ABSTRACT

Title of Dissertation: MULTISCALE, MULTITEMPORAL ASSESSMENT OF CHIMPANZEE (*Pan troglodytes)* HABITAT USING REMOTELY SENSED DATASETS

Samuel M. Jantz, Doctor of Philosophy, 2023

Dissertation directed by: Professor Matthew C. Hansen, Department of Geographical Sciences

All four sub-species of our closest living relative, the chimpanzee, are listed as endangered by the International Union for the Conservation of Nature (IUCN), and their populations continue to decline due to human activities. Effective conservation efforts require information on their population status and distribution. Traditional field surveys are expensive and impractical for covering large areas at regular time intervals, making it difficult to track population trends. Given that chimpanzees occupy a large range (2.3×10^6 km^2), new cost-effective methods and data are needed to provide relevant information on population status and trends across large geographic and time scales. The objective of this dissertation is to help fill this gap by leveraging freely available and regularly updated remotely sensed datasets to map and monitor chimpanzee habitat across their range. This research begins by first producing annual forest cover and change maps for the Greater Gombe (GGE) and Greater Mahale ecosystems (GME) in western Tanzania

using Landsat phenological metrics and machine learning methods. Canopy cover was predicted at 30-meter resolution and the Cumulative Sums (CuSum) algorithm was applied to the canopy cover time series to detect forest loss and gain events between 2000-2020. An accuracy assessment showed the CuSum algorithm was able to detect forest loss well but had more difficulty detecting gradual forest gain events. A total of 276,000 ha (+/- 27,000 ha) of gross forest loss was detected between 2000 and 2020 in the GGE and GME; however, loss was not spread equally among the two ecosystems. The results show widespread forest loss in the GME, contrasted with net forest cover gain in the GGE. Next, the annual forest cover maps, and additional derived variables, were used to train an ensemble model to predict the relative encounter rate of chimpanzee nest sightings in the GGE and GME. Model output exhibited a strong linear relationship to chimpanzee abundances and population density estimated from a recent ground survey, enabling model output to be linearly transformed into population estimates. The model predicted the two ecosystems harbor just over 3,000 individuals, which agrees with the upper limit of population estimates from ground surveys. Most importantly, the model can be applied to annually updated variables enabling the detection of potential population shifts caused by changes in landscape condition. Model output indicates a possible population reduction in portions of the GME, while the GGE is predicted to have increased its ability to sustain a larger population. Finally, Random Forests regression was used to relate predictor variables, primarily derived from Landsat data to a coarse resolution, range-wide habitat suitability map enabling the prediction of habitat suitability at 30 meter resolution. The model showed good agreement with the calibration data; however, there was considerable variation in predictive capability among the four chimpanzee sub-species. Elevation, Landsat ETM+ band 5 and Landsat derived canopy cover were the strongest predictors; highly suitable areas were

associated with dense tree canopy cover for all but the Nigeria-Cameroon and Central Chimpanzee sub-species. The model can detect changes in suitability to support monitoring and conservation planning across the chimpanzee range. Results from this dissertation highlight the promise of integrating continuously updated satellite data into habitat suitability models to detect changes through time and inform conservation efforts for chimpanzees at multiple scales.

MULTISCALE, MULTITEMPORAL ASSESSMENT OF CHIMPANZEE (*Pan troglodytes)* HABITAT USING REMOTELY SENSED DATASETS

by

Samuel M. Jantz

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Dr. Matthew C. Hansen, Chair
Dr. Christopher O. Justice
Dr. Peter V. Potapov
Dr. Matthew C. Fitzpatrick
Dr. Joseph Sullivan, Dean's Representative

Foreword

Materials from Chapters 2 and 3 are both in preparation for submission to peer-reviewed journals. Material from Chapter 4 has been published as a peer-reviewed journal article. All external contributions are identified with citations or references. All other methods, analyses, and results were developed and performed by Samuel Jantz, and all the text in this dissertation was written by Samuel Jantz. The following article has appeared in the literature:

Jantz, Samuel M, Lilian Pintea, Janet Nackoney, and Matthew C Hansen. 2016. “Landsat ETM+ and SRTM Data Provide near Real-Time Monitoring of Chimpanzee (Pan Troglodytes) Habitats in Africa.” *Remote Sensing* 8 (5): 427.

During my Ph.D. studies, I also participated in several projects that were relevant to the research topic of this dissertation. The following co-authored works have been published in peer-reviewed journals or book chapters.

Pintea, Lilian, Samuel M. Jantz, and Serge A. Wich. 2021. "From the Cloud to the Ground: Converting Satellite Data into Conservation Decisions." In: *Conservation Technology*, edited by Serge A. Wich and Alex K. Piel. Oxford, U.K.: Oxford University Press.

Bonnin, Noemie, Fiona A Stewart, Serge A Wich, Lilian Pintea, Samuel M Jantz, Rebecca Dickson, Joe Bellis, et al. 2020. “Modelling Landscape Connectivity Change for Chimpanzee Conservation in Tanzania” *Biological Conservation* 252.

Jantz, Samuel M., Lilian Pintea, Janet Nackoney, and Matthew C. Hansen. 2018. “Global Forest Maps in Support of Conservation Monitoring.” In *Satellite Remote Sensing for Conservation Action*, edited by Allison K. Leidner and Graeme M. Buchanan, 1st ed., 82–118. Cambridge University Press.

Jantz, Samuel M, Brian Barker, Thomas M. Brooks, Louise P. Chini, Qiongyu Huang, Rachel M. Moore, Jacob Noel, and George C. Hurtt. 2015. “Future Habitat Loss and Extinctions Driven by Land-Use Change in Biodiversity Hotspots under Four Scenarios of Climate-Change Mitigation.” *Conservation Biology* 29(4), 1122–1131.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background and motivation

The dramatic decline in biodiversity has become so great that it is considered an important facet of global change by itself (Walker & Steffen 1996). Biodiversity loss has multiple causes, with habitat destruction via land-cover and land-use change as the predominant driver (Purvis et al. 2000; Baillie et al. 2004; Pimm et al. 2014). Most future land-use change will likely occur in tropical regions (Sala et al. 2000), which are home to almost half of all described species (Wright 2005), including humans' closest living relatives, the chimpanzee (*Pan troglodytes*). It has been estimated that more than 70% of chimpanzee tropical forest habitats are now threatened by infrastructure development and land use change (Nellemann & Newton 2002) and has resulted in a large contraction of their range (Fig. 1.1). All four chimpanzee sub-species (Western [*P.t.versus*], Nigeria-Cameroon [*P.t. ellioti*], Central [*P.t.troglodytes*], and Eastern [*P.t. schweinfurthii*]) are classified as endangered on the International Union for the Conservation of Nature (IUCN) Red List. In fact, it has been estimated that their cumulative population has declined by more than 66% over the past 40 years (Kormos et al. 2003), and there is a decreasing trend in population for all four sub-species (Oates et al. 2008; Tutin et al. 2008; Wilson et al. 2008; Humle et al. 2016). Major threats to current populations include habitat loss from resource extraction activities and land conversion, hunting, disease, and the illegal pet trade.

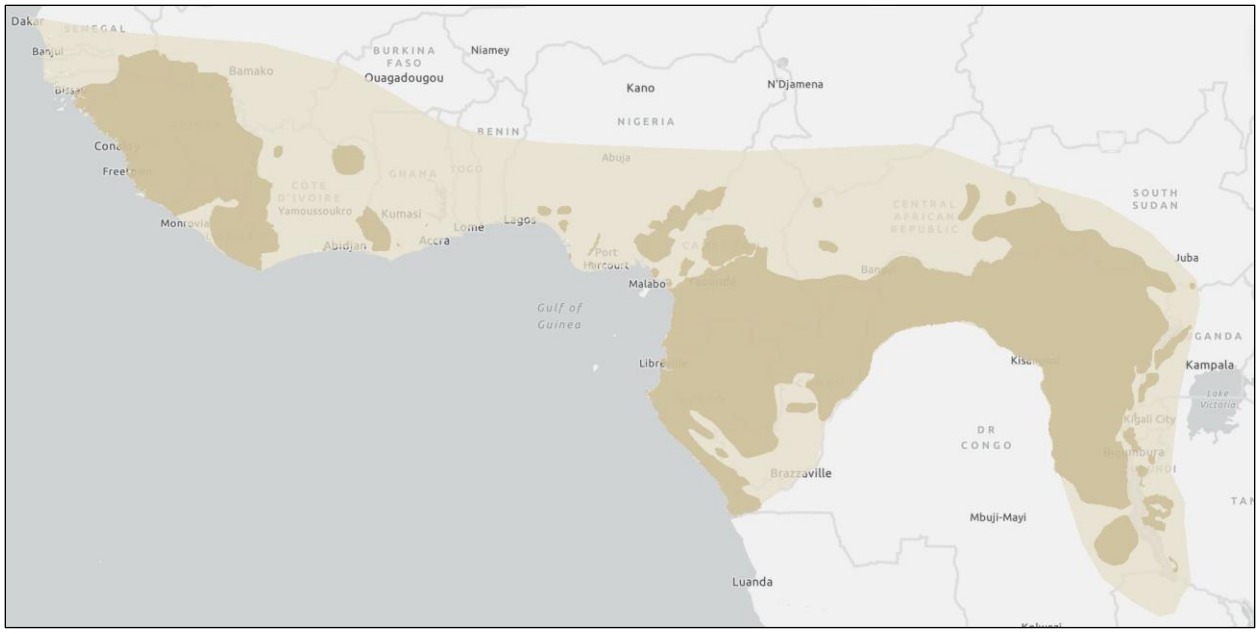


Figure 1.1 Current chimpanzee range (dark brown), and historical range (light brown). Figure credit Dr. Lilian Pintea the Jane Goodall Institute.

1.1.1 Forest Habitat Monitoring with Remote Sensing Data

Since the early 2000's, regularly updated global forest cover products have been available at a medium to low spatial resolution (≥250-m; Hansen et al. (2002)). However, finer-scale, diffuse forest disturbances that are not detectable with low resolution sensors (Hansen et al. 2008), such as conversion to scattered settlements and farms, can have severe negative consequences for many species, particularly frugivorous species (Johns & Skorupa 1987; Harcourt 1998), such as chimpanzees. Moreover, deforestation tends to force species to inhabit scattered, remnant forest patches that are embedded in a matrix of human land-uses (Haddad et al. 2015). The preservation of remnant habitat patches is a current paradigm in conservation biology (Pimm & Brooks 2013), as well as the preservation or establishment of habitat corridors to link these patches enabling the flow of individuals and genes between sub-populations (Bennett 2003). It is therefore paramount to accurately characterize habitat patches and their corridors across landscapes, which is a task that is not suitable for sensors with a coarse spatial

resolution (Benson & MacKenzie 1995; Saura 2004; Anteau et al. 2014). A sensor with a medium to high spatial resolution (<50 m) is required to detect such fine scale forest loss and land-scape heterogeneity.

For over four decades, the Landsat series of satellites has been collecting medium resolution imagery of the Earth's surface. During the majority of this time and for most of the Landsat bands, imagery has been collected at 30-m resolution. While Landsat imagery have existed since 1972, regional to global mapping using Landsat data was not possible. However, three recent advances transformed the way the scientific community is able to use and process Landsat data to develop a global Landsat-derived forest extent and change data record: (i) systematic global acquisitions, (ii) free and open access to the Landsat archive, and (iii) cloud computing and data mining algorithms. On (i) systematic global acquisitions, the Landsat 7 Enhanced Thematic Mapper Plus sensor, launched in 1999, was the first to implement a systematic global acquisition strategy. Prior to Landsat 7, images were collected inconsistently, limiting the ability to apply a standard algorithm to map global forest cover. Considering (ii) free and open access to the Landsat archive, prior to 2008, user acquisition of Landsat imagery required individual purchase, precluding continental to global-scale analyses. Scientists were limited to the data they could afford rather than the data they needed to conduct land change analyses. In 2008, the USGS formally made the Landsat archive open for free access and down-load to all users (Woodcock et al. 2008). Finally, (iii) cloud computing and data mining algorithms made it possible to process hundreds of thousands of Landsat scenes covering multiple time periods across the globe, as on an individual basis, a single Landsat image required significant computing capability to process. These advances culminated in the release of a global, annually updated forest loss product at 30 m resolution (Hansen et al. 2013). Forest loss

detected by this product refers to the removal of trees within a 30 m pixel, and information about the nature of the removal, whether by natural processes or by human activities, is not conveyed. Therefore, the product does not show deforestation per se, rather it indicates that a forest loss event was detected, and further analysis is needed to determine if the event was a conversion to some other land-use (i.e., deforestation).

The application of global forest cover and change products in the context of forest habitat monitoring has been demonstrated at regional and global scales. Tracewski et al. (2016) were the first to use the Landsat-based forest cover and loss product developed by Hansen et al. (2013) to conduct a global analysis of forest loss in Important Bird and Biodiversity Areas (IBAs) using Google Earth Engine. They found that approximately 2.5% of forest was lost across IBAs between 2000 and 2012, with the highest forest loss rates found in South America and Southeast Asia. Joshi et al. (2016) provided another application of the Hansen et al. (2013) global forest cover product that assessed forest loss within 76 landscapes that were prioritized for tiger conservation spread across 13 Asian countries. They found that within these landscapes, 7.7% of forest was lost between 2001 and 2014, much less than was anticipated, but that extensive loss occurred due to the expansion of oil palm plantations. Results of the analysis pointed to the value of satellite-based monitoring of habitat coupled with ground-based observations of tiger occurrence for prioritizing the protection of key habitats and corridors. In fact, Indian government officials from states that have tigers are committing to use the Global Forest Watch platform to monitor performance on achieving 2020 targets promoted by the Global Tiger Initiative (Joshi et al. 2016).

Jantz et al. (2018) presented a similar analysis for chimpanzees across their range. They found that over 80,000 km^2 of tree cover was lost across the entire chimpanzee range between

2001 and 2014, with nearly 9,000 km^2 (11 per cent) within protected areas. The Eastern and Western sub-species accounted for most of the forest loss, with over 34,000 km^2 and 32,000 km^2 of forest loss, respectively. They found the forest loss for the Western sub-species to be particularly troublesome with forest disappearing at a faster rate within protected areas versus outside protected areas. In Côte d'Ivoire, one-third of forest loss was inside protected area boundaries, likely a strong driver of the 90 percent reduction in chimpanzee population within Côte d'Ivoire as found by Campbell et al. (2008). Moreover, some protected areas in the country had a nearly complete removal of forest; Campbell et al. (2008) estimated a 93 per cent decline in forest cover in Marahoué National Park in Côte d'Ivoire between 2002 and 2006, which was once thought to support over 900 chimpanzees. These trends paint a bleak picture for the future of the Western sub-species since it was thought that about half of its total population was distributed within Côte d'Ivoire. Guinea and Liberia are now believed to be the remaining population strongholds for this sub-species (Kormos et al. 2003; Caldecott & Miles 2005).

1.1.2 Species Distribution Models

The Open Standards for Conservation Planning (OSCP) provides a framework to guide the development and implementation of conservation projects (The Conservation Measures Partnership 2013). A key component outlined in the OSCP is the establishment of a monitoring plan that requires indicators that are measurable, precise, consistent, and sensitive (ibid). Species Distribution Models (SDMs), also referred to as ecological niche or habitat suitability models, are useful tools that can fit easily within this framework and are able to provide much more information than just accounting for forest loss. SDMs use statistical and machine learning techniques to find empirical relationships between observed species' occurrences and environmental descriptors, such as resource, biotic and climatic factors, to estimate conditions

that are suitable for population viability (Araújo & Peterson 2012), or to predict the likelihood of occurrence (Franklin 1995; Guisan et al. 2000) in geographic space. Results from SDMs are most convincing when they are fit with both presence and absence data. However, due to the difficulty of acquiring ecologically meaningful absence data, SDMs are often constructed with so called "presence only" data (Fithian & Hastie 2013). When models are fit with presence only data, the output can be interpreted as a relative measure of environmental suitability or relative occurrence rate rather than the probability of presence (Fitzpatrick et al. 2013; Merow et al. 2013), typically with a range between 0 (unsuitable) and 1 (highly suitable).

The use of SDMs to model and map chimpanzee habitat is a recent endeavor and has been used to answer questions applicable to chimpanzee conservation and ecology. Vegetation layers derived from Landsat ETM+ data were used in (Pintea & Plumptre 2006) to model potential habitat suitability for chimpanzees in Western Tanzania at 90 meter resolution. Landsat TM derived variables were used to generate habitat distribution maps for three time periods between 1986 and 2003 for a small area in Guinea-Bissau, within the range of *P.t.verus*, where a marked decrease in habitat area and connectivity was found (Torres et al. 2010). Habitat suitability was mapped for the *P.t. schweinfurthii* range at 10 km resolution using climate, human impact, and vegetation variables to aid in the in delineation of areas for new population survey efforts (Plumptre et al. 2010). SDMs were used for *P.t.ellioti* in Nigeria and Cameroon and *P.t.troglodytes* in Cameroon using climate, topographic, human population density and canopy cover variables at 1 km resolution to investigate the role of environmental variation in the genetic dissimilarity between the two sub-species and to assess potential future impacts of climate change on their habitat suitability (Clee et al. 2015). The first range-wide study was performed by (Junker et al. 2012), who used climate, human impact and vegetation variables as

model inputs to map habitat suitability for all four chimpanzee sub-species, and all African Great Apes, at 5 km resolution for both the 1990s and 2000s and found a decrease in suitable habitat over time.

1.3 Objectives and Dissertation Structure

The over-arching goal of this dissertation is to provide conservation managers with relevant and timely information on the status of chimpanzees and their habitats. This work was done in cooperation with my advisor, GLAD colleagues, and collaborators from the Jane Goodall Institute. This work begins by focusing on the Greater Gombe and Greater Mahale Ecosystems of western Tanzania, where the Jane Goodall Institute generously shared their database of georeferenced chimpanzee nest locations, and then scales up to entire chimpanzee range. Three main objectives were identified to meet the needs of chimpanzee habitat monitoring applications.

- *Objective 1:* Create maps of annual forest cover and change at 30 m resolution for the Greater Gombe and Greater Mahale Ecosystems and perform accuracy assessments for both fractional cover and change between 2000 and 2020.
- *Objective 2:* Incorporate the products from objective 1 and georeferenced nest location data to model the relative occurrence rate of nest sightings as a function of variables derived from remote sensing data in the GME and GGE.
- *Objective 3:* Use machine learning methods to relate predictor variables derived from higher temporal and spatial resolution remotely sensed datasets to a coarse resolution chimpanzee habitat suitability map to enable the prediction of chimpanzee habitat suitability across its range at 30 m resolution.

The dissertation is organized such that each of the described objectives are addressed serially in subsequent chapters. Objective 1 is addressed in Chapter 2, Objective 2 is addressed in Chapter 3, and Objective 3 is addressed in Chapter 4.

In Chapter 2, annual forest cover and change products are produced for the GGE and GME by rescaling the global 250 m resolution MODIS VCF product to 30 m resolution with Random Forests regression and annual, Landsat based multi-temporal metrics as input. The forest maps were subsequently classified into Forest, Woodland, and Non-forest and the top of canopy relative height metric (rh100) from Global Ecosystem Dynamics Investigation (GEDI) LiDAR were used to test if the classes had significantly different height distributions. The Cumulative Sums (CuSum) algorithm was then applied to the annual canopy cover maps to detect significant loss or gain between 2000 and 2020. A manual assessment of 400 pixels was used to determine the accuracy for mapping gross forest loss between 2000 and 2020, while a 450 pixel sample was used to validate mapping forest gain.

In Chapter 3, products from Chapter 2 and geo-located nests sampled from the ground are incorporated into an Inhomogeneous Poisson Process (IPP) model to annually map the relative occurrence rate of nest sightings conditioned on environmental variables derived from remote sensing data. The output from the IPP model is then related to recent estimates of chimpanzee population abundance estimates from line transect surveys to enable the prediction of potential population shifts because of changes in landscape condition.

In Chapter 4, a coarse resolution chimpanzee habitat suitability map is downscaled to 30 m resolution, with predictor variables primarily derived from Landsat data. Random Forests was used to relate the predictor variables to the spatial distribution of habitat suitability at a coarse

scale, and then to predict habitat suitability across the chimpanzee range at 30 m resolution for different time periods.

Chapter 2 Mapping forest cover and change in the Greater Gombe and Greater Mahale ecosystems of western Tanzania

2.1 Introduction

Tanzania represents the eastern and southern edge of the chimpanzee distribution and is home to two of the longest sites of chimpanzee behavioral study: Gombe National Park (GNP) where Dr. Jane Goodall pioneered chimpanzee behavioral research in 1960 (Goodall 1986), and Mahale Mountains National Park (MMNP), where Dr. Toshisada Nishida established a permanent field station in 1965 (Nishida et al. 1983). Ground breaking research conducted at these sites helped spur the creation of the national parks that now protect Tanzanian chimpanzees and their habitats (Moyer et al. 2006; Pusey et al. 2007).

The GGE and GME are situated within the Albertine Rift montane forest and Central Zambezian Woodland ecoregions (Olson et al. 2001), and are part of the Eastern Afromontane biodiversity hotspot (Mittermeier et al. 2005). These are mosaic ecosystems with lush forests of the Congo Basin facing Lake Tanganyika in Gombe National Park (GNP) and Mahale Mountains Nation Park (MMNP). Moving east from the lake, dense forest transitions to open, drier, savanna mosaic Zambezian woodlands of East Africa, interspersed with riparian and gallery forest (TAWIRI 2018). Miombo-woodland dominated landscapes are generally species poor on which chimpanzees feed, but may be a good source of fruit and seeds during the dry season (May-October) (Moyer et al. 2006). As such, chimpanzees in these landscapes will utilize woodland for nesting, particularly on steep slopes, but always within 2 km of evergreen forest (Pintea & Plumptre 2006). The relative scarcity of food in these ecosystems results in low population densities and some of the largest home ranges among all chimpanzee populations, with an average home range size in this region of approximately 31km^2 (Bonnin et al. 2020).

Human populations place immense stress on these two ecosystems. Indeed, GNP is a small 34 km^2 strip surrounded by settlements (Fig 2.1). An analysis of satellite imagery revealed that 64% of woodland and forest was cleared or converted to farmland between 1972 to 2003 (Pintea et al. 2012). Beginning in 2005, the Jane Goodall Institute (JGI) partnered with the Tanzanian government and community stakeholders to create village land use plans with the goal of setting aside land that would allow the restoration of ecosystem services around GNP, and by 2009 the surrounding villages had assigned 26% of their village area (97km^2) to Village Forest Reserves (VFR) (Pintea et al. 2016).

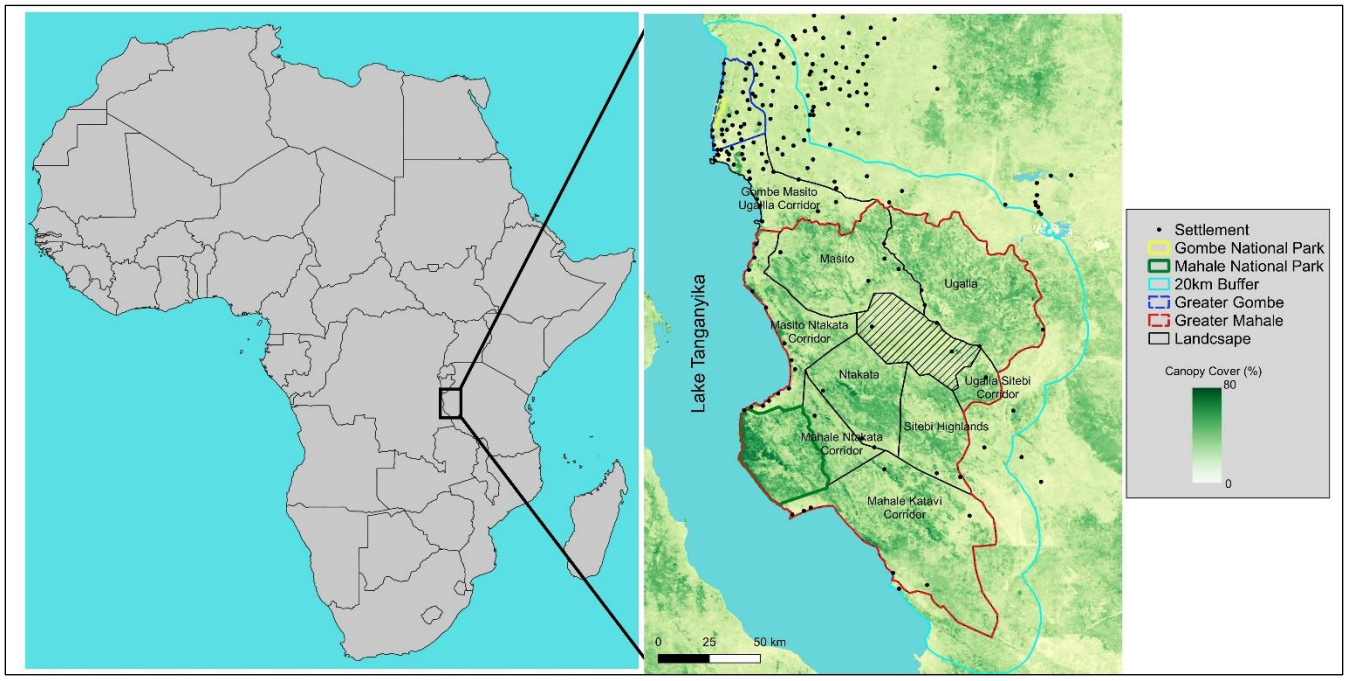


Figure 2.1 Overview of Greater Gombe and Greater Mahale Ecosystems and major landscapes. Fractional Canopy Cover from MODIS VCF. Mishamo, a semi-permanent refugee camp, is the hashed region.

In 2015, the village land-use planning process was brought to the GME (Pintea 2016). While the human population is much more dispersed in the GME, there is still tremendous pressure on chimpanzee habitat. Chimpanzees depend on riparian forests in this ecosystem for a

year round food source; however, these forests are also the preferred sites for agricultural conversion because of comparably fertile soils (Moyer et al. 2006).

Due to the reliance on forests and woodlands in these two ecosystems, an important step for conservation monitoring is to have an accurate characterization of forest cover and change in this region. In this chapter, annual forest canopy cover maps are produced at 30 m resolution by rescaling a global, coarser resolution product, then the Cumulative Sums (CuSum) algorithm is applied to detect annual forest loss and gain. The results feed directly into the next chapter for chimpanzee habitat modeling in these two ecosystems.

2.3 Data

2.3.1 Canopy cover data

Canopy cover training data are derived from the MODerate-resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF) Tree Cover dataset (MOD44B), currently in Version 6, which has a spatial resolution of 250 meters and is produced on an annual basis from 2000 onward (Dimiceli et al. 2015). MODIS VCF data have been used for several studies including quantifying forest loss in and around tropical protected areas (Defries et al. 2005), climate modelling (Lawrence & Chase 2007), quantifying global gross forest cover loss (Hansen et al. 2010) and mapping carbon emissions from tropical deforestation (Harris et al. 2012). MODIS VCF data for the period 2000-2019 for the study area were acquired from the NASA Land Processes Distributed Active Archive Center (LPDAAC) using the Application for Extracting and Exploring Analysis Ready Samples (AppEEARS Team, 2020).

2.3.2 Multi-temporal Landsat metrics

Landsat Analysis Ready Data produced by the Global Land Analysis and Discovery team at the University of Maryland (GLAD ARD) were used to create a set of annual metrics for the study area from 2000 to 2020. The GLAD ARD are 16-day composites of tiled Landsat (TM, ETM+ and OLI sensors) normalized surface reflectance from 1997 to the present updated annually and designed for land cover monitoring at global to local scales (Potapov et al. 2020). In addition to a layer describing per pixel quality, each 1˚ by 1˚ tile contains normalized surface reflectance for the blue, green, red, near-infrared (NIR), shortwave infrared band 1 (SWIR1), shortwave infrared band 2 (SWIR2) and surface brightness temperature (LST). For several reasons, including heterogeneous cloud cover and differing acquisition strategies among the constellation of Landsat satellites, it is not feasible to use the 16-day composites directly. The metrics approach increases spatial and temporal consistency by building an annual time-series of the highest quality observations and subsequently computing a set of metrics that represent salient phenological features across a landscape. A set of 456 annual phenological metrics from the GLAD ARD tiles were created as is detailed in Potapov et al. 2020 and shown in figure 2.2. The tiled Landsat data and software to create the annual phenological metrics are freely available and can be downloaded at https://glad.umd.edu/ard/home. The dataset was supplemented with 1 arc-second elevation data from Shuttle Radar Topography Mission (SRTM) (http://earthexplorer.usgs.gov), which was resampled with bilinear interpolation to match the resolution of the Landsat ARD data. Slope and aspect were then derived from the SRTM elevation data for a total of 459 features.

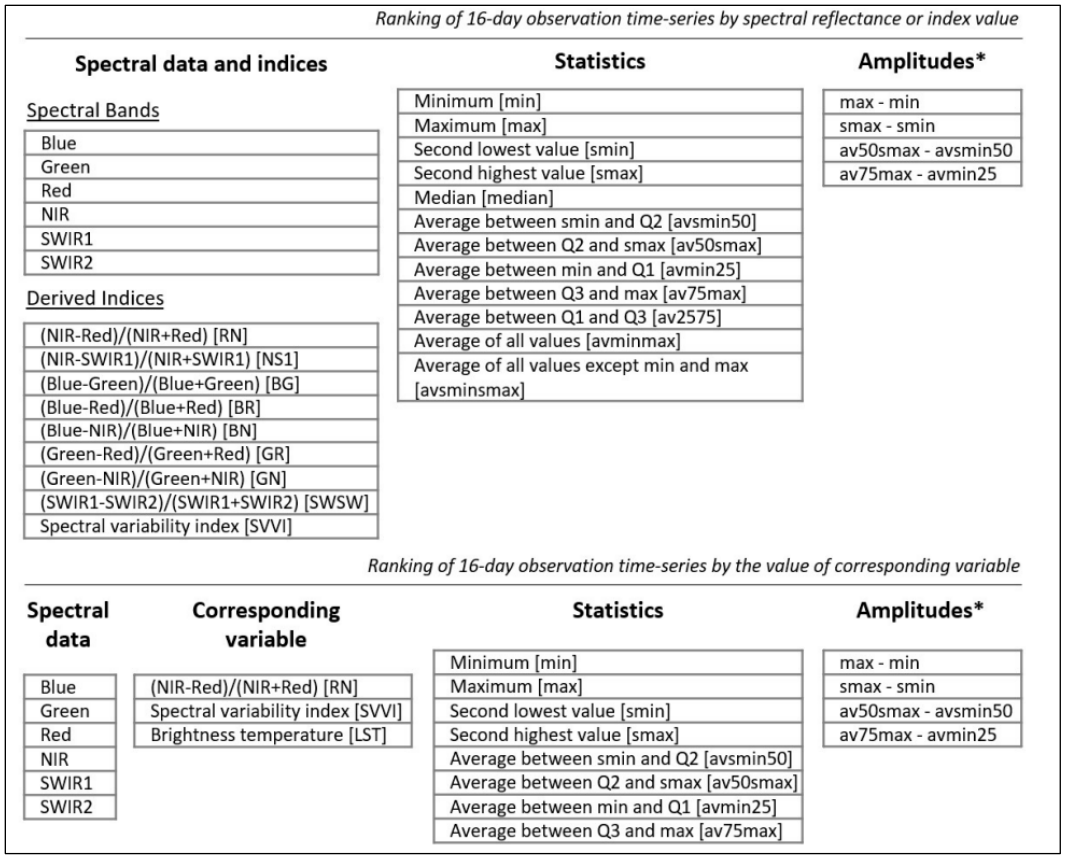


Figure 2.2 Annual metrics calculated from 16-day Landsat composites. Figure is from Potapov et al. 2020.

2.4 Methods

2.4.1 Canopy cover estimation

The 459 features were aggregated to MODIS resolution by taking the average value for all pixels within a 250 m MODIS grid-cell. This process was applied once to the SRTM based features, and annually for the Landsat based features between 2000 and 2019. Samples from all years were pooled together to create a single dataset, and the data were then randomly partitioned into five folds for cross-validation; then a Random Forest regression model was trained on each fold. Random Forests is an ensemble technique in the Classification and Regression Trees (CART) family of machine learning methods (Breiman et al. 1984). The algorithm constructs

several regression trees based on bootstrapped samples of the input dataset and at each node a random subset of features is selected, and a split is made using the feature that results in greatest reduction in mean squared error and each tree is grown to full size. Overfitting in Random Forests is avoided by averaging the predictions produced by individual trees. Random Forests have the attractive properties of being insensitive to non-normally distributed variables, non-linear relationships, and correlated predictor variables (Breiman 2001). In addition, only two primary parameters are required for Random Forest models: the number of trees constructed in each forest and the number of variables considered for splitting at each node. For each fold, 50 trees were grown, and 80% of features were considered at each node split. All 250 trees were used to predict canopy cover at Landsat 30 m resolution. The general workflow is illustrated in figure 2.3.

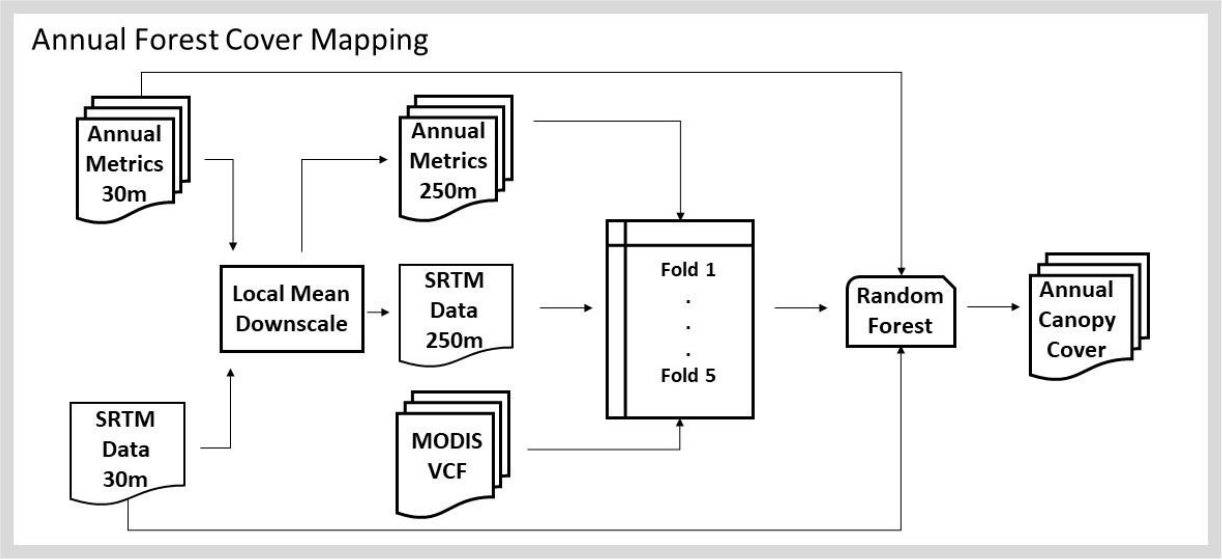


Figure 2.3 Annual canopy cover mapping workflow.

The procedure implemented in this study implicitly assumes the relationship between canopy cover, spectral, and topographic data are invariant to changes in scale, which is supported by previous studies showing that remotely sensed data are robust to changes in scale (Gao et al.

2006; Hilker et al. 2009; Feng et al. 2013). Sexton et al. (2013) produced global canopy cover maps at Landsat resolution derived from MODIS VCF data and found that the Landsat predictions could achieve higher accuracy relative to the VCF data from which they were derived. The approach in this chapter is similar but differs in the modeling algorithm, random forest rather than linear trees, and uses annual metrics rather than individual Landsat scenes.

2.4.1.1 Canopy Cover Assessment

The initial idea was to use independent reference canopy cover data from the Global Ecosystem Dynamics Investigation (GEDI) LiDAR attached to the International Space Station (ISS). GEDI is a full waveform Lidar specifically designed for measuring vegetation canopy structure and data from this mission have been used to create a suite of vegetation and aboveground biomass products (Dubayah et al. 2020). However, a preliminary analysis indicated that fractional canopy cover data from GEDI are not suitable for validation purposes. GEDI fractional cover showed extremely low agreement with the MODIS VCF data and the Landsat based 30 m estimates of cover (Fig. 2.4). The coefficient of determination (r^2) between GEDI canopy cover and MODIS canopy cover was 0.19 and only slightly higher for Landsat based canopy cover at 0.24.

The observed low correspondence with GEDI may be linked to the scale of the GEDI sensor. The GEDI footprint is 25 m, larger than most tree canopies, which may prevent accurate canopy cover characterization. Another reason may be due to the measurement of different quantities. Canopy cover from GEDI is defined as the vertical projection of all canopy material (e.g. leaves, stems etc.) onto the ground, as such, small within canopy openings are factored into the estimate (Tang & Armston 2019). On the other hand, the MODIS VCF product is the proportion of a pixel covered by trees (Hansen et al. 2002); however, Sexton et al. 2013 showed

much better correspondence with higher resolution airborne LiDAR, which suggests disagreement is linked to GEDI's relatively large footprint.

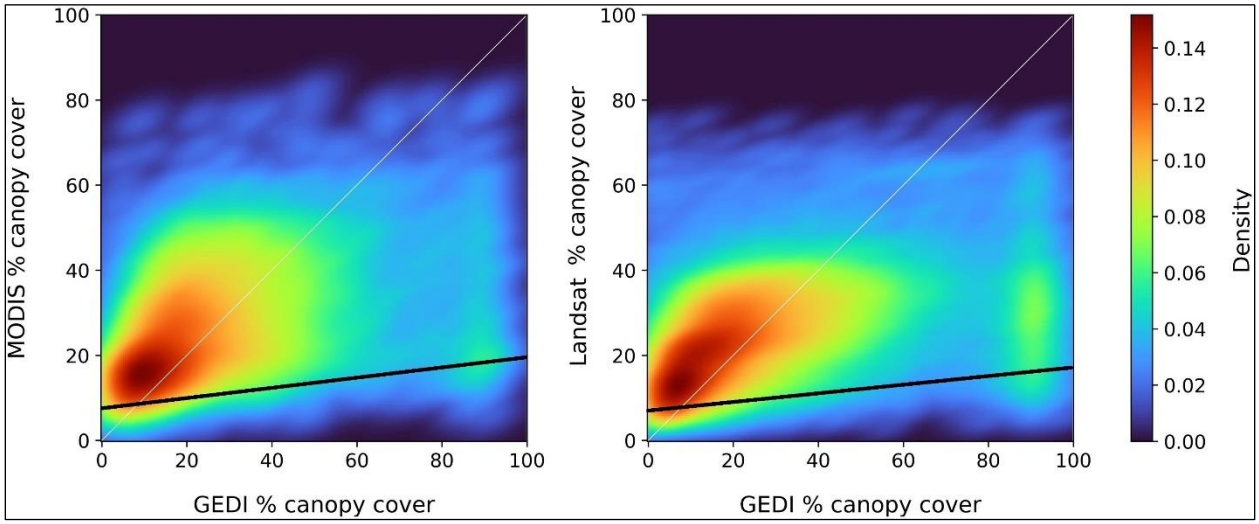


Figure 2.4 Density plots of MODIS VCF and Landsat based estimates of fractional canopy cover and GEDI estimates of fractional canopy cover for 2019. White line indicates the one-to-one line, while the black line shows the estimated linear regression.

In lieu of using GEDI fractional cover as a validation source, Landsat based canopy cover from 2020 was evaluated by creating a discretized forest type map, where forest was defined as cover greater than or equal to 50%, woodland with cover less than 50% and greater than or equal to 10%, and non-forest defined as cover less than 10%. The definition of forest in this study includes the dense canopy forests facing Lake Tanganyika in GNP and MMNP, as well as the thin strips of gallery forest interspersed within woodland. Forest types were compared using the 100^{th} (top of canopy) percentile relative height metric from GEDI L2B for the year 2020. A 2-sample Kolmogorov-Smirnov statistic was used to test for significant differences in height distributions, with the hypothesis that forest will be taller than woodland, which will be taller than non-forest. These hypotheses are based on known characteristics of the tree species in the region. For example, woodland is dominated by *Brachystegia* and *Julbernardia* species with *Brachystegia spicformis* being the first and *Julibernardia globiflora* being the second most

common species (NAFORMA 2015) and are typically between 15-20 meters tall at maturity (Orwa et al. 2009; WFO 2023). Gallery forests and the forests on the western slopes of GNP and MMNP are much more species rich, but the most visible constituents are up to 30 m tall with some *Ficus* species up to 50 m tall (Beentje & Mbago 2007). The non-forest class should be composed of non-tree vegetation, less than 5 meters, as well as bare ground. The distribution of GEDI LiDAR shots used for evaluation can be seen in Figure 2.5.

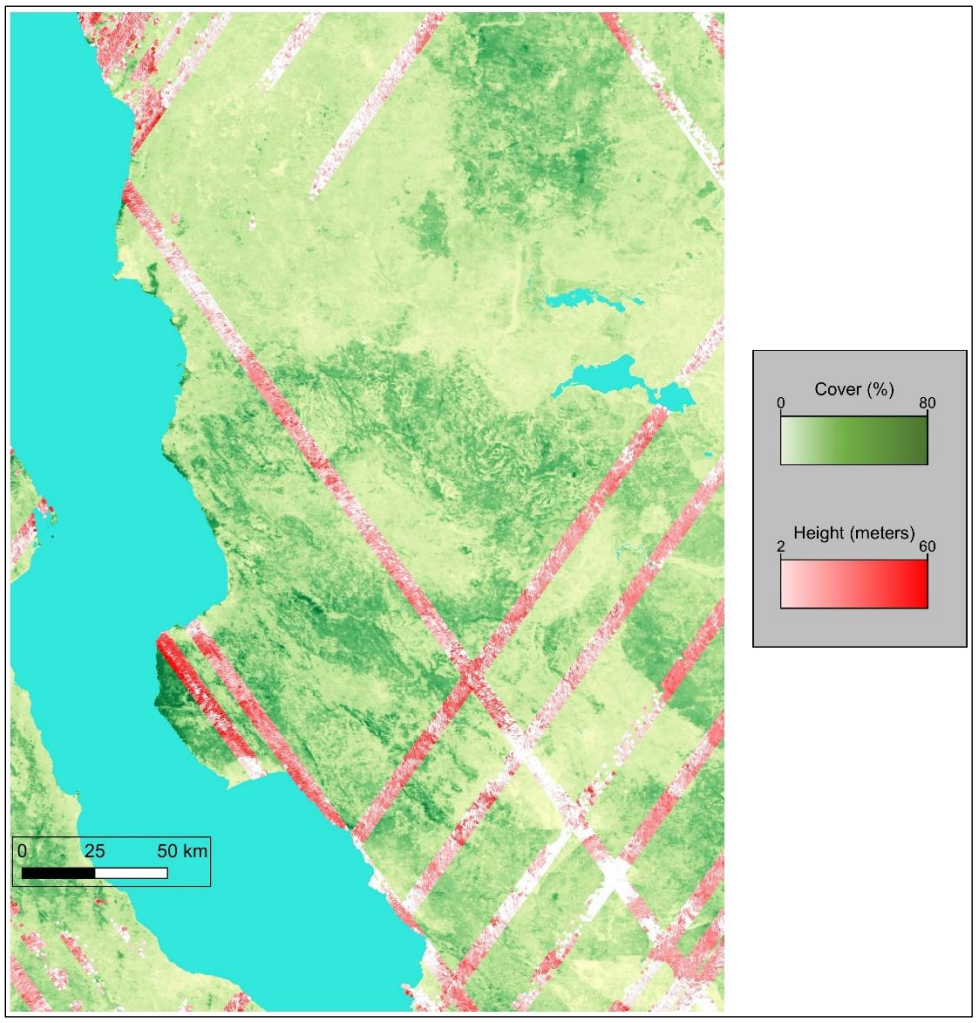


Figure 2.5 Distribution of GEDI LiDAR shots showing the rh100 metric overlayed on MODIS VCF canopy cover.

2.4.2 Forest cover change detection

A change point in a time-series can be defined as a significant shift in the distribution mean (Chen & Gupta 2000). In this chapter, change points in mean forest cover were detected with a combination of cumulative sums (CuSum) and bootstrap analysis (Page 1955, 1957; Taylor 2000). As the name may imply, CuSum is not the cumulative sum over a time-series, but rather the sum of deviations from the mean, and bootstrap analysis can be used to estimate the likelihood that the change is significant. Here, the CuSum (S) of the per pixel difference between mean canopy cover and annual canopy cover was used to determine the timing and magnitude of change.

If $C_1, C_2, \ldots C_n$ is a vector of annual canopy cover estimates from n years for a pixel and $\bar{C}$ represents the pixel mean, then the ith value in the CuSum time-series is:

$$S_i = S_{i-1} + (C_i - \bar{C})\ i = 1,2 \ldots n; and\ S_0 = 0 \qquad 2.7$$

The plot of S where no change occurs will be represented by a horizontal line that randomly fluctuates around zero; however, if a change did occur, the line will become substantially positive or negative depending on the direction of change. Before bootstrap analysis is performed, the magnitude of change, $Sdiff$, is calculated by taking the difference between the maximum and minimum of S. Bootstrap analysis can then be used to assign a significance level to a candidate change point and is calculated with the following steps:

1. Sample without replacement the original time-series, which practically results in a random reordering of the data vector.
2. Calculate CuSum (S') on the reordered data.
3. Calculate $Sdiff'$ from S'.

4. Count the number of times $Sdiff$ is greater than $Sdiff'$ and assign as X.

If N is the number of bootstrap repetitions, then the confidence level (CL) is calculated as $100\frac{X}{N}\%$. In this chapter, the number of bootstrap repetitions was set at 1000 and the CL at 90%. If the bootstrap analysis reveals a significant change in mean, the optimal change point can be estimated by minimizing the mean squared error (MSE). For a vector of pixel canopy cover estimates C with n number of elements, MSE is calculated by,

$$MSE = \sum_{1=1}^{m}(C_i - \bar{C_1})^2 + \sum_{i=m+1}^{n}(C_i - \bar{C_2})^2 \qquad 2.8$$

where $\bar{C_1} = \frac{\sum_1^m C_i}{m}$ and $C_2 = \frac{\sum_{m+1}^n C_i}{n-m}$

The value of m that minimizes MSE is the estimated change point. Bootstrap analysis and minimization of MSE can be performed iteratively on the resulting segments allowing the detection of multiple changes within a time-series. The minimum segment length for bootstrap analysis for this study was set at five years. Figure 2.6 illustrates the detection of two change points for a pixel located in an oil palm plantation, where a forest/canopy loss event was estimated to have occurred in 2009 followed by a subsequent increase around 2016. The general workflow for forest cover change detection can be seen in figure 2.7.

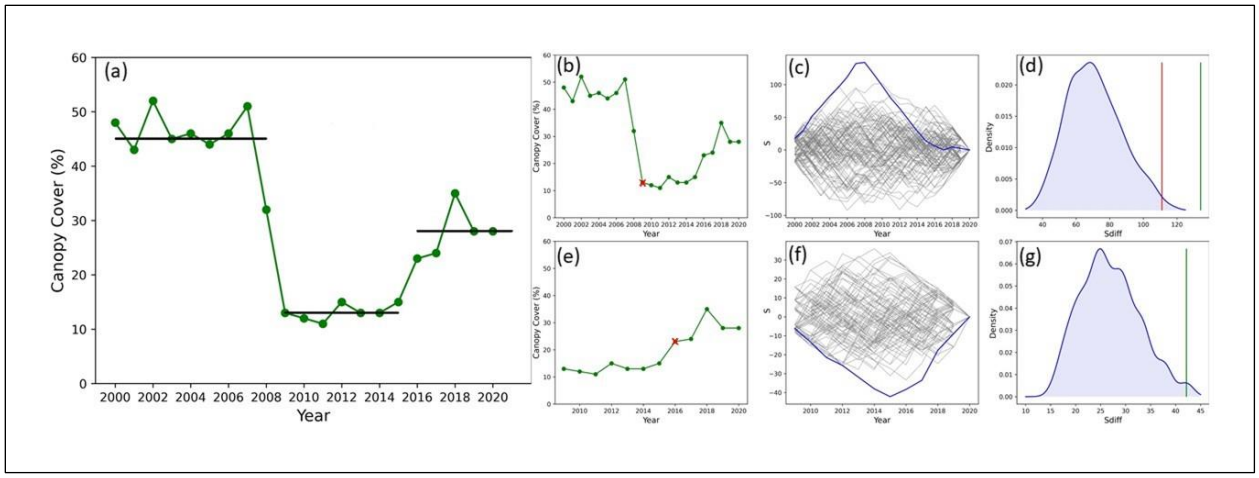


Figure 2.6 Annual canopy cover estimates and significant changes in mean (black lines) for a pixel centered at 4°56'24.5"S 29° 41' 38.7"E (a), estimate for first change point (red x) (b), sample CuSum plot (blue line) with boot-strapped samples (grey lines) (c), distribution of Sdiff with threshold value (red) and Sdiff pixel value (green) (d). (e)-(g) are the same as (b)-(d) but for the second change point. Threshold value and calculated Sdiff are equal, hence the green line obscures the red line in (g).

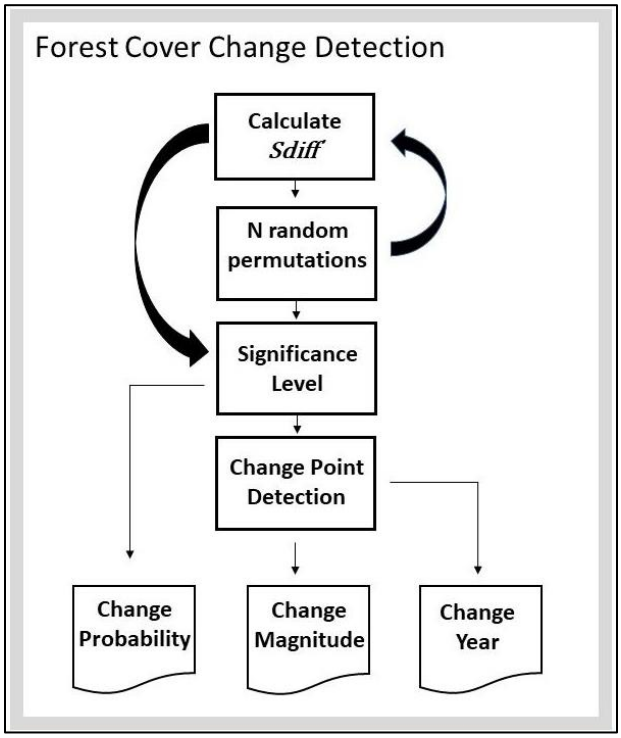


Figure 2.7 CuSum canopy cover change detection workflow.

2.4.2.1 Accuracy Assessment

Separate accuracy assessments were performed for gross forest loss and gain between 2000 and 2020. Loss is defined as a pixel having experienced a decrease in canopy cover of at

least 15% between 2000 and 2020, which results in approximately the same amount of gross forest loss as the GLAD global forest loss product for the same time period (Hansen et al. 2013). The definition of forest cover gain was slightly different since forest gain tends to be a continuous process, whereas loss is an abrupt change. For example, Figure 2.8 shows a pixel where forest cover increased by 30% between 2000 and 2020, and the CuSum algorithm detected 2 significant changes one in 2009 and another in 2016. Using the same definition for loss (i.e., a 15% change in cover), no gain would have been detected since neither of the changes are 15% or greater. Mean canopy cover shifted from 25% to 39% (14% gain), and then to 49% (10% gain) resulting in an estimated 24% increase in cover. A pixel was therefore defined to have experienced forest gain if no loss events were detected, and the total gain was 15% or greater between 2000 and 2020.

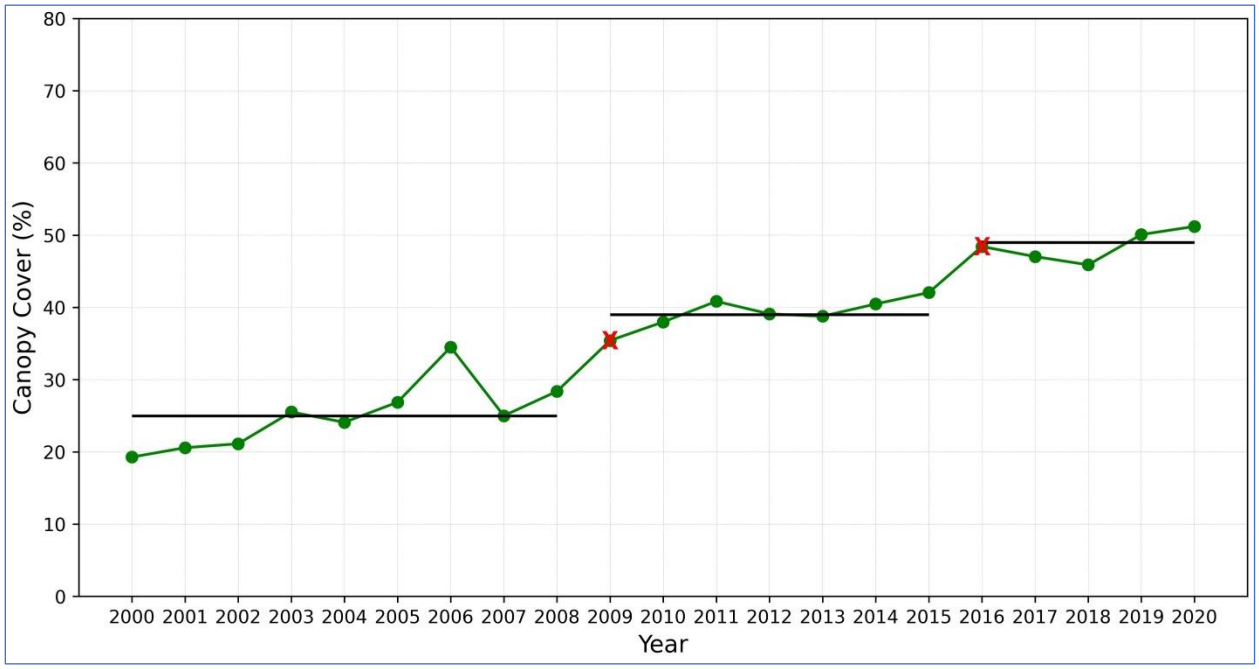


Figure 2.8 Example of canopy cover gain detection where a red X denotes when a significant change in the mean was detected.

After applying a 15% threshold, the study area was divided into three non-overlapping strata. For forest loss, this consisted of Loss, a 3-pixel buffer around Loss, and all other Land pixels. The

strata for forest cover gain validation were Gain, a 10-pixel buffer around Gain, and all other Land pixels. A larger buffer for Gain was chosen to mitigate omission error since Gain covered a very small portion of the study area (Olofsson et al. 2014) Sample pixels were randomly allocated to the strata according to optimal allocation given by,

$$n_h = n \frac{W_h S_h}{\sum W_h S_h} \tag{2.9}$$

where n is the total number of samples, n_h is the sample size for stratum h, W_h is the proportion of the study area covered by stratum by h, and S_h is the variance associated with stratum h (Cochran 1977). The variance for each stratum is unknown before the validation has been performed, therefore a reasonable estimate for the expected error must be used to calculate S_h, given by,

$$S_h = \sqrt{U_h(1 - U_h)} \tag{2.10}$$

where U_h is the user accuracy for stratum h. For targeting change, user accuracies of 0.7, 0.1, and 0.001 were assigned to the change (Loss or Gain), Buffer, and Land strata, respectively, which represents 70% user accuracy for Loss and Gain, 10% capture of change pixels in the Buffer stratum, and 0.1% omission of change pixels in the Land stratum.

An initial validation sample size of 400 pixels was chosen for each change class to be tractable in terms of time and effort. However, for validating forest gain, all sample pixels were allocated to the Buffer and Land strata. The minimum recommended number of samples for the change class is 50 (Oloffson et al. 2014), therefore 50 additional sample pixels were allocated to the Gain stratum resulting in 450 samples for forest gain validation and 400 for forest loss

validation. Samples sizes can be found in Tables 2.1 and 2.2, and the distribution of validation samples can be seen in Figure 2.9.

Table 2.1 Size, proportion, and number of samples allocated to each stratum for forest loss assessment.

Stratum	Pixels in Stratum	Stratum Weight	Samples in Stratum
Loss	1,017,610	0.02	70
Buffer	5,247,706	0.13	190
Land	35,711,569	0.85	140
Total	41,976,885	1.0	400

Table 2.2 Size, proportion, and number of samples allocated to each stratum for forest gain assessment.

Stratum	Pixels in Stratum	Stratum Weight	Samples in Stratum
Gain	43,497	0.001	50
Buffer	2,221,954	0.053	140
Land	39,711,434	0.946	260
Total	41,976,885	1.000	450

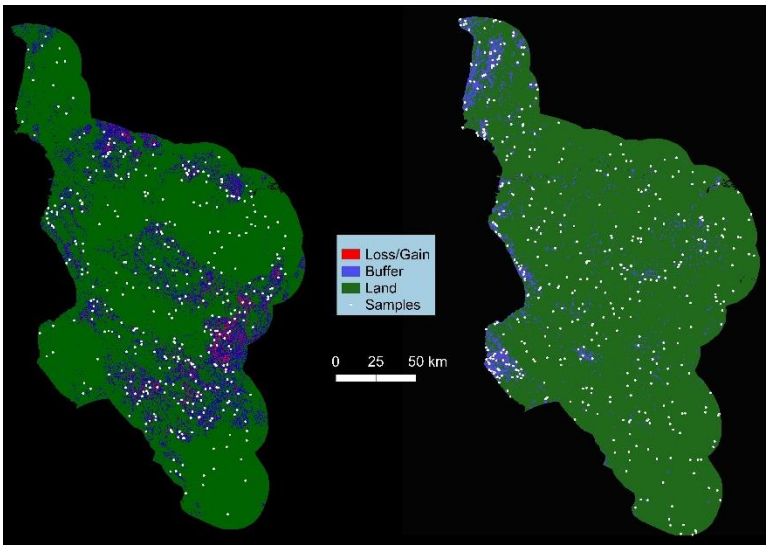


Figure 2.9 Map of forest loss strata with distribution of 400 sample pixels (left panel). Forest gain strata and distribution of 450 sample pixels (right panel).

To verify if change had occurred at the sample pixels, a visual assessment of annual interquartile mean composites for SWIR1, NIR, and SWIR2 was performed (Figure 2.7). In addition, each sample pixel was verified with high-resolution Google Earth imagery where

available. If a change was ambiguous and no high-resolution time-series imagery was available, the individual Landsat 16-day composites were examined to make a final determination.

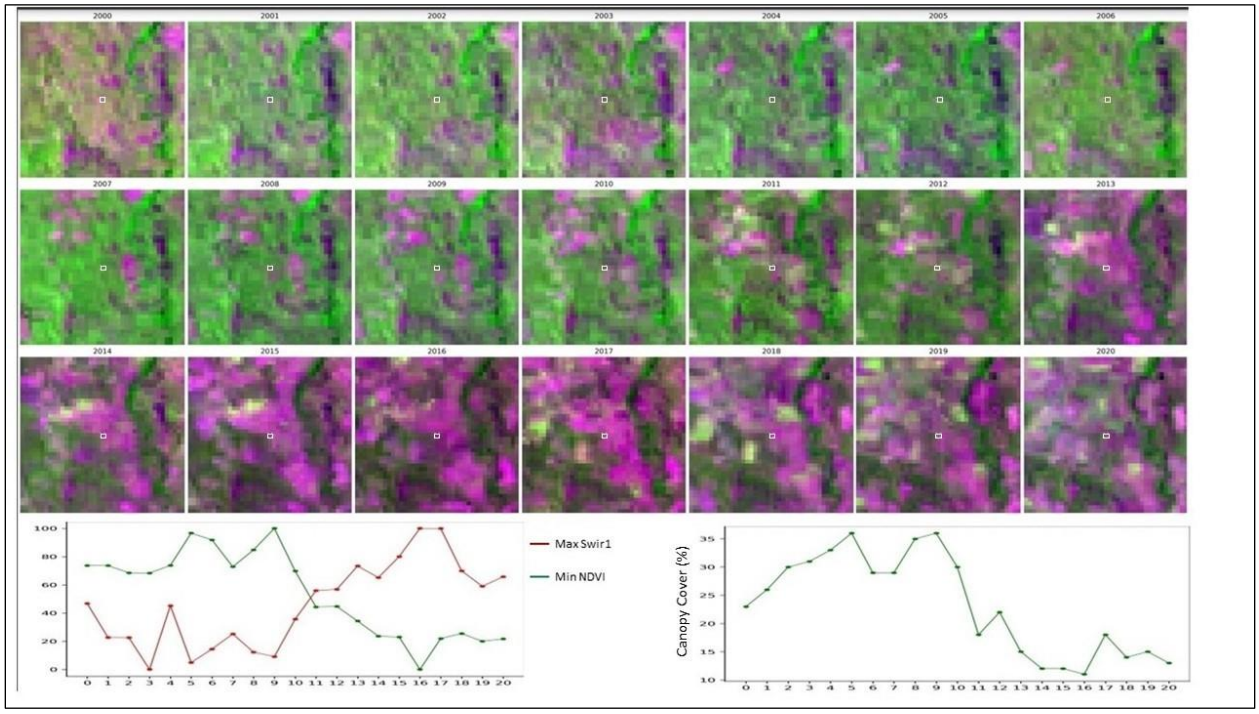


Figure 2.7 Annual composites of interquartile means for Landsat bands SWIR1/NIR/SWIR2 for a sample pixel (white square) and 300m buffer. Plots of annual minimum NDVI, maximum SWIR1, and predicted canopy cover for the target pixel are also shown.

After the validation sample was obtained, the error matrix was generated in terms of proportion of area (p_{ij}) given by,

$$p_{ij} = W_i \frac{n_{ij}}{n_{i.}} \tag{2.11}$$

where W_i is the proportion of area mapped as class i, n_{ij} is the sample count in cell ij, and $n_{i.}$ is the number of reference samples for class i (Olofsson et al. 2014). The error matrix was then used to estimate User's accuracy for class q classes (i.e., 3),

User's accuracy for class i,

$$U_i = \frac{p_{ii}}{pi.} \tag{2.12}$$

Producer's accuracy for class *j*,

$$P_j = \frac{p_{jj}}{p_{.j}} \tag{2.13}$$

In addition to the accuracy metrics, area estimated from the validation sample is reported where the area can be estimated from equation (2.14) and associated standard error from (2.15) (Olofsson et al. 2014). Finally, relative standard error was calculated from equation 2.16 to provide information on the precision of the area estimate.

$$\hat{Y} = A * \sum_{i=1}^{q} W_i \frac{n_{ik}}{n_{i.}} \tag{2.14}$$

$$SE(\hat{Y}) = A * \sqrt{\sum_{i=1}^{q} W_i^2 \frac{\frac{n_{ik}}{n_{i.}}\left(1 - \frac{n_{ik}}{n_{i.}}\right)}{n_{i.} - 1}} \tag{2.15}$$

$$RelSE = 100 * \frac{SE(\hat{Y})}{\hat{Y}} \tag{2.16}$$

In the equations, A represents the study area size and the summations are over q strata (i.e., 3) for the (k) change classes. All equations were applied separately for forest loss and gain evaluation.

2.5. Results

2.5.1 Annual canopy cover estimation

The random forest model fit the MODIS VCF data well with average RMSE of 5.52%, MAE of 4.1%, MBE of 0.53% and r^2 of 0.77 on the hold-out sample (Table 2.3), and modelled tree cover and MODIS VCF tree cover are tightly distributed along the one-to-one line (Figure 2.8). Results from cross-validation should not be interpreted as to how well the model performs at predicting canopy cover at 30 meters but is rather used as a diagnostic tool to ensure over-fitting was avoided. If the model cannot reliably predict cover at 250 meters, then it follows those predictions at 30 meters will be even less reliable.

Table 2.3 Coefficient of determination, mean bias error (MBE), mean absolute error (MAE), and root mean squared error (RMSE), and associated standard errors (SE) from results of 5-fold cross-validation.

	r^2 (SE)	MBE (SE)	MAE (SE)	RMSE (SE)
Landsat-MODIS	0.77 (0.0001)	0.53 (0.001)	4.1 (0.001)	5.52 (0.001)

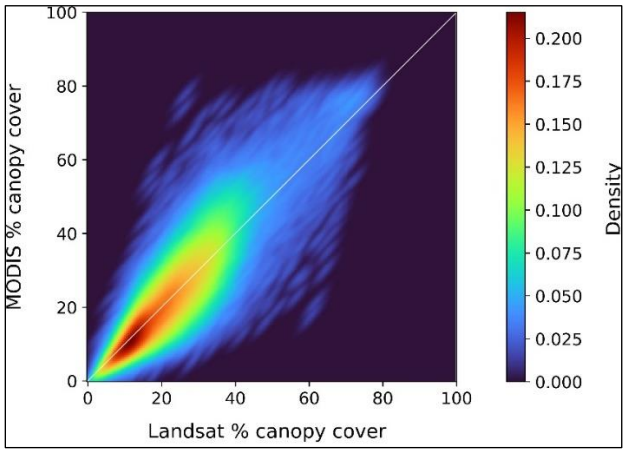


Figure 2.8 2d density plot for predicted Landsat canopy cover and MODIS VCF cover. The one-to-one line is shown in white.

The three map classes showed significantly different height distributions (Fig 2.9a). The two sample Kolmogorov-Smirnov tests confirmed the Forest class is taller than the Woodland class, which is taller than the Non-forest class with P-values well below 0.001. Median height for

the Forest class was about 25 meters, Woodland was about 12 meters shorter than Forest at 12.7 meters and Non-forest another 8 meters shorter than Woodland at under 5 meters (Fig 2.8b and Table 2.3).

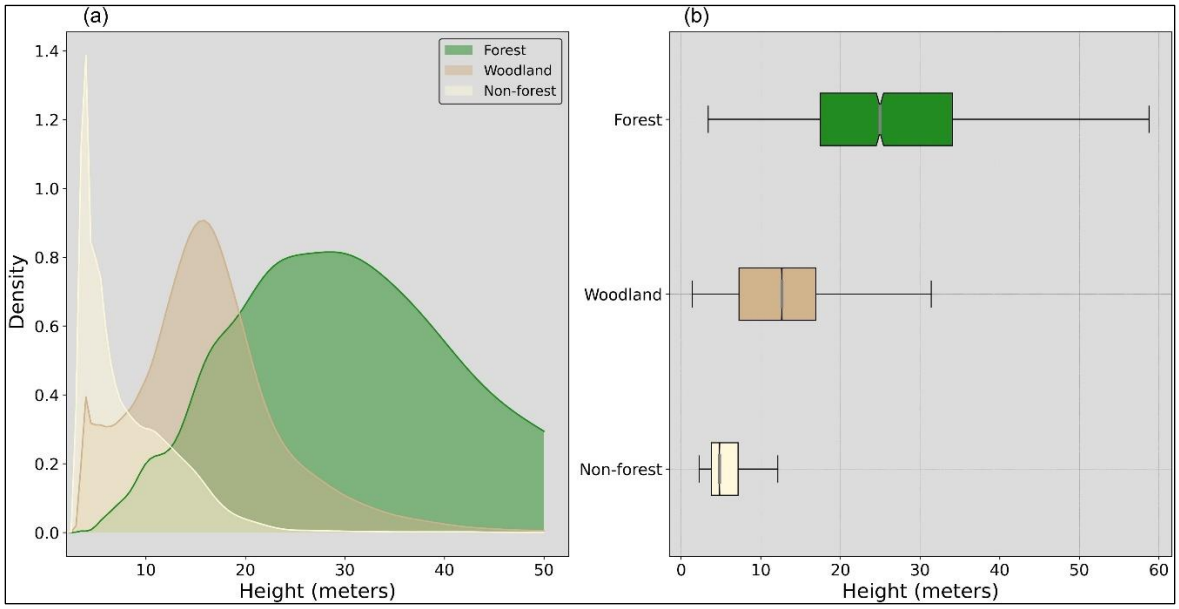


Figure 2.10 Height distribution of Forest, Woodland, and Non-forest (a). Boxplots of canopy height for Forest, Woodland, and Non-forest.

Table 2.3 Number of GEDI samples for each class (N), median, and mean tree heights for the 3 classes. Table accompanies figure 2.9.

Class	N	Median (meters)	Mean (meters)
Non-Forest	29,372	4.81	6.17
Woodland	141,973	12.65	12.94
Forest	3,313	24.97	26.71

2.5.2 Forest change detection

The initial threshold for defining forest loss was biased toward omission error (Figure 2.10), which resulted in high User's accuracy (0.83) at the expense of low Producer's accuracy (0.33) (Table 2.4). After adjusting the forest loss threshold magnitude from 15% to 12%, User's

and Producer's accuracies achieved similar values of 0.75 and 0.72, respectively (Table 2.4). The adjustment also resulted in a reduced relative standard error, dropping from 12% to just under 10%, and smaller standard errors for User's and Producer's accuracies (Table 2.4).

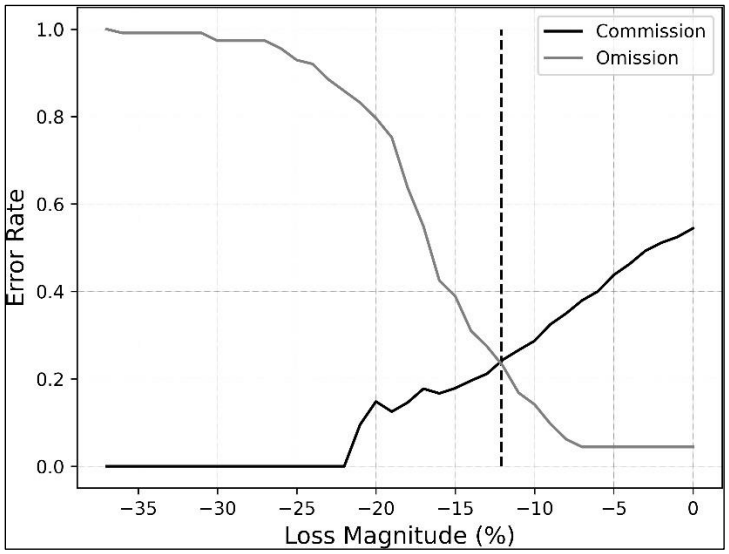


Figure 2.11 Commission and omission error rates with increasing threshold of loss magnitude for validation sample. Vertical dashed line represents threshold with equal commission and omission error.

Table 2.4 Relative standard error (RelSE), user, and producer accuracy with standard errors (SE) for the forest loss validation sample before and after adjusting the threshold to balance commission and omission error.

	RelSE (%)	User (SE)	Producer (SE)
Loss	12.01	0.83 (0.05)	0.33 (0.12)
Adjusted	9.94	0.75 (0.04)	0.72 (0.09)

Relative to forest loss detection, the CuSum algorithm showed lower accuracy and higher relative standard error for forest gain detection (Table 2.5). Relative standard error was more than twice that of forest loss prior to forest loss threshold adjustment at nearly 28%, and User's accuracy was lower at 0.74, and Producer's accuracy was quite low at just 0.2 (Table 2.5).

Table 2.5 Relative standard error (RelSE), user, and producer accuracy with standard errors (SE) for the forest gain validation sample.

	RelSE (%)	User (SE)	Producer (SE)
Gain	27.53	0.74 (0.06)	0.20 (0.06)

Compared to the GLAD global product, the CuSum algorithm detected substantially more gross woodland and forest loss between 2000-2020 (Fig. 2.11). According to the GLAD product, the GME and GGE lost about 145,000 ha of woodland and forest, the CuSum estimate prior to adjustment is just over 230,000 ha (+/- 40,000 ha), and the estimate after threshold adjustment is 276,000 ha (+/- 27,000 ha), more than 130,000 ha than the GLAD product.

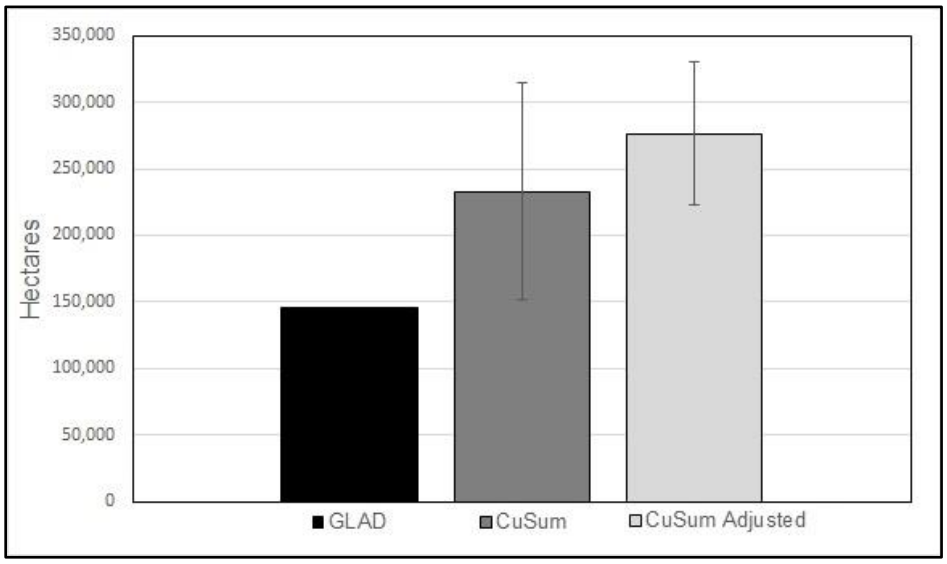


Figure 2.12 Estimates of gross loss (hectares) for dense forest and woodland from the GLAD global product and a manually interpreted sample of 400 pixels.

Forest gain could not be directly compared to the GLAD product because gain is assessed between two different time periods. The GLAD product reports forest gain between 2000 and 2012, whereas results from this study are reported for the 2000 to 2020 period. Nevertheless, the GLAD product showed 1732 ha of forest gain, or an increase of 144 ha/year for the study area between 2000 and 2012. Based on the validation sample, the CuSum algorithm

detected 12,927 (+/- 3,560 ha), or an increase of 619 ha/year. Based on these results, the CuSum algorithm detected substantially more forest gain than the GLAD product.

2.6 Discussion

Annual 30 m resolution estimates of forest cover as well annual changes (loss and gain) were produced for the GGE and GME ecosystems in this study. Cross-validated results showed that forest cover was accurately modelled at 250 m resolution; however, attempting to validate the results at 30 m against fractional cover from the GEDI LiDAR sensor indicated that GEDI is not a reliable validation source for optical estimates of canopy cover, likely due to the relatively large footprint of 25 m, which is larger than most tree canopies. Qualitatively, however, the downscaled VCF displayed good performance with respect to discriminating riparian forests. Downscaled predictions of canopy cover can resolve well the thin strips of riparian forest that are not visible in the coarser resolution training data (Fig 2.13).

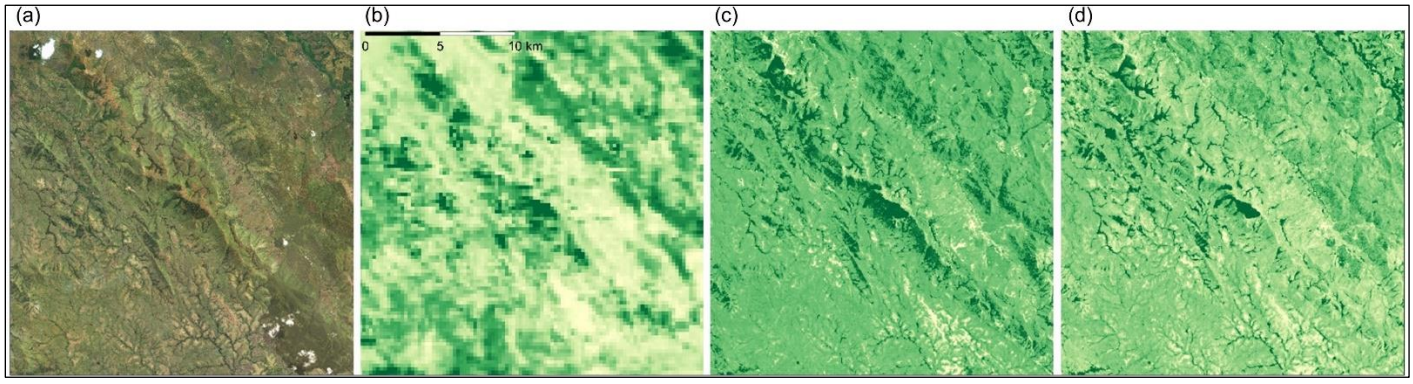


Figure 2.13 Region centered at 6° 3' 30.96"S 30° 5' 50.136"E east of MMNP. High-resolution Bing imagery (a), MODIS VCF canopy cover (b), GLAD 30 m canopy cover (c), and downscaled VCF cover (this study) (d).

Additionally, the top of canopy height metric from GEDI confirmed that the thresholds for defining Forest, Woodland, and Non-Forest were reasonably well chosen. Figure 2.14 shows the mapped classes for the year 2019, and, with exception of the western region of Mahale

Mountains National Park, Forest is generally patchily distributed within Woodland. Indeed, forest was found to occupy just 4% of the land, while woodland covers 90%.

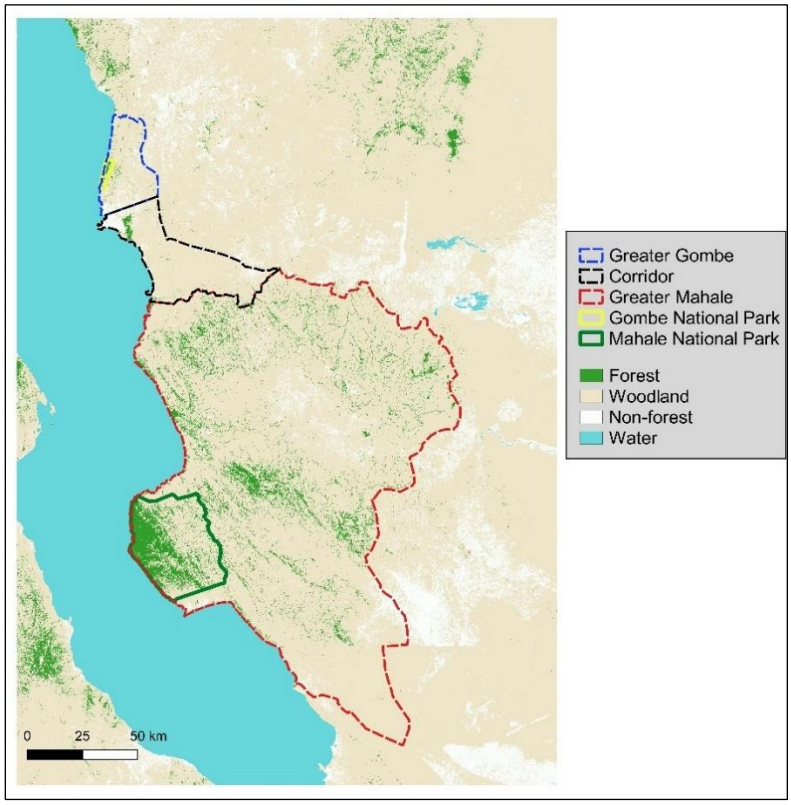


Figure 2.14 Map showing distribution of Forest, Woodland, and Non-forest in the study area in 2019.

The canopy cover product developed in this chapter demonstrated that it captures known gradients of canopy density. Figure 2.15 (Left Panel) shows an area in the northwest of MMNP covered by lowland evergreen forest known as “Kasoje” forest, which covers the flat to hilly region between Lake Tanganyika and the Mahale mountains. The mountain slopes are a mix of bamboo, bushland and montane forest, and the highest areas, above 2300 m, give way to grassland and montane bushland (Nishida 1990). A decrease in canopy density would be expected with increased elevation given this distribution of vegetation. The median values for elevation and canopy cover were calculated every 300 m within the 5000 m by 300 m transect

shown in Figure 2.15 (Left Panel) for the SRTM DEM, GLAD canopy cover product, and the downscaled VCF product. Median canopy cover values were plotted against median elevation values (Fig. 2.15 Right Panel). The downscaled VCF product shows a significantly decreasing trend with elevation ($\rho = 0.94$, $p << 0.001$) while the GLAD product does not show a trend with elevation ($\rho = 0.25$, $p = 0.36$). While this observation does not serve as validation at 30 m, it demonstrates that the product is sensitive to gradients in vegetation density that a global product is not. The Random Forests algorithm's ability to produce reasonable canopy cover estimates at 30 m resolution, despite coarser resolution training data, could be attributed to the synergistic effects of a substantial sample size that spans two decades of observations, and input features that only reflect local variation.

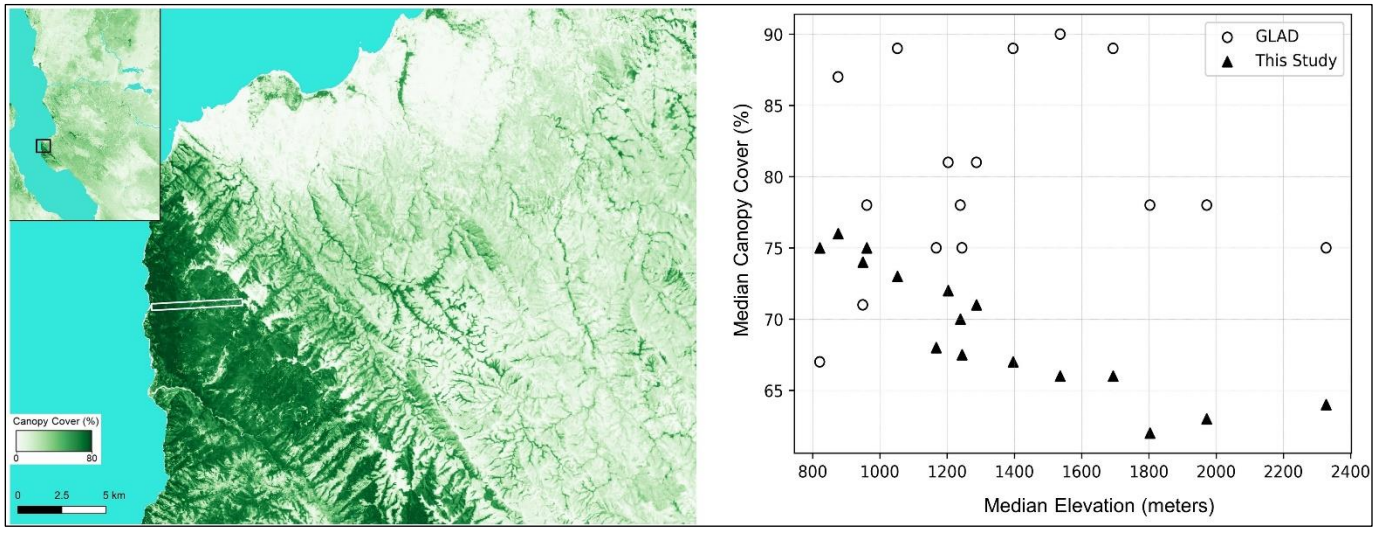


Figure 2.15 Map of Kasoje forest in MMNP with 5 km x 300 m transect (white rectangle) (Left Panel). Median canopy cover plotted against median elevation calculated every 300 m along transect (Right Panel).

The CuSum product detected substantially more forest loss compared to the GLAD product; however, both products display similar patterns of annual forest loss, gradually increasing from the year 2000 to 2013 and subsequently decreasing (Fig 2.16); however, the CuSum product shows a large spike in the year 2013. This phenomenon appears to be the effect

of including Landsat 8 data in the 16-day composites, which began in 2013 (Potapov et al. 2020). Compared to prior Landsat sensors, Landsat 8 has a higher signal to noise ratio and can detect more subtle spectral changes. The hatched region of the bars in Figure 2.16 showing the area mapped by the CuSum algorithm indicates the portion of forest loss that was detected by the CuSum algorithm and not by the GLAD product, which suddenly increases in 2013 and stays relatively high.

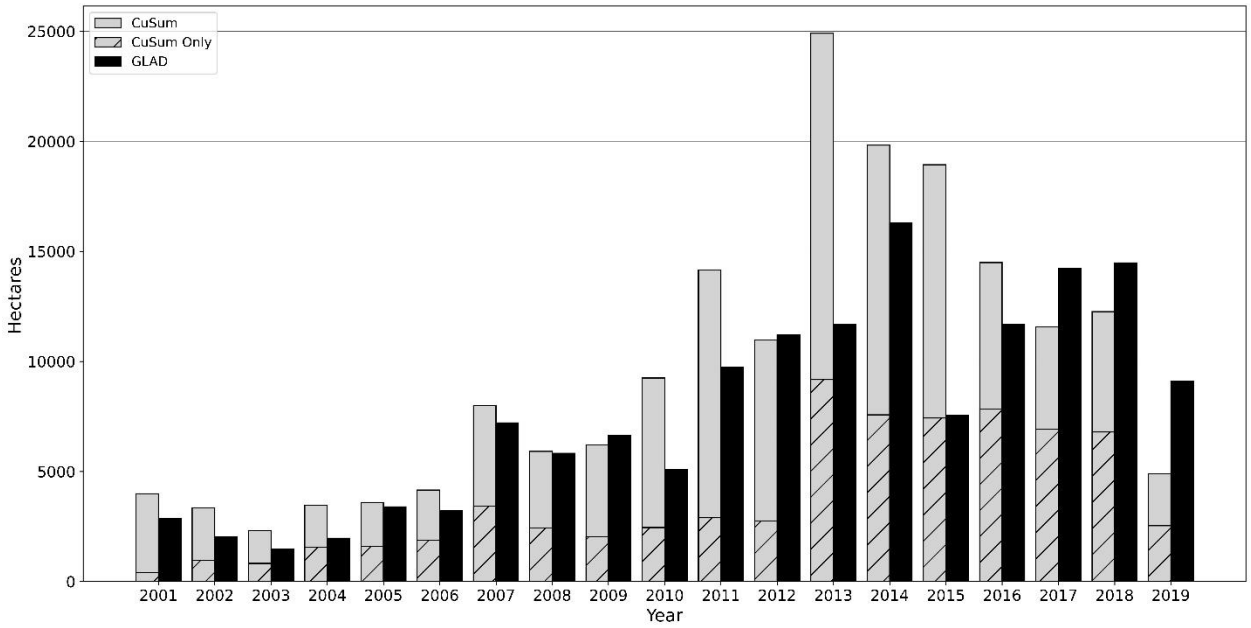


Figure 2.16 Annual loss (hectares) from the GLAD global product and the CuSum algorithm (this study). The hatched portions indicate forest loss detected by CuSum only.

The map from figure 2.17 shows where the CuSum algorithm detected forest loss and the GLAD product did not. It is apparent the CuSum product is detecting much more loss in woodlands, primarily around the periphery of the GME. Zoomed in versions of insets 1-4 of the map are displayed in figures 2.18 to 2.21 showing high resolution Google Earth imagery prior to and after loss occurred, as well as loss mapped by the GLAD and CuSum products. It is evident from the high-resolution imagery that the driver of loss in woodlands is conversion to agriculture, which agrees with what was found at the national scale, namely that agricultural

expansion is the main driver of deforestation in Tanzania (Doggart et al. 2020). Given these findings, it is likely that substantial amounts of forest loss went undetected prior to the inclusion of Landsat 8 data. The fact that the CuSum algorithm was able to accurately detect forest loss, particularly in woodlands, also serves as an implicit validation of the downscaled VCF product. Forest loss detection in woodlands requires the discrimination of change at the lower range of forest cover, from low cover (i.e., woodland) to no cover (i.e., agriculture), which requires a high-fidelity temporal signal.

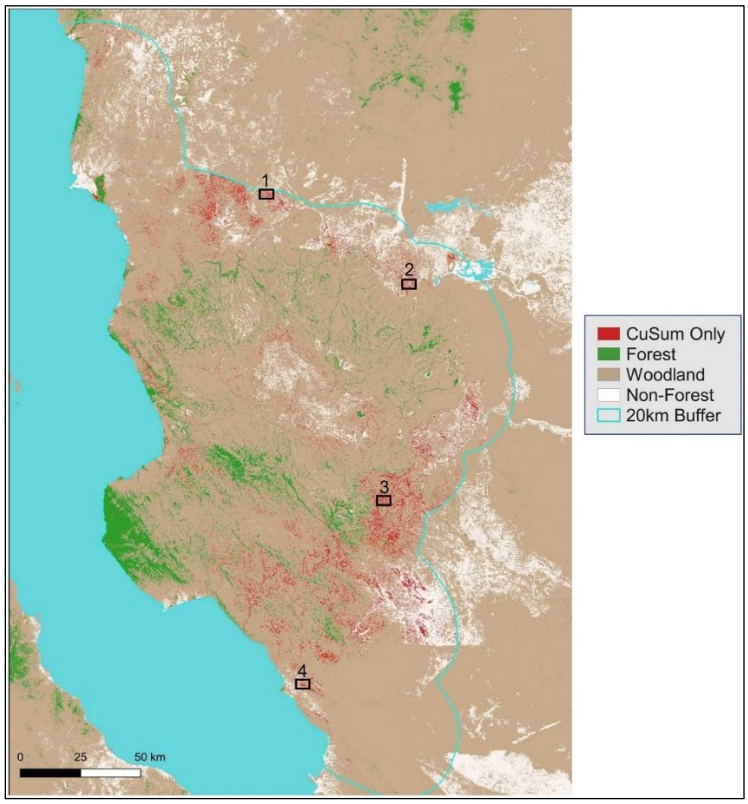


Figure 2.17 Forest loss detected by CuSum and not the GLAD product in the study area. The reader is referred to Figures 2.17-2.20 for zoomed in versions of insets 1-4.

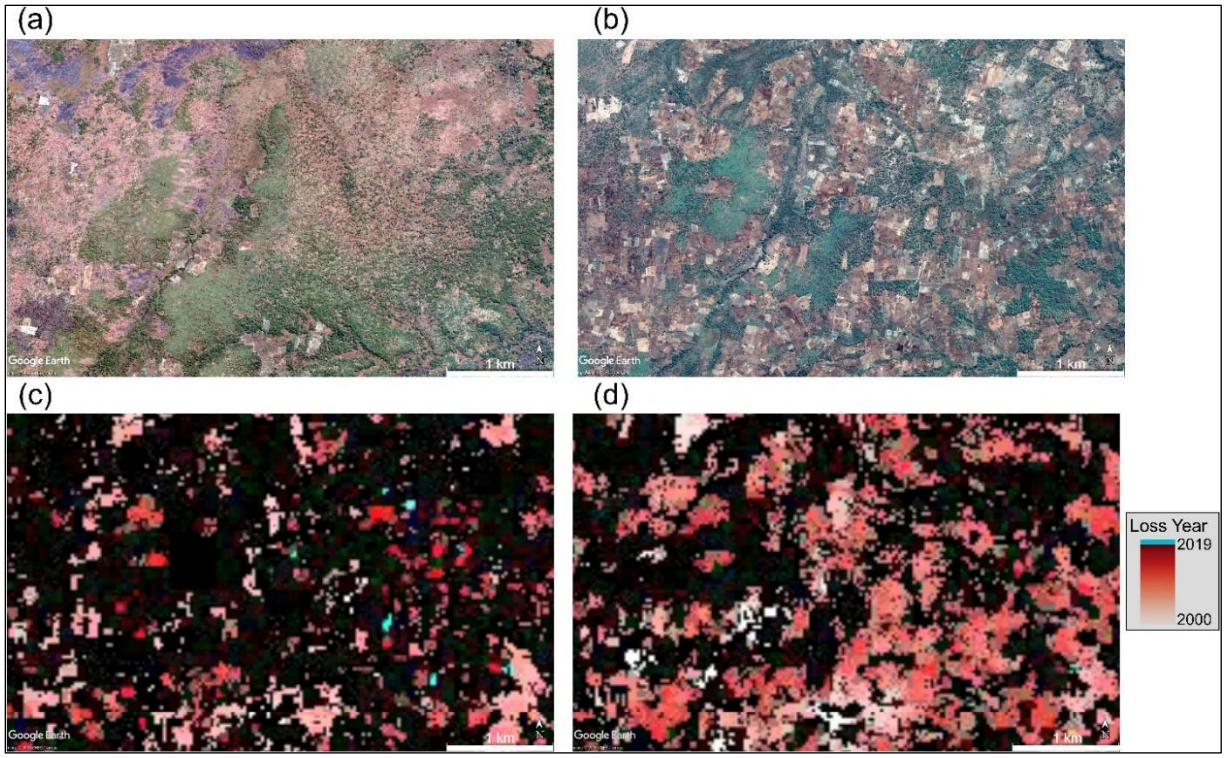


Figure 2.18 Corresponds to inset 1 of Figure 2.15. High resolution Google Earth imagery for 2005 (a) 2019 (b) GLAD forest loss (c) CuSum detected forest loss (d).

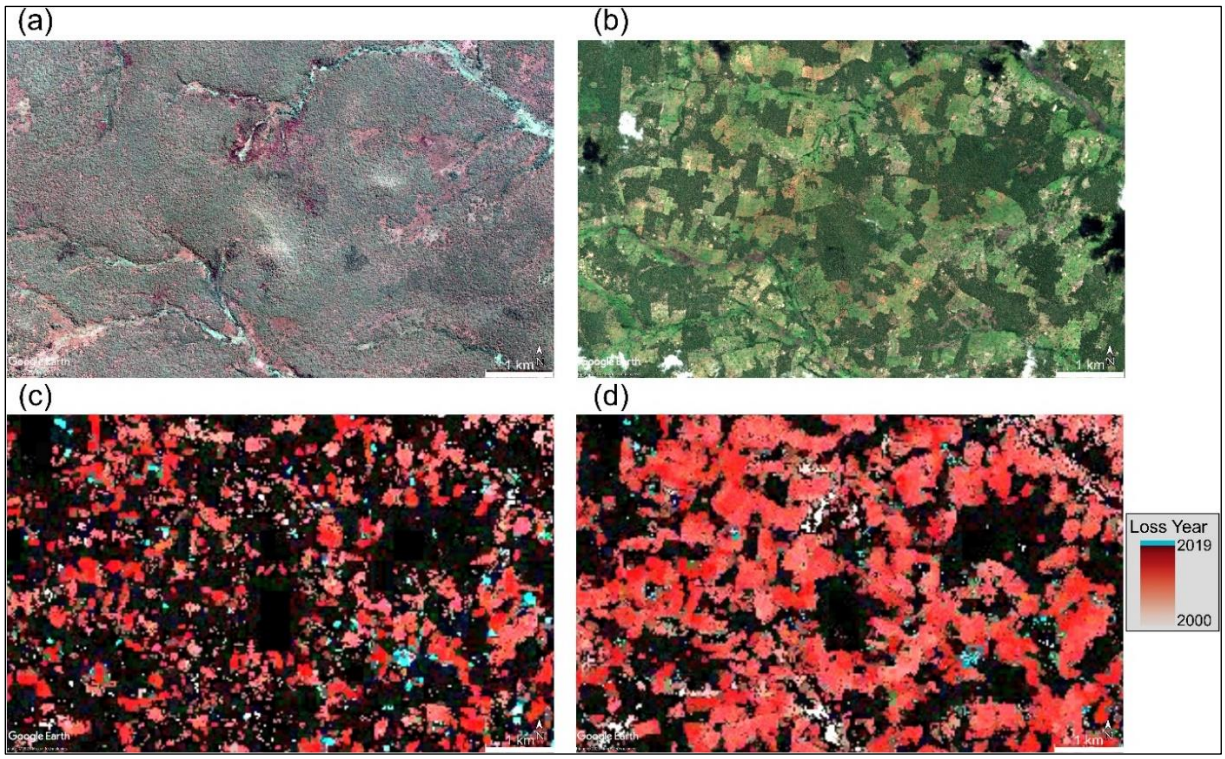


Figure 2.19 Corresponds to inset 2 of Figure 2.15. High resolution Google Earth imagery for 2009 (a) 2019 (b) GLAD forest loss (c) CuSum detected forest loss (d).

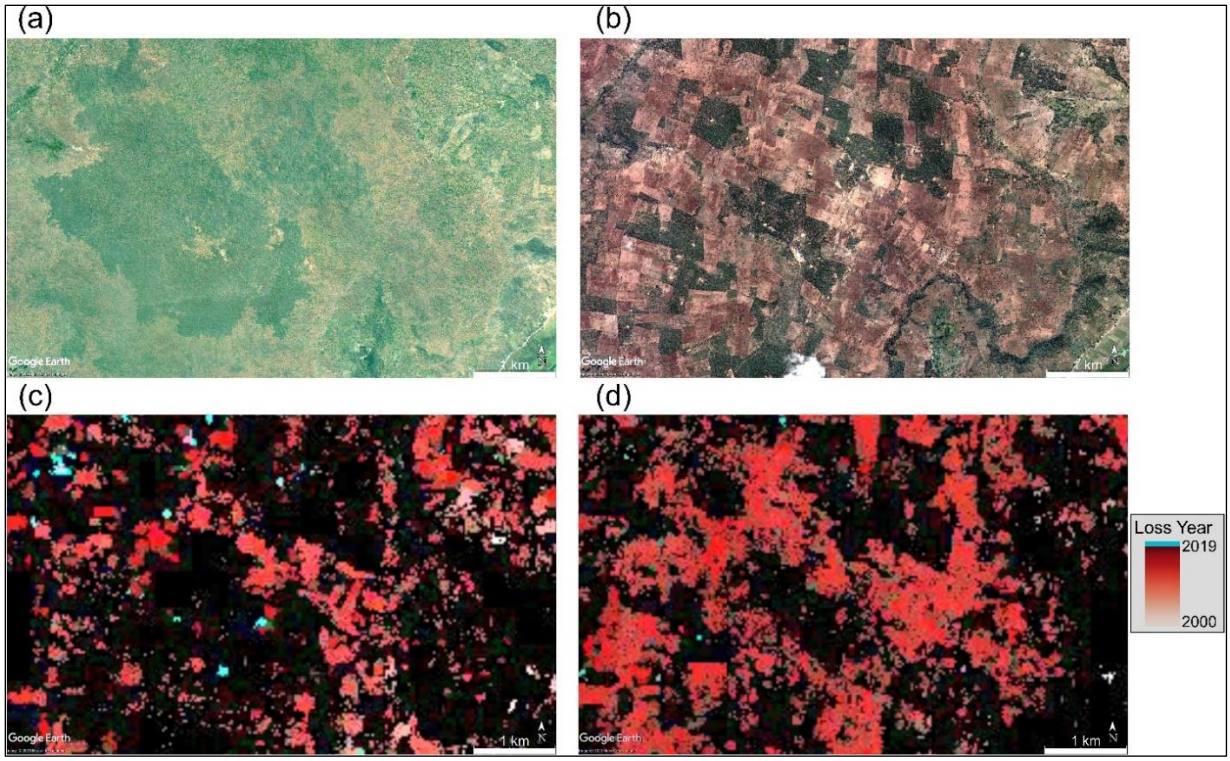


Figure 2.20 Corresponds to inset 3 of Figure 2.15. High resolution Google Earth imagery for 2010 (a) 2019 (b) GLAD forest loss (c) CuSum detected forest loss (d).

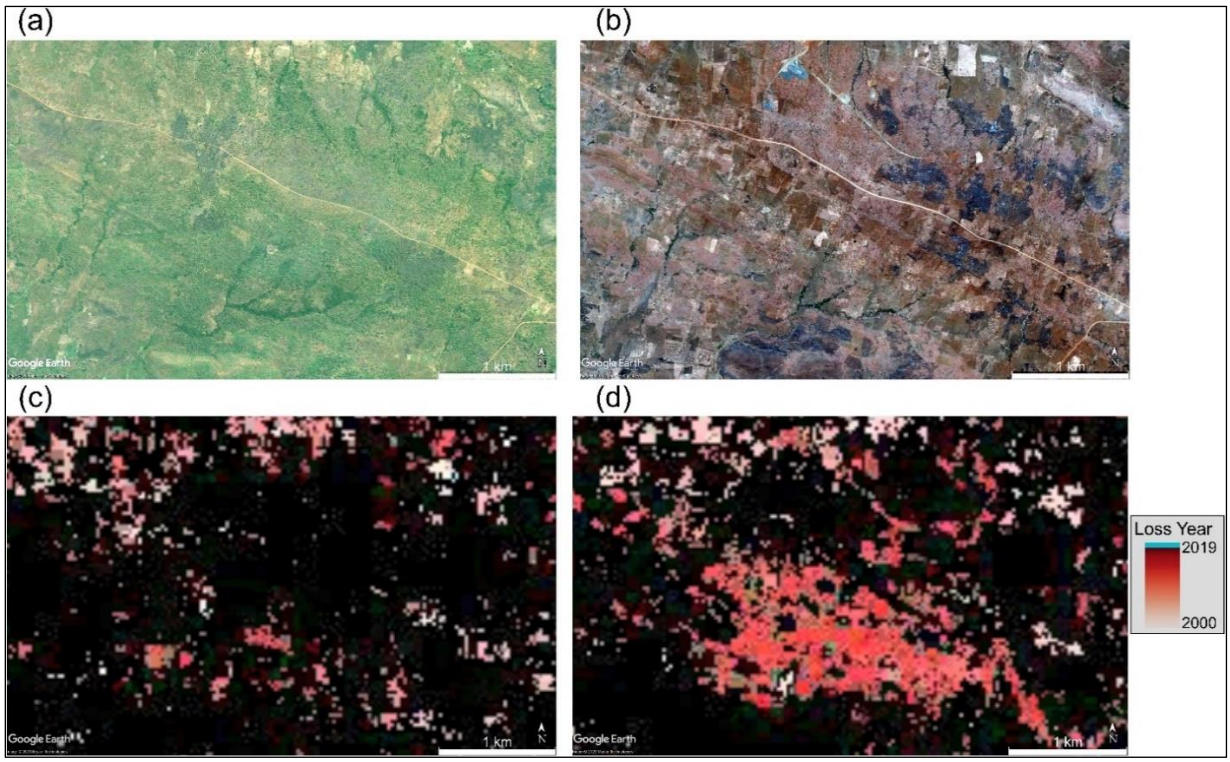


Figure 2.21 Corresponds to inset 4 of Figure 2.15. High resolution Google Earth imagery for 2010 (a) 2015 (b) GLAD forest loss (c) CuSum detected forest loss (d).

Despite dense canopy forest occupying only 4% of the study area, it is nonetheless being lost at a faster rate than woodland (Fig. 2.22 Left Panel), losing over 5% since 2000, while woodland decreased by less than 1%. Results show land-use plans in the GGE appear effective with forest net gain, and recent genetic evidence indicates that chimpanzees from GNP are dispersing into these previously barren areas (Wilson et al. 2020). However, in contrast, dense forests have declined drastically in the GME, with some village forest reserves exhibiting near-complete forest removal (Fig. 2.22 Right Panel). The current Conservation Action Plan (CAP) suggests that community members may not be familiar with the VFR bylaws and urges local governments to promote more community involvement and enforcement of village and district bylaws (TAWIRI 2018). Monitoring methods, such as those presented in this chapter, can be used to reveal the level of compliance moving forward.

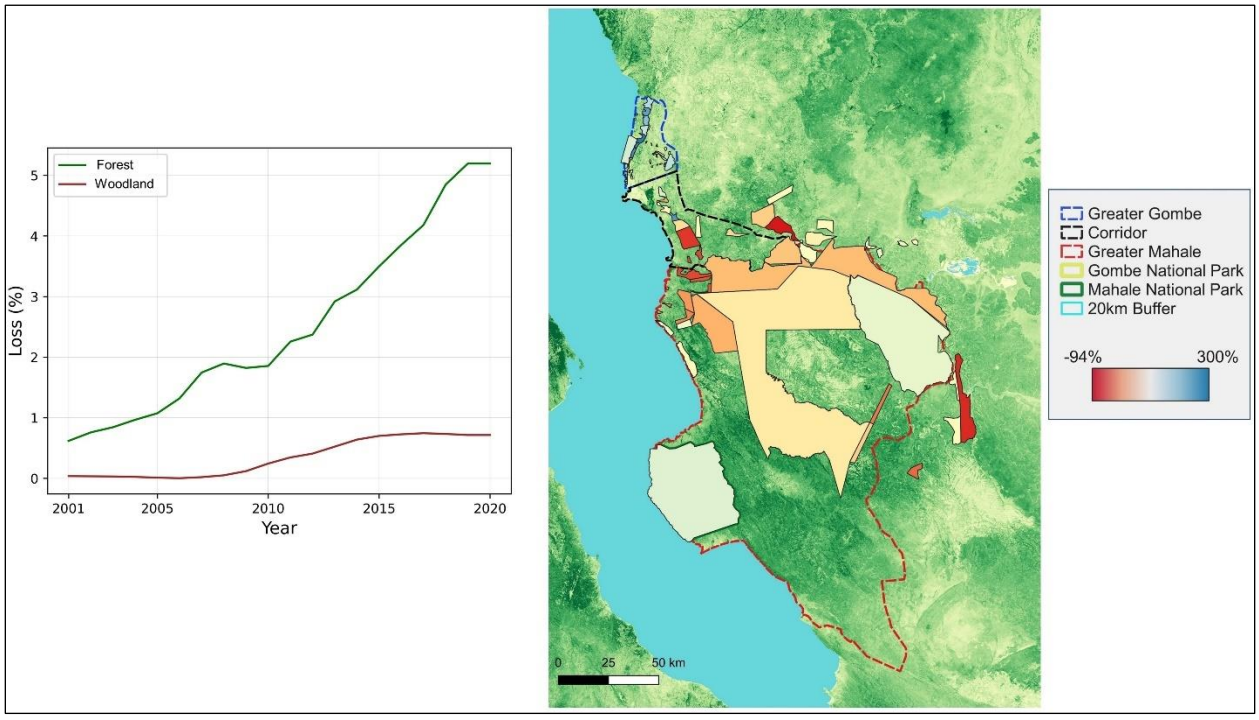


Figure 2.22 Annual percent change relative to 2000 for forest (green line) and woodland (red line) (Left Panel). Percent change of forest (canopy cover greater than 50%) for protected areas and village forest reserves between 2000 and 2020 (Right Panel).

2.7 Conclusions

In this chapter, it was shown that a Random Forests model trained on coarse (250 m) resolution data can produce annual estimates of fractional canopy cover at 30 m resolution with high fidelity, likely due to the combination of a large training sample and only allowing the algorithm to learn on spectral variability restricted to the study area. Using annual forest cover maps as input, the Cumulative Sums (CuSum) algorithm was able to accurately detect forest loss, but the accuracy for detecting forest gain events was somewhat diminished. Results show that the CuSum algorithm detected much more forest loss compared to the global GLAD product, particularly in woodlands and from 2013 onward. A plausible explanation as to why the CuSum algorithm suddenly began detecting increased forest loss is the inclusion of Landsat 8 data, which is more sensitive to subtle spectral changes than previous Landsat sensors. This can be confirmed by incorporating updated data to observe if the trend continues.

The conversion of forests to agriculture appears to be the primary driver of forest loss in the study region; however, forest loss was not distributed equally between the Greater Gombe and Greater Mahale Ecosystems. The GME showed much more loss, with some protected areas losing all dense canopy forests relative to 2000. In contrast, the GGE was predicted to have experienced a net gain in forest cover. To conserve forest habitats in the GME, local governments should ensure communities are familiar with the bylaws governing forest reserves and promote the enforcement of the regulations. The methods described in this chapter can be used to monitor the efficacy of the conservation action plan.

Chapter 3 : Linking Satellite Observations and Chimpanzee Population Survey Data for Effective Conservation Monitoring in the Greater Gombe and Greater Mahale Ecosystems

3.1 Introduction

All the estimated 2000-3000 chimpanzees of Tanzania reside in the GGE and GME, and while Gombe National Park (GNP) and Mahale Mountains National Park (MMNP) were established with the goal of protecting Tanzanian chimpanzees and their habitats , surveys have shown that the majority of chimpanzees (~75%) live outside the parks with the bulk of the population residing within the Greater Mahale Ecosystem (GME) (Moyer et al. 2006; Piel et al. 2015). An estimated 500-700 individuals reside within MMNP with an additional population of approximately 1500 living in the GME outside the park (TAWIRI 2018). Compared to the GME, GNP and the Greater Gombe Ecosystem (GGE) host a much smaller population with approximately 90 individuals inside and an additional population of about 20 outside the park (Pusey et al. 2007; Wilson et al. 2020).

Previous efforts to model chimpanzee habitat in this region have focused in the GME; however, the results were not intended for monitoring purposes. Pintea et al. (2006) used Mahalanobis distance and three environmental variables, distance to forest patches greater than 2.4 ha, elevation, and distance to steep slopes, greater than 20˚, to map areas of suitable nesting habitat in the GME, with the objective of identifying suitable habitat in areas that were not previously surveyed. Bonnin et al (2020) created a habitat suitability model using vegetation type (i.e., riparian forest, woodland, and non-forest), proportion of riparian forest, distance from riparian forest, elevation, and distance from steep slopes as input variables. Model output was used as input into a connectivity model to reveal possible reductions in habitat connectivity since

1972. Dickson et al. (2020) used the model from Bonnin et al. (2020) and applied it to projections of forest loss for 2027. The goal of the study was to demonstrate how outputs from species distribution models can be used a biodiversity indicator in the context of REDD+ projects.

Regular, exhaustive population surveys would provide decision makers with information germane to the success, or not, of implemented conservation action plans. However, conducting ground surveys requires substantial investments of time and money (Plumptre et al. 2010). An additional complicating factor is that chimpanzees are sparsely distributed in this region, which decreases the chance of detecting significant population shifts from subsequent surveys (Piel et al. 2015). In this study, variables derived from remotely sensed datasets were related to geolocated chimpanzee nests within in the framework of an Inhomogeneous Poisson Process (IPP) to estimate the relative occurrence rate of chimpanzee nests. Recent insights into model formulation have shown that framing presence only species distribution modeling problems as IPPs is equivalent to using MaxEnt (Warton & Shepherd 2010), by far the most commonly used technique, or logistic regression and its derivatives (e.g. Random Forests, Generalized Additive Models, Gradient Boosting Machines etc.) (Fithian & Hastie 2013). In addition, IPPs provide tools to guide the model fitting process, which are discussed in the Methods section. Model output from 2020 was used to derive a proportionality constant by incorporating population estimates from a recent ground survey to enable the prediction of chimpanzee numbers as a function of landscape condition. Since the model incorporates annually updated data, potential population changes can be tracked, albeit with caveats which are discussed.

3.2 Data

3.2.1 Environmental Variables

A list of variables used for habitat modeling is shown in table 3.1. The variable set was selected based on previous studies as well as known habitat needs, and most variables are based on data from optical sensors that can be annually updated to enable regular monitoring. Canopy cover was included in the set of predictor variables because measurements of forests structure, including canopy cover, were found to be strong determinants of chimpanzee habitat suitability in other studies (Plumptre et al. 2010; Junker et al. 2012; Clee et al. 2015). Torres et al. (2010) found NDVI to be useful for predicting suitable chimpanzee habitat, arguing that it served as a proxy for ecosystem productivity. Here, the interannual mean of NDVI values between the minimum and first quartile (Annual Q1 mean NDVI) was used to reflect dry season productivity, while the mean for values between the third quartile and the maximum (Annual Q3 mean NDVI) was used to reflect wet season productivity. Annual Q1 and Q3 mean NDVI were generated from the GLAD Landsat Analysis Ready Data 16-day composites described in Potapov et al. 2020.

*Table 3.1Variables used to model chimpanzee habitat. * Indicates a variable is temporally static, all others are updated annually.*

Variable	Source	Units
Canopy Cover	Chapter 2	Percent
Edge Influence (EI)	Chapter 2	Unitless
Annual Q1 mean NDVI	Chapter 2	Unitless
Annual Q3 mean NDVI	Chapter 2	Unitless
Distance to Woodland	Chapter 2	Meters
Prop. Woodland in $1km^2$	Chapter 2	Proportion
Prop. Woodland in $31km^2$	Chapter 2	Proportion
Distance to Forest	Chapter 2	Meters
Prop. Forest in $1km^2$	Chapter 2	Proportion
Prop. Forest in $31km^2$	Chapter 2	Proportion
Distance to Steep Slopes > 20˚ *	SRTM	Meters

Slope *	SRTM	Degrees
Aspect *	SRTM	Degrees
Elevation *	SRTM	Meters

In chapter 2, three different forest types were defined by various thresholds of canopy cover, namely, forest (>=50%), woodland (>=10% and <50%), and non-forest (<10%). For the forest and woodland classes, per pixel distance to the nearest patch as well as the proportion of each class within 1km^2 and 31km^2 were calculated. Areas of 1km^2 and 31km^2 were chosen to create variables that reflect the amount of each class on a local scale and within the average chimpanzee home range (i.e., 31km^2) in western Tanzania.

A landscape metric, Edge Influence (*EI*), that has been found to influence species abundances was included in the variable set (Pfeifer et al. 2017). *EI* was calculated from the canopy cover maps produced in chapter 2 and is defined by equation 3.1,

$$EI_{ij} = max\left(\sigma_{c_{ij}}, |\bar{C}_{ij} - C_{ij}|\right) \times sign\left(\bar{C}_{ij} - C_{ij}\right) \quad 3.1$$

where,

EI_{ij} = Edge Influence at pixel *ij*

$\sigma_{c_{ij}}$ = standard deviation of pixels within a 3.1km radius of pixel *ij*

$\bar{C}_{ij}$ = average canopy cover of pixels within a 3.1km radius of pixel *ij*

C_{ij} = percent canopy cover at pixel *ij*

A radius of 3.1km was used to ensure that all canopy cover variation within the home range (i.e., 31km^2) area contributed to *EI*. Absolute values for *EI* range from 0, which indicates homogenous cover (i.e., no edges) within 3.1km of a focal pixel, to 80 where a pixel is

surrounded by different habitat in a 31km^2 area. Positive values are associated with low cover (i.e., non-forest) pixels while negative values are associated with forested pixels.

Topographic features have previously been found to influence nesting site selection in western Tanzania, where nests tend to be found below 1,900 and meters and near steep slopes (Pintea & Plumptre 2006). Topographic information was therefore included from the Shuttle Radar Topography Mission (SRTM) (https://earthexplorer.usgs.gov/) Digital Elevation Model (DEM). In addition to slope, the linear distance to steep slopes (>20^o) was included, and aspect was included to investigate if chimpanzees tend to prefer nesting on a particular azimuth.

3.2.2 Presence Locations

Nest locations were provided by the Jane Goodall Institute (JGI) and cover the period between 2005-2018. Locations were opportunistically collected by park rangers and village forest monitors using smart phones and tablets to record Latitude and Longitude for each nest.

Additional nests from population survey data are from (Piel & Stewart 2014) and (Hernandez-Aguilar 2009). The combined dataset is composed of 10,000 nest locations shown in Figure 3.1.

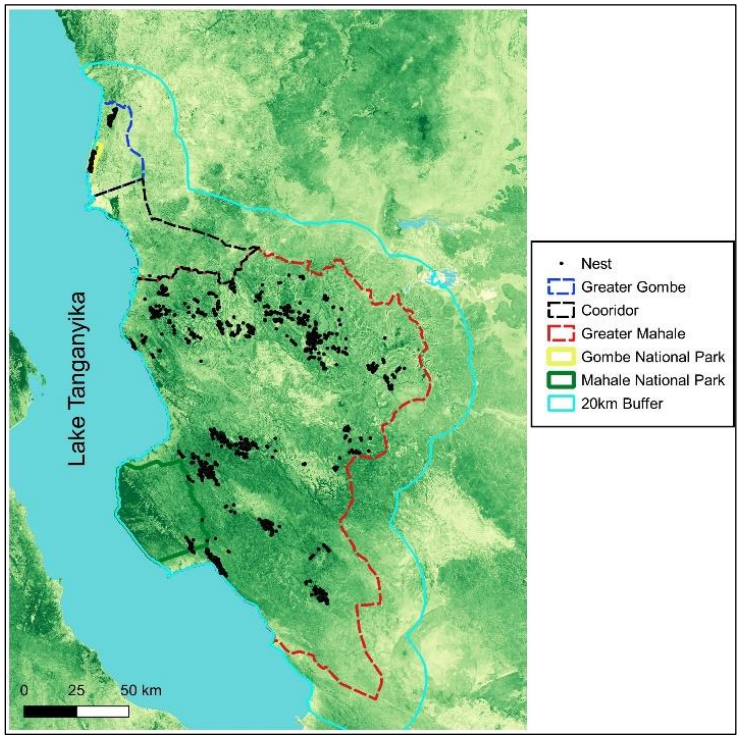


Figure 3.1 Chimpanzee nest distribution in study area.

Table 3.2 Number of nests for each year.

Year	*N nests*	*Year*	*N nests*
2005	962	2012	4565
2006	4	2013	240
2007	36	2014	522
2008	103	2015	220
2009	218	2016	135
2010	236	2017	172
2011	2452	2018	135

3.3 Methods

3.3.1 Habitat Modelling

Typical presence-only species distribution models use presence locations and so-called "pseudo-absences" or background samples with a binary response modelling method (e.g., logistic regression) to produce an index of relative habitat suitability (Ferrier et al. 2002), this methodology seems ad-hoc but has become popular, particularly with machine learning extensions like Random Forests and Boosted Regression Trees. In 2006, the MaxEnt software was released to address these perceived problems with presence-only data models. The method estimates the most uniform distribution (maximum entropy) of sampling points compared to background locations given the constraints derived from data (Phillips et al. 2006) This software has been found to make reliable predictions and has thus been widely adopted with well over 10,000 citations as of this writing. Recent statistical discoveries have demonstrated that all these models are equivalent to or an approximation of an Inhomogeneous Poisson Process (IPP) (Warton & Shepherd 2010; Aarts et al. 2012; Renner et al. 2015). However, framing the problem as an IPP provides guidance on model interpretation as well as model diagnostic tools (described in a later section).

Presence-only data can be defined as set of *n* point events $y = \{y_1 \dots y_n\}$ bound within a two-dimensional region, A, and spatially indexed at location s as $y(s)$. The intensity at $y(s)$, $\lambda(y(s))$ simplified as $\lambda(s)$, is the expected number of point events (i.e., presences) per unit area (Warton & Shepherd 2010; Renner et al. 2015). IPP models have two primary assumptions, (1) the point event locations are independent of one another, and (2) the intensity, λ, varies spatially (Diggle 2003). In the context of modeling a species' distribution, it can be assumed that λ varies

according to collocated environmental variables and can be modelled via a log-linear function with the form,

$$log\ (\lambda_i) = \beta_o + \sum_{j=1}^{k} x_{ij}\beta_j \qquad 3.2$$

where β_o is the intercept, β is the coefficient vector, and x is a matrix of environmental variables. From equation 3.2, only linear relationships can be modeled; however, the features can be expanded, for example by spline transformation, to include non-linear relationships. Modeling of non-linear relationships was enabled by expanding the features from table 3.1 by applying 1st degree splines, equivalent to hinge features in MaxEnt, using the Splinetransformer in Scikit-Learn. Ten knots were linearly spaced along the quantiles of each feature, resulting in a total 155 coefficients to be estimated, including the intercept.

Models were fit with Elasticnet penalty in the Sci-Kit learn software. Elasticnet combines L_1 and L_2 penalties, where L_1 regularization results in sparse coefficient vector (i.e. several coefficients are zero) and L_2 shrinks all coefficients but useful ones less so. In high dimensional problems, L_1 is useful for variable selection but has the pitfall of unstable solutions in the presence of correlated variables. Adding a small amount of L_2 regularization results in a stable solution, where correlated variables tend to be selected, or not, together (Zou & Hastie 2005). There are two hyperparameters for the Elastic Net, α, the proportion of L_1 to L_2 regularization, and γ, the amount of penalty. Here, α was set to 0.9 (i.e. primarily L_1) and γ determined by linearly increasing the amount of regularization until the maximum AUC was achieved on a hold-out sample.

3.3.2 Background Sample Selection

The selection of background points in presence-only modeling continues to be an ad-hoc procedure, and while there is no consensus on exactly how points should be selected, advances have been made that can guide this process. Two important components of background sampling are (1) the sampling strategy and (2) the number of points included in the model fitting process. Several studies have shown that choosing a background sample that mimics the distribution of presence locations results in the best performing models (Phillips et al. 2009; Syfert et al. 2013; Sofaer et al. 2019a) and accordingly, the selection of background samples was constrained by sampling from a bias grid. A bias grid was created by generating a Gaussian kernel density surface for the nest locations, and a random background sample was chosen using the values from the bias grid as weights (Fitzpatrick et al. 2013; Fourcade et al. 2014). The bias grid also acts to constrain background samples to areas near presence locations. Background samples taken from areas where presences have not been sampled can result in an overly optimistic assessment of model performance (Sofaer et al. 2019b).

The number of background samples selected for model fitting should account for the scale and extent of the study area (Feng et al. 2019). Recent advances in modeling techniques enable a more objective way to determine the number of samples rather than using an arbitrary number of points (e.g., 10,000). Warton & Shepherd (2010) showed that the log-likelihood of the Poisson regression model will converge when the number of background samples is sufficient (i.e., adding additional points does not contribute to anymore learning). To determine the background sample size, 10 models were fit with increasingly larger background samples and the log-likelihoods plotted against background sample size. In this case, there is not much use in fitting a model with more than approximately 100,000 background samples (Figure 3.2),

therefore 100,000 background points were sampled for each model run. For each year, background points were sampled to be proportional to the number of presence locations for that year (i.e., if 20% of the presences were recorded in 2010, then 20,000 background samples were selected from the 2010 raster layers).

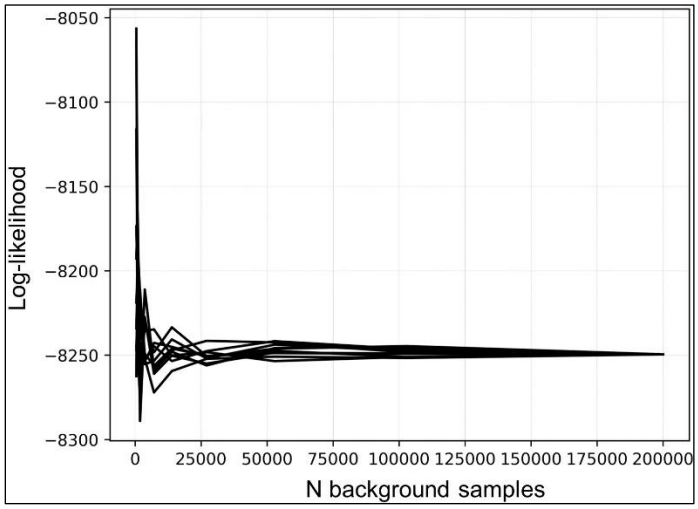


Figure 3.2 Log likelihood for 10 IPP models fit with increasing background sample sizes.

3.3.3 Spatial Bias in Nest Locations

Spatially biased datasets are common in distribution modelling where the bias tends toward easily accessible areas (e.g. roads or rivers), and model quality can be impacted because of over (under) representation of environmental conditions in regions of higher (lower) sampling effort (Kadmon et al. 2004; Anderson & Gonzalez 2011; Kramer-Schadt et al. 2013). There are several methods to correct for sampling bias; however, the simple method of spatial rarefaction has shown robust performance under different types of sampling bias (Fourcade et al. 2014).

The location data used in this chapter come from multiple sources with undocumented sampling effort and thus unknown sampling bias. Moreover, clustering is to be expected because individual chimpanzees in a group construct their nests near other group members. Ripley's

inhomogeneous K function (eq. 3.3.) was used to compare nest locations used to fit the IPP model against 99 simulations of the resultant intensity surface.

$$K(r) = \sum_{y_i}^{N} \sum_{y_{j \neq i}}^{N} \frac{I(\|y_i - y_j\| < r)}{\lambda_i \lambda_j} \quad 3.3$$

where;

r = radius (km)

N = number of points within radius r

$\| y_i - y_j \|$ = euclidean distance between points y_i and y_j

I = indicator function (i.e., 1 if condition is true, 0 otherwise).

λ_i = intensity at point y_i

Data falling above (below) the simulation envelope indicates significant clustering (dispersion) that the model does not account for at the 95% confidence level (Baddeley et al. 2000; Diggle 2003). The minimum distance between neighboring nest locations was increased, using methods from Aiello-Lammens et al. (2015), until no significant deviation from the simulation envelope was found. Figure 3.3 shows that a minimum separation of 1 km between nest locations is needed to ensure no significant clustering.

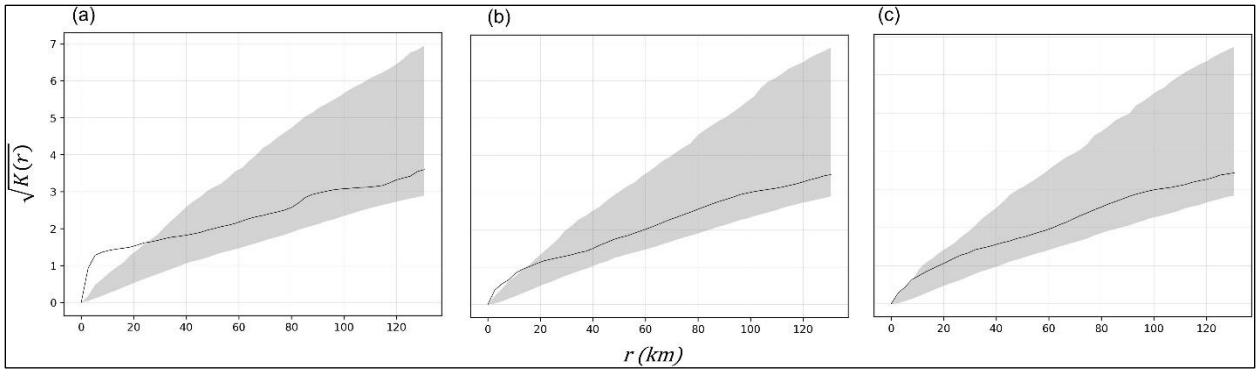


Figure 3.3 Simulation envelopes of IPP models with interpoint distances of 30m (a), 500m (b), and 1000m (c).

The equalizing effect of thinning can be seen in figure 3.4, which shows the kernel density output prior to (a) and after thinning (b). The thinning algorithm randomly selects one point from a cluster where the points are all within 1 km of each other, which means that the set of presence locations can differ with each run of the thinning algorithm. Since the thinning process involves a random selection of points, 50 sets of rarefied points were generated to fit 50 models (max = 490, min = 468, mean = 481) to include as many nest locations as possible in the model fitting process, while also accounting for time needed for model fitting. Each of the models was trained on 80% of the data, while 20% was withheld for cross-validation. Figure 3.4(c) is an example of the surface used to weight background samples for each of the 50 model runs.

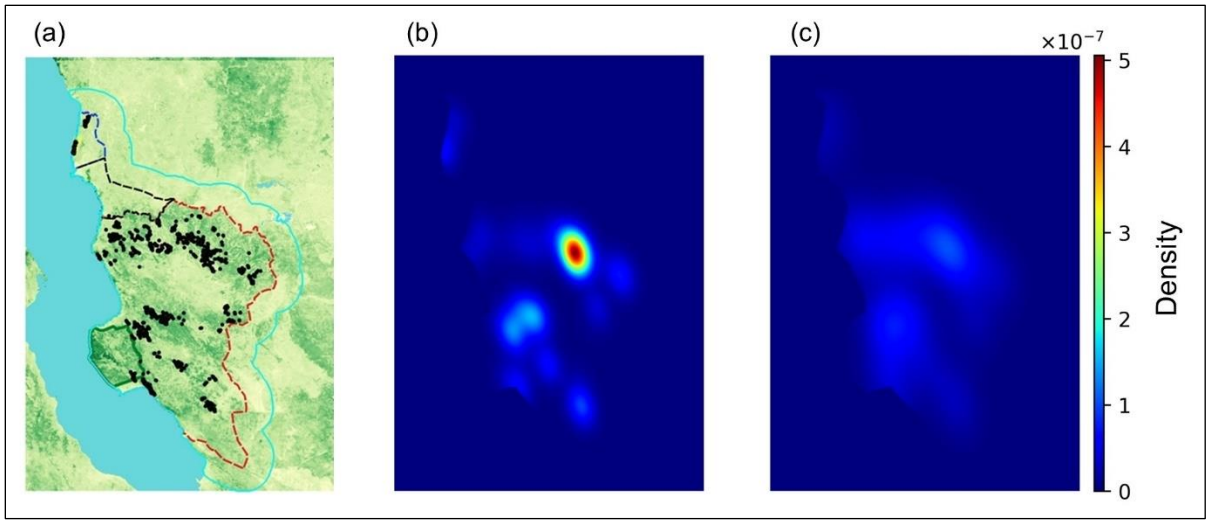


Figure 3.4 distribution of nests in the study area (a) density surface of nests in combined dataset (b), density surface of nests after thinning (c).

3.3.4 Model Prediction and Uncertainty

Once models were fit to the fifty sets of training data, described in the previous section, all models were used to predict annual intensity (λ) for the study area from 2000 to 2020. An annual intensity surface for each model for each year was created by applying the fitted coefficients to the predictor variables from table 3.1, resulting in fifty intensity surfaces per year. Model prediction uncertainty is reported as the variance around the mean of the fifty model predictions.

3.3.5 Model Evaluation

Evaluating the performance of presence only models is an active area of research and there is no single evaluation metric that provides complete information on model performance (Phillips & Elith 2010; Jiménez-Valverde 2012). Since the goal in this chapter is not to create presence/absence maps, two metrics that do not rely on the generation of a binary map were used to assess model fit. The area under the receiver operating characteristic (ROC) curve (AUC) is perhaps most widely used metric in the literature. An ROC curve is obtained by plotting

sensitivity (i.e., true positive rate) against 1-specificity (i.e., false positive rate) for all possible thresholds and the AUC is the area under the curve (Fielding & Bell 1997). However, using AUC as an evaluation metric for presence/background models has received much criticism primarily because presences and absences are equally weighted even though true absences are unknown.

A metric developed specifically for presence background modelling is the Continuous Boyce Index (CBI) (Hirzel et al. 2006), which focuses on correctly predicting presences rather than discriminating between presences and background samples, as does AUC. CBI was developed to assess models that have probabilistic output (i.e., between 0 and 1); however, the output of IPP models have a lower bound of 0 and no upper bound (see equation 3.2). After adjusting the intercept such that model output sums to 1, as opposed to the number of presence locations, a complimentary log-log transform was applied to IPP intensity values (Phillips et al. 2017). If β_o is the fitted intercept, the transformed intercept, β_o', is obtained with $\beta_o' = \beta_o - \log(n)$, where n is the number of presence locations. The transformation has no effect on model predictive performance and ensures that output from each model has the same upper and lower bounds (i.e., 0 to 1). For the remainder of this chapter, the transformed values are referred to as "suitability".

After transformation, suitability is partitioned into several bins and two frequencies are calculated for each bin. The predicted frequency of evaluation points, P_i (eq. 3.4),

$$P_i = \frac{p_i}{\sum_{j=1}^{b} p_j} \qquad 3.4$$

Where p_i is the number of evaluation points predicted to fall into suitability class i and $\sum_{j=1}^{b} p_j$ is the total number of evaluation points. The expected frequency of evaluation points, E_i (eq. 3.5),

$$E_i = \frac{a_i}{\sum_{j=1}^{b} a_j} \quad 3.5$$

Where a_i is the number of points predicted to fall into suitability class i and $\sum_{j=1}^{b} a_j$ is the number background sample points. Finally, the predicted to expected ratio, F_i, is calculated with equation 3.

$$F_i = \frac{P_i}{E_i} \quad 3.6$$

The number of bins for which F_i is calculated is dependent on bin width. Here, a bin width of 0.1 was used. Thus, the first bin spans the 0 to 0.1, and all subsequent bins were created by displacing the moving window by 0.05. For well calibrated models, F_i should be less than 1 in low suitability classes and monotonically increase to values greater than 1 for higher suitability classes. Thus, a plot of F_i versus mean habitat suitability for each class i, provides information on model calibration, where an ideal model would show F linearly increasing with suitability. In addition, cross-validation can also provide information on the confidence of model predictions along the range of suitability values by inspecting the confidence bands across suitability values. Finally, the CBI is obtained by calculating the Spearman Rank Correlation between suitability classes and *F*. Positive values indicate the model is consistent with the distribution of evaluation presence points.

3.3.6 Variable Ranking

Variables were ranked according to the permutation after calibration method (Breiman 2001; Smith & Santos 2020). After each model was fit, three metrics were calculated for the hold-out sample: the Continuous Boyce Index (*CBI*), point biserial correlation coefficient (*COR*), and area under the receiver operating characteristic curve (*AUC*). Each one of these metrics

should decrease after randomly permuting a variable that strongly influences model predictions. The general calculation is shown in equations 3.7 and 3.8.

$$D_i = P_i - P_i' \quad 3.7$$

Where,

P_i = metric value for prediction on hold-out sample

P_i' = metric value for prediction on hold-out sample after random permutation.

D_i = difference between unpermuted and permuted metric value for the *ith* variable.

The relative importance of each variable, I_i, is then calculated with eq 3.8.

$$I_i = \left(\frac{D_i}{\sum_1^k D_k} \right) x\ 100 \quad 3.8$$

3.3.7 Chimpanzee Habitat Indicator for Monitoring Populations (CHIMP)

An indicator of habitat suitability related to chimpanzee abundance was derived by linearly relating model output to ground survey estimates of chimpanzee abundance. This derivation is based on the assumption that model output is correlated with local population abundance, which has been found in several other studies (Weber et al. 2017). CHIMP values are calculated with equation 3.9.

$$CHIMP = c \sum \lambda_i \quad 3.9$$

The summation is performed over a landscape of interest and multiplied by the proportionality constant, c. The proportionality constant (i.e. the ratio of true abundance to model output), must be determined for the model output to reflect abundances across landscapes as it is

impossible for abundances to be determined from presence data alone (Fithian & Hastie 2013; Phillips et al. 2017). Abundance data used to derive the proportionality constant are from a recent line transect survey conducted by the Jane Goodall Institute (JGI), a method that has been used to estimate population densities of numerous species, including chimpanzees (Buckland 2001; Moyer et al. 2006; Piel et al. 2015). In addition to deriving the proportionality constant, a strong linear relationship between model output and abundances estimated from the line transects serves as an independent source for model validation.

The survey was conducted by randomly placing 16 1 km line transects each into 25 survey sites for a total of 400 1 km transects in the Masito-Ugalla region of the GME. Observers walked the length of each transect and recorded the perpendicular distance (up to 50 m) from all detected nests to a transect (The Jane Goodall Institute 2020). A minimum of 60 observations is needed for reasonable population density estimates; however, most of the survey sites did not reach this threshold. The sites were subsequently grouped into 6 strata based on the contiguousness of the sites, habitat types and prior survey history in the area (ibid). Strata and distribution of the transects is shown in Figure 3.5 and strata area in Table 3.2. An average population size of 716 was estimated for the 6 strata and is the value used to determine the proportionality constant for each of the 50 models. A more detailed description of how population densities and abundances are estimated see (Buckland 2001) and for chimpanzees specifically see (Moyer et al. 2006).

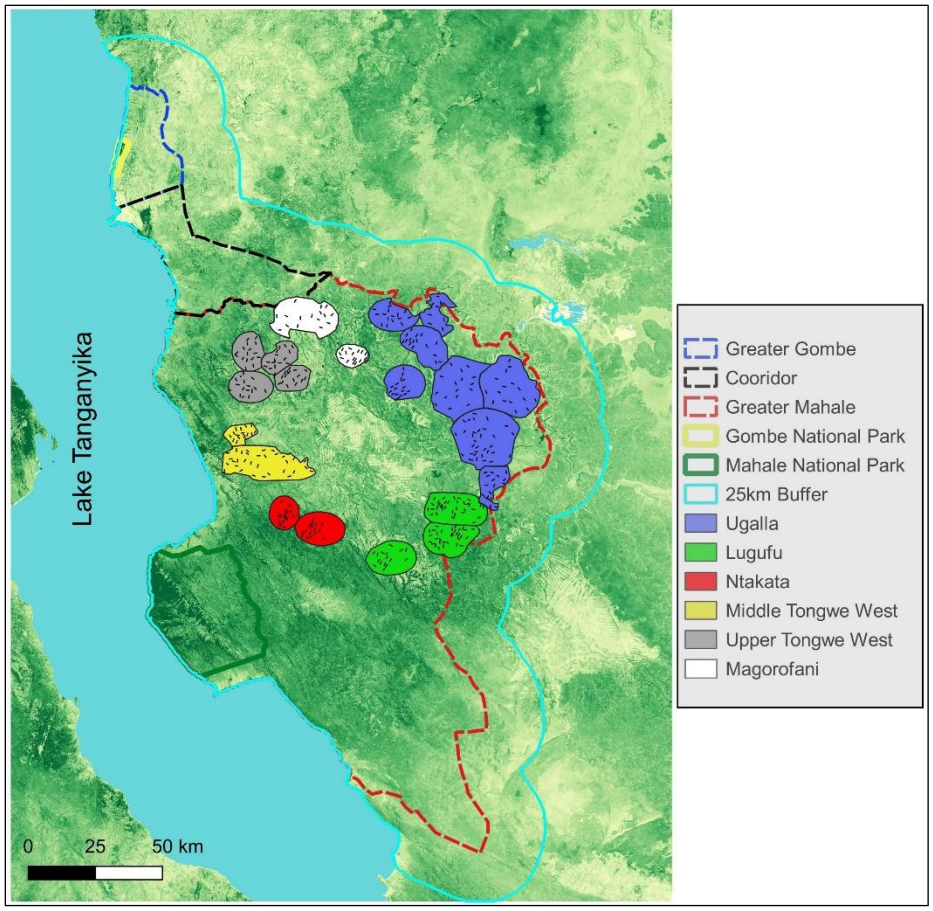


Figure 3.5 Six strata and 400 line transects used for estimating chimpanzee population density in the Masito-Ugalla landscape.

Table 3.3 Area in km^2 of six strata shown in figure 3.5

Stratum	*Ugalla*	*Lugufu Plain*	*Ntakata*	*Middle Tongwe West*	*Upper Tongwe West*	*Magorofani*
Area (km^2)	1913	634	297	410	515	370

After the proportionality constant has been determined, annual predictions can be used to track potential changes in population as a function of change in landscape condition. Here, potential population changes are reported for the major landscapes and protected areas introduced in Chapter 2, and the Mann-Kendall statistic was used to test for monotonically increasing/decreasing CHIMP values between 2000 and 2020 (Kendall & Gibbons 1990) with the pyMannKendall package (Hussain & Mahmud 2019). The Mann-Kendall statistic was

calculated for each of the 50 models, and if the average value was significant at the 0.01 level, a shift in population was considered to have occurred.

3.4 Results

The model ensemble demonstrated reasonable performance with high AUC and CBI values and small standard errors (Table 3.4), and the plot of predicted to expected ratio shows that predictive performance is nearly constant along the range of suitability values and is monotonically increasing, indicating good calibration (Figure 3.6a). CBI and AUC values for the training and testing samples are similar indicating good generalization ability (i.e. no overfitting) (Fig. 3.6).

Table 3.4 Average AUC and CBI values for training and testing data with standard errors in parentheses.

	Train	*Test*
AUC	0.88 (0.001)	0.86 (0.008)
CBI	0.98 (0.002)	0.93 (0.008)

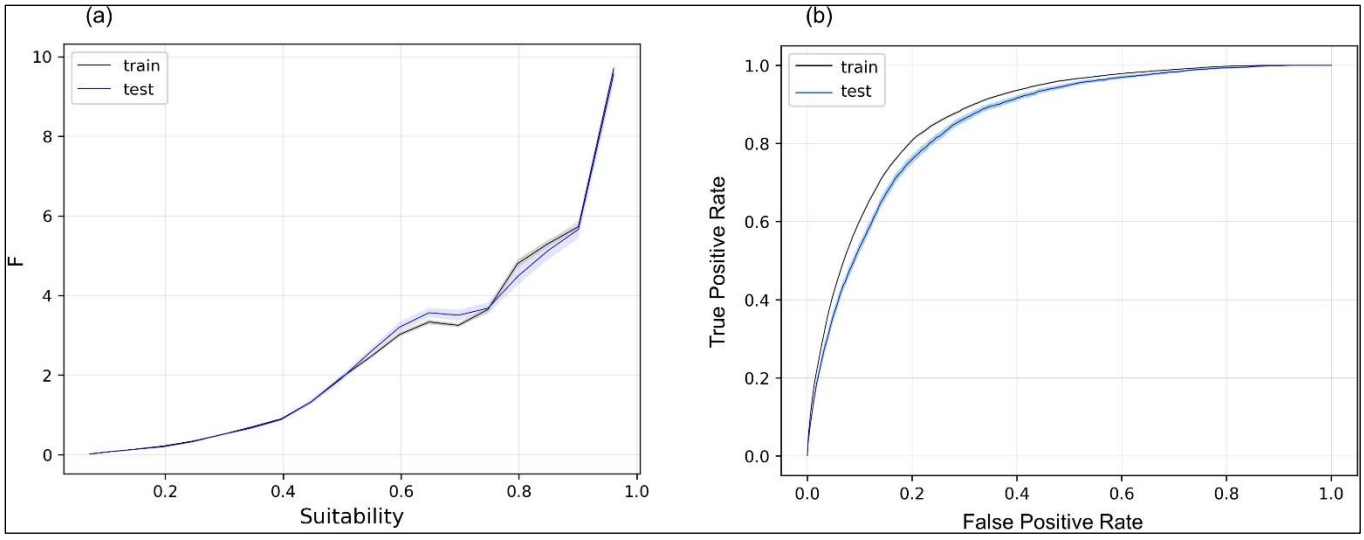


Figure 3.6 Plots of Continuous Boyce index (a), and the Receiver Operating Characteristic for train (black) and test (blue) data.

Variable rankings were similar among the 3 metrics, indeed identical for *CBI* and *COR* for the four most import variables (Fig. 3.6). Edge influence was ranked highest by *CBI* and *COR*, while distance to steep slopes was ranked highest by *AUC*, with edge influence ranked second. All metrics ranked elevation as the fourth most important variable.

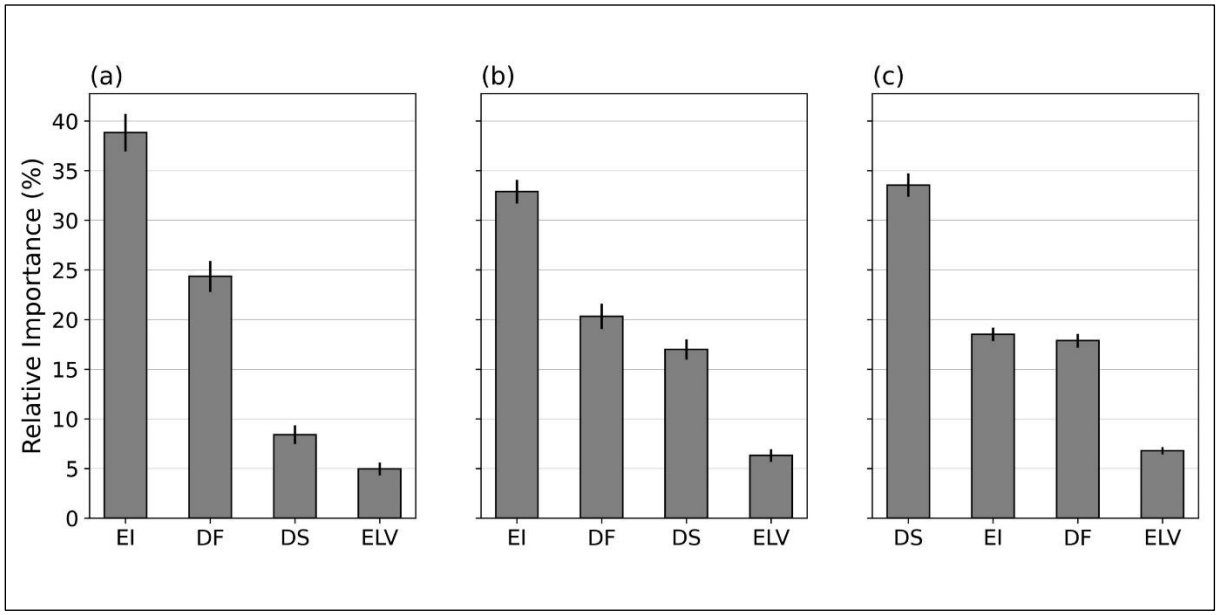


Figure 3.7 Relative importance of permute after calibration for CBI (a), COR (b), and AUC (c) for top 4 variables: edge influence (EI), distance to steep slopes (DS), distance to forest (DF), and elevation (ELV).

Predicted intensity values decline quickly with increasing distance away from forest and steep slopes and increase with elevation up to about 1600m and subsequently declines (Fig. 3.8 b-d). Predicted intensity declines rapidly with increasing edge influence (Fig 3.8a), which indicates chimpanzees preferentially, or are forced to, situate their nests in dense forest surrounded by low forest cover. Figure 3.9 (left panel) shows that, except for Mahale National Park, the highest predicted intensities are along riverine forests, which have strongly negative edge influence values (Fig 3.9 right panel).

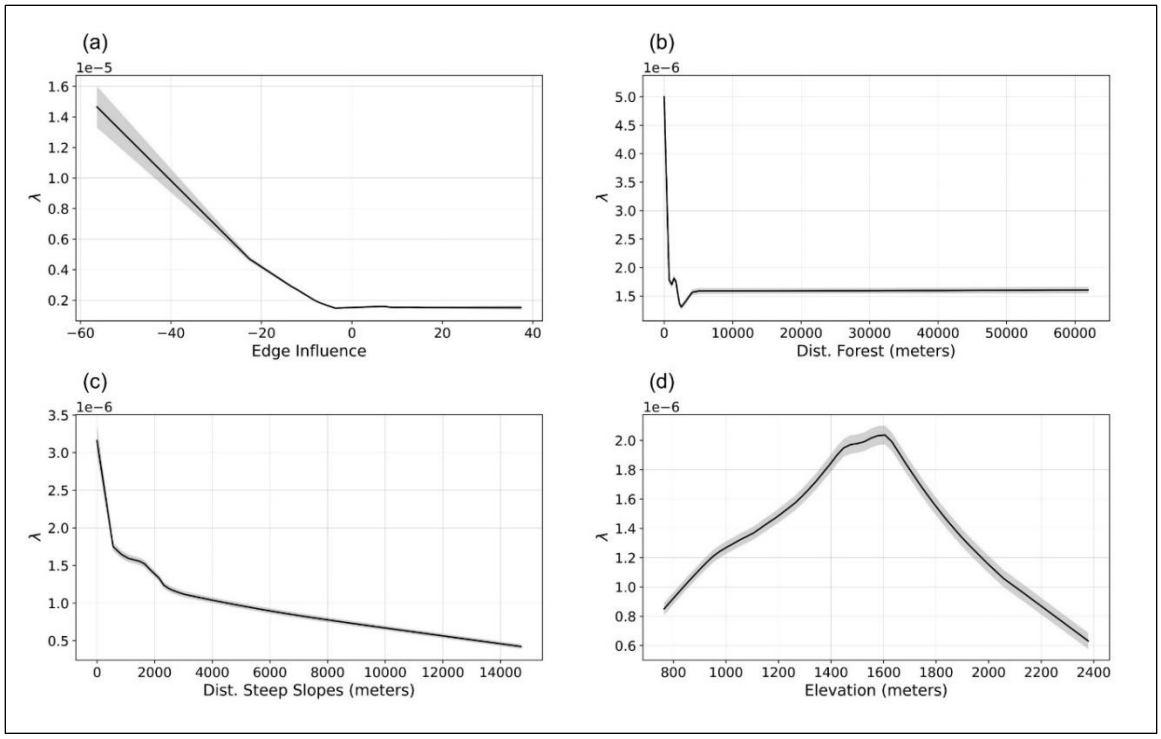


Figure 3.8 Variable response curves for 4 most influential variables with standard errors in gray.

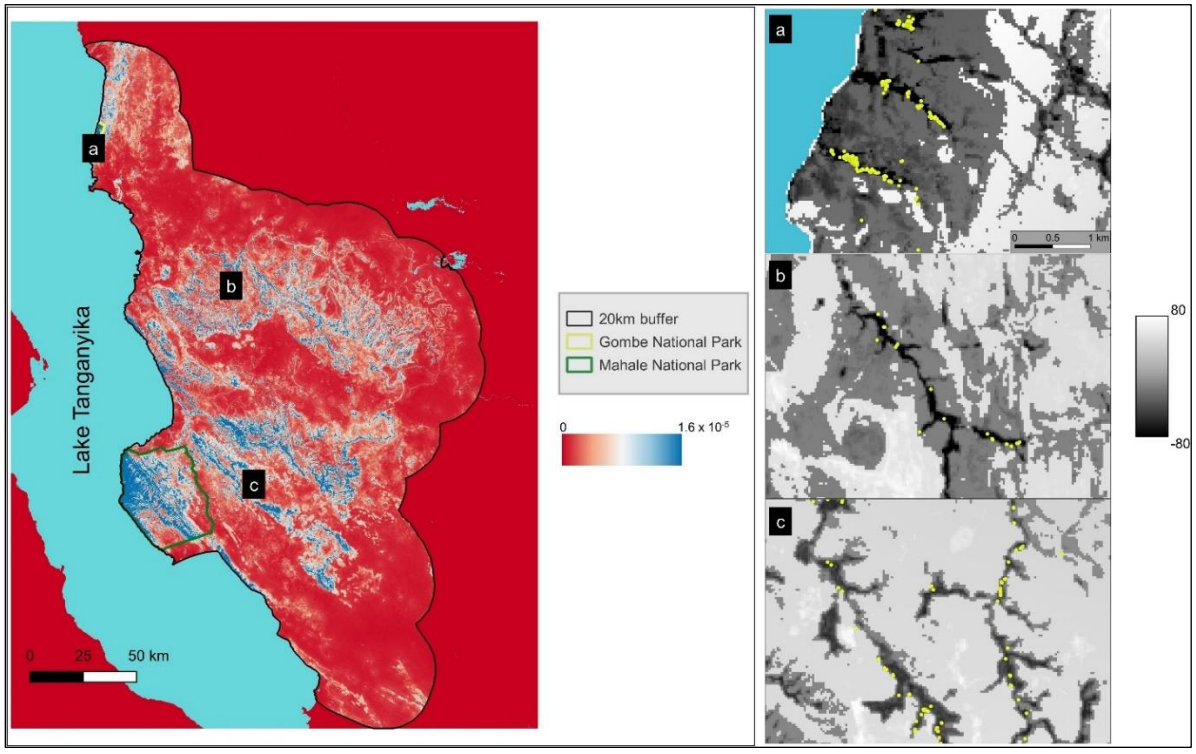


Figure 3.9 Mean intensity surface predicted from fifty models for the study area for 2015 (left panel), Edge Influence overlain with nest locations for Gombe National Park (a), Masito (b), and Ntakata (c).

CHIMP values showed a strong relationship with estimated population abundance (Fig. 3.8(b)), and a weaker but still significant relationship with estimated population density (Fig. 3.8(c)).

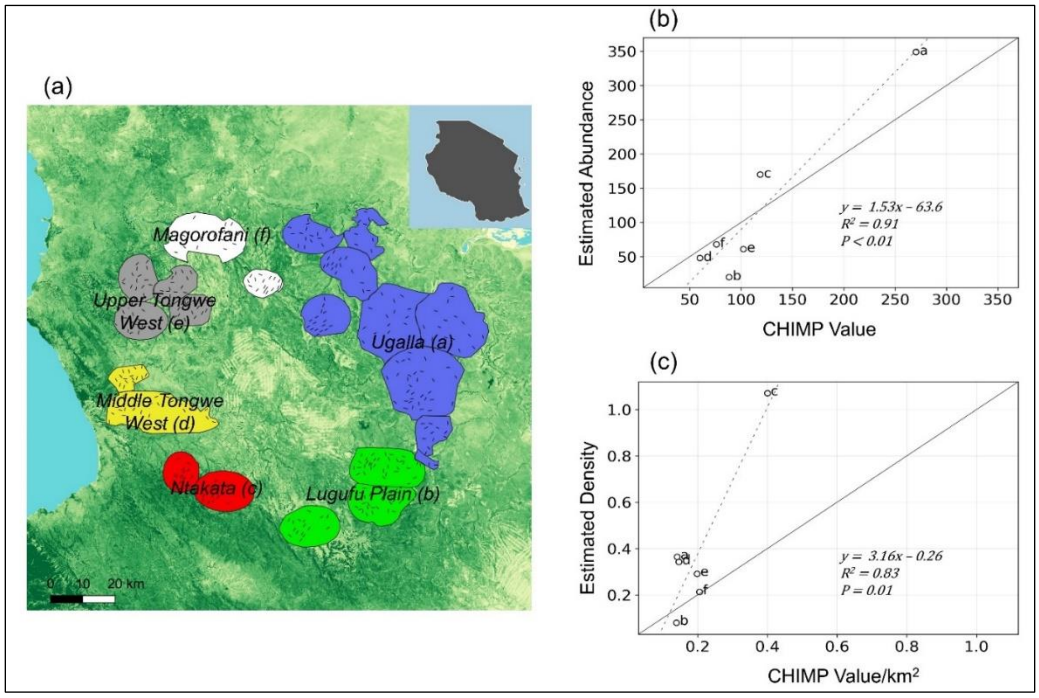


Figure 3.10 Strata and transects used for abundance estimation (a), abundance estimates versus CHIMP values for the 6 strata (b), estimated density versus CHIMP value divided by stratum area (c). Black line is one-to-one relationship and dashed line is estimated linear regression.

CHIMP values and CHIMP/km^2 for the major landscapes in the study area can be seen in table 3.5. The CHIMP value in the GME and GGE is 3000, which is only slightly higher than the upper end of survey based estimates of 3000 individuals in Tanzania (Moyer et al. 2006). The predicted CHIMP value for GNP is 34, almost one-third fewer than the estimated population of 90 (Pusey et al. 2007); however, the CHIMP value for GNP and GGE is 129, which is only slightly higher than the estimated population of 110 (Wilson et al. 2020). The estimated chimpanzee population for MMNP is between 500-700 (TAWIRI 2018), and the CHIMP value of 524 is well within that range.

Table 3.5 CHIMP values, CHIMP/km^2, and 95% confidence intervals for the major landscapes and the whole study area for 2020.

Landscape	*CHIMP*	*95% CI*	*CHIMP/km^2*	*95% CI*
Gombe-Masito-Ugalla Corridor	73	± 2	0.05	0.001
Gombe National Park	34	1	0.93	0.031
Greater-Gombe	95	2	0.13	0.003
Mahale-Katavi Corridor	348	5	0.08	0.001
Mahale National Park	524	11	0.35	0.007
Masito	416	2	0.16	0.001
Masito-Ntakata Corridor	252	3	0.16	0.002
Ntakata	342	2	0.20	0.001
Sitebi Highlands	299	4	0.20	0.002
Ugalla	450	2	0.13	0.001
Ugalla-Sitebi Corridor	59	1	0.13	0.001
Mahale-Ntakata Corridor	154	2	0.19	0.002
Total CHIMP / Mean CHIMP/km^2	3046	19	0.21	0.062

A significant decrease in CHIMP value was predicted for Masito and the Gombe Masito Ugalla Corridor (Fig. 3.11), while significant increases were predicted for Greater Gombe and the Sitebi Highlands (Fig. 3.11). Several village forest reserves were predicted to have significant reductions in CHIMP value, primarily in the Masito region (Fig. 3.12), while the forest reserves in Greater Gombe were either stable or had significant increases.

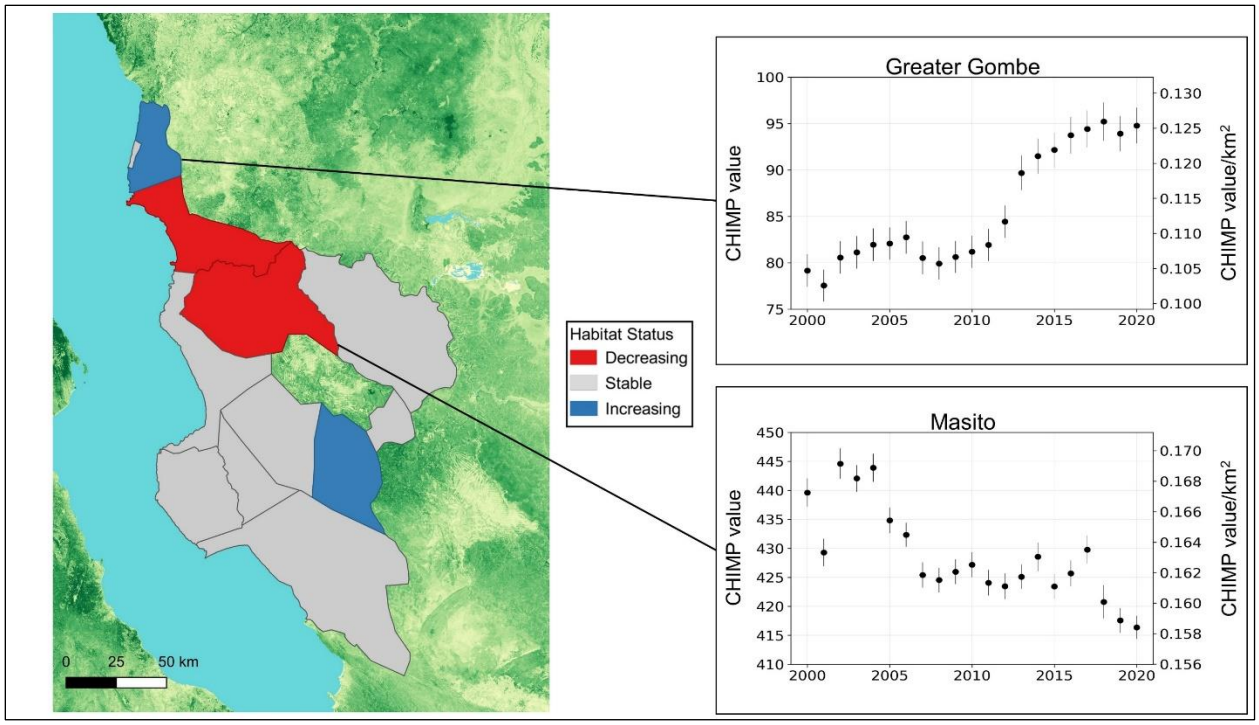


Figure 3.11 Significant negative trends in CHIMP value in red and positive in blue for landscapes in the study area (Left Panel). Example plots of CHIMP values for Greater Gombe and Masito with 95% confidence intervals between 2000 and 2020 (Right Panel).

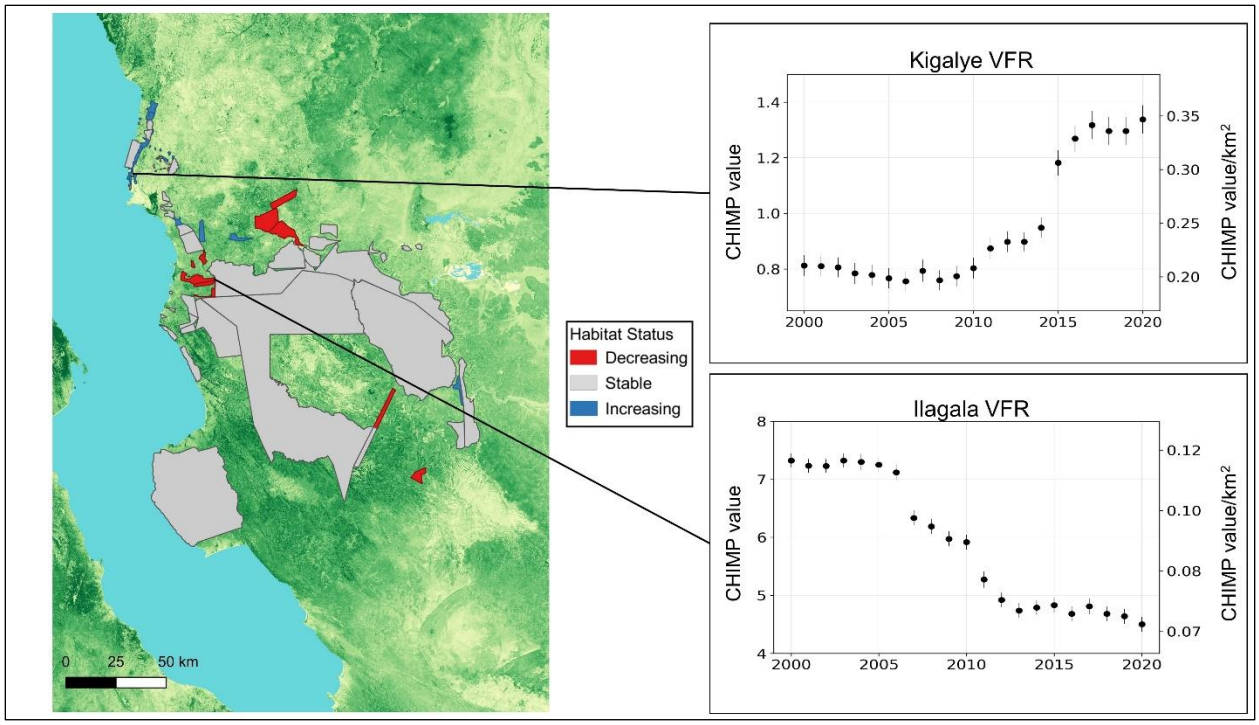


Figure 3.12 Significant negative trends in CHIMP value in red and positive in blue for protected areas and village forest reserves (VFR) (Left Panel). Example plots of CHIMP values for Kigalye VFR and Illagala VFR with 95% confidence intervals between 2000 and 2020 (Right Panel).

3.5 Discussion

Determining the validity of presence-only species distribution models is a challenging task, primarily due to data limitations. Scientists typically want to include as many presences as possible in their models, therefore comparison against an independent dataset is quite rare. Here, multiple lines of evidence show the model output in this study is a good representation of chimpanzee habitat. The cross-validation results showed that the model was not over-fit and well calibrated. These types of discrimination metrics (e.g. AUC) are usually the only reported source of model performance in a large portion of the SDM literature. Spatial dependence, which if not accounted for can result in inflated values of model performance (Dormann et al. 2007), was addressed with Ripley's Inhomogeneous K function, which is also an indicator of model goodness of fit (Warton & Shepherd 2010). Moreover, the response curves of the highest ranked

variables agree with known chimpanzee habitat preferences and landscape configuration. Most importantly, model output was shown to be well correlated with independent estimates of chimpanzee abundance from line transect surveys.

While the model output displayed a strong correlation with survey estimates of chimpanzee density and abundance, it is important to exercise caution when interpreting temporal changes. Notably, a landscape exhibiting a substantial decrease in CHIMP value is more likely to signify an actual population shift than a landscape with a significant increase in CHIMP value. Increases in CHIMP value do not necessarily indicate an increase in chimpanzee abundance; rather, they indicate an improved landscape capable of supporting a larger population. However, the true state of the population can only be determined through data collected from ground surveys. Encouragingly, recent genetic evidence has shown that chimpanzees are dispersing from GNP to the once barren forests north of the park (Wilson et al. 2020), which supports the results presented here. Conversely, a substantial decrease in CHIMP value suggests that the landscape can no longer sustain its previous chimpanzee population, necessitating either migration to alternative habitats or the unfortunate possibility of population decline. A significant decrease in CHIMP value was predicted for the Masito landscape, which agrees with results from Piel et al. (2015) who found a decrease, albeit insignificant, in mean population density estimated from repeated ground surveys.

3.6 Conclusions

This study incorporated annually updated remote sensing data to generate models of chimpanzee habitat at 30 m resolution for the GGE and GME, and abundance data from ground surveys was related to model output to enable the prediction of population size as a function of landscape condition. While model output was well correlated with abundance estimates from

ground surveys, model output should not serve as an alternative for continued population surveys. The relationship between model output and abundance was derived from a single year, therefore the accuracy of predicted changes in population are unknown but does agree with what has been reported in the literature. Data from planned future surveys in the same area could be used to determine if the model faithfully predicts population shifts.

Chapter 4 Employing Remotely Sensed Datasets to Map and Monitor Range-Wide Chimpanzee (*Pan troglodytes*) Habitat Suitability

4. Introduction

Changes in forest cover affect the delivery of important ecosystem services including climate regulation, carbon storage, water quality and supply, and biodiversity richness (Vitousek 1997; Avissar & Werth 2005; Jantz et al. 2014; Pimm et al. 2014). More than one-third of global forest cover has been lost (Defries 2012); rates of forest loss in the tropics, Earth's main reservoir for terrestrial biodiversity, are rising (Myers et al. 2000; Mittermeier et al. 2005; Kim et al. 2015). Forest loss, whether from fire, wood harvest, or clear cutting for agriculture, remains a major threat to species survival (Brooks et al. 1999a; Brook et al. 2006; Hilton-Taylor et al. 2008). In several studies, conservation scientists have employed the species area relationship to estimate the number of species that could eventually become extinct due to the loss of forest habitat (Pimm & Askins 1995; Brooks et al. 1999b; Strassburg et al. 2012). Due to recent advancements in computer processing power, it is now possible to go beyond using assumed area relationships to directly map and monitor physical attributes that are critical for habitat quality such as forest structure, which includes canopy cover and height, as well as forest loss and gain dynamics.

The use of SDMs to model and map chimpanzee habitat suitability is a recent endeavor and has been used to answer questions applicable to chimpanzee conservation and ecology. Vegetation layers derived from Landsat ETM+ data were used in Pintea and Plumptre (2006) to model potential habitat suitability for chimpanzees in Western Tanzania at 90 m resolution.

Landsat TM derived variables were used to generate habitat distribution maps for three time periods between 1986 and 2003 for a small area in Guinea-Bissau, within the range of *P.t. verus*, where a marked decrease in habitat area and connectivity was found (Torres et al. 2010). Habitat suitability was mapped for the *P.t. schweinfurthii* range at 10 km resolution using climate, human impact and vegetation variables to aid in the delineation of areas for new population survey efforts (Plumptre et al. 2010). MaxEnt was used to model *P.t. ellioti* habitat in Nigeria and Cameroon and *P.t. troglodytes* habitat in Cameroon using climate, topographic, human population density and canopy cover variables at 1-km resolution to investigate the role of environmental variation in the genetic dissimilarity between the two sub-species and to assess potential future impacts of climate change on their habitat suitability (Clee et al. 2015). The first range-wide study was performed by Junker et al. 2012, who used climate, human impact, and vegetation variables as model inputs to map habitat suitability for all four chimpanzee sub-species, and all African Great Apes, at 5-km resolution for both the 1990s and 2000s and found a decrease in suitable habitat over time.

In this chapter, previous efforts to map chimpanzee habitat suitability were extended based on remotely sensed datasets that cover the entire range of all four chimpanzee sub-species and are regularly updated. Due to the inaccessibility of presence locations across the entire chimpanzee range, a coarse resolution habitat suitability map was rescaled to 30 m resolution using variables derived from remotely sensed datasets as input into a Random Forests regression model. The results are relevant to conservation planning and measuring success as they portray finer scale habitat variation over several time periods and can be used in monitoring plans by providing updated habitat status at regular time periods.

4.3 Data

4.3.1 Study Area

The study area covers the ranges of the four chimpanzee sub-species, which span central and western Sub-Saharan Africa over an approximately 2.3 x 10^6 km^2 area (Caldecott & Miles 2005) (Figure 4.1), similar in size to one-third of the contiguous United States land area. Chimpanzees occupy a wide variety of forest habitats across their range, including humid evergreen forests, mosaic woodlands, deciduous forest, and dry savanna woodlands (Kortlandt 1983). Their habitats also cover a large elevation range, from sea-level in West Africa up to 2,600 meters in East Africa (Kortlandt 1983). For our study, we used geographic range extents for each of the sub-species from the IUCN Red List. We buffered each range by 100km to cover any suitable habitat that may fall outside the delineated range.

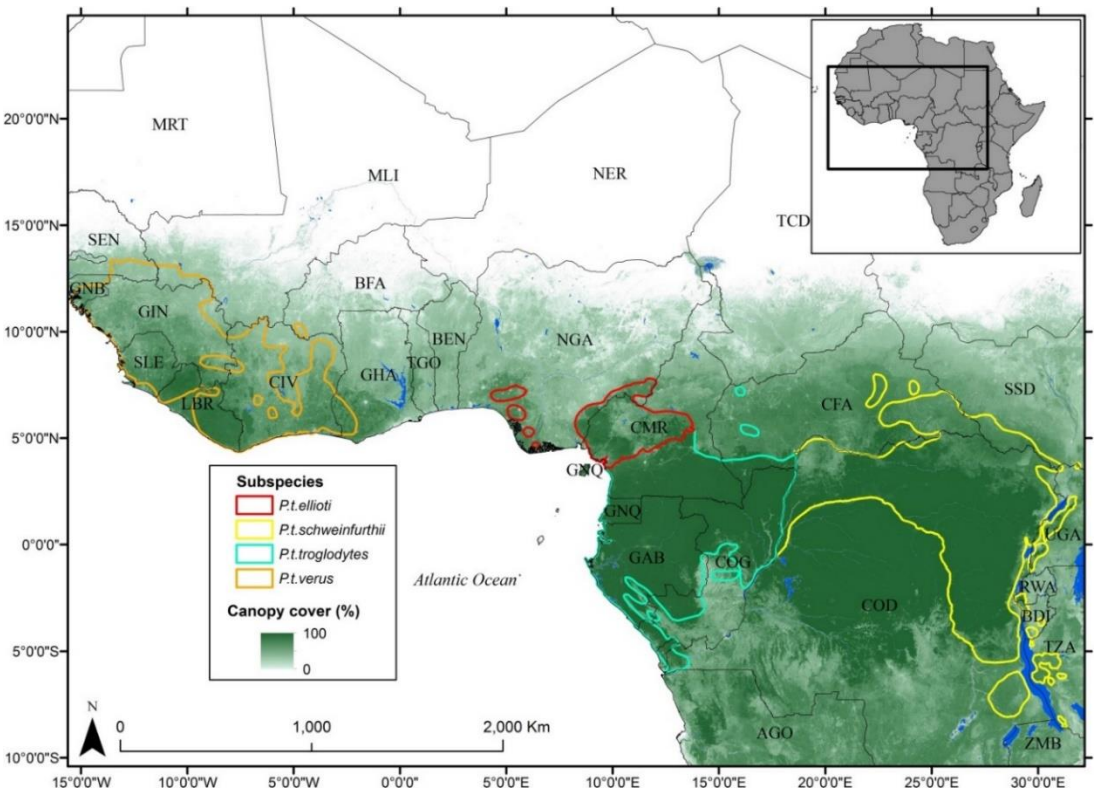


Figure 4.1 Overview of study area with IUCN Red List ranges of the four chimpanzee sub-species overlaid on top of a map of percent canopy cover from Hansen 2013.

4.3.2 Habitat Data

Data on the spatial distribution of habitat suitability for most of the study area are from (Junker et al. 2012), who used location data from the IUCN/Species Survival Commission A.P.E.S. (Ape Populations, Environments and Surveys) database (http://apes.eva.mpg.de), which serves as a central repository for all ape related survey data. Unfortunately, the survey data are not publicly available. In lieu of training a habitat suitability model with survey data, a regression model was trained to relate satellite derived features to the output from Junker et al. (2012). Their estimates of habitat suitability are therefore foundational for this study.

A two-stage approach was used to develop the HSMs in (Junker et al. 2012). First, they related environmental variables to chimpanzee presence-only data to create suitability maps using MaxEnt (Phillips et al. 2006), a software package widely used to generate habitat

suitability models where presence only data are available. MaxEnt uses records of presence only locations and a set of environmental predictors across a defined landscape. The software extracts background samples, where presence is unknown, to contrast against the known presence locations. The software then generates models using the principle of maximum entropy, which asserts that models should be chosen that are similar to prior expectations but are also consistent with the data (ibid). The initial distribution is usually uniform in geographic space (i.e., all cells in a landscape are equally likely to contain an individual), but other distributions are possible based on prior knowledge of the sampling design. Parameters are estimated for exponential models and the final model selected is most similar to the prior distribution (i.e., uniform) and whose predictions match the mean and variance of the environmental data at observed locations (Merow et al. 2013).

Location data covering the dry forests and miombo woodlands of western Tanzania were not included in the APES database; therefore, these important chimpanzee habitats were not reflected in the output from found in (Junker et al. 2012). In order to include these habitats in the model, a habitat suitability model created by (Pintea & Plumptre 2006) specifically for this region (Figure 4.2B) was incorporated. Mahalanobis distance was used to relate chimpanzee location data, collected as part of biodiversity surveys conducted by (Moyer et al. 2006), to three environmental predictors: distance to the nearest forest patch, distance to steep slopes (i.e., slopes greater than 20˚) and elevation (ibid). In the context of habitat suitability modeling, the Mahalanobis distance represents the squared distance, in feature space, of the vector of values for the environmental predictors at un-sampled locations to the mean vector of values for the same predictors at sampled locations (Farber & Kadmon 2003). Mahalanobis distances are approximately Chi-square distributed with n degrees of freedom where n is the number predictor

variables, therefore a location that has a vector identical to the mean vector will have a typicality of 1.0 while larger distances will approach a typicality of 0 (Clark et al. 1993). In order for the prediction from this model to be on the same scale as (Junker et al. 2012), Mahalanobis distances were converted to typicality probabilities by assuming a Chi-square distribution with three degrees of freedom. The suitability map was then coarsened from its native 90m resolution to approximately 5km resolution and combined with the 2000s prediction from Junker et al. (2012).

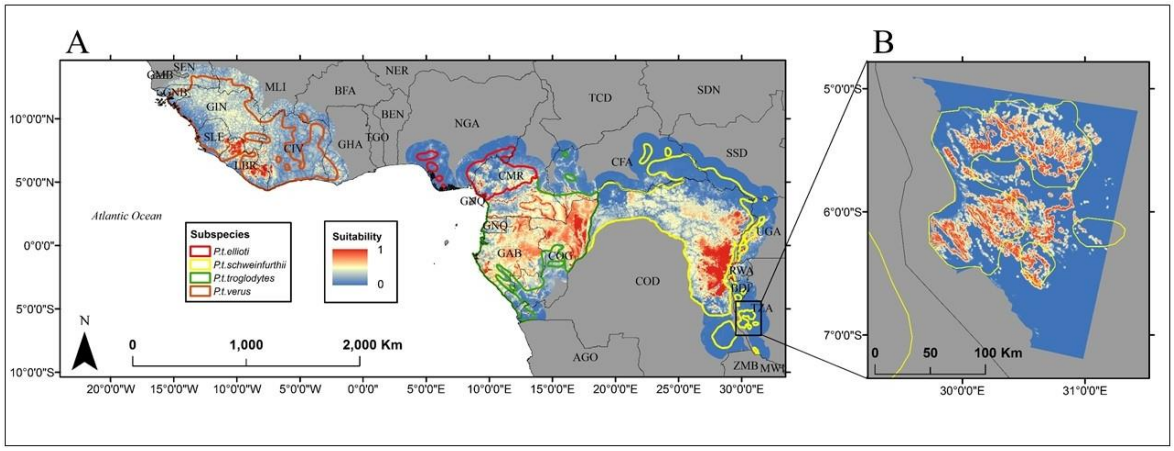


Figure 4.2 (A) Habitat suitability data used for Random Forest calibration. The map for the 2000s time period from Junker et al., (2012) and (B) suitability map created specifically for the miombo woodlands of western Tanzania from Pinta et al. (2006).

4.3.3 Satellite Data

Spectral reflectance, forest structure and forest loss data were from a recent processing and synthesis of the Landsat ETM+ data archive performed by Hansen et al. (2013). This dataset consists of global annual composites of percent canopy cover, canopy height, and top of canopy reflectance for Landsat ETM+ bands 3, 4, 5 and 7 as well as annual data of forest loss from 2000 to 2012. The global products are in a geographic coordinate system and were disaggregated into 10˚ x 10˚ tiles with a resolution of 0.00025 decimal degrees (~30m). To reduce the effects of noise and to fill data gaps due to persistent cloud cover, the annual composites were divided into four time periods by applying a three-year median to the canopy cover, canopy height, and

spectral band layers, resulting in datasets representing 2000-2003, 2004-2006, 2007-2009, and 2010-2012 time periods. The method for generating Landsat based canopy cover and forest loss data can be found in (Hansen et al. 2013), however the canopy height data were not included as part of the initial data release. Global tree height maps were generated on an annual time scale using a regression tree model that related height measurements from GLAS LiDAR shots to Landsat time-series metrics. Model error was quantified by using a random 90% subset of the available GLAS data to train the model while the remaining 10% was used for testing model performance. Root mean square error and mean absolute error for the tropics was 8.1 meters and 5.9 meters, respectively (Tyukavina et al. 2015; Hansen et al. 2016).

Topographic data for the study region were from the Shuttle Radar Topography Mission (SRTM) 3 arc-second resolution dataset and locations of rivers were from the SRTM Waterbodies Dataset (SWBD). The SWBD was created as a by-product of the finished SRTM DEM and includes water bodies detectable at approximately 30m resolution at the time of the SRTM shuttle flight, circa February 2000 (USGS 2003). The SWBD is distributed in vector format and was rasterized to ~30m resolution to match the resolution of the Hansen et al. (2013) dataset. The elevation and slope layers were also resampled to ~30m to match the resolution of the Hansen et al., (2013) dataset.

4.3.4 Environmental Variables and Calibration Data

A full list of predictor variables used to construct the habitat suitability models is shown in Table 4.1. The set of predictor variables were created based on the variables used to generate the HSMs in Junker et al., (2012) and variables used in other studies on chimpanzee habitat suitability. Canopy cover and canopy height were included in the set of predictor variables because measurements of forest structure (e.g. percent canopy cover) were found to be important

in the determination of chimpanzee habitat suitability in several previous studies (Plumptre et al. 2010; Junker et al. 2012; Clee et al. 2015). Spectral metrics, namely NDVI, were found to be an important predictor in Torres et al., (2010), which they argued serves as a proxy for ecosystem productivity. For this reason, all Landsat ETM+ bands and band derivatives from the Hansen et al. (2013) dataset were included because several studies have found an association between spectral data and forest structure metrics (Helmer et al. 2010; Baccini et al. 2012; Wilkes et al. 2015).

Distance to the nearest forest patch was used as a predictor variable both in the habitat suitability models for western Tanzania (Pintea & Plumptre 2006) and in (Torres et al. 2010) and was therefore included in the set of predictor variables. In this study, forest was defined as 40% canopy cover with a minimum patch size of 0.5 ha, which is a combination of definitions used by the United Nations Environment Programme (UNEP) and Food and Agriculture Organization (FAO) for closed forests (Singh et al. 2001; Food and Agriculture Organization of the United Nations 2006). A threshold of 40% or greater was applied to the canopy cover layers and patches smaller than 5 Landsat pixels or approximately 0.5 ha were removed. These layers were used to calculate the minimum per pixel distance to the nearest forest patch for all four time periods.

Human impacts on chimpanzee habitat were found to be important in Plumptre et al. (2010), Junker et al. (2012), and Clee et al. (2015) who used some form of gridded human population as a predictor variable in their habitat models. Deforestation and forest fragmentation are the direct result of human activity and are known to adversely affect biodiversity (Vitousek 1994). In this study, forest loss and fragmentation metrics were used as proxies for human impact on chimpanzee habitat. Forest edge layers were created for each time period by calculating the per pixel distance from the interior of each forest patch to the nearest forest edge.

In addition, the proportion of forest edge within a buffer of 1km and 25km was calculated for each pixel. A buffer of 1km was chosen to capture the influence of local scale fragmentation while the 25km diameter represents the upper limit of a chimpanzee's home range (Goodall 1986; Jones et al. 1996; Nowak 1999).

The forest loss data from Hansen et al (2013) was used to create four binary loss/no-loss layers for 2000-2003, 2004-2007, 2008-2009, and 2010-2012 time periods. To clean the data, pixels that were labeled as forest-loss in non-forest pixels were removed and the same minimum size to define forest patches (i.e. 0.5 ha) was used. These layers were used to calculate the per pixel distance to the nearest forest loss patch and the proportion of loss in 1km and 25km diameters, respectively.

Topographic information including elevation, slope and distance to steep slopes were used as predictor variables because previous studies have found they factor into the prediction of chimpanzee habitat suitability. Chimpanzee nests were generally found to be absent in forests and woodlands at elevations greater than approximately 1,900m (Pintea et al., 2006, Plumbtre et al., 2010). They also found that chimpanzees may have preferred to nest on steep slopes, with a grade greater than 20˚. Slope was found to be a prominent factor in the prediction of habitat suitability in Clee et al. (2015), particularly for the *P.t.ellioti* sub-species. Finally, the rasterized SWBD dataset was used to calculate the per pixel distance to rivers because Junker et al. 2012 found proximity to rivers to be a significant predictor of habitat suitability in all their models.

All predictor variables for the 2000-2003 time period were resampled from 30m to 5km to match the spatial resolution of the calibration dataset. Values for the habitat suitability layer and predictor layers were extracted for each of the four sub-species ranges resulting in 18,000, 71,500, 45,000, 48,000 samples for *P.t.ellioti*, *P.t.schweinfurthii*, *P.t.troglodtyes* and *P.t. verus*,

respectively. Fifty boot-strapped training and testing datasets were generated for each sub-species by randomly sampling 75% of each dataset for model calibration and 25% for model validation. An additional fifty training and testing datasets were generated that represent the entire chimpanzee range by combining training and testing datasets from each sub-species. Habitat suitability was then predicted from the range-wide, 5km resolution calibration map using a suite of remotely sensed derived metrics (Table 4.1).

*Table 4.1List of predictor variables used as input to Random Forest regression models. * indicates variable was included in final model. See section 4.3.4.1 for details on variable selection.*

Variable	Units	Source	Abbreviation
Landsat ETM+ band 3	% reflectance	Hansen et al. 2013	B3
Landsat ETM+ band 4	% reflectance	Hansen et al. 2013	B4
Landsat ETM+ band 5	% reflectance	Hansen et al. 2013	B5*
Landsat ETM+ band 7	% reflectance	Hansen et al. 2013	B7
Normalized band4/band3	Unitless	Hansen et al. 2013	NormB4/B3
Normalized band4/band5	Unitless	Hansen et al. 2013	NormB4/B5
Normalized band4/band7	Unitless	Hansen et al. 2013	NormB4/B7*
band3/band5	Unitless	Hansen et al. 2013	B3/B5*
band3/band7	Unitless	Hansen et al. 2013	B3/B7
band5/band7	Unitless	Hansen et al. 2013	B5/B7
Canopy cover	Percent	Hansen et al. 2013	CC*
Canopy height	Meters	Hansen et al. 2013	HT*
Distance to forest	Meters	Hansen et al. 2013	DF
Distance to forest loss	Meters	Hansen et al. 2013	DL
Forest loss in 1km buffer	Proportion	Hansen et al. 2013	L1K
Forest loss in 25km buffer	Proportion	Hansen et al. 2013	L25K*
Distance to forest edge	Meters	Hansen et al. 2013	DE*
Forest edge in 1km buffer	Proportion	Hansen et al. 2013	E1K
Forest edge in 25km buffer	Proportion	Hansen et al. 2013	E25K*
Distance to rivers	Meters	SWBD	DR*

Distance to steep slopes	Meters	SRTM	DS*
Elevation	Meters	SRTM	EL*
Slope	Degrees	SRTM	SLP*

4.4 Methods

4.4.1 Habitat Suitability Estimation

Random Forests regression was used to estimate range-wide chimpanzee habitat suitability with the Sci-kit Learn module in the python software environment (Pedregosa & Varoquaux 2011). Random Forests is an ensemble technique in the Classification and Regression Trees (CART) family of machine learning methods (Breiman et al. 1984). Random Forests works by constructing several regression trees based on bootstrapped samples of the input dataset and each node is split using the best from a random subset of predictors. The final prediction from a Random Forest is the aggregate prediction from all trees. Random Forests have the attractive properties of being insensitive to non-normally distributed variables, non-linear relationships, and co-linear predictor variables (Breiman 2001). In addition, only two primary parameters are required for Random Forest models; the number of trees constructed in each forest and the number of variables to consider at each node split.

Random Forests regression models were fit using satellite data from 2000-2003 to have a close temporal match with Junker et al. (2012). The climate, demographic and topographic variable data used in Junker et al. (2012) are not updated regularly and could presumably be used to predict habitat suitability over a wide temporal range. Canopy cover metrics, on the other hand, were from AVHRR and MODIS continuous field estimates representing forest in the first few years of the 2000s decade.

While Junker et al. (2012) created a separate model for each sub-species, a general, range-wide model was investigated to determine if it could reasonably predict habitat suitability for each of the sub-species. An initial Random Forests regression model was constructed for each bootstrapped training sample for each sub-species as well as a general, range-wide model from the pooled sub-species data. Predictions were made from each model on the test data and Root Mean Square Error (RMSE) was recorded for each bootstrapped sample. A preliminary analysis indicated that using a range-wide model to predict suitability for each of the sub-species' ranges did not result in a substantial increase in error for any of the sub-species (Figure 4.3), the remainder of this chapter will therefore focus on the development of a general, range-wide model.

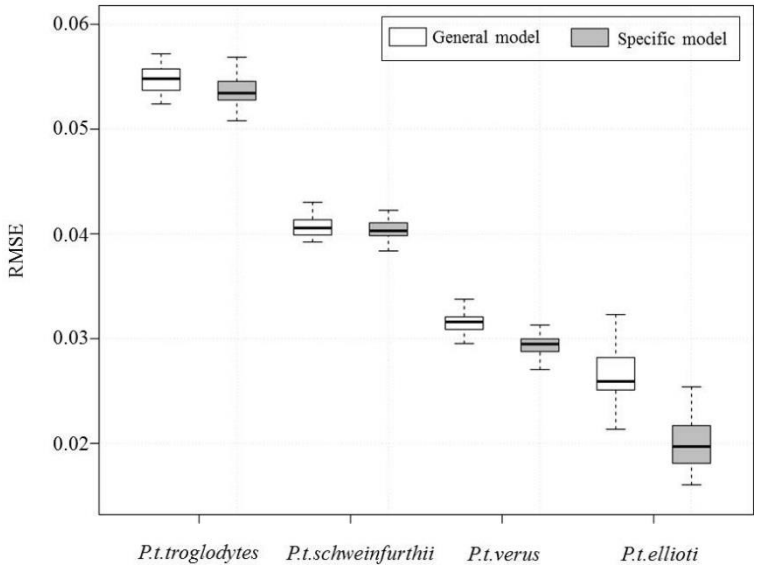


Figure 4.3 Boxplots of 50-fold cross-validated RMSE of a general, range-wide model (white) and four models trained on data for each of the individual sub-species (grey).

4.4.2 Selection of Predictor Variables

To create the most parsimonious model possible, highly correlated variables were removed in addition to variables that contributed little to the predictive capability of the model. A correlation matrix was created for all predictor variables to determine the coefficient of

determination for each combination of variables. For each pair of highly correlated variables (r^2>0.9), the coefficient of determination was calculated between each variable and habitat suitability and the predictor variables with the highest coefficient of determination were retained. Canopy height and percent canopy cover were found to have a correlation coefficient greater than 0.9; however, both variables were retained because of their recognizable ecological importance to chimpanzee habitat. Next, recursive feature elimination was used to remove variables that had little influence on the model's predictive power. Recursive feature elimination operates by first ranking all variables by a measure of importance. Here, variables were ranked by the mean increase in RMSE when the values for a particular variable were randomly permuted (Strobl et al. 2007). The variable that resulted in the smallest increase in RMSE was dropped and a new model was created, and new ranks were then calculated for each variable. This process continued until RMSE reached a minimum (Granitto et al. 2006). The lowest RMSE occurred when a subset of 12 variables was used to construct Random Forest models. Mean RMSE for models created using a subset of variables (0.041 ± 0.002) was slightly lower but not significantly different than the mean RMSE (0.042 ± 0.003) from models created using the full set of variables.

4.4.3 Habitat Suitability Prediction

All samples from the calibration dataset were used to train the final Random Forests model of habitat suitability. It was found that out of bag (i.e., the sample withheld during the construction of trees) error stabilized when a forest was created with 250 trees and 7 of the variables considered at each node split. The final Random Forest model was used to create range-wide habitat suitability maps for the 2000-2003, 2004-2006, 2007-2009 and 2010-2012 time periods at the 30m Landsat pixel scale.

4.5 Results

Similar spatial patterns of habitat suitability were found between the calibration data (Figure 4.4a) and Random Forest model predictions (Figure 4.4b). Moreover, chimpanzee habitat suitability was modeled reasonably well compared to the calibration data, explaining 82% (±0.2%) of the variance for the entire chimpanzee range (Table 4.2). However, there was considerable variation in the model's predictive capability among the four sub-species. The model was able to explain 35% (±1.95%), 89% (±0.18%), 66% (±0.47%), and 73% (±0.44%) of the variance for *P.t. ellioti*, *P.t. schweinfurthii*, *P.t. troglodytes*, and *P.t. verus*, respectively (Table 4.2). Compared to the calibration data, the model tended to underestimate habitat suitability (Figure 4.5 a,b). In general, areas where suitability was underestimated occurred at low elevations with sparse canopy cover, while the opposite was true for overestimated areas. Linear error patterns could be seen within the range of *P.t. troglodtyes* (Figure 4.5A), which resulted from a spatial mismatch of rivers between this study and Junker et al. (2012) due to differing river network datasets.

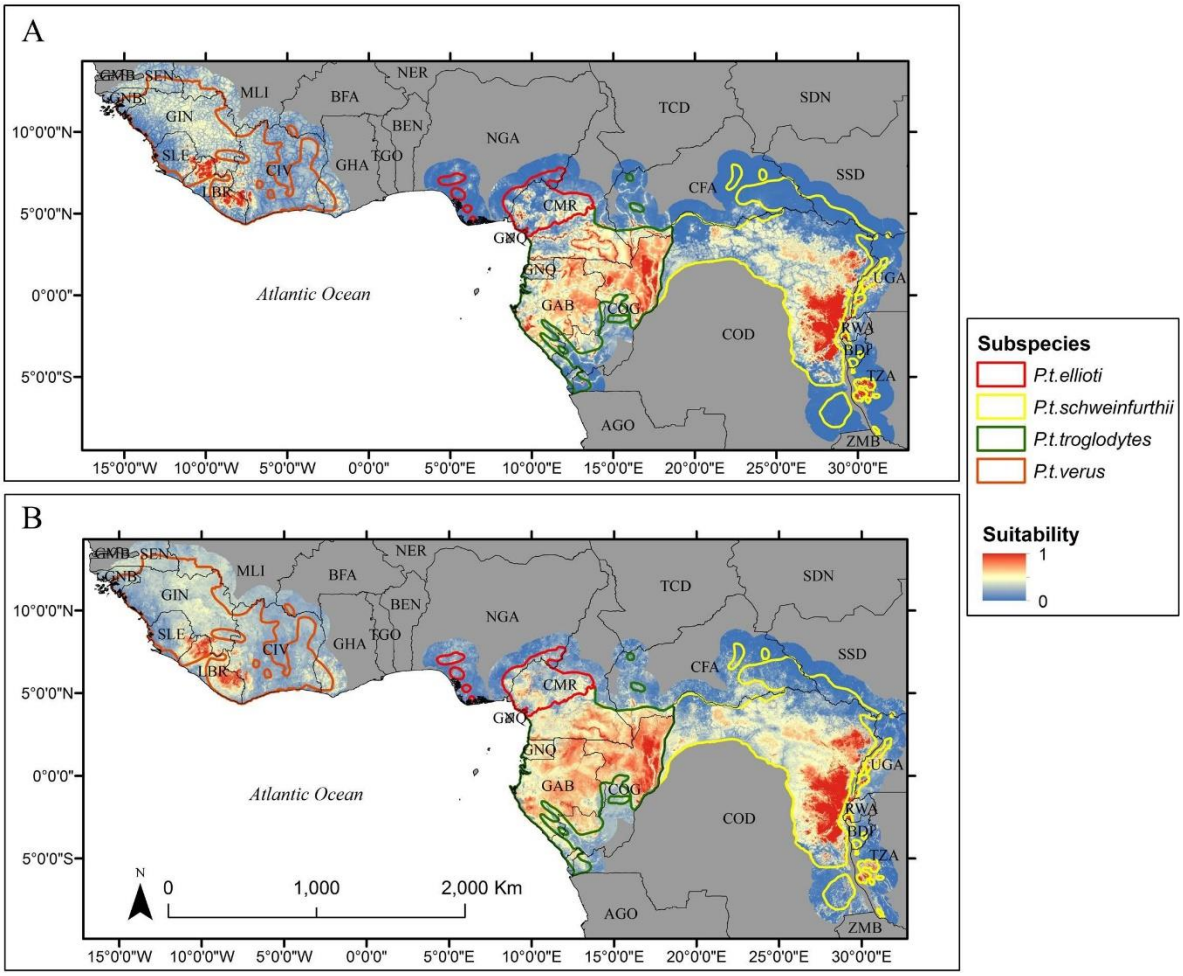
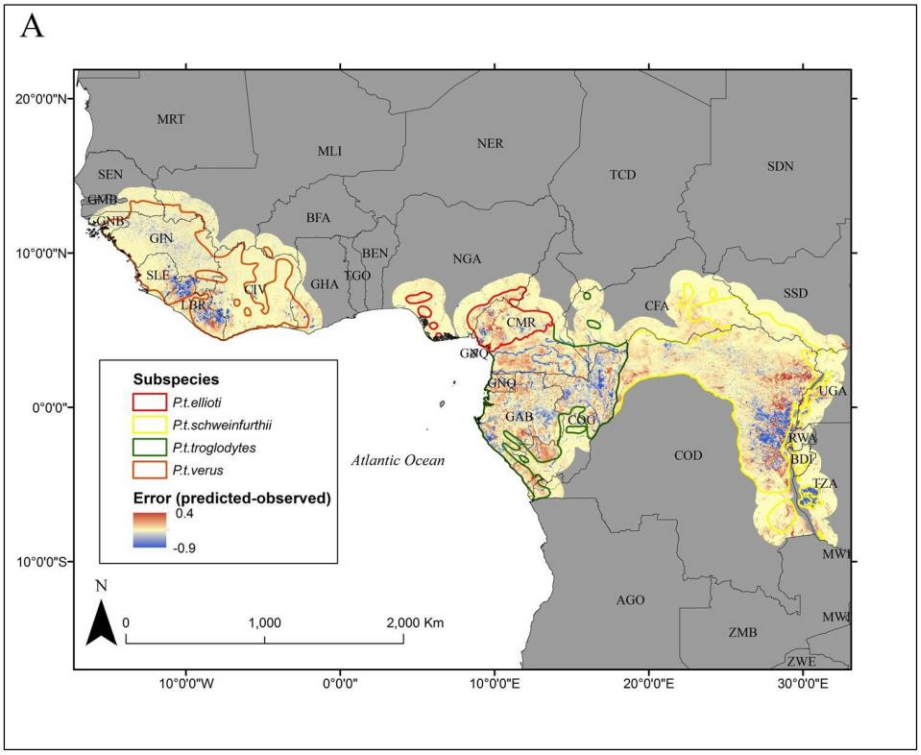


Figure 4.4 Spatial distribution of suitability values for 5km resolution map (A), and Random Forests prediction at 30m resolution (B).


A
20°0'0"N
10°0'0"N
0°0'0"
10°0'0"S
MRT
MLI
NER
TCD
SDN
SEN
GMB
GNB
GIN
BFA
BEN
NGA
SLE
LBR
CIV
GHA
TGO
CFA
SSD
CMR
GNQ
GAB
COG
COD
UGA
RWA
BDI
TZA
MWI
AGO
ZMB
ZWE
Atlantic Ocean
Subspecies
P.t.ellioti
P.t.schweinfurthii
P.t.troglodytes
P.t.verus
Error (predicted-observed)
0.4
-0.9
N
0
1,000
2,000 Km
10°0'0"W
0°0'0"
10°0'0"E
20°0'0"E
30°0'0"E

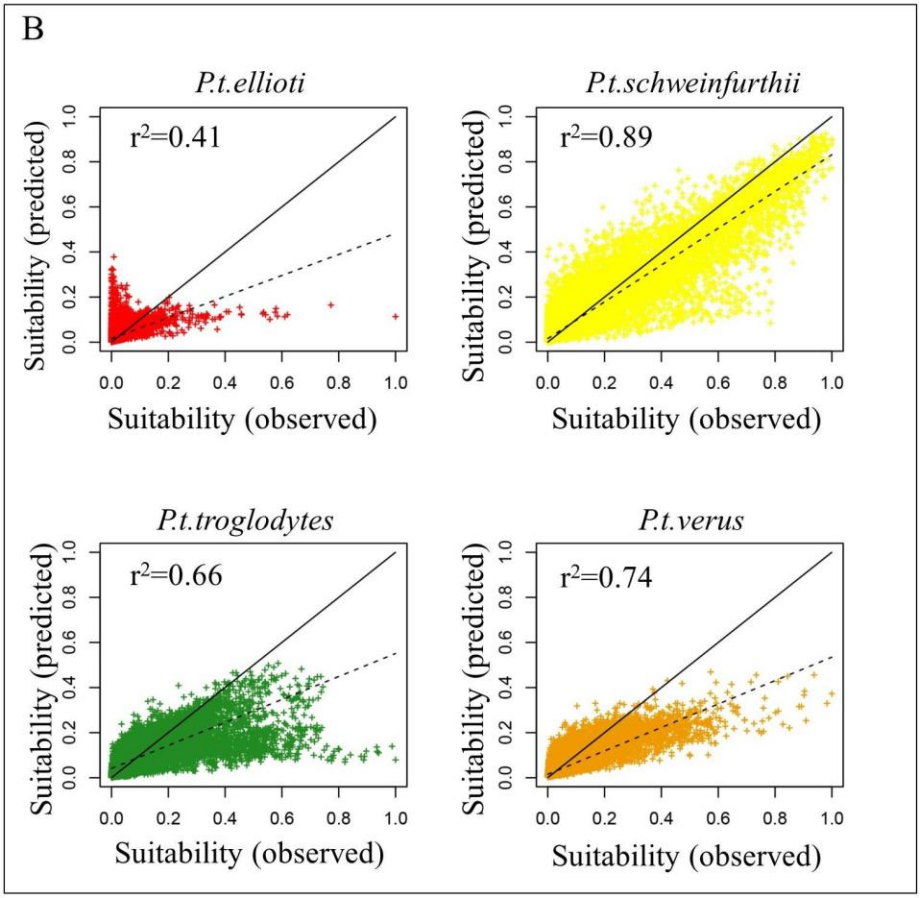

B
P.t.ellioti
r²=0.41
P.t.schweinfurthii
r²=0.89
P.t.troglodytes
r²=0.66
P.t.verus
r²=0.74
Suitability (predicted)
Suitability (observed)
0.0
0.2
0.4
0.6
0.8
1.0

Figure 4.5 Spatial distribution of differences between the 5km resolution map used for model calibration and the 30m map after coarsening to 5km resolution (A), distribution of suitability values of 5km calibration map (x-axis) and coarsened 30m map (y-axis) for each sub-species (B). The one-to-one line is shown in solid black and results of regressing values from the coarsened 30m map onto values from the 5km map for each sub-species are shown by the dashed black line.

Table 4.2 Mean correlation coefficients with standard errors reported for the entire chimpanzee range as well as for each sub-species.

Range	Mean r^2	Standard Error
Range-wide	0.82	0.1
P.t.ellioti	0.35	0.99
P.t.schweinfurthii	0.89	0.09
P.t.troglodytes	0.66	0.24
P.t.verus	0.73	0.22

Elevation, Landsat ETM+ band 5, and Landsat derived canopy cover were the strongest predictor variables in our model (Figure 4.6). A significant association between habitat suitability and bioclimatic variables was found in the suitability models for all four sub-species in Junker et al. (2012). The bioclimatic variables used in Junker et al. (2012) were taken from the WorldClim set of global climate layers, which were created by interpolating weather station data using latitude, longitude, and elevation as independent variables (Hijmans et al. 2005). It is therefore probable that elevation acted as a proxy for suitable climate conditions in the Random Forest model. Shortwave infrared reflectance is highly correlated with forest/non-forest, where forests have lower reflectance due to absorption properties of live vegetation and light extinguishing effects of tall forest canopies. The value of band 5 in discriminating Congo Basin Forest from non-forest was illustrated in Hansen et al., (2008). Trees in the CART family have been shown to exhibit a bias toward variables that have many cut-points (i.e., variables with many unique values) (Strobl et al. 2007). Band 5 is tightly correlated with canopy cover (r^2 = 0.83) and canopy height (r^2 = 0.82) in the calibration data. Canopy cover values were all integers between 0 and 100, while canopy height values were all integers between 0 and 23. On the other

hand, band 5 values were continuous on the interval 0 to 0.7, which provided the algorithm with many more possible cut-points relative to canopy cover and canopy height. A reasonable explanation is that short-wave infrared reflectance is driving the model as a surrogate for forest structure.

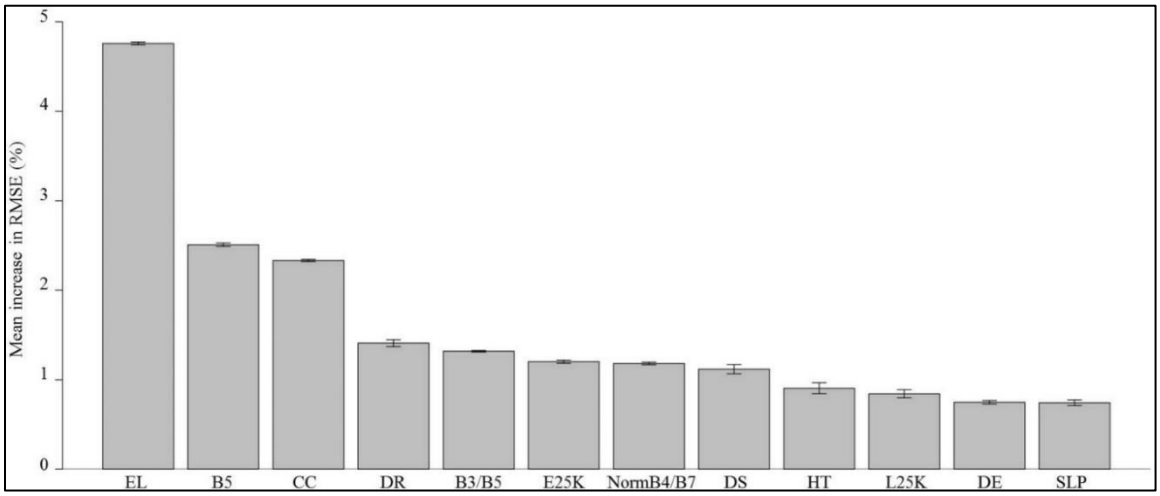


Figure 4.6 Relative importance of 12 variables selected for final Random Forest model after recursive feature elimination. Confidence intervals (95%) were calculated using bootstrap sampling (N=50). Variable importance was measured by the mean increase in RMSE % when values for a variable were randomly permuted.

4.6 Discussion

This study represents the first attempt to map temporal dynamics of range-wide chimpanzee habitat suitability at a medium-spatial resolution. The model demonstrated sensitivity to environmental change, which is a key feature for monitoring applications. The area depicted in Figure 4.7 is situated in the Republic of Congo, which historically has not experienced any substantial human disturbance but recently has seen the most rapid increase in the establishment of logging roads among all Central African nations (Hansen et al. 2008). The model was able to detect the fine-scale changes in habitat suitability due to human activities in the area (Figure 4.7B–E), whereas the coarse scale model was unable to resolve any fine-scale variation (Figure 4.7F). These results highlight the value of integrating continuously updated

variables derived from satellite remote sensing into habitat suitability models for regular monitoring and decision support. The Landsat suite of satellites has provided the longest continuing record of standardized data of Earth's land cover at 30-m resolution; metrics derived from these data will be crucial for the development of standardized habitat suitability models that can be repeated and integrated into long-term monitoring plans.

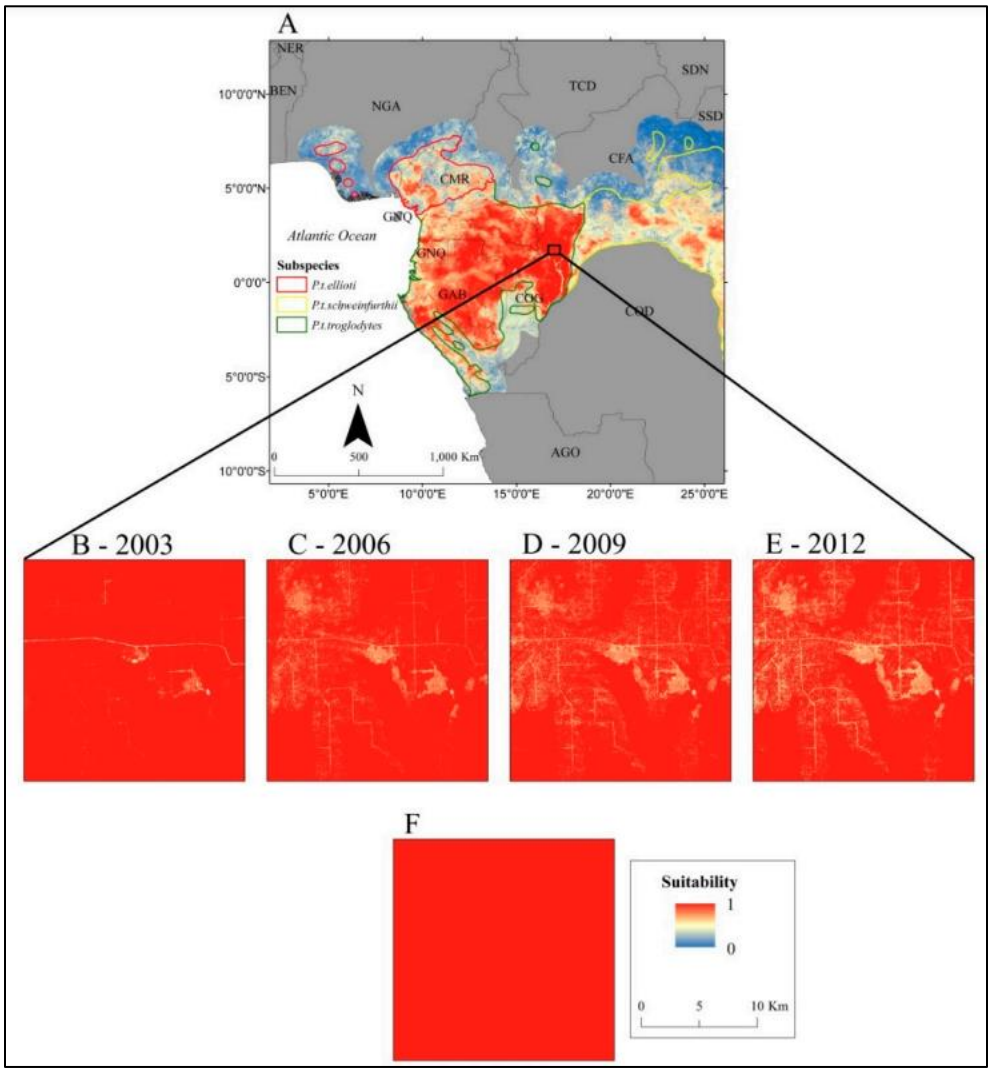


Figure 4.7 Suitability values predicted by a Random Forests model for Central Africa (A), an area experiencing selective logging circa 2003 (B), circa 2006 (C), circa 2009 (D), circa 2012 (E) and the area as depicted in the 5-km resolution calibration data (F).

Previous studies have reported successful efforts to model chimpanzee habitat suitability, including the two sources from which the model was calibrated. Direct comparison between studies is difficult because, so far, there are only a few published studies and they generally do not have overlapping study areas or represent habitat at substantially different spatial scales. All studies except for Pintea et al. (2006) utilized the MaxEnt software to generate their models, which reports model accuracy based on the Area Under the Curve of the Receiver Operating Characteristic (AUC). AUC characterizes the performance of a model at all possible thresholds by a single number that represents the probability that a known presence location will be ranked higher than a random background location (Phillips et al. 2006). Typical values for AUC are between 1.0–0.5 where a value of 1.0 represents a perfect fit while 0.5 indicates the model is no better than a random expectation (Phillips et al. 2006). The first study to use Landsat data to model chimpanzee habitat suitability was by Torres et al. (2010) and reported an average AUC value greater than 0.9 indicating a good model fit. In addition, they were also able to detect changes in habitat suitability between two time periods. While their study area was much smaller than what is considered in this chapter, a 2700 km^2 region in Guinea-Bissau, it lends support to the approach of using Landsat data to detect changes in chimpanzee habitat suitability. Similar AUC values were reported in Junker et al. (2012), the primary calibration data source for the Random Forests model, which ranged from 0.86 to 0.93. Junker et al. (2012) were able to compare their habitat suitability model for *P.t. schweinfurthii* to the results from Plumptre et al. (2010) and found large differences in the spatial distribution of habitat suitability for this sub-species. However, the maximum AUC reported by Plumptre et al. (2010) was 0.67, which suggests the results from Junker et al. (2012) are a better representation of habitat for this sub-species. Due to a lack of calibration data, important miombo woodland habitat in Western

Tanzania was predicted as unsuitable in Junker et al. (2012), which is the reason that model output from Pintea et al. (2006) was included as calibration data. While Pintea et al. (2006) did not provide AUC values for their model, they did assess model predictions with a 25% excluded portion of the nest locations that was not used to create the Mahalanobis distance model and an independent dataset from historical surveys. They found 92% of nests that were not used in model calibration and 84% of nests from historical surveys fell into the moderate to high suitability range. This indicates that the model was able to predict suitable nesting sites, but unfortunately does not provide any information about the model's commission error.

While the Random Forest model performed well across the chimpanzee range, it did show variation in its predictive capability. Compared to the calibration data, it underestimated regions of high suitability in the *P.t. ellioti* and *P.t. troglodytes* ranges, which is likely due to missing predictor variables. Dense canopy cover was found to be an important predictor of habitat suitability, however; Junker et al. (2012) did not find canopy cover to be a significant predictor variable in their models for these two sub-species. Both sub-species are found in central Africa, which is thought to be the last bastion of contiguous high quality chimpanzee habitat. While these large tracts of undisturbed forest are likely suitable places for chimpanzees to live, research has found a catastrophic decline in populations due to intense bush-meat hunting and disease (Walsh et al. 2003). High value tree species are selectively logged in the region, which leaves forests largely intact; however, the creation of logging roads opens up previously isolated forests, increasing the frequency of human contact and allowing easier access for hunters (Walsh et al. 2003; Laporte et al. 2007). An absence of chimpanzees in these intact forests is likely the reason why habitat suitability is low in these areas and incorporating information on the proximity to logging roads and camps would benefit model accuracy.

It is important to note that inaccurate predictions by the model could also arise from errors in the input datasets as well as errors in the datasets used to calibrate the model. Tree canopy cover was an important predictor variable in the model and was derived from Landsat ETM+ spectral data (Hansen et al. 2013). Canopy cover estimates from satellite data can be influenced by the composition of understory vegetation, satellite signal saturation and phenological noise (Sexton et al. 2013; Ozdemir 2014). Moreover, tree canopy cover from Hansen et al. (2013) does not discriminate between forest types which could lead to predicting suitable habitat in intensely managed forests. Habitat suitability models in Junker et al. (2012) were generated using bioclimatic variables from the WorldClim dataset and considerable uncertainty surrounds the estimated climate conditions in Africa due to an extremely sparse network of weather stations found there (Laporte et al. 2007). Since elevation likely served as a proxy for climate in our model, errors in Junker et al. (2012) associated with poorly estimated climate conditions will be present in the model.

An additional source of model uncertainty could come from the assumption that habitat suitability is insensitive to changes in spatial resolution. However, our model used only continuous variables, and the scaling properties of continuous data have been known for a long time (Stevens 1946). Reflectance and canopy cover, which figure prominently in our estimates of habitat suitability, have been demonstrated to be insensitive to changes in resolution (Gao et al. 2006; Sexton et al. 2013). In addition, Guisan et al. (2007) showed that coarsening the grain size of predictive variables in a species distribution model did not significantly impact model performance. We further evaluated model sensitivity to changes in spatial resolution by averaging all pixels of the 30-m suitability model output within the extent of the 5-km grid-cells from the map used to calibrate the model. Figure 4.5A portrays the difference between averaged

predictions from our model at 5-km resolution and the map used for calibration (observed), where areas of underestimation are seen in blue while overestimates are seen in red. Figure 4.5B depicts the relationship of suitability values from the calibration data (x-axis) and average predicted suitability from the Random Forests model in 5-km grid-cells for each sub-species. The correlation coefficients between the estimated habitat suitability and suitability from the calibration map were nearly identical to the relationships found from the cross-validation error assessment (Table 4.2), which suggests that error scales with spatial resolution.

4.7 Conclusions

In this study, we extended previous efforts to create maps of range-wide chimpanzee habitat suitability by using Random Forest regression to relate predictor variables derived from higher temporal and spatial resolution remotely sensed datasets to a coarse resolution chimpanzee habitat suitability map. We were able to map habitat suitability reasonably well, explaining 82% (0.2%) of the variance in the calibration data based on 50-fold cross-validation; however, there was considerable variation in predictive ability among the four different chimpanzee sub-species. Our model accounted for 35% (1.95%), 89% (0.18%), 66% (0.47%) and 73% (0.44%) of the variance for *P.t. ellioti* (Nigeria-Cameroon Chimpanzee), *P.t. schweinfurthii* (Eastern Chimpanzee), *P.t. troglodytes* (Central Chimpanzee) and *P.t. verus* (Western Chimpanzee), respectively. We found that highly suitable areas were associated with dense tree canopy cover for all but the Nigeria-Cameroon and Central Chimpanzee sub-species. A recent proliferation of logging roads within the ranges of these sub-species has resulted in increased hunting pressure and disease transmission, causing severe declines in chimpanzee populations; therefore, seemingly undisturbed tracts of forest are devoid of chimpanzees. Incorporating information on the proximity to logging roads and camps may therefore benefit

future modeling efforts. An important aspect of our approach that lends itself to conservation planning is the value of integrating continuously updated variables derived from remote sensing into habitat suitability models to detect and map change in suitability through time. With recent increases in data processing capabilities, developing such abilities to dynamically model and map changes in habitat suitability will be crucial for monitoring applications and the development of management plans that consider long-term trends in human-induced land cover change.

Chapter 5 Conclusion

5.1 Summary of Research

The studies in this dissertation have focused on leveraging medium-spatial resolution time-series remote sensing data to support regular monitoring of chimpanzee habitats for conservation applications. The products developed from this research provide the highest resolution and up to date maps of chimpanzee habitat across its range in sub-Saharan Africa. The research begins by focusing on the most southern and eastern extent of the chimpanzee range in western Tanzania in Chapters 2 and 3, then scaling up to cover the entire African range in Chapter 4. In this chapter, research summaries from this dissertation are presented and implications for conservation and future research are drawn.

In chapter 2, annual 30 m resolution forest canopy cover maps for the Greater Gombe and Greater Mahale ecosystems in Tanzania were developed by downscaling a global 250 m product using Landsat-derived spectral metrics and Random Forest regression. The Cumulative Sums algorithm was applied to the annual canopy cover maps to detect forest loss and gain between 2000-2020. Validation showed the method performed well for mapping forest loss but had lower accuracy for forest gain. Results revealed widespread forest loss in the Greater Mahale Ecosystem, including within many protected areas, primarily driven by agricultural expansion. In contrast, the Greater Gombe Ecosystem showed an overall forest cover increase, suggesting community forest management has been effective. Monitoring methods developed in this chapter can help monitor the success of strengthening village forest reserve bylaw enforcement to conserve forest habitat in the Greater Mahale Ecosystem.

In Chapter 3, a model was developed to enable annual monitoring of potential chimpanzee populations shifts as a function of landscape change in western Tanzania. An ensemble of fifty models were trained by relating environmental variables, with most variables generated from forest cover products in Chapter 2, to geolocated chimpanzee nests within an Inhomogeneous Poisson Process framework. Cross-validation results indicate the ensemble model was well calibrated and that relationships between variables and model output reflect what known about chimpanzee ecology in the region. Most importantly, model output showed a strong correlation with independent chimpanzee abundance estimates from line transect surveys, enabling the prediction of population size as a function of landscape condition. Results revealed potential declines in chimpanzee numbers for the Masito landscape and some village forest reserves in the GME due to habitat degradation, whereas increases were predicted in Greater Gombe due to successful community forest management. Though caution is needed interpreting predicted trends, particularly for population increases, the modeling approach allows regular monitoring of chimpanzee populations that can inform conservation priorities.

In Chapter 4, "off-the-shelf" remote sensing products were used as input into a Random Forests regression model to downscale the output from a coarse resolution habitat suitability model. The modeled relationship was used to make predictions of habitat suitability over the entire chimpanzee range at 30-meter resolution. The model predicted habitat suitability well with an r^2 of 0.82 (± 0.002) based on 50-fold cross validation; however, there was considerable variation in the predictive capability among the four sub-species modeled individually. The influence of several variables derived from Landsat ETM+ were tested, which included metrics of forest canopy and structure, for four three-year time periods between 2000 and 2012. Elevation, Landsat ETM+ band 5 and Landsat derived canopy cover were the strongest

predictors; highly suitable areas were associated with dense tree canopy cover for all but the Nigeria-Cameroon and Central Chimpanzee sub-species. These two sub-species are likely more impacted by disease and hunting, which is not reflected in the satellite signal. Because the models were sensitive to such temporally based predictors, the results highlight the value of integrating continuously updated variables derived from satellite remote sensing into habitat suitability models for regular monitoring over large geographic areas.

5.2 Conservation Implications

Building capacity for satellite image processing presents a variety of challenges, as it requires advanced technical expertise and substantial financial resources, in the form of personnel, and computer hardware and software (Palumbo et al. 2016) This can be a significant burden for conservation organizations on shoestring budgets to bear individually. The work for this dissertation was carried out on a single desktop computer with free and open-source software, and all the satellite data are freely available as well. As such, the work described in this dissertation can serve as a template for generating spatially explicit products relevant for conservation planning and monitoring.

The outputs from chapter 3 could aid in planning future chimpanzee population surveys in the Masito-Ugalla Landscape. Past surveys have failed to conclusively determine population trends, with Piel et al. (2015) finding a decrease compared to a 2007 survey and The Jane Goodall Institute (2020) finding an increase compared to a 2017 survey. However, neither was outside the margin of error, so even the direction of change is unclear. Given widespread habitat loss across the GME, a declining population would be reasonably expected in the Masito-Ugalla landscape. Techniques borrowed from the remote sensing community could be leveraged to increase the precision of population estimates and determine the presence of a population trend.

For example, Song et al. (2021) accurately estimated the expansion of soybean area in South America by combining satellite imagery with sample field data. They divided South America into a grid of 20 km by 20 km blocks and assigned each block a stratum of high, medium, or low based on soybean area mapped from satellite data. From this, they calculated strata means and variances that enabled estimating the standard error of soybean area for different allocations of strata and sample sizes. They ultimately settled on a random selection of 25 blocks from each stratum (75 total) and randomly sampled 20 pixels in each block that field teams then visited to determine the presence of soybean. The data collected by the field teams was then used to estimate soybean area for each stratum, and ultimately an estimate for all of South America. A similar stratified sampling strategy could be applied to population surveys in the Masito-Ugalla landscape. The landscape could be divided into a regular grid, with each grid-cell assigned the average CHIMP value within it to construct strata and calculate means and variances. The number of strata, block sample sizes, and number of transects placed in each block would depend on available resources, but sufficiently sampling of higher variance strata should be emphasized to maximize the precision of the population estimate. The means and variances from the strata could then be combined to produce a population estimate that should have smaller error margins than estimates from previous surveys.

Products from this dissertation are relevant to REDD+ (Reducing emissions from deforestation and forest degradation) projects, many of which are also registered under the Climate, Community and Biodiversity Standard (CCB), committing to the generation and reporting of positive impacts in select community and biodiversity indicators. Biodiversity monitoring on many REDD projects has been restricted to simple assessments of species' presence/absence. Using chimpanzees as a case study, Dickson et al. 2020 presented an example

of how species distribution models can provide more detailed information for biodiversity monitoring through a small project in the Ntakata mountains east of Mahale Mountains National Park. They demonstrated that preserving forest for carbon credits has the co-benefit of maintaining suitable chimpanzee habitat. The output products from Chapter 3 could also serve as an indicator but would provide more detailed information. Specifically, the modeling results could indicate the potential population size that was not displaced due to REDD project implementation rather than the area of suitable habitat that was preserved. The output products from Chapter 4 could be used in exactly the same way as described in Dickson et al. (2020), but since the maps cover the entire chimpanzee range, they could be leveraged for numerous REDD projects across Africa.

5.3 Future Research Implications

Recent genetic evidence indicates the chimpanzees of Gombe National Park in Tanzania may be part of a larger meta-population that spans north to Burundi, Rwanda, and Uganda (Bonnin et al. 2023), and occasional migrations of female chimpanzees into Gombe from unknown origins support this idea (Walker et al. 2017). The chimpanzee habitats in these countries resemble Tanzania's landscapes of rift valley forests and savanna woodlands interspersed with riverine forests (McLennan 2008; Plumptre et al. 2010). Anthropogenic pressures like forest clearing for agriculture also threaten the chimpanzees across these regions. While the approximately 700 chimpanzees in Burundi and Rwanda are relatively well protected in Kibira and Nyungwe National Parks, Uganda's estimated 5,000 chimpanzees face widespread habitat loss, where large forest tracts have been cleared for cash crops (McLennan 2008; Van Dijk et al. 2021). Methods outlined in chapters 2 and 3 could be applied to these regions to characterize and monitor chimpanzee habitats. In addition, new variables could be added to the

variable set from Chapter 3 to potentially improve predictive performance. An important factor for nest site selection is tree height (Fruth & Hohmann 1996), and Potapov et al. (2021) demonstrated that canopy height can be accurately modeled from GEDI LiDAR shots using the same Landsat metrics applied to canopy cover mapping in Chapter 2. Establishing corridors connecting these populations could facilitate gene flow within the larger meta-population. Bonnin et al. (2020) already used connectivity modeling to identify important corridors in the Gombe-Masito-Ugalla Ecosystem. Their analysis used inputs similar to what was produced in Chapter 3; therefore, this type of analysis could be expanded across borders to identify important corridors to connect and conserve chimpanzee populations in East Africa.

The likelihood of negative future outcomes needs to be considered for the continued survival of wild chimpanzees, with a focus on future climate impacts and land-use pressures placed on African Tropical forests. While tropical forest loss in South America and Southeast Asia is thought to be driven by export markets for agricultural and forest products (DeFries et al. 2010), Sub-Saharan Africa does not seem to be following this trend (Fisher 2010; Kissinger et al. 2012) where subsistence agriculture and the local demand for forest products such as timber and charcoal are the primary drivers of forest loss. Deforestation trends related to increasing local populations will likely continue as the population growth rate in Africa is not expected to slow over this century (Gerland et al. 2014). In addition to the impacts of continued population growth, chimpanzees and their forest habitats could soon begin to experience the impacts of agro-industrial plantations. The economic incentive for industrial oil palm plantations has led to tremendously high deforestation rates in southeast Asia and concomitant impacts on biodiversity (Wilcove et al. 2013) and there is increasing evidence that this dynamic may soon begin to impact Africa's tropical forests, leading to the destruction of chimpanzee forest habitats. Recent

major land deals have been made which set aside large tracts of land for new oil palm plantations in the Congo Basin (Feintrenie 2014; Wich et al. 2014) and a recent analysis has shown that large portions of unprotected, highly suitable great ape habitat are in areas conducive for oil palm plantations (Wich et al. 2014). Given these circumstances, a scenario-based planning process could be implemented to anticipate where investments can be made to mitigate, as much as possible, the continued loss of chimpanzees and their habitats.

Glossary

APES	Ape Populations, Environments and Surveys
ARD	Analysis Ready Data
AUC	Area Under the Curve
AVHRR	Advanced Very High Resolution Radiometer
CAP	Conservation Action Plan
CART	Classification and Regression Trees
CBI	Continuous Boyce Index
CC	Canopy Cover
CHIMP	Chimpanzee Habitat Indicator for Monitoring Populations
CI	Confidence Interval
CL	Confidence Level
COR	Point Bi-serial Correlation Coefficient
DE	Distance to Forest Edge
DEM	Digital Elevation Model
DF	Distance to Forest
DL	Distance to Forest Loss
DR	Distance to River
DS	Distance to Steep Slopes
EI	Edge Influence
ELV	Elevation
ETM+	Enhanced Thematic Mapper Plus
FAO	Food and Agriculture Organization
GEDI	Global Ecosystem Dynamics Investigation
GGE	Greater Gombe Ecosystem
GLAD	Global Land Analysis and Discovery
GLAS	Geoscience Laser Altimeter System
GME	Greater Mahale Ecosystem
GNP	Gombe National Park
HT	Canopy Height
IPP	Inhomogeneous Poisson Process
ISS	International Space Station
IUCN	International Union for the Conservation of Nature
JGI	the Jane Goodall Institute
LPDAAC	Land Processes Distributed Active Archive Center
LST	Land Surface Temperature
MAE	Mean Absolute Error
MBE	Mean Bias Error
MMNP	Mahale Mountains National Park
MODIS	Moderate Resolution Imaging Spectroradiometer

MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
OLI	Operational Land Imager
OSCP	Open Standards for Conservation Planning
REDD	Reducing Emissions from Deforestation and degradation
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
SDM	Species Distribution Model
SE	Standard Error
SLP	Slope
SRTM	Shuttle Radar Topography Mission
SWBD	SRTM Waterbodies Dataset
TAWIRI	Tanzania Wildlife Research Institute
TM	Thematic Mapper
UNEP	United Nation Environment Programme
USGS	United States Geological Survey
VCF	Vegetation Continuous Fields
VFR	Village Forest Reserve

Bibliography

Aarts G, Fieberg J, Matthiopoulos J. 2012. Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods in Ecology and Evolution **3**:177–187.

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography **38**:541–545.

Anderson RP, Gonzalez I. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. Ecological Modelling **222**:2796–2811.

Anteau MJ, Wiltermuth MT, Sherfy MH, Shaffer TL. 2014. Measuring and predicting abundance and dynamics of habitat for piping plovers on a large reservoir. Ecological Modelling **272**:16–27. Elsevier B.V.

Araújo M, Peterson A. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology **93**:1527–1539.

Avissar R, Werth D. 2005. Global Hydroclimatological Teleconnections Resulting from Tropical Deforestation. Journal of Hydrometeorology **6**:134–145.

Baccini A et al. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nature Climate Change **2**:182–185. Nature Publishing Group.

Baddeley AJ, Møller J, Waagepetersen R. 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica **54**:329–350.

Baillie JEM, Hilton-taylor C, Stuart SN. 2004. 2004 IUCN Red List of Threatened Species. A Global Species Assessment. IUCN, Gland, Switzerland and Cambridge, UK.

Beentje HJ, Mbago FM. 2007. A GUIDE TO THE FIG TREES OF WESTERN TANZANIA WITH SPECIAL EMPHASIS ON GOMBE AND MAHALE NATIONAL PARKS. Journal of East African Natural History **96**:1–26.

Bennett AF. 2003. Linkages in the landscape: The role of corridors and connectivity in wildlife conservation. Page 254. IUCN, Gland, Switzerland and Cambridge, UK. Available from http://books.google.com/books?hl=en&lr=&id=hkkXEpcWOOgC&oi=fnd&pg=PR9&dq=Linkages+in+the+Landscape+The+Role+of+Corridors+and+Connectivity+in+Wildlife+Conservation&ots=1-UvjYV2Ri&sig=reC408WOMYf1LHAZAkrh4w_VjKk.

Benson BJ, MacKenzie MD. 1995. Effects of sensor spatial resolution on landscape structure parameters. Landscape Ecology **10**:113–120.

Bonnin N et al. 2020. Modelling landscape connectivity change for chimpanzee conservation in Tanzania | LJMU Research Online. Biological Conservation **252**. Available from http://researchonline.ljmu.ac.uk/id/eprint/13890/.

Bonnin N et al. 2023. Barriers to chimpanzee gene flow at the south-east edge of their distribution. Molecular Ecology **32**:3842–3858.

Breiman L. 2001. Random forests. Machine learning **45**:5–32.

Breiman L, Friedman J, Stone CJ, Olsen RA. 1984. Classification and Regression Trees. Chapman and Hall/CRC, Washington, DC.

Brook B, Bradshaw C, Koh L, Sodhi N. 2006. Momentum Drives the Crash: Mass Extinction in the Tropics1. Biotropica **38**:302–305.

Brooks T, Tobias J, Balmford A. 1999a. Deforestation and bird extinctions in the Atlantic forest. Animal Conservation **2**:211–222.

Brooks TM, Pimm SL, Oyugi JO. 1999b. Time Lag between Deforestation and Bird Extinction in Tropical Forest Fragments. Conservation Biology **13**:1140–1150.

Buckland ST. 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford University Press. Available from https://books.google.com/books?id=9-LiOqoKE9QC.

Caldecott J, Miles L. 2005. World Atlas of the Great Apes and Their Conservation. UNEP World Conservation Monitoring Centre, Cambridge, UK.

Campbell G, Kuehl H, N'Goran Kouamé P, Boesch C. 2008. Alarming decline of West African chimpanzees in Côte d'Ivoire. Current Biology **18**:R903–R904.

Chen J, Gupta AK. 2000. Parametric Statistical Change Point Analysis. Springer Science + Business Media, New York.

Clark JD, Dunn JE, Smith KG. 1993. A Multivariate Model of Female Black Bear Habitat Use for a Geographic Information System. The Journal of Wildlife Management **57**:519–526.

Clee PRS et al. 2015. Chimpanzee population structure in Cameroon and Nigeria is associated with habitat variation that may be lost under climate change. BMC Evolutionary Biology:1–13.

Cochran WG (William G. 1977. Sampling techniques. John Wiley and Sons, New York, Chichester, Toronto, Brisbane, Singapore.

Defries R. 2012. Why Forest Monitoring Matters for People and the Planet. Pages 1–12 in Achard F, Hansen MC, editors. Global Forest Monitoring from Earth Observation. CRC Press, Boca Raton, London, New York.

Defries R, Hansen A, Newton AC, Hansen MC. 2005. Increasing Isolation of Protected Areas in Tropical Forests over the past Twenty Years. Ecological Applications **15**:19–26.

DeFries RS, Rudel T, Uriarte M, Hansen MC. 2010. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. Nature Geoscience **3**:178–181.

Dickson R, Baker M, Bonnin N, Shoch D, Rifkin B, Stewart FA, Piel AK. 2020. Combining Deforestation and Species Distribution Models to Improve Measures of Chimpanzee Conservation Impacts of REDD: A Case Study from Ntakata Mountains, Western Tanzania. Forests **11**:1195.

Diggle PJ. 2003. Statistical Analysis of Spatial Point PatternsSecond. Arnold, London.

Dimiceli C, Carroll M, Sohlberg R, Kim DH, Kelly M, Townshend JRG. 2015. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.

Doggart N, Morgan-Brown T, Lyimo E, Mbilinyi B, Meshack CK, Sallu SM, Spracklen DV. 2020. Agriculture is the main driver of deforestation in Tanzania. Environmental Research Letters **15**:034028.

Dubayah R et al. 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Science of Remote Sensing **1**:100002. Elsevier Ltd.

F. Dormann C et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography **30**:609–628.

Farber O, Kadmon R. 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecological Modelling **160**:115–130.

Feintrenie L. 2014. Agro-industrial plantations in Central Africa, risks and opportunities. Biodiversity and Conservation **23**:1577–1589.

Feng M, Sexton JO, Huang C, Masek JG, Vermote EF, Gao F, Narasimhan R, Channan S, Wolfe RE, Townshend JR. 2013. Global surface reflectance products from Landsat: Assessment using coincident MODIS observations. Remote Sensing of Environment **134**:276–293.

Feng X, Park DS, Walker C, Peterson AT, Merow C, Papeş M. 2019. A checklist for maximizing reproducibility of ecological niche models. Nature Ecology and Evolution **3**:1382–1395. Nature Publishing Group.

Ferrier S, Watson G, Pearce J, Drielsma M. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales . I . Species-level modelling:2275–2307.

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation **24**:38–49.

Fisher B. 2010. African exception to drivers of deforestation. Nature Geoscience **3**:375–376. Nature Publishing Group.

Fithian W, Hastie T. 2013. Finite-sample equivalence in statistical models for presence-only data. the Annals of Applied Statistics **7**:1917–1939.

Fitzpatrick MC, Gotelli NJ, Ellison AM, Itzpatrick MACF, Otelli NIJG. 2013. MaxEnt versus MaxLike: empirical comparisons with ant species distributions. Ecosphere **4**:Article 55.

Food and Agriculture Organization of the United Nations. 2006. Global Forest Resources Assessment 2005, Main Report. Progress Towards Sustainable Forest Management. Food and Agricultural Organization of the United Nations, Rome.

Fourcade Y, Engler JO, Rödder D, Secondi J. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PloS one **9**:e97122.

Franklin J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography **19**:474–499.

Fruth B, Hohmann G. 1996. Nest building behavior 1n the great apes: the great leap forward? Pages 225–240 in McGrew WC, Marchant LF, Nishida T, editors. Great Ape Societies. Cambridge University Press, Cambridge, UK.

Gao F, Masek J, Schwaller M, Hall F. 2006. On the Blending of the MODIS and Landsat ETM + Surface Reflectance. IEEE Transactions on Geoscience and Remote Sensing **44**:2207–2218.

Gerland P et al. 2014. World population stabilization unlikely this century. Science **346**.

Goodall J. 1986. The Chimpanzees of Gombe: Patterns of Behavior. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, Massachusetts.

Granitto PM, Furlanello C, Biasioli F, Gasperi F. 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemometrics and Intelligent Laboratory Systems **83**:83–90.

Guisan A et al. 2007. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions **13**:332–340.

Guisan A, Zimmermann NE, Zimmerman NE. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147–186.

Haddad NM et al. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances **1**:1–9.

Hansen MC et al. 2013. High-resolution global maps of 21st-century forest cover change. Science (New York, N.Y.) **342**:850–3.

Hansen MC, DeFries RS, Townshend JRG, Sohlberg R, Dimiceli C, Carroll M. 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. Remote Sensing of Environment **83**:303–319.

Hansen MC, Potapov PV, Goetz SJ, Turubanova S, Tyukavina A, Krylov A, Kommareddy A, Egorov A. 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. Remote Sensing of Environment **In press**:2–13. Elsevier B.V.

Hansen MC, Roy DP, Lindquist E, Adusei B, Justice CO, Altstatt A. 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. Remote Sensing of Environment **112**:2495–2513.

Hansen MC, Stehman SV, Potapov PV. 2010. Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences of the United States of America **107**:8650–8655.

Harcourt A. 1998. Ecological Indicators of Risk for Primates, as Judged by Species' Susceptibility to Logging. Pages 56–79 in Caro T, editor. Behavioral Ecology and Conservation Biology. Oxford University Press, Oxford.

Harris NL, Brown S, Hagen SC, Saatchi SS, Petrova S, Salas W, Hansen MC, Potapov PV, Lotsch A. 2012. Baseline map of carbon emissions from deforestation in tropical regions. Science **336**:1573–1576.

Helmer EH, Ruzycki TS, Wunderle JM, Vogesser S, Ruefenacht B, Kwit C, Brandeis TJ, Ewert DN. 2010. Mapping tropical dry forest height, foliage height profiles and disturbance type and age with a time series of cloud-cleared Landsat and ALI image mosaics to characterize avian habitat. Remote Sensing of Environment **114**:2457–2473. Elsevier B.V.

Hernandez-Aguilar RA. 2009. Chimpanzee nest distribution and site reuse in a dry habitat: implications for early hominin ranging. Journal of Human Evolution **57**:350–364. Elsevier.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965–1978.

Hilker T, Wulder M a., Coops NC, Linke J, McDermid G, Masek JG, Gao F, White JC. 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sensing of Environment **113**:1613–1627. Elsevier Inc.

Hilton-Taylor Craig, Pollock CM, Chanson JS, Butchart SHM, Oldfield TEE, Katariya V. 2008. State of the world's species. Pages 15–41 in Vie J-C, Hilton-Taylor Craig, Stuart SN, editors. Wildlife in a changing world:An analysis of the 2008 IUCN Red List of Threatened Species. IUCN, Gland, Switzerland. Available from http://www.iucn.org/dbtw-wpd/html/RL-2009-001/cover.html.

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling **199**:142–152.

Humle T, Boesch C, Campbell G, Junker J, Koops K, Kuehl HS, Sop T. 2016. Pan troglodytes ssp. verus. Available from http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T15935A5323101.en.

Hussain Md, Mahmud I. 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. Journal of Open Source Software **4**:1556.

Jantz P, Goetz S, Laporte N. 2014. Carbon stock corridors to mitigate climate change and promote biodiversity in the tropics. Nature Climate Change **4**:138–142.

Jantz SM, Pintea L, Nackoney J, Hansen MC. 2018. Global Forest Maps in Support of Conservation Monitoring. Pages 82–118 in Leidner AK, Buchanan GM, editors. Satellite Remote Sensing for Conservation Action, 1st edition. Cambridge University Press. Available from https://www.cambridge.org/core/product/identifier/9781108631129%23CN-bp-4/type/book_part (accessed November 19, 2022).

Jiménez-Valverde A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography **21**:498–507.

Johns AD, Skorupa JP. 1987. Responses of Rain-Forest Primates to Habitat Disturbance : A Review. International Journal of Primatology **8**:157–191.

Jones C, Jones C, Jones J, Wilson D. 1996. Pan Troglodytes. Mammalian Species **529**:1–9.

Joshi AR et al. 2016. Tracking changes and preventing loss in critical tiger habitat. Science Advances **2**:1–8.

Junker J et al. 2012. Recent decline in suitable environmental conditions for African great apes. Diversity and Distributions **18**:1077–1091.

Kadmon R, Farber O, Danin A. 2004. EFFECT OF ROADSIDE BIAS ON THE ACCURACY OF PREDICTIVE MAPS PRODUCED BY BIOCLIMATIC MODELS. Pages 401–413 Ecological Applications.

Kendall MG (Maurice G, Gibbons JD. 1990. Rank correlation methods5th ed. E. Arnold, London.

Kim D, Sexton JO, Townshend JR. 2015. Accelerated deforestation in the humid tropics from the 1990s to the 2000s. Geophysical Research Letters **42**:3495–3501.

Kissinger G, Herold M, De Sy V. 2012. Drivers of Deforestation and Forest Degradation: A Synthesis Report for REDD+ Policy Makers. Page 48 Lexeme Consulting. Vancouver, Canada.

Kormos R, Boesch C, Bakarr MI, Butynski TM. 2003. West African Chimpanzees. Status Survey and Conservation Action Plan. IUCN/SSC Primate Specialist Group, Gland, Switzerland and Cambridge, UK.

Kortlandt A. 1983. Marginal habitats of chimpanzees. Journal of Human Evolution **12**:231–278.

Kramer-Schadt S et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Diversity and Distributions **19**:1366–1379. Blackwell Publishing Ltd.

Laporte NT, Stabach J a, Grosch R, Lin TS, Goetz SJ. 2007. Expansion of industrial logging in Central Africa. Science (New York, N.Y.) **316**:1451.

Lawrence PJ, Chase TN. 2007. Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). Journal of Geophysical Research: Biogeosciences **112**.

McLennan MR. 2008. Beleaguered Chimpanzees in the Agricultural District of Hoima, Western Uganda. Primate Conservation **23**:45–54.

Merow C, Smith MJ, Silander J a. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography **36**:1058–1069.

Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks TM, Mittermeier CG, Lamoreux J, da Fonseca G a. 2005. Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrerestrial Ecoregions. Cemex, Mexico City.

Moyer D et al. 2006. Surveys of Chimpanzees and other Biodiversity in Western Tanzania. Page 65. Wildlife Conservation Society, Jane Goodall Institute, University of California San Diego.

Myers N, Mittermeier R a, Mittermeier CG, da Fonseca G a, Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853–8.

NAFORMA. 2015. National Forest Resources Monitoring and Assessment of Tanzania Mainland: Main Results. Ministry of Natural Resources and Tourism, Tanzania.

Nellemann C, Newton A. 2002. The Great Apes - the road ahead. A Globio perspective on the impacts of infrastructural development on The Great Apes. Page 30. United Nations Environment Programme.

Nishida T. 1990. The Chimpanzees of the Mahale Mountains: Sexual and Life History Strategies. University of Tokyo Press, Tokyo.

Nishida T, Wrangham R, Goodall J, Uehara S. 1983. Local differences in plant feeding habits of chimpanzees between the Mahale mountains and Gombe National Park, Tanzania. Journal of Human Evolution **12**:467–480.

Nowak R. 1999. Walkers Mammals of the World. Johns Hopkins University Press, Baltimore and London.

Oates JF, Dunn A, Greengrass E, Morgan BJ. 2008. Pan troglodytes ssp. ellioti. The IUCN Red List of Threatened Species 2008. Available from http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T40014A10301774.en.

Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA. 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment **148**:42–57. Elsevier Inc.

Olson DM et al. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience **51**:933.

Orwa C, Matua A, Kindt R, Jamnadass R, Anthony S. 2009. Agroforestree Database:a tree reference and selection guide version 4.0. Available from http://www.worldagroforestry.org/sites/treedbs/treedatabases.asp.

Ozdemir I. 2014. Linear transformation to minimize the effects of variability in understory to estimate percent tree canopy cover using RapidEye data. GIScience & Remote Sensing **1603**:37–41.

Page ES. 1955. A Test for a Change in a Parameter Occurring at an Unknown Point. Biometrika **42**:523–527.

Page ES. 1957. On Problems in which a Change in a Parameter Occurs at an Unknown Point. Biometrika **44**:248–252.

Palumbo I, Rose RA, Headley RMK, Nackoney J, Vodacek A, Wegmann M. 2016. Building capacity in remote sensing for conservation: present and future challenges. Remote Sensing in Ecology and Conservation:1–9.

Pedregosa F, Varoquaux G. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12**:2825–2830.

Pfeifer M et al. 2017. Creation of forest edges has a global impact on forest vertebrates. Nature **551**:187–191. Nature Publishing Group.

Phillips SJ, Anderson RP, DudÍk M, Schapire RE, Blair ME. 2017. Opening the black box: an open-source release of Maxent. Ecography:1–7.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231–259.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological applications : a publication of the Ecological Society of America **19**:181–97.

Phillips SJ, Elith J. 2010. POC plots: calibrating species distribution models with presence-only data. Pages 2476–2484 Ecology.

Piel A, Stewart F. 2014. Census and conservation status of chimpanzees (Pan troglodytes schweinfurthii) across the Greater Mahale Ecosystem. Page 74. The Nature Conservancy.

Piel AK, Cohen N, Kamenya S, Ndimuligo SA, Pintea L, Stewart FA. 2015. Population status of chimpanzees in the Masito-Ugalla Ecosystem, Tanzania. American Journal of Primatology **77**:1027–1035.

Pimm S, Askins R. 1995. Forest losses predict bird extinctions in eastern North America. Proceedings of the National Academy of Sciences of the United States of America **92**:9343–9347.

Pimm SL, Brooks T. 2013. Conservation: Forest fragments, facts, and fallacies. Current Biology **23**:R1098–R1101. Elsevier.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science **344**. Available from http://www.sciencemag.org/cgi/doi/10.1126/science.1246752.

Pintea L, Mtiti E, Mavanza, M, Abdallah F, Kashula A, Mjungu D, Collins A, Kamenya S. 2016. 20 Years and Counting: Adaptive Management of Chimpanzee Habitats in the Greater Gombe Ecosystem, Tanzania. The Jane Goodall Institute. Available from https://docs.google.com/document/d/1WXt0tSDKfEIbraoG4jcAL7hVx0MDEap2Lfdtu50a3FM/edit?usp=sharing.

Pintea L, Plumptre AJ. 2006. Prediction of suitable habitat for chimpanzees using remote sensing and GIS. Pages 47–54 in Moyer D et al., editors. Surveys of chimpanzees and other biodiversity in Western Tanzania. Wildlife Conservation Society, the Jane Goodall Institute, U.S. Fish and Wildlife Service.

Pintea L, Pusey AE, Wilson ML, Gilby IC, Collins DA, Kamenya S, Goodall J. 2012. Long-term changes in the ecological factors surrounding the chimpanzees of Gombe National Park. Page in Plumptre AJ, editor. Impacts on Biodiversity and Ecosystems. Nova Science Publishers, New York.

Plumptre A et al. 2010. Eastern Chimpanzee (Pan Troglodytes Schweinfurthii): Status Survey and Conservation Action Plan, 2010-2020. IUCN/SSC Primate Specialist Group, Gland, Switzerland. Available from http://books.google.com/books?hl=en&lr=&id=LQheCyr4j2sC&oi=fnd&pg=PA1&dq=Eastern+Chimpanzee+(Pan+troglodytes+schweinfurthii)+Status+and+Survey+Conservation+Action+Plan+2010-2020&ots=RCQaWAF16B&sig=sSVXgoRWI89P8bSnZL1_c8R9fGc (accessed January 31, 2013).

Potapov P et al. 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment **253**:112165.

Potapov P, Hansen MC, Kommareddy I, Kommareddy A, Turubanova S, Pickens A, Adusei B, Tyukavina A, Ying Q. 2020. Landsat analysis ready data for global land cover and land cover change mapping. Remote Sensing **12**.

Purvis a, Jones KE, Mace GM. 2000. Extinction. BioEssays : news and reviews in molecular, cellular and developmental biology **22**:1123–33.

Pusey AE, Pintea L, Wilson ML, Kamenya S, Goodall J. 2007. The contribution of long-term research at Gombe National Park to chimpanzee conservation. Conservation biology : the journal of the Society for Conservation Biology **21**:623–34.

Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI. 2015. Point process models for presence-only analysis. Methods in Ecology and Evolution **6**:366–379.

Sala OE et al. 2000. Global biodiversity scenarios for the year 2100. Science (New York, N.Y.) **287**:1770–1774. Department of Ecology and Instituto de Investigaciones Fisiológicas y Ecológicas vinculadas a la Agricultura, Faculty of Agronomy, University of Buenos Aires, Avenida San Martín 4453, Buenos Aires 1417, Argentina. sala@ifeva.edu.ar.

Saura S. 2004. Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. Landscape Ecology **19**:197–209.

Sexton JO et al. 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS Vegetation Continuous Fields with lidar-based estimates of error. International Journal of Digital Earth:130321031236007.

Singh A, Shi H, Zhu Z, Foresman T. 2001. An assessment of the status of the world's remaining closed forests. UNEP, Nairobi, Kenya.

Smith AB, Santos MJ. 2020. Testing the ability of species distribution models to infer variable importance. Ecography **43**:1801–1813.

Sofaer HR et al. 2019a. Development and Delivery of Species Distribution Models to Inform Decision-Making. BioScience **69**:544–557. Oxford University Press.

Sofaer HR, Hoeting JA, Jarnevich CS. 2019b. The area under the precision-recall curve as a performance metric for rare binary events. Methods in Ecology and Evolution **10**:565–577. British Ecological Society.

Song X-P et al. 2021. Massive soybean expansion in South America since 2000 and implications for conservation. Nature Sustainability **4**:784–792.

Stevens SS. 1946. On the Theory of Scales of Measurement. Science **103**:677–680.

Strassburg BBN, Rodrigues ASL, Gusti M, Balmford A, Fritz S, Obersteiner M, Turner RK, Brooks TM. 2012. Impacts of incentives to reduce emissions from deforestation on global species extinctions. Nature Climate Change **2**:350–355.

Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC bioinformatics **8**:25.

Syfert MM, Smith MJ, Coomes DA. 2013. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. PLoS ONE **8**:55158.

Tang H, Armston J. 2019. Algorithm Theoretical Basis Document (ATBD) for GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics. Page 39. 1.0. University of Maryland, College Park, College Park, MD.

TAWIRI. 2018. Tanzania Chimpanzee Conservation Action Plan 2018-2023. Page 60. Tanzania Wildlife Research Institute. Available from https://tawiri.or.tz/wp-content/uploads/2018/07/Tanzania-Chimpanzee-Conservation-Action-Plan-2018.pdf.

Taylor W. 2000. Change-Point Analysis: A Powerful New Tool For Detecting Changes. Available from http://www.variation.com/cpa/tech/changepoint.html (accessed March 28, 2022).

Team A. 2020. Application for Extracting and Exploring Analysis Ready Samples (AppEEARS). NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, USA. Available from https://lpdaacsvc.cr.usgs.gov/appeears/.

The Conservation Measures Partnership. 2013. Open Standards for the Practice of Conservation, Version 3.0. Page 51 Open Standards for the Practice of Conservation.

The Jane Goodall Institute. 2020. ESTIMATING AND ASSESSING CHIMPANZEE POPULATIONS AND THREATS IN THE MASITO UGALLA ECOSYSTEM FINAL CHIMPANZEE SURVEY REPORT. Page 27. The Jane Goodall Institute.

Torres J, Brito JC, Vasconcelos MJ, Catarino L, Gonçalves J, Honrado J. 2010. Ensemble models of habitat suitability relate chimpanzee (Pan troglodytes) conservation to forest and landscape dynamics in Western Africa. Biological Conservation **143**:416–425. Elsevier Ltd.

Tracewski Ł, Butchart SHM, Donald PF, Evans M, Fishpool LDC, Graeme M B. 2016. Patterns of twenty-first century forest loss across a global network of important sites for biodiversity. Remote Sensing in Ecology and Conservation **2**:37–44.

Tutin CEG, Baillie JEM, Dupain J, Gatti S, Maisels F, Stokes EJ, Morgan DB, Walsh PD. 2008. Pan troglodytes ssp. troglodytes. The IUCN Red List of Threatened Species 2008. Available from http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T15936A5323646.en.

Tyukavina A, Baccini A, Hansen MC, Potapov PV, Stehman SV, Houghton RA, Krylov a M, Turubanova S, Goetz SJ. 2015. Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. Environ. Res. Lett. Environ. Res. Lett **10**:74002–74002. IOP Publishing.

Van Dijk K, Cibot M, McLennan MR. 2021. Chimpanzees (*Pan troglodytes*) adapt their nesting behavior after large-scale forest clearance and community decline. American Journal of Primatology **83**:e23323.

Vitousek PM. 1994. Beyond Global Warming: Ecology and Global Change. Ecology **75**:1861–1876.

Vitousek PM. 1997. Human Domination of Earth's Ecosystems. Science **277**:494–499.

Walker B, Steffen W, editors. 1996. Global Change and Terrestrial Ecosystems. Press Syndicate of the University of Cambridge, New York.

Walker KK, Rudicell RS, Li Y, Hahn BH, Wroblewski E, Pusey AE. 2017. Chimpanzees breed with genetically dissimilar mates. Royal Society Open Science **4**:160422.

Walsh P, Abernethy K, Bermejo M. 2003. Catastrophic ape decline in western equatorial Africa. Nature **422**:611–614.

Warton DI, Shepherd LC. 2010. Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. The Annals of Applied Statistics **4**:1383–1402.

Weber MM, Stevens RD, Diniz-Filho JAF, Grelle CEV. 2017. Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. Ecography **40**:817–828.

WFO. 2023. Julbernardia globiflora (Benth.) Troupin. Available from http://www.worldfloraonline.org/taxon/wfo-0000169101 (accessed September 1, 2023).

Wich SA, Garcia-Ulloa J, Kühl HS, Humle T, Lee JSH, Koh LP. 2014. Will oil palm's homecoming spell doom for Africa's great apes? Current biology **24**:1659–63.

Wilcove DS, Giam X, Edwards DP, Fisher B, Koh LP. 2013. Navjot's nightmare revisited: Logging, agriculture, and biodiversity in Southeast Asia. Trends in Ecology and Evolution **28**:531–540. Elsevier Ltd.

Wilkes P, Jones SD, Suarez L, Mellor A, Woodgate W, Soto-Berelov M, Haywood A, Skidmore AK. 2015. Mapping Forest Canopy Height Across Large Areas by Upscaling ALS Estimates with Freely Available Satellite Data. Remote Sensing **7**:2–25.

Wilson ML et al. 2008. Pan troglodytes ssp. schweinfurthii. The IUCN Red List of Threatened Species 2008. Available from http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T15937A5324021.en.

Wilson ML et al. 2020. Research and conservation in the greater Gombe ecosystem: Challenges and opportunities. Biological Conservation **252**. Elsevier Ltd.

Woodcock CE et al. 2008. Free access to Landsat imagery. Science (New York, N.Y.) **320**:1011.

Wright SJ. 2005. Tropical forests in a changing environment. Trends in ecology & evolution **20**:553–60.

Zou H, Hastie T. 2005. Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society. Series B **67**:301–320.