. In LULC change modelling, these criteria are also referred to as ‘explanatory’ variables, as these *explain* the components of the causal relationships determining the land cover dynamics and they form a critical prerequisite for developing a realistic land change model. The explanatory variables sum up the ‘knowledge’ that the model will use to simulate future land cover scenarios (Allan et al., 2008). The driving forces, influencing change in the case study area is based on spatial analysis and a literature review of similar case studies. Each driver variable was tested for its potential explanatory value using Cramer’s V scores. Cramer’s V is a coarse statistic measure of the strength of association or dependency between two variables and it ranges from 0,0 to 1,0 in values. Generally, variables

with a total Cramer's V score higher than 0,15 are considered useful and those with a score over 0,4 are considered good (Eastman, 2016a).

In choosing explanatory variables, the processes contributing to land cover change needs to be visualised in the form of a spatial dataset representing the underlying changes, at a spatial resolution consistent with the land cover maps. GIS data sets were identified to describe the transitions in the case study area and geo-processing was performed to derive spatial datasets to, either directly or as a proxy, explain the underlying changes for each transition. According to Eastman (2016b) variables cannot be categorical and thus needs to be continuous and quantitative.

The drivers that were used in this study include;

1. *Elevation (Digital Elevation Model (DEM))*
2. *Aspect (Asp)*
3. *Slope*
4. *Evidence Likelihood (EL)*
5. *Distance from artisanal mines (D_am)*
6. *Distance from disturbance (D_disturb)*
7. *Distance from cities (D_cities)*
8. *Distance from forests (D_forest)*
9. *Distance from mining concessions (D_mining)*
10. *Distance from roads (D_roads)*
11. *Distance from waterways (D_water)*

Basic metadata and the explanatory potential for each driver measured in Cramer's V scores, in total, and for each land class, can be seen from Table 7 below. The temporal resolution of the datasets can also be seen from the table. The temporal resolution is important as it is critical to identify spatial datasets that reflect the conditions at the time of the earlier land cover map (2010), and as seen from the table, the majority of the explanatory drivers used in this study derives from the immediate period before or after 2010.

Elevation, aspect and slope are considered geophysical limitations and these are commonly used in LULC modelling literature (Allan et al., 2008; Gibson et al., 2018; Mishra et al., 2014; Paegelow et al., 2007; Teresa et al., 2015). These types of drivers act as physical limitations or incentives for certain transitions to occur. For example, water bodies will not expand into areas with high slopes (Gibson et al., 2018) and gentle slopes are generally considered more appropriate for agriculture and urban areas (Wondie et al., 2011). The aspect (i.e. the direction of the slope) influence the relative amount and intensity of sunlight and thus influence the suitability for certain types of vegetation and agriculture (Wondie et al., 2011). Elevation determines the

distribution of vegetation (Wondie et al., 2011) and generally tend to influence urban growth (Shade et al., 2019).

Evidence likelihood is a way to transform a categorical variable into a continuous surface, based on the relative frequency of pixels belonging to the different categories within the areas of change (MIRICI, 2018). In this study, **evidence likelihood** is a quantitative measure of the frequency of change between urban areas and cropland (also called disturbance) and all other land classes from 2010-2015. Thus, it represents the relative frequency of which the different land cover classes occurred in the areas that transitioned to urban or cropland. Essentially it answers the question of each category of the variable, “*How likely is it that you would have a value like this if you were an area that would experience change?*” (Eastman, 2016a), meaning that it established the suitability of each pixel to transform into urban areas or cropland. This variable aims to explain the geospatial processes that determine urban expansion and agricultural intensification.

The distance drivers represent the proximity of pixels to forces that either constraints or incentivise land cover changes. As reviewed in the introduction in section 1, mining activities is one of the primary driver of deforestation within the Virunga area “*Artisanal mining operations are unregulated and often occur in riparian zones, removing forest and vegetation cover to process the mineral soil* (Institute for Environmental Security, 2008)”. Accordingly, there is a documented relationship between deforestation and mining operations, thus, distance from **artisanal mines** and **distance from mining concessions** are included as proxy drivers of forest conversion, the rationale being that the closer in proximity a forested area is to known mining operations, the more likely it is to be deforested. Likewise, these drivers will likely positively correlate with an increase of open land, urban areas and cropland nearer to the mining concessions. **Distance to disturbance** is a spatial driver made from extracting Euclidian distances from areas which were urban or cropland in 2010. The hypothesis is that future anthropogenic disturbance is believed to be closer to areas of existing disturbance, and thus distances to existing disturbance are believed to be closely correlated with urbanisation processes and agricultural expansion. This relationship was tested by extracting the areas that transitioned to cropland or urban areas between 2010 and 2015 and correlating the frequency of changes with the distance to disturbance layer for 2010 (Figure 16). As can be seen from the figure, there is a very sharp decline in the frequency of change when moving further away from disturbed areas in 2010, until the point where nearly no change occurs when moving further than 2 km away from the disturbed areas. This means that anthropogenic disturbance is positively correlated with distance from disturbed areas.

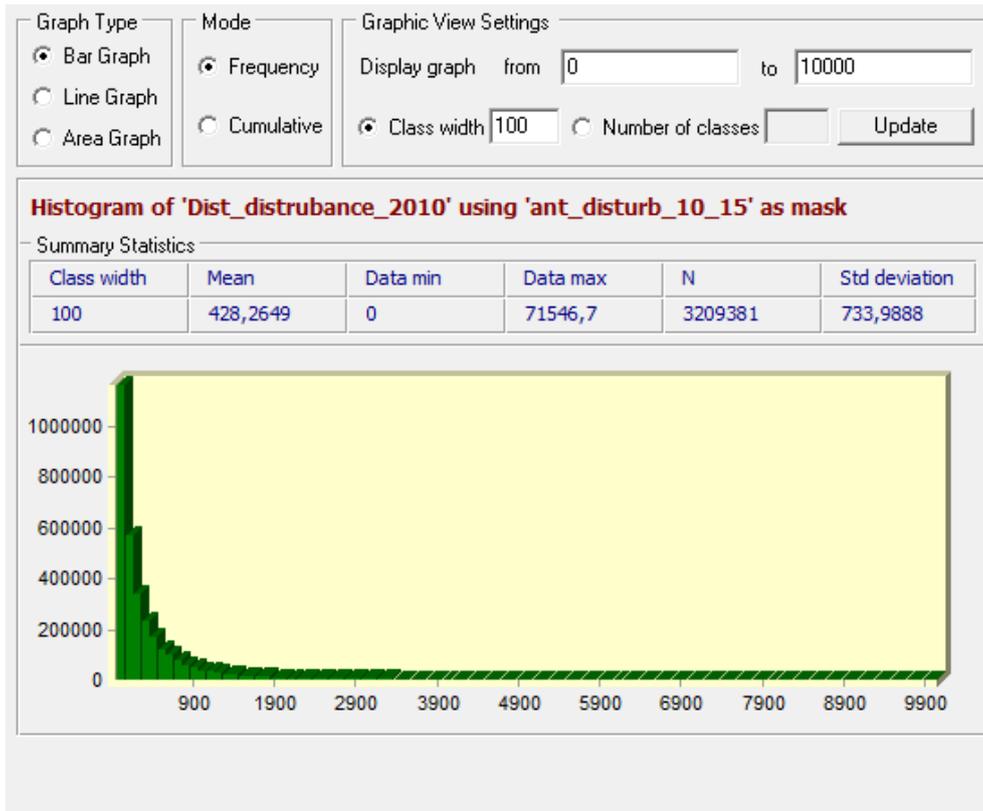


Figure 16 Histogram depicting the correlation between the distance (in meters) from disturbed areas in 2010 and the actual disturbance between 2010 and 2015

The closer a pixel is to urban areas, the more likely it is to be transformed, due to urbanisation processes, and thus *distance from cities* explains the relationship between urban expansion and LULC changes. Furthermore, as reviewed in Institute for Environmental Security (2008), deforestation patterns in Virunga NP are correlated with urban areas as a result of firewood collection and agricultural expansion. *Distance from forests* is a measure of the Euclidian distances to forest areas in 2010 and is among a number of transitions, believed to be a causal driver of deforestation processes as deforestation is likely to be closer to the edge of the existing forest areas, than in the middle of the deep forest. *Distance to roads* and *distance to waterways* represents drivers of accessibility and as reviewed in Allan et al. (2008), forest conversion to agriculture and plantations occur when roads and rivers provide easy access and water for irrigation.

A description of how the GIS data for each explanatory driver was derived and pre-processed is provided in the next section 2.2.2.1.1.

Table 7 Description of potential explanatory variables and associated Cramer's V scores

Variable	DEM	Asp	Slope	EL	D_am	D_disturb	D_cities	D_forests	D_mining	D_roads	D_water	
Data origin	SRTM 90m Digital Elevation Database v4 ²			Land cover 2010 + 2015	International Peace Information Service (IPIS) ³	Land cover 2010	World Resources Institute ⁴	Land cover 2010	World Resources Institute ⁵			
Data format	Raster (GeoTiff)			Raster (GeoTiff)	Shapefile (points)	Raster (GeoTiff)	Shapefile (points)	Raster (GeoTiff)	Shapefile (polygons)	Shapefile (lines)	Shapefile (lines)	
Coordinate system	WGS 84			EPSG:4051	WGS 84	EPSG:4051	WGS 84	EPSG:4051	WGS 84			
Spatial resolution (m)	90 m cell resolution resampled to a 30 m resolution			30 m	1: 50 000 vector scale converted to 30 m cell resolution	30 m	1: 50 000 vector scale converted to 30 m cell resolution	30 m	1: 50 000 vector scale converted to 30 m cell resolution			
Temporal resolution	2008			2010-2015	2009-2016	2010	2009	2010	2013	2009	2009	
Geoprocessing	Reproject	Reproject; computed from DEM		Computed from land cover maps	Reproject; clip to AOI; Euclidian distance from all artisanal mines (rasterize)	Reclassify boolean (urban/cropland); Euclidian distance from disturbed areas	Reproject; clip to AOI; Euclidian distance from all cities (rasterize)	Reclassify boolean (forest); Euclidian distance from forest areas	Reproject; clip to AOI; Euclidian distance from all mining concessions (rasterize)	Reproject; clip to AOI; Euclidian distance from all roads (rasterize)	Reproject; clip to AOI; Euclidian distance from all waterways (rasterize)	
Cramer's V	<i>Forest</i>	0,52	0,21	0,26	0,68	0,17	0,40	0,17	0,43	0,14	0,26	0,27
	<i>Water</i>	0,61	0,83	0,57	0,42	0,32	0,68	0,54	0,82	0,27	0,57	0,30
	<i>Urban</i>	0,14	0,07	0,06	0,10	0,08	0,04	0,06	0,04	0,09	0,09	0,07
	<i>Cropland</i>	0,42	0,32	0,34	0,29	0,31	0,52	0,20	0,32	0,31	0,34	0,29
	<i>Open land</i>	0,23	0,17	0,24	0,65	0,20	0,15	0,18	0,30	0,24	0,11	0,08
	Overall	0,42	0,42	0,33	0,48	0,22	0,42	0,28	0,46	0,22	0,32	0,20

² <http://srtm.csi.cgiar.org/>

³ <http://geo.ipisresearch.be/geoserver/web/>

⁴ <http://www.wri.org> - downloaded from GeoData at Tufts (<https://geodata.tufts.edu/>)

⁵ <http://www.wri.org> - downloaded from GeoData at Tufts (<https://geodata.tufts.edu/>)

2.2.2.1.1 Pre-processing of explanatory drivers

This section will provide a brief description of the geospatial data acquisition and processing steps used to derive the spatial datasets for each of the 11 explanatory variables. Maps of the final processed datasets can be seen from Figure 20.

DEM

The DEM data retrieved consists from Version 4 of the ‘SRTM digital elevation dataset’ and was exported from Google Earth Engine using the JavaScript code presented in Figure 17 below. To execute this code, first, a shapefile of the AOI was uploaded to the Google Earth Engine ‘assets’ repository and imported as a variable ‘*framed*’. Subsequently, the code imports the data variable (CGIAR/SRTM90_V4 is the Image ID for the SRTM V.4 dataset in Google Earth Engine) and clips this to the framed object (AOI). Lastly, the data is exported using the function ‘*Export.image.toDrive*’, setting the parameters for scale to 30 and CRS to EPSG:4051. Setting these parameters will ensure that the dataset is resampled from its native 90 m resolution to 30 m and reprojected into an EPSG:4051 coordinate system.

```
Imports (1 entry)
  var AOI: Table users/icemads/AOI
1 Map.centerObject(AOI,6)
2 //CGIAR/SRTM90_V4
3 var dem = ee.Image('CGIAR/SRTM90_V4').clip(AOI);
4
5 Export.image.toDrive({
6   image: dem.select('elevation'),
7   description: 'DEM_30m',
8   folder: 'Master Thesis',
9   scale: 30,
10  region: AOI.geometry().bounds(),
11  maxPixels: 24818397458,
12  crs: 'EPSG:4051'
13 });
14
```

Figure 17 JavaScript code to acquire DEM data from Google Earth Engine

Slope and aspect

Google Earth Engine was used to process the SRTM digital elevation dataset to retrieve Slopes in degrees. These were retrieved and exported using the JavaScript code presented in Figure 18 below. Similarly, to the DEM, this code frames the data retrieval to the extent of the AOI and imports the DEM as a variable. Next, the variables slope and aspect are created, using the inherent Google Earth Engine algorithm ‘*ee.Algorithms.Terrain*’ with the DEM variable as input data. This algorithm derives three separate bands; slope, aspect and hill shade from a DEM dataset. The ‘*Export.image.toDrive*’ function is applied, using each of the variables and setting the parameters for scale to 30 and CRS to EPSG:4051. The ‘Slope’ output band is selected as the export image for slope and the ‘aspect’ output band for aspect.

```

Imports (1 entry)
  var AOI: Table users/icemads/AOI
1 Map.centerObject(AOI,6)
2 //CGIAR/SRTM90_V4
3 var dem = ee.Image('CGIAR/SRTM90_V4').clip(AOI);
4 var slope = ee.Algorithms.Terrain(dem);
5 var aspect = ee.Algorithms.Terrain(dem)
6
7 Export.image.toDrive({
8   image: slope.select('slope'),
9   description: 'slope_30m',
10  folder: 'Master Thesis',
11  scale: 30,
12  region: AOI.geometry().bounds(),
13  maxPixels: 24818397458,
14  crs: 'EPSG:4051'
15 });
16
17 Export.image.toDrive({
18  image: aspect.select('aspect'),
19  description: 'aspect_30m',
20  folder: 'Master Thesis',
21  scale: 30,
22  region: AOI.geometry().bounds(),
23  maxPixels: 24818397458,
24  crs: 'EPSG:4051'
25 });

```

Figure 18 JavaScript code to process DEM data to acquire datasets for slopes and aspect

Evidence likelihood

The evidence likelihood variable was created by using the integrated variable transformation utility within TERRSET's LCM. 'Evidence Likelihood' is chosen in the transformation type and the 'Ant_disturb_10_15' raster file is chosen as the 'transition layer'. The 'Ant_disturb_10_15' layer is made by mapping all changes to cropland or urban land between 2010 and 2015, using the output classification from the change analysis of transition types (Figure 15), and reclassifying these changes into a Boolean map of: no change (0) or change (1). The 'input variable' is the earlier land cover map for 2010 and the output variable name is set to 'Evidence likelihood'. Ticking the categorical box, the resulting layer is a quantitative variable made from a categorical Boolean variable of change/no change, describing the relative frequency with which different land cover classes occurred in the areas that transitioned to cropland or urban areas.

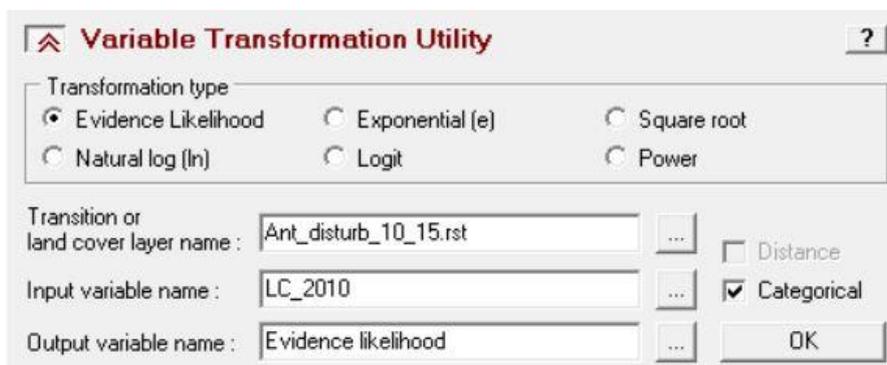


Figure 19 User interface of the Variable Transformation Utility in TERRSET LCM

Distance artisanal mining

The point dataset on the location of artisanal mining operations in the Democratic Republic of Congo derives from the International Peace Information Service and was acquired from their Geoserver⁶. The file was downloaded in a Shapefile format, complete with attribute tables covering individual subclasses, such as the name of the site, visit date and mineral mined. Following retrieval, the dataset was reprojected into EPSG:4051 and clipped to the extent of the AOI using ArcMaps ‘*Clip*’ tool along with a vector file covering the extent of the AOI. Subsequently, Euclidean distances were derived from the point dataset, using ArcMaps ‘*Euclidean distance*’ tool. The output cell size of the raster file was set to 30 m and the AOI raster was applied as a mask in the environment settings.

Distance from disturbance

The distance from disturbance describes the distance from cropland or urban areas in 2010. It was created using two tools within TERRSET’s GIS analysis package, namely ‘*Reclass*’ and ‘*Distance*’. The ‘*reclass*’ tool was used to reclassify the earlier land cover map for 2010 into a Boolean, urban/cropland (1) and non-urban/cropland (0), categorical raster layer. Consequently, the ‘*distance*’ tool was used to derive Euclidian distance to the areas classified as urban or cropland in 2010.

Distance from cities

The point dataset on the location of all major and minor cities in the Democratic Republic of Congo derives from the World Resources Institute and was acquired through Tuft University’s geoportal (GeoData at Tufts)⁷. The file was downloaded in a Shapefile format, complete with attribute tables covering individual subclasses, such as the name of city, province and district. Following retrieval, the dataset was reprojected into EPSG:4051 and clipped to the extent of the AOI using ArcMaps ‘*Clip*’ tool along with a vector file covering the extent of the AOI. Subsequently, Euclidean distances were derived from the point dataset, using ArcMaps ‘*Euclidean distance*’ tool. The output cell size of the raster file was set to 30 m and the AOI raster was applied as a mask in the environment settings.

Distance from forests

The distance from forests describes the distance from existing forest areas in 2010. It was created using two tools within TERRSET’s GIS analysis package, namely ‘*Reclass*’ and ‘*Distance*’. The ‘*reclass*’ tool was used to reclassify the earlier land cover map for 2010 into a Boolean, forest (1) and non-forest (0), categorical raster layer. Consequently, the ‘*distance*’ tool was used to derive Euclidian distance to the areas classified as forest in 2010.

⁶ <http://geo.ipisresearch.be/geoserver/web/>

⁷ <https://geodata.tufts.edu/>

Distance from mining concessions

The mining concessions data, containing the location of all registered mining concessions in the Democratic Republic of Congo, derives from the World Resources Institute and was acquired through Tuft University's geoportal (GeoData at Tufts). The file was downloaded in a Shapefile format, complete with attribute tables covering individual subclasses, such as licence holder of the concession, resource type and area size. Following retrieval, the dataset was reprojected into EPSG:4051 and clipped to the extent of the AOI using ArcMaps 'Clip' tool along with a vector file covering the extent of the AOI. Subsequently, the clipped dataset was rasterized using Arcmap's tool 'Polygon to Raster'. Lastly, Euclidean distances were derived using the 'Euclidean distance' tool. The output cell size of the raster file was set to 30 m and the AOI raster was applied as a mask in the environment settings.

Distance roads

The line dataset on roads in the Democratic Republic of Congo derives from the World Resources Institute and was acquired through Tuft University's geoportal (GeoData at Tufts). The file was downloaded in a Shapefile format, complete with attribute tables covering individual subclasses, such as road classes (Locale, Provinciale and Nationale). Following retrieval, the dataset was reprojected into EPSG:4051 and clipped to the extent of the AOI using ArcMaps 'Clip' tool along with a vector file covering the extent of the AOI. Subsequently, Euclidean distances were derived from the line dataset, using ArcMaps 'Euclidean distance' tool. The output cell size of the raster file was set to 30 m and the AOI raster was applied as a mask in the environment settings.

Distance to waterways

The line dataset on waterways in the Democratic Republic of Congo derives from the World Resources Institute and was acquired through Tuft University's geoportal (GeoData at Tufts). The file was downloaded in a Shapefile format, complete with attribute tables covering individual subclasses, such as waterway class and length. Following retrieval, the dataset was reprojected into EPSG:4051 and clipped to the extent of the AOI using ArcMaps 'Clip' tool along with a vector file covering the extent of the AOI. Subsequently, Euclidean distances were derived from the line dataset, using ArcMaps 'Euclidean distance' tool. The output cell size of the raster file was set to 30 m and the AOI raster was applied as a mask in the environment settings.

Explanatory variables

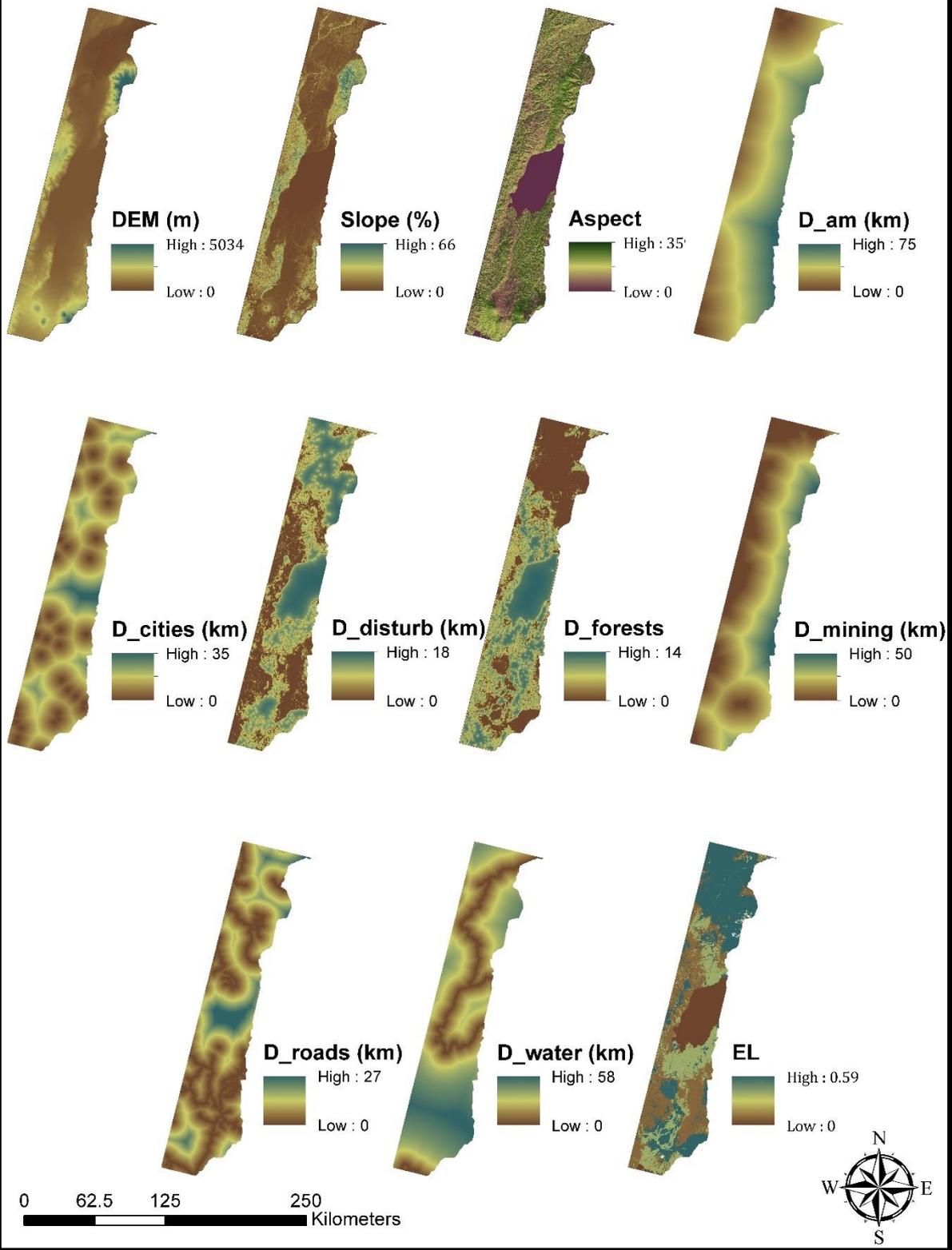


Figure 20 Processed explanatory variable datasets used as input for the MLP modelling

2.2.3 Modelling Transition Potential – MLP calibration

The LCM module allows three different approaches to creating transition potential maps based on the individual sub-models and associated explanatory variables: MLP neural network, logistic regression, or a similarity-weighted instance-based machine learning tool (SimWeight) (Eastman, 2016a). According to Eastman (2016), MLP performs stronger than the two alternatives when modelling non-linear relationships between land cover change and explanatory variables. In contrast to both SimWeight and logistic regression, MLP can also run multiple transitions per sub-model, making it more flexible and dynamic when multiple transition types are modelled. Furthermore, as MLP has fewer parameter settings and functions through an automatic and dynamic learning process, it is also more user-friendly, and consequently, MLP was chosen for this study. Operationally, within LCM, MLP creates a random sample of cells that transitioned and a sample of cells that persisted and use half of the samples to train the model and develop multivariate functions (adjusting the weights) to predict the potential for change based on the value of the conditions at each location (Allan et al., 2008). The other half of the subset sample of cells that transitioned and persisted is used to test the performance of the model (validation). When launched in the LCM, the MLP starts to train on the training samples and starts to operate in automatic mode, whereby it iteratively adjust parameters to increase the performance of the model (Eastman, 2016a). The automatic mode monitors and modifies the start and end learning rates for a dynamic and continuously adjusted learning process (Eastman, 2016a). While the automatic mode was used to modify and adapt the start and end learning rates and select the number of hidden layer nodes, the other parameters of the MLP were set as follows as suggested by Eastman, (2016a): Momentum factor (0,5); Sigmoid constant (1,0); sample size per class (10000 or the minimum cells which transitioned within the sub-model, if less than 10000).

As reviewed in section 2.2.2, the 12 major transitions which occurred in the period between 2010 and 2015 were grouped together in 6 different sub-models, namely; Abandonment/reclamation, Afforestation, Agricultural intensification, Deforestation, Natural dynamics and Urban intensification. The next step in modelling the transition potential was to assign the explanatory variables to each sub-model. Variables can be added to the model either as static, meaning that they don't change over time, such as slope, or dynamic, meaning that they do change over time, such as proximity to roads (assuming dynamic road development). Static variables are unchanging over time and express aspects of basic suitability for transitions under consideration, while dynamic variables are time-dependent, such as proximity to existing forest areas or road networks, and are recalculated during the course of a future land cover simulation (Eastman, 2016a). In this study DEM, slope, aspect, EL, D_am, D_mining and D_water was used as static variables, while D_disturb, D_cities, D_forests and D_roads were designated as dynamic variables.

An iterative approach was used to establish the most appropriate, and accurate, combination of driver variables for each sub-model, while avoiding overfitting. Each sub-model was fitted with all 11 explanatory drivers to

being with, and an iterative approach was used to remove the driver with the least explanatory potential, while assessing the accuracy score and skill level of the model after each iteration. The accuracy score provides a value in percentage which indicates how well the model is able to predict the changes that happened between 2010 and 2015, accounting for both change and persistence. The skill measure compares the number of correct predictions, minus those attributable to random guessing, to that of a hypothetical perfect prediction (Gibson et al., 2018). Thus, the skill measure provides an indication of how the explanatory drivers will explain past changes. The skill is measured on a scale from -1 to 1, where values less than 0 indicates that the model performs worse than what would be expected by random guessing, 0 indicates that the model performs as well as random guessing while values between 0 and 1 indicates that the performance of the model exceeds what is expected by pure chance.

After each iteration of calibrating individual sub-models using MLP, a report about the nature of the model performance is created (see example in Appendix C). This provides critical information on the overall accuracy and skill of the model, the skill measure broken down by component (transition & persistence type) and the explanatory power of each variable. Section 3 of the report entitled '*Sensitivity of Model to Forcing Independent Variables to be Constant*' provides information on the sensitivity of the explanatory variables. As seen in Figure 21 below, sub-section 3) in this section entitled '*Backwards Stepwise Constant Forcing*' provides information on the accuracy and skill measure of the model when holding one or more variables constant. In step 1, the variable with the lowest negative effect on the skill is held constant, and this provides information on the explanatory potential of this variable. If the accuracy and skill of the model don't decrease by much, when holding the variable constant, this suggests that the variable has little value and can be removed (Eastman, 2016a). On each iteration of the calibration of each sub-model, the variable with the least explanatory potential was removed until a combination of 5-6 of the variables with the strongest explanatory potential was left under each sub-model.

3) Backwards Stepwise Constant Forcing

Model	Variables included	Accuracy (%)	Skill measure
With all variables	All variables	72.25	0.6300
Step 1: var.[5] constant	[1,2,3,4,6,7,8,9,10]	72.25	0.6300
Step 2: var.[5,10] constant	[1,2,3,4,6,7,8,9]	72.23	0.6298
Step 3: var.[5,10,7] constant	[1,2,3,4,6,8,9]	72.06	0.6275
Step 4: var.[5,10,7,9] constant	[1,2,3,4,6,8]	69.86	0.5981
Step 5: var.[5,10,7,9,4] constant	[1,2,3,6,8]	65.48	0.5397
Step 6: var.[5,10,7,9,4,8] constant	[1,2,3,6]	62.50	0.4999
Step 7: var.[5,10,7,9,4,8,3] constant	[1,2,6]	54.55	0.3940
Step 8: var.[5,10,7,9,4,8,3,1] constant	[2,6]	47.77	0.3036
Step 9: var.[5,10,7,9,4,8,3,1,6] constant	[2]	29.40	0.0587

Figure 21 Extract from the calibration report indicating accuracy scores and skill measure of the model when holding variables constant.

Consequently, the final selected variables were loaded into the sub-model structure to execute the final iteration of the MLP training. The final skill measure and accuracy rate of each model calculated through MLP is summarized in Figure 22 below and the explanatory drivers used under each sub-model and selected performance scores is provided in Table 8.

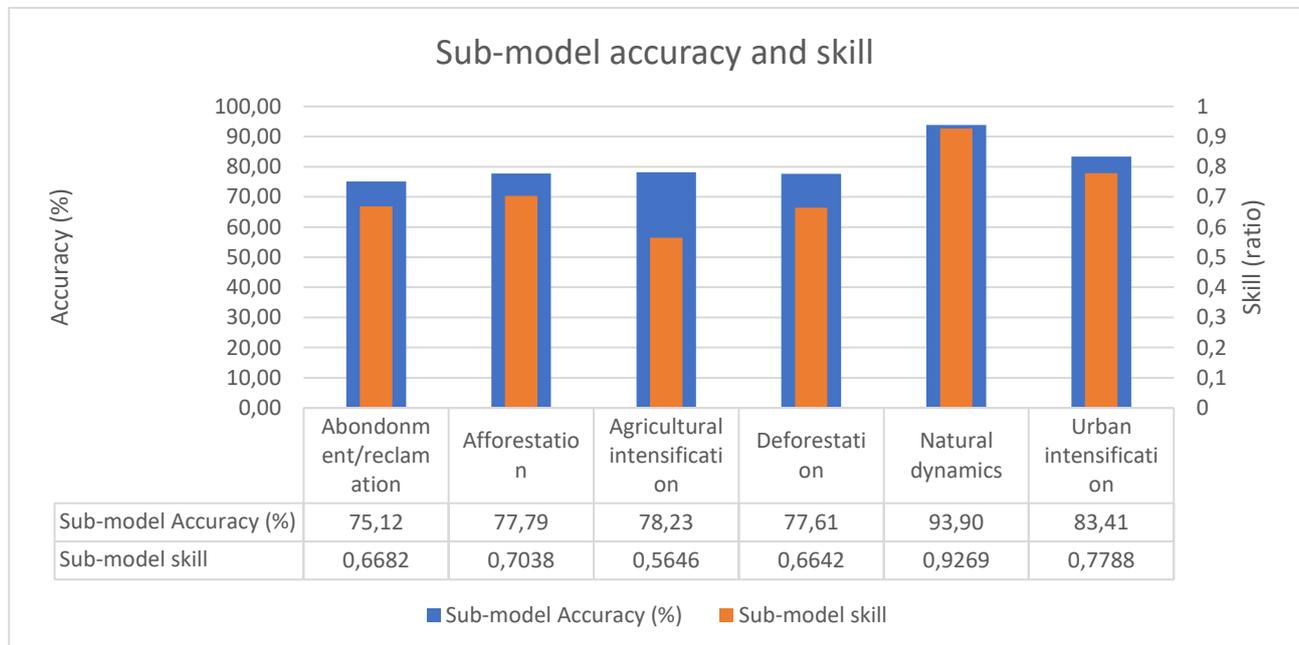


Figure 22 Sub-model accuracy and skill measure from MLP

As can be seen from the figure, the accuracy and skill measure reveal some disparity between the level of confidence of the transition modelling under each sub-model, however overall the values are fairly consistent,

ranging from 75 % to 93 %. Abandonment/reclamation has the lowest accuracy score (75,12 %), followed by deforestation (77,61 %), afforestation (77,79 %) and agricultural intensification (78,23 %). Agricultural intensification, however, has the lowest skill measure of all the sub-models (0,56). Natural dynamics and urban intensification performed best, with accuracies of 93,90 % and 83,41 % respectively. The skill measure of these two sub-models was also the highest among all six, with 0,93 for natural dynamics and 0,78 for urban intensification.

The outcome of the transition potential modelling is a series of transition potential maps, describing the suitability for each of the 12 major transitions included in the sub-models. These maps can be seen in Figure 23 below.

Table 8 Sub-models included in MLP with associated explanatory variables and selected performance indicators

Sub-model	Explanatory variables	Transition/Persistence class	Class measure (ratio)	skill	Sub-model accuracy	Sub-model skill	RMS	
							Training	Testing
Abandonment/reclamation	DEM; Slope; D_am; D_cities; D_mining; D_water	Urban to Openland	0.8134		75.12 %	0.6682	0.2980	0.3071
		Cropland to Openland	0.5741					
		Persistence: Urban	0.7401					
		Persistence: Cropland	0.5398					
Afforestation	DEM; Slope; EL; D_disturb; D_forests; D_water	Cropland to Forest	0.5181		77.79 %	0.7038	0.2751	0.2737
		Openland to Forest	0.8918					
		Persistence: Cropland	0.6536					
		Persistence: Openland	0.7515					
Agricultural intensification	DEM; D_am; D_disturb; D_mining; D_roads; D_water	Openland to Cropland	0.5961		78.23 %	0.5646	0.3899	0.3906
		Persistence: Openland	0.5329					
Deforestation	DEM; D_am; D_disturb; D_mining; D_roads; D_water	Forest to Cropland	0.6103		77.61 %	0.6642	0.3358	0.3369
		Forest to Openland	0.8300					
		Persistence: Forest	0.5516					
Natural dynamics	DEM; Slope; EL; D_forests; D_water	Forest to Water	0.9848		93.90 %	0.9269	0.1207	0.1281
		Water to Forest	0.9096					
		Openland to Water	0.8707					
		Persistence: Forest	0.8677					
		Persistence: Water	0.9849					
		Persistence: Openland	0.9441					
Urban intensification	Slope; EL; D_am; D_cities; D_mining; D_roads	Cropland to Urban	0.8664		83.41 %	0.7788	0.2536	0.2489
		Openland to Urban	0.6564					
		Persistence: Cropland	0.8294					
		Persistence: Openland	0.7630					

Transition potentials

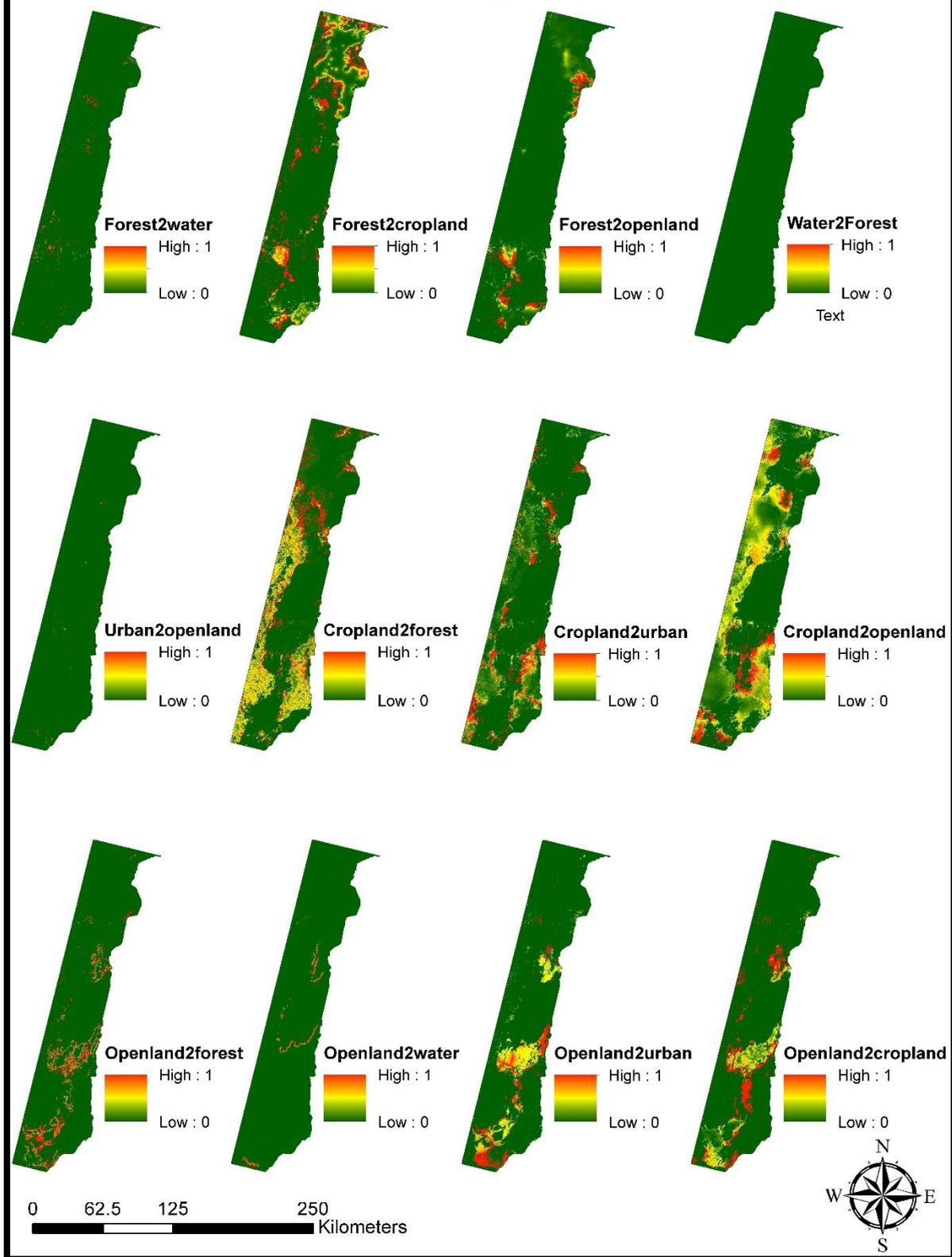


Figure 23 Transition potentials

2.2.4 Change prediction and model validation

Following the transition sub-model development, the 12 transition potential maps were used as input in a Markov Chain model to simulate future LULC changes. The Markov Chain determines the amount of change using the earlier and later land cover map along with a pre-specified future year (Eastman, 2016a). The Markov module produces a transition probability matrix, seen in Figure 24 below, which is a matrix which records the probability of each landcover class to change into every other land class category. It also creates a transition areas matrix which is a record of the number of pixels that are expected to change from each land cover class over the specified time frame (Eastman, 2016a). Finally, the Markov Chain creates a set of conditional probability images which reports the probability of a land cover type to be found at each pixel after the specified prediction date (Eastman, 2016a). However, as the matrices only determine the quantity of change, the transition potential (suitability) maps are utilized within the Markov analysis to spatially allocate changes in order to make a land cover prediction for a future year (Eastman, 2016a).

Given :	Probability of changing to :				
	Forest	Water	Urban	Cropland	Openland
Forest	0.3169	0.0017	0.0014	0.6271	0.0528
Water	0.0019	0.9953	0.0000	0.0017	0.0011
Urban	0.0208	0.0016	0.4025	0.2786	0.2965
Cropland	0.1029	0.0005	0.0034	0.8022	0.0910
Openland	0.0701	0.0035	0.0268	0.6451	0.2545

Figure 24 Markov Chain transition probability matrix

Consequently, Markov Chain analysis was used to make a LULC prediction for 2019 and subsequently validated by using the actual 2019 land cover map for comparison. Yearly recalculation stages were assigned in the model to specify the frequency of which the dynamic variables are recalculated in the model. This means that the D_disturb, D_cities, D_forests and D_roads explanatory variables are updated in the model every year until the prediction year.

Figure 25 below shows the actual and the predicted land cover map for 2019. A visual inspection indicates that the predicted land cover map, overall, looks fairly similar to the actual land cover map, however there are localised discrepancies where the model failed to predict changes/persistence, for example in the mid-west where the simulation predicted cropland to replace large open land areas, when in actuality it didn't.

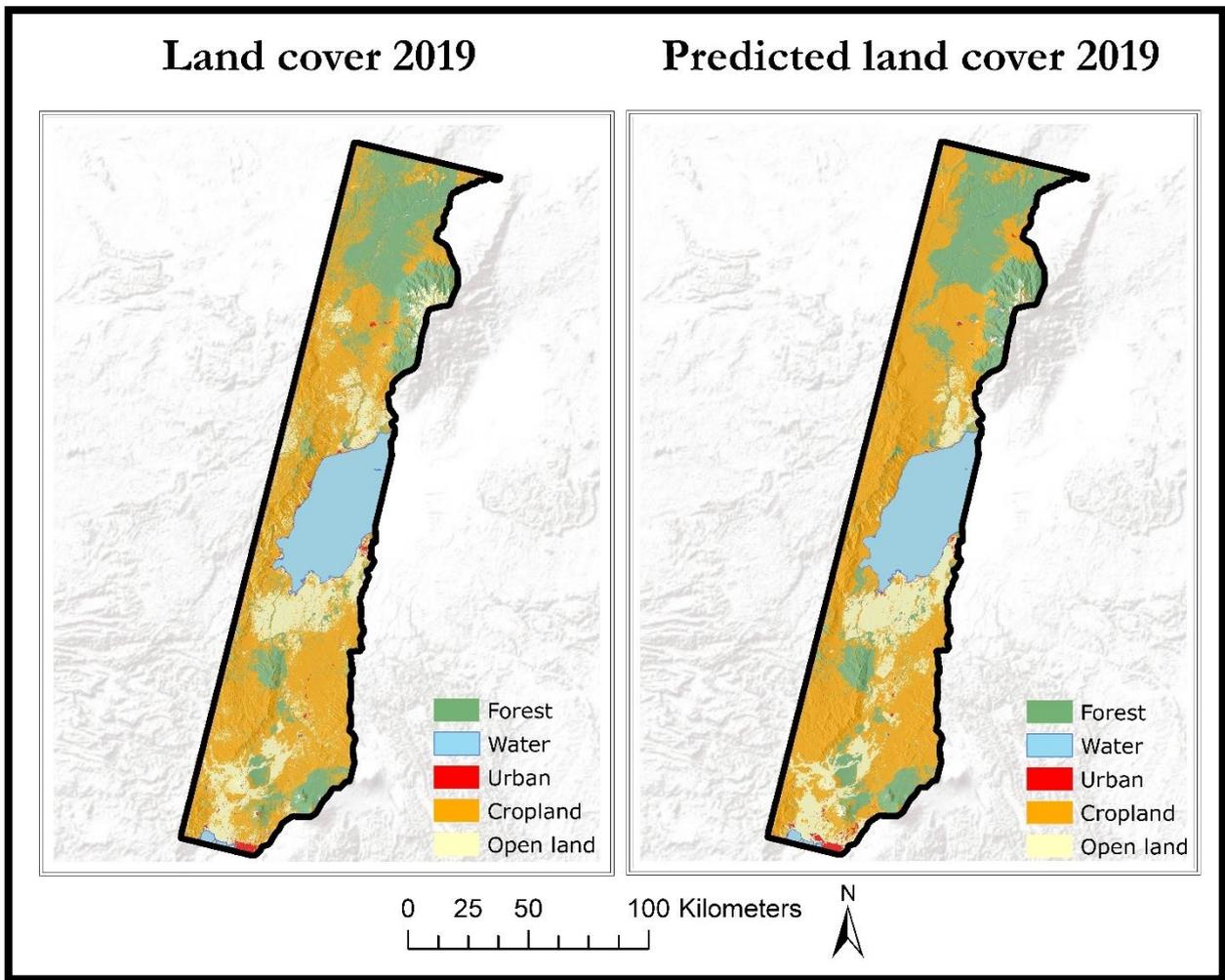


Figure 25 Actual land cover map for 2019 versus the predicted 2019 land cover map

However, a visual inspection will not allow for a comprehensive assessment of the subtle differences and overall accuracy of the prediction, and thus statistical analysis was used to compare the actual and simulated land cover maps for 2019, using Kappa Index of Agreement (KIA) scores. The *validate* module in TERRSET provides a comprehensive means to measure the agreement between two categorical images by calculating various KIA and related statistics (the full report of the analysis is included as Appendix D). The KIA scores can be used to test the agreement between a ‘comparison’ map and a ‘reference map’, both in terms of the quantity of cells in each land cover category and the agreement in terms of location of these cells (Eastman, 2016a). The Kappa Standard (K_{standard}) is equivalent to kappa and indicates the proportion of correctly assigned pixels versus the proportion that is correct by chance. The Kappa for no information (K_{no}) indicates the overall agreement between the simulated and reference map (Eastman, 2016a). The Kappa for location (K_{location}) is a measure of the spatial accuracy in the overall landscape, due to the correct assignment of values in each category between the simulated and reference map (Eastman, 2016a). The Kappa for stratum-level location ($K_{\text{locationStrata}}$) is a measure of the spatial accuracy within preidentified strata, and it indicates how well the grid

cells are located within the strata (Eastman, 2016a). The combination of K_{standard} , K_{no} , K_{location} and $K_{\text{locationStrata}}$ scores allows for a comprehensive assessment of the overall accuracy both in terms of location and quantity. All KIA scores range from 0 to 1 (or 0 % to 100%), where 0 indicates that agreement is equal to agreement due to chance and 1 (or 100 %) indicates perfect agreement.

The K scores of the simulated 2019 land cover map ‘*comparison*’ compared with the actual 2019 land cover map ‘*reference*’ is provided in Table 9 below.

Table 9 K scores for 2019

K INDICATORS	2019
K_{STANDARD}	0,8828
K_{NO}	0,9224
K_{LOCATION}	0,9001
$K_{\text{LOCATIONSTRATA}}$	0,9001

The statistics from the k scores shows that K_{no} is 0,9224, K_{location} is 0,9001, $K_{\text{locationStrata}}$ is 0,9001 and the overall K_{standard} is 0,8828. According to Zadbagher et al. (2018) a model is valid if the overall Kappa (K_{standard}) score exceeds 70 % (or 0,7). The K_{standard} score, close to 90 %, is a very strong indicator of the overall accuracy and performance of the model, and the remaining k scores, all exceeding 85 %, indicate that there are almost no, or very small quantification and location errors between the predicted and the actual land cover map for 2019. Thus, the simulation has a strong ability to predict both the quantity and the locations of change.

3 Results

The validated model described in the previous section 2 was used to make a land cover prediction for each year, starting with the year 2020 up until and including the year 2030, which coincides with the critical landmark year for the achievement of the UN SDG's. The compilation of land cover predictions from 2020 through to 2030 can be seen from Figure 26, while the predicted land cover for 2030 is presented in Figure 27.

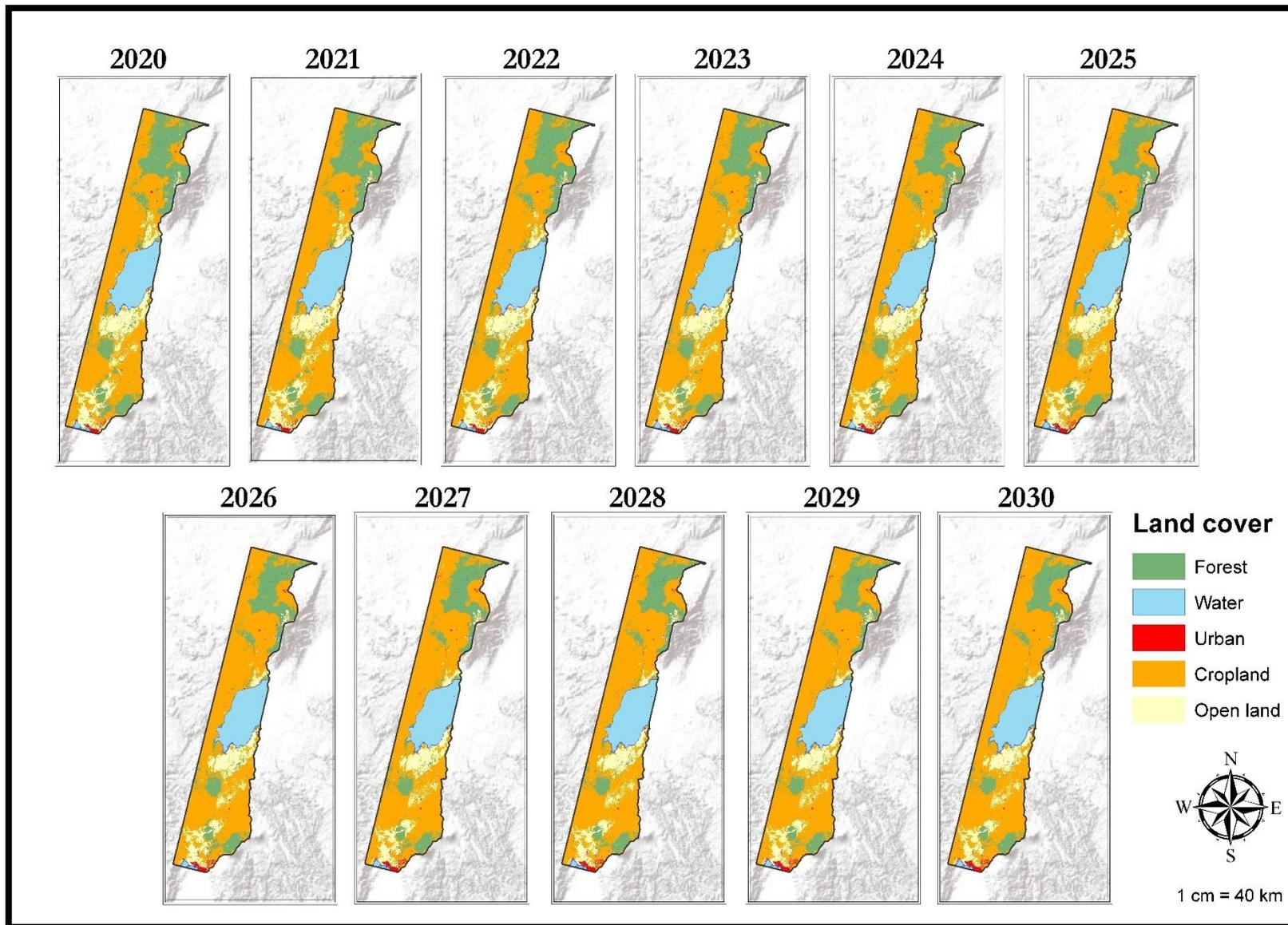


Figure 26 Predicted land cover maps from 2020 to 2030

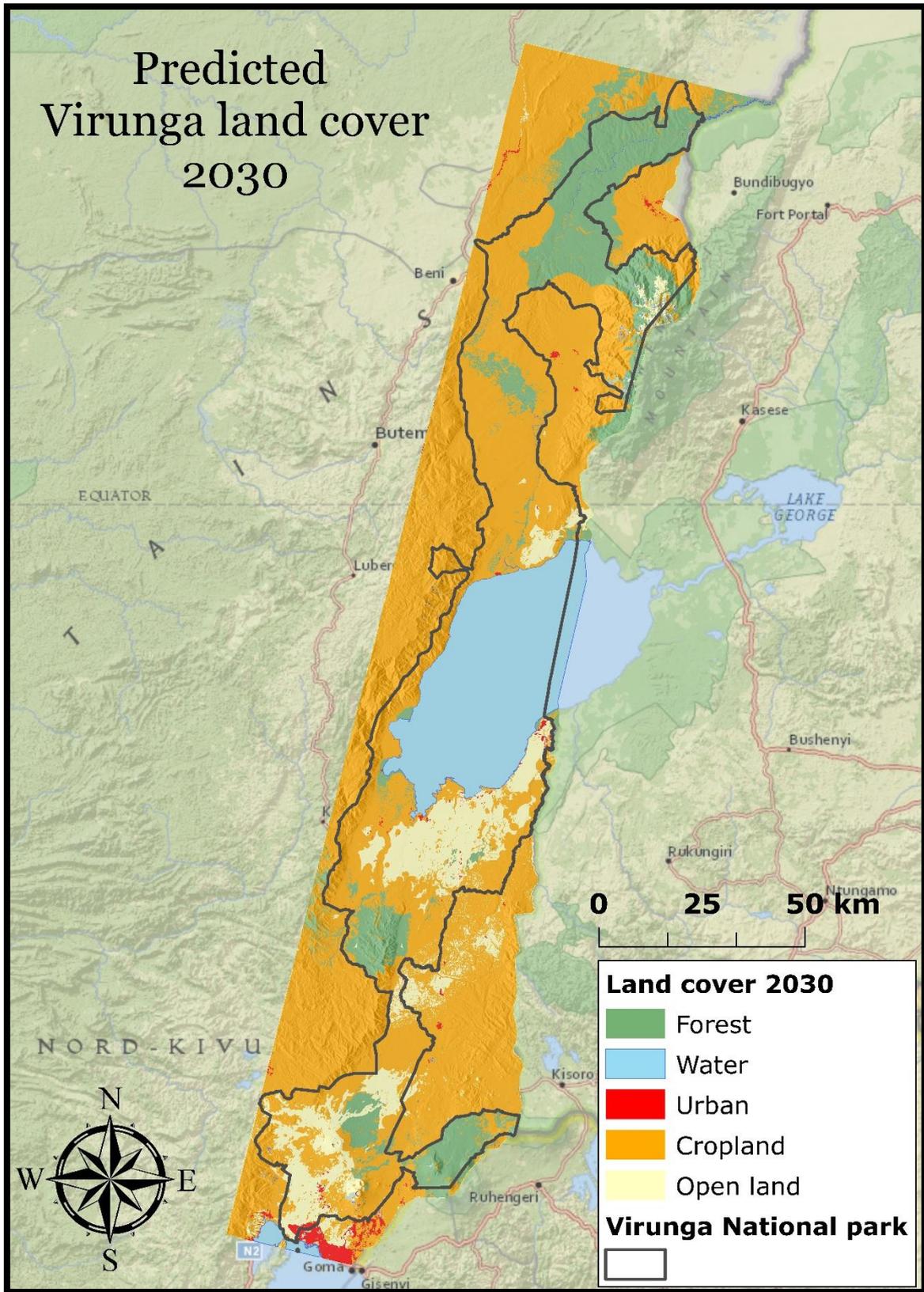


Figure 27 Predicted 2030 land cover in Virunga

The series of land cover predictions covering the whole period from 2020 to 2030, and the final land cover map for 2030 presented in Figure 27 clearly illustrates that the model predicts continuous cropland expansion, primarily at the expense of forest areas and existing open lands. The model also predicts continuous urban development, particularly around existing settlements. The collective change per class in total, and per cent change per year, is illustrated in Figure 28 below. As depicted in the graph, the forest cover will continue to decrease throughout the 10-year period, with an average annual loss of 4,21 % and a total area loss of 1085 km², from 3104 km² in 2020 to 2019 km² in 2030. Water coverage will, as expected remain largely the same, gaining a negligible average of 0,04 % per year. Urban expansion and development of new settlements will continue, gaining an average annual of 3,44 %. The total urban area is predicted to increase by 38 km², from 95 km² in 2020 to 133 km² in 2030, and looking at the predicted land cover map, most of this is expected to be as a result of urban sprawl around the main city of Goma in south-eastern Virunga. As also visually apparent, cropland expansion will continue throughout the 10-year period, gaining an average annual area of 1,83 %, and a total area gain of 1522 km², from 7636 km² to 9161 km² class coverage. Along with forest areas, open land/grassland zones are expected to decrease the most, by 2,96 % per year, losing a total of 482 km² in the 10-year period, from 1857 km² in 2020 to 1375 km² in 2030.

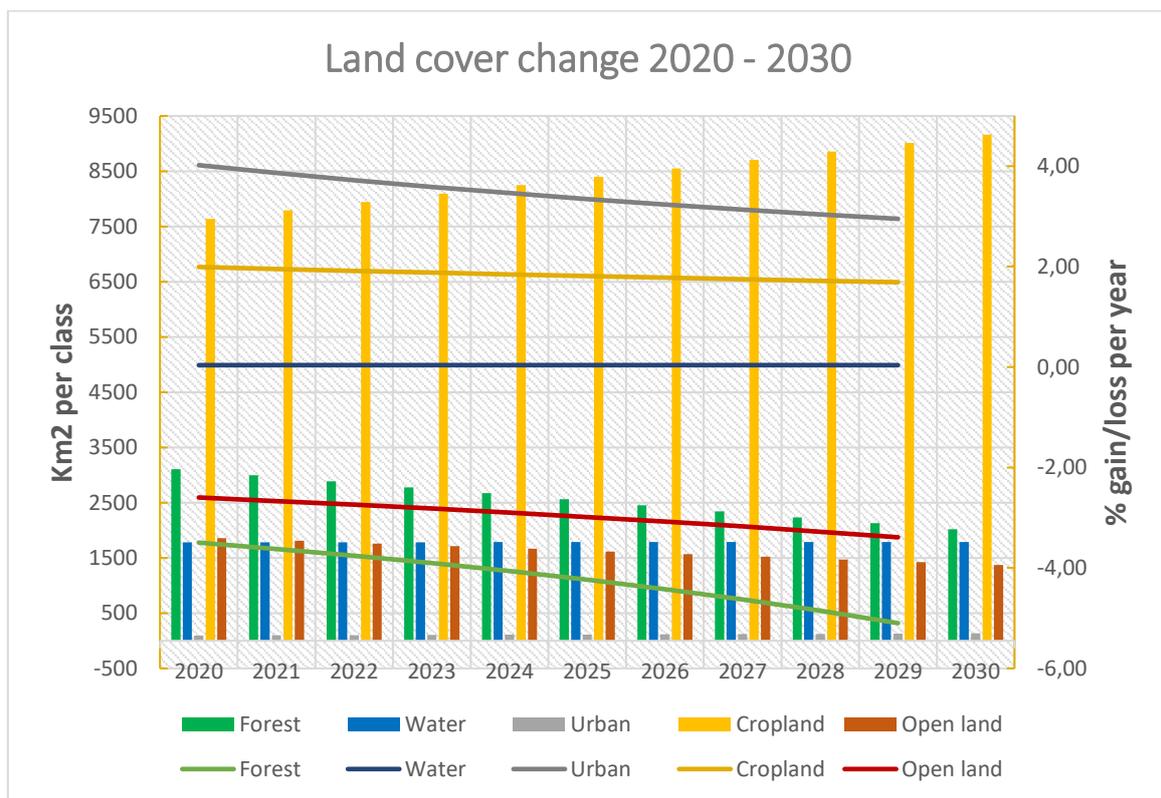


Figure 28 Predicted land cover change between 2020 and 2030, in % yearly (gain/loss) and total annual area coverage in km² per class

4 Discussion and conclusion

Understanding LULC changes, transitions, landscape risks and dynamics is paramount in order to inform policies, planning interventions and actions aiming to ensure sustainable development in all dimensions (economic, social and environmental) conforming to the objective of the UN SDG's and the 2030 agenda for sustainable development. In this study, a combined MLP-Markov Chain approach has been used to simulate future land cover changes in the period from 2020 to 2030, in a case study area covering the Virunga NP in the Democratic Republic of the Congo, and its immediate vicinity. Two simulations were carried out. The first (2019) was used for model validation and accuracy assessment, and the second (2030) was used to predict landscape change in the Virunga NP catchment. The assessment of the spatial patterns of LULC change derived through a change analysis of historical trends combined with the development of a plausible future land cover scenario for the Virunga catchment will help to improve the understanding of the land system and establish cause-effect relationships between driver variables and land cover dynamics. Thus, the LULC change model aims to contribute to informing policy responses aiming to support sustainable land management and landscape planning decisions within the Virunga NP.

As an empirical-statistical model, the LULC model developed in this study predicts a future land cover state, based on a business as usual scenario. Past land cover changes within the Virunga catchment has been largely linked with charcoal production and cropland expansion, which have impeded conservation efforts and put critical pressure on the ecological integrity of the landscape and its biodiversity. By cross-tabulating two land cover maps for 2010 and 2015, this study aimed to quantify past land cover changes and identify spatial trends of change. It concludes that forest conversion into cropland is the most common and frequent type of landcover change, contributing to the majority of the total net forest loss of 28,7 % between 2010 and 2015. The most significant forest loss occurred around the perimeter of the forest areas in the northern sector of the case study area, however, the forest areas just north of the North Kivu provincial capital of Goma, also experienced substantial losses. While a 318,9 km² forest cover gain was also identified in the change assessment, these gains cannot be qualified in existing literature, and as these gains are largely located within, or close by, existing cropland areas, they may likely be the result of misclassified pixels, possibly classifying plantation development as forest. While the urban cover is the least predominant land class type within the case study area, the cross-tabulation indicated that the urban land cover quadrupled between 2010 and 2015. The majority of the urban gain, however, is associated with significant urban development around Goma, located in the south-eastern corner of the case study area. Another major land cover change results from the conversion of open land/grassland areas into cropland. In total, open land areas were reduced by 30,8 % in the period from 2010 to 2015 and the majority of these transitions was located just north of Lake Edward.

If unchecked and unregulated, the LULC change model developed in this study indicates that the landscape within and outside the NP will continue to change dramatically in the next 10 years. While Figure 28 in the previous section quantified the collective amount of changes, per land class per year, from 2020 to 2030, Table 10 below quantifies the projected land changes from 2019 to 2030, between land classes. As can be seen from the cross tabulation, forests are attributed with the most significant land class loss in the period from 2019-2030, and almost all forest loss (1579,5 km² of a total of 1651,1 km²) is associated with cropland expansion. Open land is also projected to experience significant land class loss (1162,8 km²) the majority of which (1081,8 km²) is also attributed to cropland expansion. While urban areas are projected to continue to be the minority land class in the Virunga NP catchment, the total urban area cover is projected to almost double, from 68,5 km² to 132,7 km² in the 11-year period. Most of this is attributed to the conversion of cropland and open land areas. Thus, the majority of all transitions, gains and losses, are for the most part attributed with the expansion of agricultural lands, and largely at the expense of forest areas.

Table 10 Cross tabulation between actual 2019 land cover and the simulated land cover for 2030

		LC_2019					
		Forest	Water	Urban	Cropland	Open land	Total (km²)
Simulated LC_2030	Forest	1707,2	0,2	0,5	289,7	15,8	2019,1
	Water	4,8	1766,3	0,3	3,6	13,9	1789,6
	Urban	0,9	0,0	45,7	34,0	51,3	132,7
	Cropland	1579,6	0,5	15,7	6469,5	1081,8	9161,3
	Open land	65,8	0,2	6,3	308,3	991,3	1374,9
	Total (km²)	3358,3	1767,2	68,5	7105,1	2154,0	39605,0

Addressing forest loss is a primary component of conservation efforts and land management planning within the Virunga NP catchment. Therefore, it is critical to determine not just the amount of forest cover loss, but also the spatial extent and location of forest dynamics. Figure 29 below illustrates the spatial location of the dynamics of forest land, and as seen from the figure, forest loss is largely concentrated in the northern part of the case study area, and particularly the north-eastern margin of the NP. This change is consistent with past deforestation patterns, which has historically been more predominant in the north where larger and more remote forest areas are located, and literature (i.e. (Jones, 2018)) indicate that illegal slashing of old growth forest to produce carbonized wood has been particularly predominant in the northern sector. This is largely caused by rebel groups operating near the city of Beni, supplying local villages and larger cities in the outskirts of the national park with charcoal (Jones, 2018). Besides charcoal, army groups have also been known to transport illegal timber along the Kamango Route, linking the Democratic Republic of the Congo with Uganda, causing further forest loss and fragmentation (Jones, 2018). Conversion of cleared forest areas and slashing of trees to plant subsistence crops, such as cassava and maize, is another primary driver of forest loss, particularly in the south (Jones, 2018). While forest loss is also expected to continue in the southern part of the NP,

particularly just north of Goma, the high montane forests to the north-east of Goma is predicted to remain largely intact, likely protected by its high altitude and steep terrain, making the area less accessible and thus less likely to be logged.

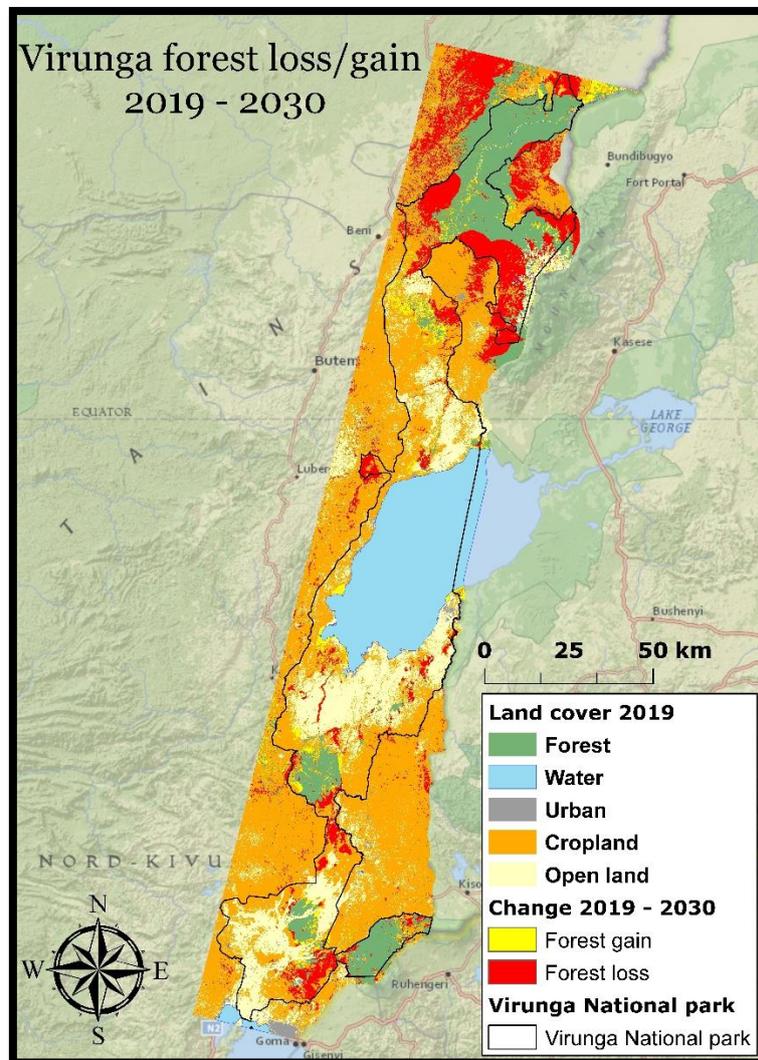


Figure 29 Spatial location of forest loss/gain from 2019 to 2030

4.1 Policy response options, planning interventions and SDG implementation

“The Democratic Republic of Congo is at a major crossroads: after a decade of little progress, the country must rise to the challenge of the SDG targets in a context of state fragility, high poverty, demographic growth, and urbanization (World Bank, 2017)”

As the LULC change model developed in the context of this study is an empirical-statistical projection of past changes into the future, the outcome represents likely LULC changes as a reflection of a business as usual scenario. Thus, the results help to understand the intrinsic drivers of change while providing valuable information on possible future LULC configurations in the Virunga catchment and thus an indication of the

causes and consequences of land-use change. In the absence of reformed regulatory policies, legal frameworks and planning interventions, Virunga NP will continue to be threatened by encroachment and deforestation, primarily caused by cropland expansion and persistent conflict. The high population density and a continuous population growth believed to be around 3 % in the Virunga region (Kayijamahe, 2008), will inevitably result in fewer resources outside the Virunga NP, which will ultimately put more pressure within the park, resulting in further damaging human impact. The large-scale deforestation and conversion to agriculture caused by human activities will severely alter the integrity of the landscape and cause strong negative impacts on biodiversity and soil degradation, while undermining the natural resource foundation on which the local livelihoods depend. The formulation of adequate spatial policies in the Virunga catchment must balance the competing needs for land to feed the accelerating population and provide energy and resources, while reducing the loss of ecosystems and biodiversity. The SDGs provide the blueprint for such policy planning and interventions, aiming to balance prosperity for both people and the planet.

The direct exploitation of resources and expansion of cropland activities is intimately linked with the economic situation of the people (Kayijamahe, 2008), and thus in order to protect the biodiversity and integrity of the Virunga NP, policies should aim to improve and strengthen the economic security and livelihoods of the people living in its vicinity. The Virunga NP exists between the extremes of economic poverty and natural wealth, which has made it a target for all of those who aim to profit from its resources. A 2013 report by the World Wildlife Fund (WWF) entitled ‘The economic value of Virunga national park’ concluded that the “*direct use of Virunga’s ecosystem could generate US\$348 million per year and help diversify DRC’s economy* (WWF, 2013)”. The main direct contributors to this value are tourism (US\$235 million), fisheries (US\$90 million) and hydropower (US\$10 million), while another US\$63.8 million, primarily attributed to carbon sequestration and erosion control, can be generated through the provision of ecosystem services (WWF, 2013). If sustainably managed, the outstanding natural value of the Virunga NP could contribute significantly to the local economy, while providing livelihoods for 45.000 people through the provision of job opportunities (WWF, 2013). Thus, policies should aim to strengthen conservation action by creating an alternative economy which incorporates and enables the surrounding communities from a thriving and well managed national park, while embracing the framework of the SDGs.

Incentivising alternatives to charcoal:

As mentioned previously, the vast deforestation in the northern and southern sectors of the park, visualized in the simulated 2030 LULC map, is largely believed to be a reflection of illegal charcoal production and land clearing for agricultural expansion. The major demand for charcoal is located within the major villages, refugee camps and the capital city of Goma in particular. As the majority of the population in Goma rely on charcoal for their entire energy consumption, the prediction of a total clearing of the forest just north of Goma is highly probable and inherently linked with charcoal production and cropland expansion. Electricity is recognized to

have substantial benefits for poverty reduction, health and education, and thus access to electricity should be incentivised and subsidized. Realising the US\$10 million potential of hydropower in Virunga NP alone, would not only contribute to providing job opportunities and tax revenue, but more importantly, release pressure on forests to obtain charcoal. Furthermore, the affordability and availability of modern cooking fuels and practices could be subsidised through regulatory reforms, i.e. reducing costs on kerosene stoves and cylinders (Crawford et al., 2008), or through the establishment of micro-credit systems. As evident from the LULC change model cropland expansion cause the majority of the land transformation and will continue to grow. Thus, adopting measures to support the development of sustainable biomass production initiatives, i.e. by improving linkages to agriculture, animal husbandry, agroforestry, etc. could be another approach to reduce the dependency on charcoal. Such policy initiatives would not only contribute to promote conservation action, and thus contribute to realizing SDG 15 (*Life on land*), but also contribute to the realisation of multiple SDG's, including SDG 1 (*No poverty*), SDG 3 (*Good health and well-being*), SDG 8 (*Decent work and economic growth*), SDG 13 (*Climate action*) and SDG 7 which aims to ensure access to affordable, reliable, sustainable and modern energy for all⁸.

Community development:

Land grabbing for subsistence agriculture has been another primary driver of change, historically, and unregulated and illegal encroachment has threatened the fringes of the Virunga NP. The LULC change model predicts that vast expanses of the NP will be subject to cropland expansion, at the expense of forests, savannahs and grassland, in 2030. To counteract the infringement, enforcement of existing legislation needs to be strengthened while at the same time community development efforts should aim to build capacity to pave the way for an alternative, and more sustainable, livelihood options for the increasing population. Community-based planning and management is undoubtedly a cornerstone of conservation action and SDG implementation, as local communities are effectively custodians of their environment. Consequently, the local communities should be involved in the wider planning framework in order to maximise the development potential and environmental benefits. Thus, in order to contribute to the conservation of the NP and reduce land grabbing, economic development in the region, communal development projects and community involvement should be promoted, e.g. expanding the fragmented and desolate road infrastructure in order to improve market access, and thus increase revenue potential from agricultural and artisanal productions. Other communal development projects could support the promotion of alternative income generating activities, such as ecotourism development or educational programmes which could facilitate access to the tourism industry, such as free public park ranger or guide training programmes. Depending on the nature of community development programmes, successful implementation of initiatives such as those outlined above could potentially contribute to the realisation of SDG 1 (*No poverty*), SDG 3 (*Good health and well-being*), SDG 4

⁸ <https://sustainabledevelopment.un.org/sdg7>

(Quality education), SDG 8 (*Decent work and economic growth*), SDG 9 (*Industry, Innovation and Infrastructure*), SDG 11 (*Sustainable cities and communities*) and SDG 15 (*Life on land*).

Utilising the LULC change model to gain intergovernmental support and mobilise resources:

While underpinning the need for reformative action to counteract the impact of deforestation and land degradation in Virunga, it is vital to realize that the majority of the policies and actions suggested will require significant investments. Accordingly, the Democratic Republic of the Congo will, to some extent, be relying on support and engagement from donor countries in order to forge strong bilateral relationships through which investments can be sourced and policies framed. Furthermore, collective international support can be forged using the framework of existing Multilateral Environmental Agreements (MEA)'s in order to better integrate conflict-concerns into the implementation and priorities and attain earmarked funding for targeted capacity building and conservation activities. For this purpose, the LULC change model and the simulated land cover for 2030 is not only an effective policy support tool to inform spatial planning and policy-making, but also a vital instrument which can be used for lobbying activities in order to gather support for conservation and poverty reduction activities and strategies at the intergovernmental level. Insight into a probable future LULC scenario within one of the most biodiverse world heritage sites in the world, which indicates that most of the forest resources within the NP will be gone by 2030, may provide further traction to support collective action and mobilisation of resources to preserve the integrity of the park and the biodiversity within it. The fortification of these bilateral and multilateral relationships will be vital in order to mainstream and finance conservation actions across sectoral policies, contributing to sustainable energy production, poverty reduction, education, health etc., thus underpinning a coordinated strategy providing political and economic governance while increasing human capacity and wellbeing. While potentially contributing to realise the majority of the SDG's, development and revitalisation of global partnerships to strengthen the implementation of the SDG's is the overall objective of SDG 17 (*Partnerships for the Goals*).

4.2 Reproducibility of the study

Detailed accounts of the software packages (including version numbers), scripts, datasets, workflows and step by step methodological guidelines should allow anyone with the same system setup and dependencies to run the analysis again, re-creating the results or use it as a guiding framework for replicating it in future research aiming to quantify and qualify future land cover change. While the script for the land cover classification generically apply to any case study area, upon collection of locally applicable training datasets, replication of the model in other settings and contexts is possible upon collection of relevant datasets for explanatory variables. Thus, the approach can be replicated in other regions to compare differences and similarities in future LULC patterns and predictions.

However, even so, and while careful elaboration of the experiment artefacts – datasets, pre-processing steps, parameters, software components, source code, etc. should allow for independent validation and reproducibility of the specific results of this study, most operations of machine learning algorithms involve some degree of randomisation, making them particularly elusive in terms of replicability. The script for the land classification makes use of random forests to classify the input image. and while random forests are considered highly accurate (Rodriguez-Galiano et al., 2012; Suthaharan, 2016), the process of building the trees in the ensemble is random. Furthermore, the process of splitting the training data geometries into 500 sample points, is random, and thus an exact reproducibility of the classification results relies on a perfectly harmonised training dataset, which for this study was impossible to obtain as in situ data sampling was impossible. The element of some degree of “randomness” in the land cover classification is critical in terms of reproducibility, as the land cover maps are the foundation of the LULC change model, and thus different results in the classification will likely result in a different prediction of a future scenario.

4.3 Sensitivity analysis

While reproducibility of the results of this study is inherently imperative, replicability and improvement of the design are equally important. This is largely facilitated through the identification and realisation of limitations and sensitivities in the project design.

All models are simplifications of the real world, and as such, they are inherently subject to potential errors as they depend on the data and assumptions applied. The results of the LULC change model developed in this context is affected by several factors, such as the accuracy of the image classification for developing the land cover maps, selection of land cover classes, filtering processes, data aggregation and data availability, selection of explanatory variables, etc. While due diligence to existing literature and a careful selection of the methodological framework can alleviate the impact of some of these errors, no research method is perfect, and all come with certain trade-offs.

4.3.1 Human factors and temporal variations

The land cover classification was conducted using two-year temporal composites of Landsat images (2008-2010 for 2010, 2013-2015 for 2015 and 2017-2019 for 2019), in order to reduce NoData values caused by clouds. While these years were selected in order to use the most recent reflection of land change dynamics in the Virunga catchment for the calibration of the model, land change rates are volatile, varying inter-annually and at short time periods (UN-DESA, 2012). However, as detailed yearly historic accounts of land change dynamics and unpredictable non-linear shifts in the Virunga catchment are limited in literature, an inherent risk lies in having projected land change extremes, rather than norms, due to the relatively short interval between the two calibration datasets. For example, model calibration cannot account for non-linear shifts such as those caused by, sudden conflict, climatic events, economic fluctuations, political shifts and natural disasters, and thus if, e.g. the period between 2008 and 2010 represented an atypical period of extreme

agricultural development rates, this pattern will have been projected into the future. Thus, the past may not always provide the best indicator of the future.

Similarly, the model is also limited by its inability to include human behaviour, climate extremes and specific policies, all of which are major drivers of LULC change (UN-DESA, 2012). Accordingly, unexpected events and impacts caused by sudden inflows of refugees from neighbouring countries, natural disasters, shifting perceptions of political opportunity and risk, changing governments, land use reforms, etc. will all considerably alter the dynamics of LULC change, thus shifting the trajectory of development.

4.3.2 The Modifiable Area Unit Problem (MAUP)

Geographical space is continuous and thus there is not perfect discontinuity on the surface of the earth (Wong, 2008). In geographic modelling a raster surface is usually used to mimic the continuity of the earth's surface, however, in the context of this study, a boundary is used to demarcate a case study area. This represents an analytical issue coined the Modifiable Area Unit Problem (MAUP) and it refers to the fact that these boundaries represent an artificial construct, and thus a spatial aggregation at a smaller or larger scale will inevitably alter the results of the analysis (Wong, 2008). As in all other spatially disaggregated geographical models, the LULC change model developed in this study is subject to the MAUP problem. This means that the same data used in the context of this study would likely yield different results if aggregated in a different way. The boundaries of the case study area for this study was purposefully demarcated to include a small landmass outside of the Virunga NP, in order to reflect processes operating outside the borders of the park, but ultimately affects LULC changes within it. However, the demarcation of the case study area is still an arbitrary construct, and thus if the data had been aggregated in another way, the results may have been vastly different, i.e. if larger infrastructure (i.e. highways, road networks, airports, etc.) and large cities (i.e. the capital of Kinshasa) lying outside of the case study area would have been included. This would have affected the training of the model and the calculation of the LULC changes. Thus, the spatial changes occurring within the case study area is shaped and formed by various external human, environmental and socio-economic processes which cannot be demarcated by artificial boundaries.

4.4 Conclusion

The Virunga catchment in the eastern part of the Democratic Republic of the Congo is subject to dramatic deforestation rates and land grabbing, causing significant changes to the land cover dynamics in one of the most biodiverse regions of Africa. In order to inform conservation actions and management practices to protect the diversity and integrity of the Virunga NP, while developing sustainable land managing policies and socio-economic reforms it is vital to understand the drivers and dynamics of LULC changes.

This study was successfully able to use a combination of cloud processing platforms (Google Earth Engine), GIS software (ArcGIS) and LULC modelling tools (LCM in TerrSet) to simulate future deforestation and land change patterns in the Virunga catchment. It provides a good understanding of the predicted LULC changes, under a status quo scenario, over the next ten years, and thus presents an effective policy support tool for decision makers and administrative bodies aiming to strengthen SDG implementation while preserving park resources.

The LULC model predicted that the largest shift between classes is attributed with the conversion of forest areas into cropland and the overall general trend is a significant increase in cropland with a net gain of more than 2000 km². The increase in cropland is primarily located in the north of the Virunga catchment where a substantial proportion of the remaining forest areas is predicted to be replaced by cropland. The primary drivers of deforestation were identified as elevation, distance to artisanal mines and mining concessions and distance to cropland and cities, distance to roads and distance to water. These drivers all reflect the inherent relationship between accessibility to forested areas and proximity to human activities, which is consistent with literature and consistent with the hypothesis that charcoal production and land clearing for mining, urban expansion and subsistence agriculture are the primary contributors to deforestation within the Virunga NP.

5 Bibliography

- Al-Ahmadi, F., & Hames, A. (2009). Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas, Kingdom of Saudi Arabia مقارنة طرق أربعة لاستخلاص تصنيف الأرضي الغطاء لاستخدامات الأراضي وصور من الأقمار الصناعية الخام. *Journal of King Abdulaziz University-Earth Sciences*, 20(1), 167–191. <https://doi.org/10.4197/ear.20-1.9>
- Allan, R., Förstner, U., Salomons, W., Paegelow, M., & Olmedo, M. T. C. (2008). *Modelling Environmental Dynamics: Advances in Geomatic Solutions*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Alphan, H., Doygun, H., & I Unlukaplan, Y. (2008). Post-classification comparison of land cover using multitemporal Landsat and ASTER imagery: The case of Kahramanmaraş, Turkey. *Environmental Monitoring and Assessment*, 151, 327–336. <https://doi.org/10.1007/s10661-008-0274-x>
- Andersen, F. (2018). Virunga National Park, the heart of darkness as UNESCO World Heritage. *Continents Manuscripts*, (11). <https://doi.org/10.4000/coma.2773>
- Beysolow II, T. (2017). *Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning and Implementing Deep Learning Models Using R* (T. Beysolow II, ed.). <https://doi.org/10.1007/978-1-4842-2734-3>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brownlee, J. (2016). Crash Course On Multi-Layer Perceptron Neural Networks. Retrieved April 3, 2019, from Deep Learning website: <https://machinelearningmastery.com/neural-networks-crash-course/>
- Crawford, A., & Bernstein, J. (2008). *MEAs, Conservation, and conflict - A case study of Virunga National Park, DRC* (Vol. 1). Retrieved from https://www.iisd.org/pdf/2008/meas_cons_conf_virunga.pdf
- Eastman, R. J. (2016a). *TerrSet Geospatial Monitoring and Modeling System - Manual*. Retrieved from www.clarklabs.org
- Eastman, R. J. (2016b). *Terrset Geospatial Monitoring and Modelling System - Tutorial*. CLARKS LABS.
- ESRI. (2019). ArcGIS Desktop. Retrieved April 1, 2019, from <https://www.esri.com/en-us/store/arcgis-desktop>
- 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(7). <https://doi.org/10.1016/j.heliyon.2018.e00693>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- GOOGLE. (2019). Earth Engine Code Editor. Retrieved April 4, 2019, from <https://developers.google.com/earth-engine/playground>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guerrero, G., Masera, O., & Mas, J.-F. (n.d.). Land use / Land cover change dynamics in the Mexican highlands: current situation and long term scenarios. In *Modelling Environmental Dynamics* (pp. 57–76). https://doi.org/10.1007/978-3-540-68498-5_2

- Harris, J. R., & Grunsky, E. C. (2015). Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Computers and Geosciences*, 80, 9–25. <https://doi.org/10.1016/j.cageo.2015.03.013>
- Institute for Environmental Security. (2008). *Mining, forest change and conflict in the Kivus, eastern Democratic Republic of Congo - Outcome of a short study within the IES-ESPA programme*. Retrieved from http://www.envirosecurity.org/espa/PDF/Mining_forest_change_and_conflict_in_the_Kivus.pdf
- Jones, B. (2018). Deforestation surges in Virunga National Park in the wake of violence. *MONGABAY*. Retrieved from <https://news.mongabay.com/2018/10/deforestation-surges-in-virunga-national-park-in-the-wake-of-violence/>
- Kayijamahe, E. (2008). *Spatial modelling of mountain gorilla (Gorilla beringei beringei) habitat suitability and human impact Virunga Volcanoes Mountains, Rwanda, Uganda and Democratic Republic of Congo*. Retrieved from https://webapps.itc.utwente.nl/librarywww/papers_2008/msc/nrm/kayijamahe.pdf
- Kulkarni, A. D., & Lowe, B. (2016). *Random Forest Algorithm for Land Cover Classification International Journal on Recent and Innovation Trends in Computing and Communication Random Forest Algorithm for Land Cover Classification*. Retrieved from http://scholarworks.uttyler.edu/compsci_fach<http://hdl.handle.net/10950/341><http://www.ijritcc.org>
- Kumar, L., & Mutanga, O. (2018). Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. *Remote Sensing*, 10(10), 1509. <https://doi.org/10.3390/rs10101509>
- Liping, C., Yujun, S., & Saeed, S. (2018). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques-A case study of a hilly area, Jiangle, China. *PloS One*, 13(7), e0200493. <https://doi.org/10.1371/journal.pone.0200493>
- Mas, J., Kolb, M., Paegelow, M., Teresa, M., Olmedo, C., & Houet, T. (2014). Environmental Modelling & Software Inductive pattern-based land use / cover change models : A comparison of four software packages. *Environmental Modelling and Software*, 51, 94–111. <https://doi.org/10.1016/j.envsoft.2013.09.010>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- MDN. (2019). JavaScript? Retrieved April 1, 2019, from About JavaScript website: https://developer.mozilla.org/en-US/docs/Web/JavaScript/About_JavaScript
- Meiyappan, P., Dalton, M., Neill, B. C. O., & Jain, A. K. (2014). Spatial modeling of agricultural land use change at global scale. *Ecological Modelling*, 291, 152–174. <https://doi.org/10.1016/j.ecolmodel.2014.07.027>
- Ming, D., Zhou, T., Wang, M., & Tan, T. (2016). Land cover classification using random forest with genetic algorithm-based parameter optimization. *Journal of Applied Remote Sensing*, 10(3), 035021. <https://doi.org/10.1117/1.jrs.10.035021>
- MIRICI, M. E. (2018). LAND USE/COVER CHANGE MODELLING IN A MEDITERRANEAN RURAL LANDSCAPE USING MULTI-LAYER PERCEPTRON AND MARKOV CHAIN (MLP-MC). *Applied Ecology and Environmental Research*, 16(1), 467–486. https://doi.org/10.15666/aeer/1601_467486
- Mishra, V., Rai, P., & Mohan, K. (2014). Prediction of land use changes based on land change modeler (LCM) using remote sensing: A case study of Muzaffarpur (Bihar), India. *Journal of the Geographical Institute Jovan Cvijic, SASAZbornik Radova Geografskog Instituta Jovan Cvijic, SANU*, 64(1), 111–127. <https://doi.org/10.2298/ijgi1401111m>

- National Research Council. (2014). *Advancing Land Change Modeling: Opportunities and Research Requirements*. <https://doi.org/10.17226/18385>
- Noszczyk, T. (2018). Human and Ecological Risk Assessment : An International A review of approaches to land use changes modeling. *HUMAN AND ECOLOGICAL RISK ASSESSMENT*, 0(0), 1–29. <https://doi.org/10.1080/10807039.2018.1468994>
- Paegelow, M., Teresa, M., & Olmedo, C. (2007). *Possibilities and limits of prospective GIS land cover modelling — a compared case study: Garrotxes (France) and Alta Alpujarra Granadina*. 8816. <https://doi.org/10.1080/13658810500076443>
- Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156–168. <https://doi.org/10.1016/j.rse.2016.10.010>
- Pérez-Vega, A., Mas, J.-F., & Ligmann-Zielinska, A. (2012). Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest. *Environmental Modelling & Software*, 29(1), 11–23. <https://doi.org/https://doi.org/10.1016/j.envsoft.2011.09.011>
- Phiri, D., & Morgenroth, J. (2017). *Developments in Landsat Land Cover Classification Methods : A Review*. <https://doi.org/10.3390/rs9090967>
- Rainer, H., Lanjouw, A., Kayitare, A., Rutagarama, E., Sivha, M., Asuma, S., & Kalpers, J. (2001). *Beyond Boundaries: Transboundary Natural Resource Management for Mountain Gorillas in the Virunga-Bwindi Region*.
- Reed, Reed, R. D., & Marks, R. J. (2014). MLP Representational Capabilities. In *Neural Smithing : Supervised Learning in Feedforward Artificial Neural Networks*. Retrieved from https://sfx.aub.aau.dk/sfxaub?ctx_enc=info%3Aofi%2Fenc%3AUTF-8&ctx_tim=2019-04-03T12%3A58%3A23IST&ctx_ver=Z39.88-2004&req.language=dan&rft_id=info%3Aid%2Fprimo.exlibrisgroup.com%3Aprimo3-Article-vlebooks&rft.artnum=&rft.atitle=&rft.au=Reed%2C+Russell&rft
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Sammut, C., & Webb, G. I. (Eds.). (2010). Markov Process. In *Encyclopedia of Machine Learning* (p. 646). https://doi.org/10.1007/978-0-387-30164-8_516
- Shade, C., & Kremer, P. (2019). Predicting Land Use Changes in Philadelphia Following Green Infrastructure Policies. *Land*, 8(2), 28. <https://doi.org/10.3390/land8020028>
- Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification. In *Integrated Series in Information Systems* (Vol. 36). <https://doi.org/10.1007/978-1-4899-7641-3>
- Teresa, M., Olmedo, C., Gilmore, R., Jr, P., & Paegelow, M. (2015). Environmental Modelling & Software Comparison of simulation models in terms of quantity and allocation of land change. *Environmental Modelling and Software*, 69, 214–221. <https://doi.org/10.1016/j.envsoft.2015.03.003>
- UN-DESA. (2012). *Sustainable land use for the 21st century*. Retrieved from <https://sustainabledevelopment.un.org/content/documents/1124landuse.pdf>
- UNEP-WCMC, & IUCN. (2019). Virunga National Park. Retrieved May 10, 2019, from The World Database on

Protected Areas (WDPA) website: <https://www.protectedplanet.net/>

- UNESCO. (2018). Virunga National Park (Democratic Republic of the Congo). Retrieved May 9, 2019, from State of Conservation website: <http://whc.unesco.org/en/soc/3815>
- UNESCO. (2019). Virunga National Park. Retrieved May 9, 2019, from World Heritage List website: <http://whc.unesco.org/en/list/63>
- United Nations. (2013). Democratic Republic of Congo (DRC). Retrieved May 3, 2019, from <http://unhabitat.org/drc/>
- Wondie, M., Teketay, D., Melesse, A., & Schneider, W. (2011). Relationship between Topographic Variables and Land Cover in the Simen Mountains National Park, a World Heritage Site in Northern Ethiopia. *International Journal of Remote Sensing*.
- Wong, D. (2008). The Modifiable Areal Unit Problem (MAUP). In *The SAGE Handbook of Spatial Analysis*. Retrieved from https://books.google.dk/books?id=phEgXfbCU_YC
- World Bank. (2017). *WASH Poor in a Water-Rich Country: A Diagnostic of Water, Sanitation, Hygiene, and Poverty in the Democratic Republic of Congo*. Retrieved from www.worldbank.org/water
- World Resources Institute. (2019). Global Forest Watch. Retrieved May 10, 2019, from www.globalforestwatch.org
- WWF. (2013). *THE ECONOMIC VALUE OF VIRUNGA NATIONAL PARK REPORT*. Retrieved from http://awsassets.panda.org/downloads/the_economic_value_of_virunga_national_park_lr_2.pdf
- Yang, Z. R. (2010). Multi-layer Perceptron. In Z. R. Yang (Ed.), *Machine Learning Approaches to Bioinformatics* (pp. 133–153). https://doi.org/10.1142/9789814287319_0010
- Yee, A. (2017, September 14). In Africa's Oldest Park, Seeking Solutions to a Destructive Charcoal Trade. *YaleEnvironment360*. Retrieved from <https://e360.yale.edu/features/in-africas-oldest-park-seeking-solutions-to-a-destructive-charcoal-trade>
- Zadbagher, E., & Becek, K. (2018). *Modeling land use / land cover change using remote sensing and geographic information systems : case study of the Seyhan*.

6 Appendices

Appendix A.

Source code for the 2019 land cover classification in Google Earth Engine

```
Map.centerObject(AOI, 9);
Map.addLayer(AOI, {}, 'aoi');
//For Landsat surface reflectance product cloud masking
function maskclouds(image) {
  var cloudShadowBitMask = 1 << 3; // cloud shadow
  var cloudsBitMask = 1 << 5; // cloud
  var qa = image.select('pixel_qa');
  var date = image.get('system:time_start');
  var mask = qa.bitwiseAnd(cloudShadowBitMask).eq(0)
    .and(qa.bitwiseAnd(cloudsBitMask).eq(0));
  var ndvi =
image.normalizedDifference(['B5', 'B4']).multiply(10000).rename('NDVI');
  return
image.addBands(ndvi).updateMask(mask).divide(10000).set('system:time_start', date);
}
//Landsat 8 image collection
var L8collection = ee.ImageCollection('LANDSAT/LC08/C01/T1_SR')
  .filterDate('2017-01-01', '2019-03-23')
  .filterBounds(AOI)
  .filter(ee.Filter.lt('CLOUD_COVER', 35))
  .map(maskclouds);
print (L8collection);
var testimage = L8collection.median().clip(AOI);

Map.addLayer(testimage.select(['B5', 'B4', 'B3']), {min:0, max:0.4}, 'false
color');
Map.addLayer(testimage.select(['B4', 'B3', 'B2']), {min:0, max:0.4}, 'true color
color');
// Subsample training polygons with random points
// this ensures all classes have same sample size
// also EE can't handle too many cells at once
var trainingLayers = [forest, water, city, cropland, openland2];
var n = 500;
// loop over training layers
for (var i = 0; i < trainingLayers.length; i++) {
  // sample points within training polygons
  var pts = ee.FeatureCollection
    .randomPoints(trainingLayers[i].geometry(), n);
  // add class
  var thisClass = trainingLayers[i].get('class');
  pts = pts.map(function(f) {
    return f.set({class: thisClass});
  });
  // extract raster cell values
  var training = testimage.sampleRegions(pts, ['class'], 30);
  // combine training regions together
  if (i === 0) {
    var trainingData = training;
  } else {
    trainingData = trainingData.merge(training);
  }
}
```

```

    }
  }
  print (trainingData);

  //// classify with random forests
  // use bands 1-7+NDVI
  var bands = ['B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'NDVI'];
  // fit a random forests model
  var classifier = ee.Classifier.randomForest(500)
    .train(trainingData, 'class', bands);
  // produce the land cover map
  var classified = testimage.classify(classifier);
  var p = ['00ff00', 'ff0000', '000000', '0000ff', 'orange',];
  // display
  Map.addLayer(classified, {palette: p, min: 1, max: 5}, 'classification');

  //Accuracy assessment
  //Test the classifiers' accuracy. (data, y, X)
  var trainingClassifier = classifier.train(training, 'class', bands);
  //Separate validation
  var testingsep =
  forestvali.merge(watervali).merge(cityvali).merge(croplandvali).merge(openlandvali);
  // Add reducer output to the Features in the collection.
  testingsep = testimage.sampleRegions(testingsep, ['class'], 30);
  //print (testingsep)
  var validation_sep = testingsep.classify(trainingClassifier);
  //print (validation_sep)
  var errorMatrix_sep = validation_sep.errorMatrix('class', 'classification');
  //print('Error Matrix:', errorMatrix_sep);
  var ft = ee.FeatureCollection([ee.Feature(null, {'Accuracy':
  errorMatrix_sep.accuracy(), 'Producer
  Accuracy':errorMatrix_sep.producersAccuracy(), 'User
  Accuracy':errorMatrix_sep.consumersAccuracy(), 'Kappa': errorMatrix_sep.kappa(),
  'Error Matrix':errorMatrix_sep.array()})]);

  // Define customization options.
  var options = {
    title: 'Landsat 8',
    hAxis: {title: 'Wavelength (micrometers)'},
    vAxis: {title: 'Reflectance'},
    lineWidth: 1,
    pointSize: 4,
    series: {
      0: {color: '00FF00'}, // forest
      1: {color: '0000FF'}, // water
      2: {color: 'FF0000'}, // city
      3: {color: 'orange'}, // openland1
      4: {color: 'grey'}, // cropland
      5: {color: 'yellow'}, // openland2
    }
  };

  // Define a list of Landsat 8 wavelengths for X-axis labels.
  var wavelengths = [0.44, 0.48, 0.56, 0.65, 0.86, 1.61, 2.2, 2.5];
  // Create the chart and set options.
  var spectraChart = ui.Chart.image.regions(
    testimage.select(bands), trainingLayers, ee.Reducer.mean(), 30, 'class',
    wavelengths)
    .setChartType('ScatterChart')

```

```
.setOptions(options);

// Display the chart.
print(spectraChart);

Export.table.toDrive({collection: ft, description: 'accu_2018',
  fileNamePrefix: 'accu_2018', folder: 'Master thesis', selectors: ['User
Accuracy', 'Producer Accuracy', 'Accuracy', 'Kappa', 'Error Matrix']});
// Export the image, specifying scale and region.
Export.image.toDrive({
  image: classified,
  description: 'VirungaLC_2018',
  scale: 30,
  folder: 'Master thesis',
  region: AOI.geometry().bounds(),
  maxPixels: 2091108075,
  crs: 'EPSG:4051'
});
```

Appendix B. Land cover maps confusion matrixes

Table 11 Confusion matrix and accuracy - 2010 land cover map

Error matrix for 2010 land cover classification							
		Ground truth reference					
		Forest	Water	City	Cropland	open land	Total
Clas/postersified	Forest	48	0	0	2	0	50
	Water	0	50	0	0	0	50
	City	0	0	42	0	8	50
	Cropland	2	0	1	45	2	50
	open land	0	0	1	2	47	50
	Total	50	50	44	49	57	250

Table 12 Confusion matrix and accuracy - 2015 land cover map

Error matrix for 2015 land cover classification							
		Ground truth reference					
		Forest	Water	City	Cropland	open land	Total
Classified image	Forest	50	0	0	0	0	50
	Water	0	50	0	0	0	50
	City	0	0	45	0	5	50
	Cropland	5	0	0	41	4	50
	open land	0	0	1	0	49	50
	Total	55	50	46	41	58	250

Table 13 Confusion matrix and accuracy - 2019 land cover map

Error matrix for 2019 land cover classification							
		Ground truth reference					
		Forest	Water	City	Cropland	open land	Total
Classified image	Forest	43	0	0	7	0	50
	Water	0	50	0	0	0	50
	City	0	0	46	2	2	50
	Cropland	0	0	0	44	6	50
	open land	0	0	1	2	47	50
	Total	43	50	47	55	55	250

Land Change Modeler MLP Model Results

(Created: 04-05-2019 13:39:28)

1. General Model Information

1) Input Files

Independent variable 1	DEM_b1
Independent variable 2	Dist_artmining_new_b1
Independent variable 3	Dist_mining_new_b1
Independent variable 4	Dist_roads_new_b1
Independent variable 5	Dist_distrubance_2010
Independent variable 6	Dist_waterways_new_b1
Training site file	LCM_v2_Train_Deforestation

2) Parameters and Performance

Input layer neurons	6
Hidden layer neurons	6
Output layer neurons	3
Requested samples per class	10000
Final learning rate	0.0001
Momentum factor	0.5
Sigmoid constant	1
Acceptable RMS	0.01
Iterations	10000
Training RMS	0.3358
Testing RMS	0.3369
Accuracy rate	77.61%
Skill measure	0.6642

3) Model Skill Breakdown by Transition & Persistence

Class	Skill measure
Transition : Forest to Cropland	0.6103
Transition : Forest to Openland	0.8300
Persistence : Forest	0.5516

2. Weights Information of Neurons across Layers

1) Weights between Input Layer Neurons and Hidden Layer Neurons

Neuron	h-Neuron 1	h-Neuron 2	h-Neuron 3	h-Neuron 4	h-Neuron 5	h-Neuron 6
i-Neuron 1	4.5626	0.2243	-9.2037	4.2459	-12.5762	-1.0072
i-Neuron 2	2.5363	2.1855	-2.9245	-0.8296	7.8734	-5.3826
i-Neuron 3	-3.3068	-2.6368	2.6611	-0.3019	-2.4268	-5.7239
i-Neuron 4	-1.8815	0.0703	-0.6131	-0.7668	-2.9260	3.8677
i-Neuron 5	-18.1908	16.9675	-18.2358	7.6296	7.2956	1.9399
i-Neuron 6	-8.2517	-1.5462	9.0971	-17.2515	12.6814	-3.0591

2) Weights between Hidden Layer Neurons and Output Layer Neurons

Neuron	o-Neuron 1	o-Neuron 2	o-Neuron 3
h-Neuron 1	12.1323	-9.7654	-13.2859
h-Neuron 2	-10.8397	13.1835	-5.1108
h-Neuron 3	1.4191	8.8686	-7.8621
h-Neuron 4	-6.4247	-9.7666	14.9483
h-Neuron 5	2.7476	-13.9790	7.7220
h-Neuron 6	3.7543	-13.7349	6.9767

3. Sensitivity of Model to Forcing Independent Variables to be Constant

1) Forcing a Single Independent Variable to be Constant

Model	Accuracy (%)	Skill measure	Influence order
With all variables	77.61	0.6642	N/A
Var. 1 constant	68.23	0.5234	2
Var. 2 constant	71.55	0.5733	5
Var. 3 constant	68.71	0.5306	3
Var. 4 constant	75.40	0.6309	6 (least influential)
Var. 5 constant	70.87	0.5631	4
Var. 6 constant	66.12	0.4919	1 (most influential)

✘

2) Forcing All Independent Variables Except One to be Constant

Model	Accuracy (%)	Skill measure
With all variables	77.61	0.6642
All constant but var. 1	37.17	0.0575
All constant but var. 2	30.89	-0.0366
All constant but var. 3	29.84	-0.0524
All constant but var. 4	33.46	0.0019
All constant but var. 5	33.46	0.0019
All constant but var. 6	39.01	0.0852

✘

3) Backwards Stepwise Constant Forcing

Model	Variables included	Accuracy (%)	Skill measure
With all variables	All variables	77.61	0.6642
Step 1: var.[4] constant	[1,2,3,5,6]	75.40	0.6309
Step 2: var.[4,5] constant	[1,2,3,6]	66.42	0.4964
Step 3: var.[4,5,1] constant	[2,3,6]	50.91	0.2636
Step 4: var.[4,5,1,6] constant	[2,3]	42.32	0.1348
Step 5: var.[4,5,1,6,3] constant	[2]	30.89	-0.0366

```

Number of total runs      :      1
Multi-resolution VALIDATE :      Categorical Image Comparison
=====
Comparison image file    :      LC_2019_b1_new
Reference image file     :      Prediction_2030_stage_4
Strata/Mask image file   :      N/A

Number of valid strata:   1; Number of valid categories:   6

//Beginning of run:      1|
Resolution scale:       1 x 1

```

Classification agreement/disagreement
According to ability to specify accurately quantity and allocation

Information of Allocation	Information of Quantity		
	No[n]	Medium[m]	Perfect[p]
Perfect[P(x)]	P(n) = 0.5098	P(m) = 0.9894	P(p) = 1.0000
PerfectStratum[K(x)]	K(n) = 0.5098	K(m) = 0.9894	K(p) = 1.0000
MediumGrid[M(x)]	M(n) = 0.4738	M(m) = 0.9354	M(p) = 0.9299
MediumStratum[H(x)]	H(n) = 0.1667	H(m) = 0.4483	H(p) = 0.4492
No[N(x)]	N(n) = 0.1667	N(m) = 0.4483	N(p) = 0.4492

```

AgreementChance = 0.1667
AgreementQuantity = 0.2817
AgreementStrata = 0.0000
AgreementGridcell = 0.4870
DisagreeGridcell = 0.0541
DisagreeStrata = 0.0000
DisagreeQuantity = 0.0106

```

```

      Kno = 0.9224
      Klocation = 0.9001
      KlocationStrata = 0.9001
      Kstandard = 0.8828

```

```
//Ending of run:      1
```