

## RESEARCH ARTICLE

# Development of spatial models and maps for tree species diversity and biomass in a miombo ecosystem, western Tanzania

Adrienne B. Chitayat<sup>1</sup>  | Matthew Lewis<sup>2</sup>  | Moses Anyelwisye<sup>3</sup> | Gabriel S. Laizer<sup>4</sup> | Fiona A. Stewart<sup>5,6,7</sup>  | Serge A. Wich<sup>5</sup>  | Alex K. Piel<sup>6,7</sup> 

<sup>1</sup>Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup>Great Apes & Gibbons Program, Arcus Foundation, Cambridge, UK

<sup>3</sup>Independent Researcher, Arusha, Tanzania

<sup>4</sup>Independent Researcher, Morogoro, Tanzania

<sup>5</sup>School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

<sup>6</sup>Department of Anthropology, University College London, London, UK

<sup>7</sup>Department of Human Origins, Max Planck Institute of Evolutionary Anthropology, Leipzig, Germany

## Correspondence

Adrienne B. Chitayat, Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands.  
Email: [chitayata@gmail.com](mailto:chitayata@gmail.com)

## Funding information

Arcus Foundation; Salk Institute for Biological Studies

Co-ordinating Editor: Hannes Feilhauer

## Abstract

**Aim:** Miombo, a prominent dry forest formation, holds ecological importance for both humans and wildlife. Trees are a driving force behind miombo dynamics, thus, spatially explicit metrics of tree cover are essential for evaluating habitat characteristics, resource availability, and environmental change. We developed predictive models and maps of tree species diversity and biomass within a previously undescribed landscape.

**Location:** Mahale Mountains National Park (MMNP), Greater Mahale Ecosystem (GME), Tanzania.

**Methods:** We created models of tree density, basal area, tree species richness, and tree diversity according to the Shannon Diversity Index. We created a predictive model using an ensemble modeling approach using plot-based data from MMNP and predictor variables derived from satellite data associated with climate, habitat structure, plant productivity, and topography. We assessed predictor importance across models and produced maps based on model predictions and compared them to land cover type and protective status.

**Results:** Results revealed strong positive correlations between tree metrics ( $r \geq 0.70$ ) and substantial overlap in the selection and relative importance of predictors. Canopy height was the most important predictor across models, followed by climate and topography predictors associated with energy. Predictors derived from the soil-adjusted vegetation index were also valuable. Model performances ranged from  $R^2$  values of 0.45 to 0.55, with tree density performing best. Maps show high tree species diversity and biomass in protected areas.

**Conclusions:** This study and the maps it produced provide a baseline for land management and future modeling efforts in the GME. Our results highlight the contribution of a wide variety of environmental predictors and the importance of a select few. We confirmed the importance of the current protected area network where conservation efforts align, and help sustain, an abundance and diversity of trees. Current and

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Applied Vegetation Science* published by John Wiley & Sons Ltd on behalf of International Association for Vegetation Science.

historical disturbance-related predictors should be considered to address remaining unexplained variance.

#### KEYWORDS

basal area, biomass, ensemble model, Mahale, miombo, satellite data, species diversity, species richness, tree cover, tree density

## 1 | INTRODUCTION

Miombo is one of Africa's most valuable, threatened, and extensive dry forest formations (Timberlake & Chidumayo, 2011; Ribeiro, 2015). Stretching from Tanzania to Zimbabwe, the miombo ecoregion boundary is primarily determined by an interaction between topography and climate and occurs within a zone that experiences extreme fluctuations in rainfall. Its distinct biodiversity and vast distribution have garnered international recognition, and it is included in the world's top five list of most valuable wilderness areas (Mittermeier et al., 2003). Miombo provides a myriad of services (e.g., nutrient cycling, erosion control) and resources (e.g., food, wood fuels), supporting the approximately 150 million people who live on these lands and in nearby urban settlements (Ryan et al., 2016). This ecosystem is highly threatened by anthropogenic land use activities and climate change, dramatically affecting species distributions and ecosystem processes (Verhegghen et al., 2022). Dry forests like miombo now surpass humid forests (Sunderland et al., 2015) (that have been the subject of far more scientific inquiry and resource allocation) (Schröder et al., 2021) in deforestation rate.

Miombo ecosystems are dominated by tree species from the family Fabaceae in the genera *Brachystegia*, *Julbernardia*, and *Isoberlinia* (Frost, 1996). While they are primarily composed of woodland, these ecosystems are a mosaic of vegetation types, e.g., grasslands, shrublands, dense forests, that vary in size and composition according to environmental factors like soil, water, and topography and anthropogenic factors like land use (Timberlake & Chidumayo, 2011). The dynamics of miombo are driven by the dynamics of trees, which strongly influence vegetation structure, water and nutrient cycling, and fire regimes (Frost, 1996). Data detailing tree cover characteristics such as tree distribution, density, and diversity can help researchers describe land cover types and identify resource-rich areas, as well as assess the health and productivity of the ecosystem to the benefit of both wildlife and humans (Torres & Lovett, 2013; Ribeiro, 2015; Ryan et al., 2016). The importance and vulnerability of miombo and its resources has garnered attention (Syampungani et al., 2009), but inter-site variability warrants expanding research into understudied regions and localities (FAO, 2019; Schröder et al., 2021).

One such understudied region is the Greater Mahale Ecosystem (GME) in western Tanzania. The Tanzania miombo represents the most northern and eastern boundary of the miombo ecoregion, yet it is largely separated from most of the ecoregion due to the expansion of cropland (Timberlake & Chidumayo, 2011). The GME (20,000 km<sup>2</sup>) is part of this ecoregion and the Albertine Rift, and

encompasses crucial habitat for threatened species and plays a vital role in regional biodiversity and ecosystem services (Plumptre et al., 2007; Bietsch et al., 2016). The land is managed under a variety of designations that include protected and unprotected areas. Mahale Mountains National Park (MMNP), the only national park within the GME, exhibits low human activity and protects key habitat for the biodiversity of the region (Plumptre et al., 2007; Chitayat et al., 2021). Outside MMNP the landscape faces threats from accelerated human population growth and expanding land use practices (e.g., agriculture, charcoal production) that are driving habitat loss and degradation and threatening biodiversity in the region (Pintea, 2012; Thomsen et al., 2023). Tree cover within the GME has yet to be comprehensively evaluated and described, and because of the importance of tree cover in miombo dynamics, this is a significant gap in our understanding of this landscape and the broader miombo ecosystem. One reason this has not yet been accomplished is due to the physical and logistical challenges involved in surveying the vast miombo woodlands of the GME, compounded by the fact that historical research efforts in western Tanzania have primarily concentrated on chimpanzees at established long-term field sites (Nakamura, 2012; Wilson, 2012; Piel et al., 2019).

Satellite remote sensing technology has emerged as an invaluable tool for alleviating the limitations of ground surveys, providing spatial-temporal data that can be used to identify and describe biophysical landscape attributes like trees (Kerr & Ostrovsky, 2003). Moreover, as satellite data is often collected continuously and is freely available to the public, it provides accessible, up-to-date information that can be easily integrated into monitoring frameworks. Spatial predictive models can be used to connect satellite data with field data to make broad-scale predictions that expand the utility of in-situ observations (McNellie et al., 2021; Barreras et al., 2023). These predictions can be used to generate continuous maps that serve as valuable communication tools and can also inform land management decisions, guide conservation efforts, and facilitate infrastructure development projects. Spatial models can also aid in the identification of current and historic landscape conditions, uncovering patterns that aid our understanding of ecosystem dynamics and predicting changes that may arise (Jinga & Palagi, 2020).

The GME is vital for biodiversity, especially threatened species, in the miombo ecoregion. Here, we aim to fill a crucial gap in our understanding of tree cover in the GME with the development of important tools that can describe tree cover characteristics related to species diversity and biomass in a miombo. We conducted field surveys in MMNP where the minimal human impact allows us to

focus on non-anthropogenic sources of tree cover variability. Our objectives were to (1) create predictive models of tree cover characteristics related to tree species diversity and biomass and (2) generate continuous, high-resolution maps for the GME. We present a baseline assessment on the development of tree cover spatial models for the GME and discuss landscape patterns revealed by model predictions concerning land cover land use types and protected areas (PAs).

## 2 | METHODS

### 2.1 | Study area

The GME is an ecologically rich area that boasts a variety of distinct species and those of conservation concern, including eastern chimpanzees (*Pan troglodytes schweinfurthii*), African elephants (*Loxodonta africana*), and mninga trees (*Pterocarpus angolensis*) (Plumptre et al., 2007; Thomsen et al., 2023). Situated alongside Lake Tanganyika, the health of this ecosystem is inherently connected to the lake, which is significant for its ecosystem services and its diverse aquatic life, hosting 300 endemic species of fish (Sweke et al., 2013). The GME exhibits topography consisting of broad valleys interspersed by steep mountains and flat plateaus, ranging in elevation from 780 to 2460m above sea level (Carvalho et al., 2022). MMNP protects 1517km<sup>2</sup> of land and encompasses the GME's highest peaks. The region is highly seasonal, experiencing a rainy season from October to mid-May and a dry season from mid-May to September. Annual precipitation ranges from 900 to 2100mm, and temperatures vary from 11 to 38°C.

### 2.2 | Response variables – Tree cover characteristics

We collected tree data in 463 plots across MMNP from March 2018 to January 2019 (Figure 1) as part of an extensive study on chimpanzees (Chitayat et al., 2021). Plots were placed along transects at 12 randomly selected sites where we established eight transects, each measuring 1 km in length. These transects featured plots measuring 5 m × 100m, spaced at 100m intervals. Within each plot, all trees with a diameter at breast height (1.3m) ≥ 10cm were identified and measured.

We calculated tree cover characteristics at the plot level that were used as response variables in the development of four separate predictive models: (1) tree density, (2) basal area of trees, (3) tree species richness, and (4) tree species diversity according to the Shannon Diversity Index. Tree density and basal area are both useful biomass proxies (Torres & Lovett, 2013): density provides information about the number of trees per unit area, while basal area quantifies the volume of trees per unit area according to the diameter at breast height. Similarly, species richness and the Shannon index offer different measures of species diversity: species richness counts the number

of species per unit area, while the Shannon index considers both the number of species and their relative abundances in its calculation (Shannon, 1948; Whittaker, 1972). We selected these characteristics because they provide valuable and complementary information about forest structure and composition. Non-parametric Spearman rank correlation coefficient tests were conducted to establish the relationship between all pairs of tree cover characteristics. We tested for spatial autocorrelation using Moran's *I* test and found no statistically significant spatial autocorrelations ( $p > 0.05$ ) for any response variables (Zuur et al., 2010).

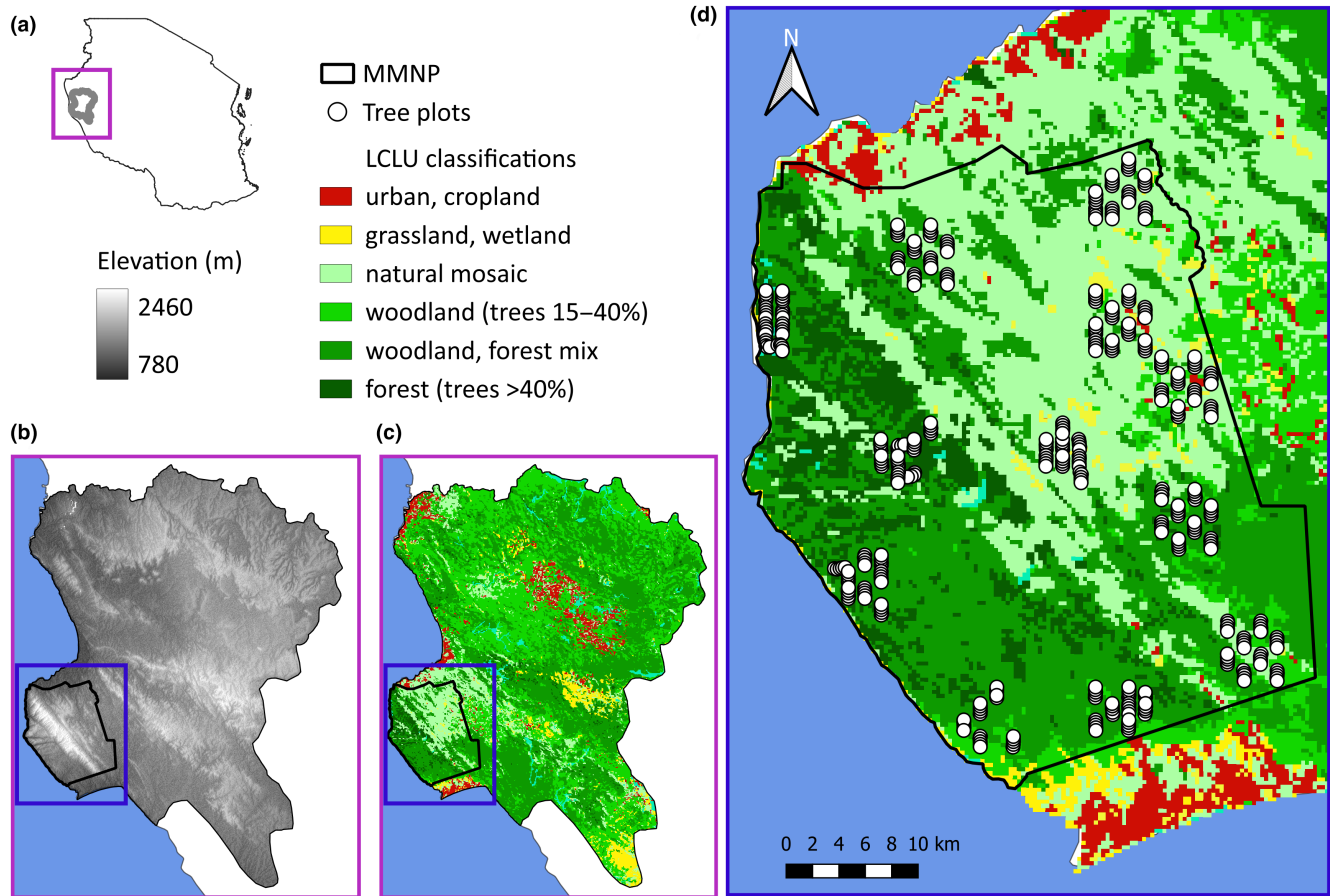
### 2.3 | Predictor variables – Satellite remote-sensing data

We utilized environmental predictor variables derived from remote-sensing satellite data to build models of tree cover characteristics in the GME. Tropical dry forests can be challenging to model using satellite data owing to the pronounced temporal and spatial fluctuations of these ecosystems (Bastin et al., 2017; Verhegghen et al., 2022); however, incorporating a wide array of biotic and abiotic predictors can help improve the quality of model outputs (Slik et al., 2010; Zellweger et al., 2015). The predictors were categorized into four groups: (1) climate, (2) habitat structure, (3) plant productivity, and (4) topography (Table 1).

We tested several predictors derived from climatic data, including mean annual temperature and precipitation from the period of 1979–2013 (Karger et al., 2017) and mean annual actual and potential evapotranspiration from the period of 1950–2000 (Zomer et al., 2006). These variables influence the availability of crucial resources like energy and water (White, 1983; Ribeiro et al., 2020) and are often used for modeling vegetation in miombo (Pearson et al., 2006; Slik et al., 2010; Zellweger et al., 2015).

Canopy heights (m) were used to represent habitat structure (Potapov et al., 2021). This variable has demonstrated previous success for modeling tree cover characteristics (Wolf et al., 2012; Knapp et al., 2020).

To represent plant productivity, we included a variety of vegetation indexes (VI) that were derived from Sentinel-2. We used L1C data for its higher spatial resolution and a cloud-masking technique was applied to remove pixels affected by clouds and cirrus using the QA60 band (bits 10 and 11). While surface reflectance data (L2A) are commonly used, atmospheric correction is not always required for classification applications, and vegetation indices such as the normalized difference vegetation index (NDVI) and soil-adjusted vegetation index (SAVI) reduce atmospheric effects (Song et al., 2001; Verhegghen et al., 2022). Researchers commonly utilize VIs to model trees because their capacity to quantify greenness is effective for assessing plant abundance and variability (Pettorelli et al., 2005; Timberlake et al., 2010; Barati et al., 2011; Mutowo & Murwira, 2012; Zellweger et al., 2015; Cavada et al., 2017; Mayes et al., 2017). For each VI, we calculated summary statistics using all of the 2018 values, the study period, which served as our predictors



**FIGURE 1** Locator map of the study area in western Tanzania. Maps show (a) the location of the Greater Mahale Ecosystem (GME) in Tanzania, (b) a digital elevation model of the GME, (c) land cover types in the GME, and (d) the location of vegetation plots sampled across Mahale Mountains National Park (MMNP).

(Cabacinha & de Castro, 2009). For example, maximum (max) value was the highest observed value from 2018 and the mean value was the mean of all 2018 values.

Topographic predictors were used to describe topographic position and heterogeneity (Amatulli et al., 2018; Mattivi et al., 2019) and obtained from the Shuttle Radar Topography Mission (SRTM). These features can influence tree cover characteristics through associations with local microclimates, soil properties, and water accumulation (de Castilho et al., 2006; Thuiller et al., 2006; Engelbrecht et al., 2007; Homeier et al., 2010; Slik et al., 2010; Fricker et al., 2015).

Satellite data were collected at resolutions ranging from 10 to 1000m (see Table 1). Data were resampled, scaling either up or down, to a uniform resolution of 100m using conventional bilinear interpolation (Divišek & Chytrý, 2018). This was done to mitigate potential challenges from spatial resolution disparities like data compatibility and model complexity (Moudrý et al., 2019).

## 2.4 | Predictor variable selection process

To create the most parsimonious predictive models, we applied a variable selection process to remove predictors that provided little

predictive capacity or were highly correlated based on a correlation coefficient of  $r \geq |0.7|$  (Jantz et al., 2016). First, we utilized recursive feature elimination to remove predictors of low importance (David et al., 2022). Then, using Pearson product-moment and Spearman rank correlation coefficient tests, we identified pairs of predictors that were highly correlated (Zuur et al., 2010) and retained the predictor with higher relative importance according to recursive feature elimination rankings. Finally, we tested for multicollinearity using a variance inflation factor (VIF) set to 10 or more. This variable selection process was used for each response variable and subsequent model. We only included the selected predictors in future analyses and the building of ensemble models. A spatial evaluation of the calibration range of selected predictors showed that our training data were able to capture the majority of conditions across the GME but that coverage varies between predictors (Appendix S1).

## 2.5 | Ensemble model development

To build and verify the models, each data set was divided into training (80%) and testing (20%) subsets. We employed an ensemble modeling approach that combines multiple individual base

TABLE 1 Environmental predictor variables used to develop ensemble models of tree cover characteristics.

Predictor variable	Description	Mean value (range)	Model	Source
<b>Climate</b>				
Rain	Mean annual precipitation (mm) <sup>a</sup>	1369 (1042–1828)	BA	CHELSEA
Temperature	Mean monthly temperature (°C) <sup>a</sup>	21.5 (15.7–23.7)	R, SDI	CHELSEA
AET	Mean annual actual evapotranspiration <sup>b</sup>	982 (392–1334)	D, BA, R, SDI	CGIAR-CSI
PET	Mean annual potential evapotranspiration <sup>b</sup>	1709 (1351–1862)	BA	CGIAR-CSI
<b>Habitat structure</b>				
Canopy height	Canopy height (m) <sup>c</sup>	9.3 (0–25.9)	D, BA, R, SDI	GLAD
<b>Productivity</b>				
NDVImax	Maximum (max) normalized difference vegetation index (NDVI) <sup>d</sup>	0.78 (0.69–0.84)	BA	S2
NDVImean	Mean NDVI	0.34 (0.26–0.45)	BA	S2
RGImean	Mean red–green index (RGI) <sup>e</sup>	0.11 (0.01–0.25)	D, BA, R, SDI	S2
RGIstdv	Standard deviation (stdv) of RGI	0.09 (0.04–0.14)	D, BA, R, SDI	S2
RRmean	Mean reflectance ratio (RR) <sup>e</sup>	1.29 (0.74–1.59)	D, BA, R, SDI	S2
RRstdv	Stdv of RR	0.20 (0.07–0.99)	BA	S2
SAVImax	Max soil-adjusted vegetation index (SAVI) <sup>f</sup>	1.17 (1.03–1.26)	D, R, SDI	S2
SAVImean	Mean SAVI	0.51 (0.38–0.78)	D, R, SDI	S2
SAVIstdv	Stdv SAVI	0.10 (0.03–0.17)	D, BA, R, SDI	S2
SLAVImax	Max specific leaf area vegetation index (SLAVI) <sup>g</sup>	1.69 (1.31–2.28)	D, BA, R, SDI	S2
<b>Topography</b>				
Elevation	Elevation (m a.s.l.) <sup>h</sup>	1204 (797–2212)	D	SRTM
Ruggedness	Terrain ruggedness index <sup>h</sup>	2.0 (0.7–6.5)	D	SRTM
Roughness	Topographic roughness index <sup>h</sup>	15.6 (2.6–53.7)	BA, R, SDI	SRTM
TPI	Topographic position index <sup>h</sup>	0.04 (–2.34–3.55)	BA	SRTM
TWI	Topographic wetness index <sup>i</sup>	7.2 (4.7–12.7)	BA	SRTM

Note: Individual models were developed for tree density (D), stand basal area (BA), tree species richness (R), and tree diversity according to the Shannon Diversity Index (SDI). Data sources include the CGIAR-CSI Global-Aridity and Global-PET Database (CGIAR-CSI) (1000m), Climatologies at high resolution for the Earth land surface areas (CHELSA) (1000m), Copernicus SENTINEL-2 (S2) (10m), Global Land Analysis and Discovery (GLAD) (30m), and Nasa's Shuttle Radar Topography Mission (SRTM) (30m).

<sup>a</sup>Karger et al. (2017).

<sup>b</sup>Zomer et al. (2006).

<sup>c</sup>Potapov et al. (2021).

<sup>d</sup>Pettorelli et al. (2005).

<sup>e</sup>Cavada et al. (2017).

<sup>f</sup>Mutowo and Murwira (2012).

<sup>g</sup>Barati et al. (2011).

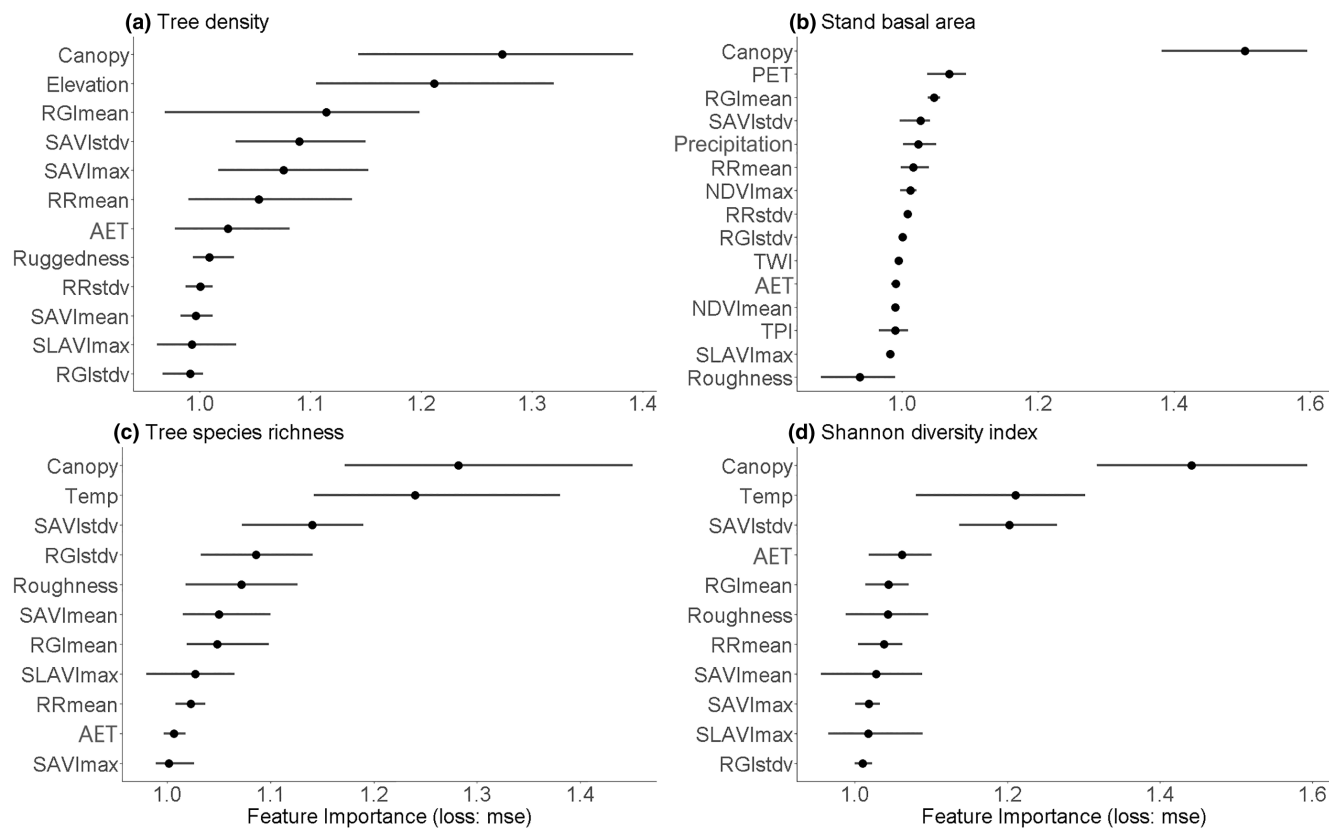
<sup>h</sup>Amatulli et al. (2018).

<sup>i</sup>Mattivi et al. (2019).

models into a single predictive ensemble model that capitalizes on the collective strength of its components to improve model robustness and reduce uncertainty (Van Der Laan et al., 2007). Ensemble models were developed separately for each tree cover characteristic using the package *SuperLearner* and a 10-fold cross-validation risk. Three base modeling methods were used to develop ensembles: a regression method (multivariate adaptive regression splines, MARS) and two machine-learning methods (gradient boosted model, GBM, and random forest, RF). All methods account for non-linear relationships and automatically

consider interactions between variables while avoiding overfitting (Leathwick et al., 2006; Zellweger et al., 2016).

For our ensemble models, we included the default model and several tuned models of each base modeling method, which can increase ensemble model performance by leveraging the strengths of different and diverse model configurations (Polley & Van Der Laan, 2010). We tuned base models according to critical hyperparameters that can enhance prediction accuracy and reliability (Appendix S2) (Yates et al., 2023). Tuning was performed using a 10-fold cross-validation to ensure robust model performance and avoid



**FIGURE 2** Importance of predictor variables used to model individual tree cover characteristics in the Greater Mahale Ecosystem. Plots show a 95% confidence interval around the mean squared error (MSE).

overfitting. The optimal values were determined through a grid search over specified ranges, allowing for systematic exploration of parameter combinations.

We present three statistical metrics to evaluate the performance of our ensemble models: the correlation coefficient ( $r$ ) depicting the relationship between predicted and observed values, the percentage of explained variance ( $R^2$ ), and the percentage of root mean squared error derived from a 10-fold cross-validation resampling procedure with data shuffling. Higher  $r$  and  $R^2$  values and lower root mean squared error values signify superior model performance. We assessed the importance of predictor variables by measuring their contributions to reducing cross-validated risk through mean squared error.

## 2.6 | Mapping spatial model predictions

We used ensemble models to generate spatially explicit predictions and produce individual maps of tree cover characteristics. Maps showing the uncertainty of model predictions were also generated for each tree cover characteristic by quantifying the standard deviation of prediction results from all individual models within the ensemble. We created a composite map aggregating all tree cover characteristics by summing each raster cell's normalized predicted values, scaled from 0 to 1. Then, we evaluated the distribution of

composite values against land cover designations as determined by the Climate Change Initiative and European Space Agency (CCI/ESA) land cover map for 2020 (Zanaga et al., 2021). All analyses were performed in R (R Core Team, 2022) and QGIS 3.6 (QGIS Development Team, 2020).

## 3 | RESULTS

### 3.1 | Ensemble models of tree cover characteristics

Tree cover characteristics varied between plots, ranging from 0 to 37 for tree density (mean =  $9 \pm 7$ ), 0 to 5.1 for basal area (mean =  $0.5 \pm 0.5$ ), 0 to 16 for tree species richness (mean =  $5 \pm 3$ ), and 0 to 2.6 for Shannon index (mean =  $1.2 \pm 0.7$ ). Pairwise correlation tests revealed all tree cover characteristics to be highly correlated with each other ( $r = 0.70 - 0.97$ ), with the strongest relationship between tree species richness and the Shannon index.

Each model included at least one predictor variable from each category (e.g., climate). Several predictors were selected for basal area that were absent from other models (e.g., precipitation, NDVImean). Other than these predictors for basal area, the selection and relative importance of predictor variables showed substantial overlap among response variables. This was especially true between species richness and the Shannon index models that included the

same predictors and displayed the same top three most important (Figure 2). Canopy height consistently demonstrated the highest mean importance across ensemble models. Temperature, potential evapotranspiration, and elevation emerged as the second most important predictors across models. These predictors also exhibited strong positive correlations with each other that prevented their inclusion within the same model.

Performance scores varied between ensemble models of tree cover characteristics, with  $R^2$  scores ranging from 0.45 to 0.55 and our model for tree species density performing best across metrics (Table 2). Additionally, predicted values were all significantly correlated with their observed values ( $p < 0.001$ ).

### 3.2 | Maps of tree cover characteristics in the GME

We expanded model predictions to produce maps of tree biomass and diversity in the GME (Figure 3), which exhibited generally low levels of uncertainty (Figure 4). Higher uncertainty values are predominantly observed in areas with known human settlements along the coast of Lake Tanganyika and the north-central region of the GME around the Mishamo refugee camp, as well as along the highest elevation peaks in MMNP. Similarities and divergences between tree cover characteristics can be readily observed in our composite map (Figure 5), such as high-value areas along the western side of MMNP. Nearly half of the GME is under some protective status (49%), where the majority of the highest and lowest composite values occur, 64% and 66%, respectively. Composite values were disproportionately distributed across land cover types, with results showing that sites dominated by stretches of forest represent only 1% of the landscape but hold 10% of the highest 25% of composite values (Table 3). Similarly, results showed that the lowest 25% of composite values occur in primarily non-tree-dominated, natural or human-modified land cover types, such as grasslands and croplands. Additionally, the land cover classification system was sometimes inaccurate within the study area. For example, some misclassifications occurred in areas that are known to be occupied by human settlements and croplands but were classified as land cover types such as grasslands or shrublands.

## 4 | DISCUSSION

We modeled and made predictions of tree species diversity and biomass for the GME by using an ensemble modeling framework,

available remotely sensed satellite data, and plot-based survey data of trees in MMNP. Individual ensemble models achieved moderate performance scores, with our models of tree density and tree species richness performing best. High correlations between observed and predicted values for all response variables suggest that models could generalize well on unseen data. Maps that show the extension of model-based predictions allow for visualizing tree cover characteristics across the GME landscape.

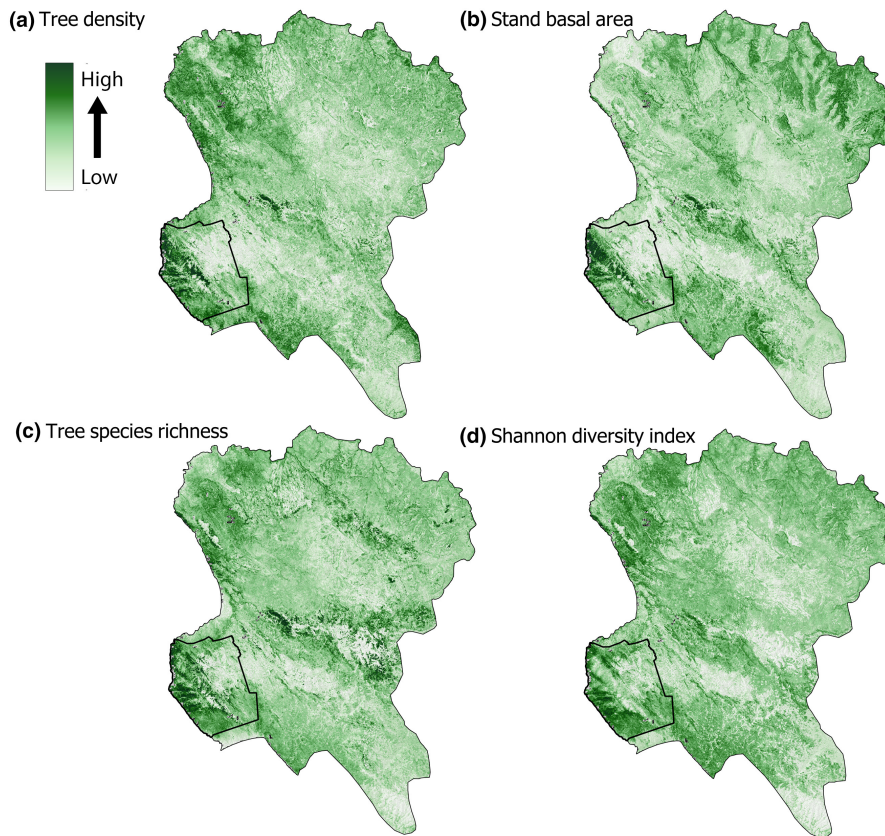
### 4.1 | Relative importance of predictor variables

The diverse environmental predictors we tested showed substantial overlap during the predictor selection process and in relative importance among models. This likely stems from the strong positive correlations between tree cover characteristics similarly affected by environmental factors like soil conditions (Homeier et al., 2010). Canopy height was the most important predictor for all models, which aligns with previous research showing that canopy height can account for a significant portion of the observed variability in tree species diversity and biomass metrics in tropical forests on local scales (Wolf et al., 2012; Knapp et al., 2020). The importance of canopy height may also correspond to structural contrasts between GME vegetation types. Some areas, like grasslands, are not tree-dominated and exhibit low tree biomass; conversely, areas where trees dominate the landscape display discernible canopy structures. Ogawa et al. (2007) observed differences in canopy height between vegetation types in the GME, showing that forested sites, predominantly located in valley bottoms, have a greater density of tall trees than woodlands. Similar findings have been documented for other sites, showing tree height, basal area, and richness decrease from valleys to plateaus (Homeier et al., 2010; Rodrigues et al., 2020). This pattern may stem from local differences in soil water and nutrient availability that alter plant growth and can impact species coexistence (Ryan & Yoder, 1997; Homeier et al., 2010). Canopy height can vary across different vegetation types under the characteristics of the species found therein, which display minimal overlap in MMNP (Chitayat et al., 2021).

The strong positive correlations between temperature, potential evapotranspiration, and elevation and their concurrent high importance across models suggest that a common underlying factor like atmospheric energy availability may indirectly drive their value (Hawkins et al., 2003; Fisher et al., 2020). Energy availability is known to influence plant growth, reproduction (Dong et al., 2012),

**TABLE 2** Performance metrics of individual ensemble models for tree cover characteristics related to tree species diversity and biomass.

Response variable	Correlation coefficient ( $r$ )	Root mean squared error (RMSE)	Explained variance ( $R^2$ )
Tree density	0.7	16%	0.55
Tree stand basal area	0.7	19%	0.47
Tree species richness	0.7	19%	0.54
Tree species diversity (Shannon Diversity Index)	0.7	23%	0.45



**FIGURE 3** Predictions of tree cover characteristics across the Greater Mahale Ecosystem. Predictions were generated from individual ensemble models developed using plot-based vegetation data from Mahale Mountains National Park (outlined in black) and satellite-derived predictors. The color gradient legend applies to all maps.

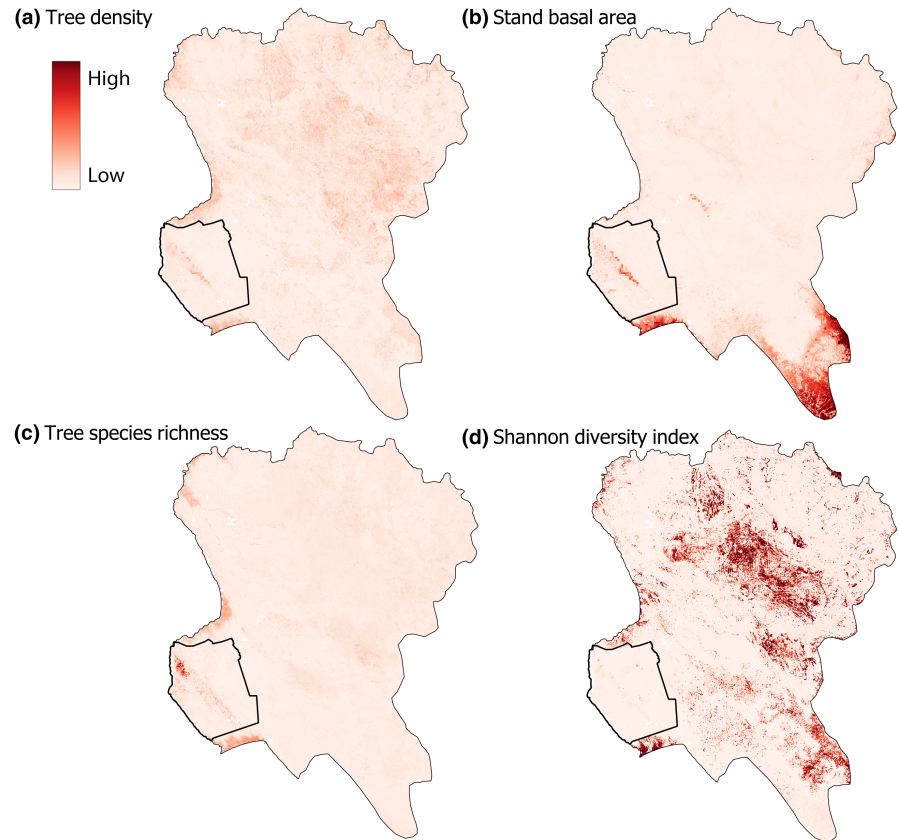
and species diversity (Fisher et al., 2010). Conversely, water availability is often shown to drive plant dynamics in the tropics (Hawkins et al., 2003; Siyum, 2020); however, water-related predictors were mostly excluded from model development (i.e., topographic wetness index [TWI], precipitation). Precipitation was included in our model of basal area but demonstrated only modest importance toward predictions. This aligns with the results of (Barreras et al., 2023), which showed that mean precipitation held relatively low importance in ensemble models for tree density and height, especially in dry ecosystem types. Actual evapotranspiration was the only water-related predictor chosen across response variables, but it consistently displayed only modest importance. Actual evapotranspiration's broad connection to water and energy dynamics might render it a somewhat simplistic index of the environmental factors affecting plant growth (Fisher et al., 2010), allowing it to be widely applicable but restricting its explanatory power. However, the spatial resolution of our climatic data scale may have also influenced the effectiveness of these predictors in describing tree cover dynamics. These data were upscaled from a coarser resolution of 1000m to 100m, which redistributes the data across a finer grid but does not add any new information. Coarse data have clear benefits for model development, reducing model size and complexity, but coarse climatic data can also introduce errors when there is a discrepancy between the climate used in analysis and that experienced by organisms and the habitat on the ground (Bütikofer et al., 2020). Particularly in heterogeneous and mountainous regions where local microclimate conditions can differ dramatically, fine-scale data sets may be more

useful. Future modeling efforts may benefit from exploring the use of higher-resolution climatic data, diminishing the risk of missing critical small-scale variations.

Multiple SAVI-based predictors were selected and showed importance among models. SAVI provides advantages over other VIs given the GME's high seasonality and predominantly open vegetation, leading to significant spatial and temporal soil exposure and reduced vegetative coverage (Huete, 1988). Moreover, SAVI's ability to mitigate the influence of soil brightness is valuable in fire-impacted landscapes, like the GME, where exposed and charred soil leads to distinct soil-vegetation contrasts (White & Swint, 2014; Meng et al., 2017). In Tanzania, fire events occur annually even in PAs, and are caused mainly by anthropogenic ignitions for activities such as farm preparation, cattle grazing, hunting, honey harvesting, and charcoal production (Kikula, 1986). PAs like MMNP are also subjected to prescribed burns managed by authorities, which are generally employed to protect the ecosystem under a fire management plan. In miombo, fire plays a pivotal role in shaping tree species composition and is often linked to a decline in tree biomass (Frost, 1996; Chidumayo, 2013).

The importance of VI-based predictors differed between models of tree species diversity and models of tree biomass. RGI (red-green index) mean was found to be more important in models of tree biomass, likely due to its association with tree crown size which is related to tree size (Cavada et al., 2017). In addition, when predicting the biomass of long-lived species, the mean may be a more relevant metric as it smooths out the short-term fluctuations and

**FIGURE 4** Maps showing the level of uncertainty of ensemble model predictions for tree cover characteristics across the Greater Mahale Ecosystem, with Mahale Mountains National Park outlined in black. The color gradient applies to all maps and provides the relative level of prediction uncertainty (standard deviation) from low to high.



extremes inherent in metrics such as standard deviation or maximum. Conversely, SAVI standard deviation encapsulates temporal variation in vegetation greenness and demonstrates greater importance in models of tree species diversity. Temporal variations reflect seasonal changes influenced by tree species compositions, distributions, and phenological characteristics (e.g., deciduous). This predictor gains significance by capturing these differences and the occurrence of disturbance events like fires, which often result in dramatic changes to greenness, and connecting them to spatial differences in tree diversity.

#### 4.2 | Spatial predictions across the GME

Model predictions demonstrated spatial similarities between tree cover characteristics, revealing overlapping areas of high or low values. This may reflect stages of ecological succession as the GME experiences annual fires (Naftal et al., 2022) and successional dynamics play a crucial role in shaping vegetation structure and composition (Ouyang et al., 2016), particularly in areas recovering from or experiencing disturbance events (Kalaba et al., 2013; Saito et al., 2014). The observed pattern between tree characteristics may also indicate a potential functional relationship between diversity and biomass, whereby diverse tree communities encourage tree growth through more efficient resource utilization (Grossman et al., 2018). Similar patterns can be expanded to other taxa, with tree diversity and abundance identified as key drivers of taxon-level (e.g., bats,

birds, insects) and forest-associated biodiversity in previous studies (Harvey et al., 2006; Ampoorter et al., 2020). Further investigation on this topic is required as the relationship between biomass and biodiversity is often complex and non-linear, at times exhibiting a hump-shaped (unimodal) pattern that reveals a decline in biodiversity at the highest biomass levels (Graham & Duda, 2011). In a woodland-forest mosaic in eastern Tanzania, Shirima et al. (2015) observed a monotonically increasing trend between tree species richness and biomass in miombo woodland, but a unimodal pattern in montane forest. This suggests that there could be an optimal level of biomass that maximizes species diversity in montane forests but not miombo woodland (Shirima et al., 2015).

Forests, particularly non-riverine forests that stretch out over large areas, are rare in the GME. We observe that high tree species diversity and biomass occur disproportionately more in forest and mixed woodland-forest land covers. The highest concentration of high-value predictions can be found along the lakeshore of MMNP, which encompasses large swaths of miombo woodland and the largest forest block in the region. This finding supports the well-established importance of northwestern MMNP, which is known to support relatively unique species assemblages and high densities of species like chimpanzees (Plumptre et al., 2007; Chitayat et al., 2021). PAs in East Africa were historically established with large, charismatic mammal species in mind, yet these efforts can lead to broad advances that extend beyond the original intention, such as biodiversity conservation or carbon sequestration (Banda et al., 2006; Dickson et al., 2020). Outside MMNP, we see high values clustered within the Ntakata

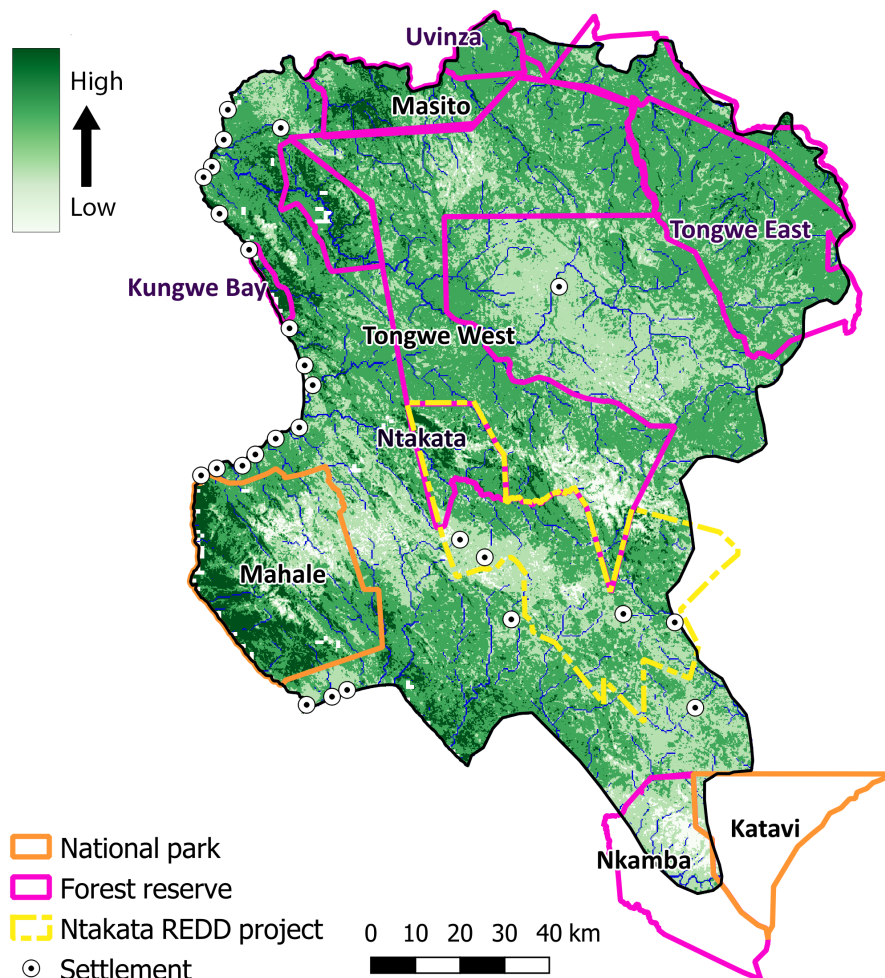


FIGURE 5 Composite map of tree species diversity and biomass predicted across the Greater Mahale Ecosystem. Values are based on the summed predictions of individual ensemble models.

TABLE 3 Percentage of composite values of tree cover characteristics occurring in different land cover types.

Land cover type	% GME landscape	Tree cover composite values			
		Low quartile 1	Med-low quartile 2	Med-high quartile 3	High quartile 4
Cropland, human settlement	5%	10%	14%	2%	1%
Grassland, wetland	5%	32%	7%	3%	4%
Natural mosaic	10%	32%	16%	6%	3%
Woodland (15%–40% trees)	40%	14%	41%	42%	26%
Woodland, forest mix	39%	12%	22%	46%	56%
Forests ( $\geq 40\%$ trees)	1%	0%	0%	1%	10%

Note: Land cover types may encompass multiple vegetation types, for example, natural mosaics can include bamboo thicket, shrubland, and miombo woodland. Composite values are divided into equal interval quartiles, with quartile 1 containing the lowest 25% of values and quartile 4 the highest 25% of values calculated across the GME.

forest reserve. This area also possesses significant ecological importance, providing sanctuary for endangered and endemic species and corridor habitat for elephants and chimpanzees beyond the boundaries of national parks (Plumptre et al., 2007; Lindsey et al., 2020). This reserve is part of the Ntakata Mountains REDD project, spearheaded by Carbon Tanzania (Dickson et al., 2020).

The composite map highlights the importance of the PA network in the GME for conserving tree species diversity and

biomass. However, there is clear variation both within and between PAs likely stemming from multiple factors. This includes the natural heterogeneity of the GME and occurrence of vegetation types that display low tree cover, such as grasslands, and thus low composite values. Additionally, there are differences in when PAs were established, the history of land use across the GME, and PA designation (e.g., national park, village forest reserve), which dictates restrictions on human activity and often the availability

of resources to enforce those restrictions. This results in a spectrum of conservation outcomes, showcasing the value of long-established and strictly regulated PAs like MMNP, as well as those occurring in areas with a history of sustainable land use practices like Ntakata forest reserve. Comparatively, those with less restrictive protections or lacking resources may fail to prevent habitat loss and degradation (Itoh et al., 2012; Thomsen et al., 2023), or may simply require more time to benefit from their recent changes to their status and resources, such as the Tongwe West Forest reserve.

### 4.3 | Model improvements

The performance scores of our models demonstrate unexplained variance that may result from the scale of the predictors used or a lack of predictors related to current and historic environmental disturbances. After MMNP was gazetted in 1985, authorities prohibited human activities except those related to tourism and research. However, illegal activities (e.g., wood harvesting, medicinal plant collection) and human encroachment into the park can be challenging to manage and vary over time. Miombo demonstrates the ability to recover from human activities, but the process can take decades, and recurrent disturbances can significantly impede the recovery rate (Williams et al., 2008; Chidumayo, 2013; Kalaba et al., 2013). Because of these dynamics, Ribeiro (2015) argue that land use data and land cover recovery rates are valuable for modeling vegetation patterns in miombo.

Future models may also benefit from the inclusion of fire-related predictors that go beyond the scope of the VI predictors we included. Fire has a well-documented influence on tree cover in miombo woodlands (Frost, 1996) and studies indicate that fire characteristics like frequency, intensity, and timing can have varied effects on trees and tree communities (Ryan & Williams, 2011; Chidumayo, 2013; Mapaure, 2013; Saito et al., 2014; Tarimo et al., 2015; Buramuge et al., 2023). While fire has the potential to open up areas for colonization and promote diversity through the establishment of new species, frequent and intense fires can reduce diversity and biomass by favoring fire-tolerant species and suppressing tree growth because of mortality or damage to the stems, bark, and roots (Ribeiro, 2015; Buramuge et al., 2023). Moreover, disturbance impacts can be synergistically influenced by additional factors like large herbivores (Mapaure, 2013; Ivory & Russell, 2016). Improved model accuracy may require the integration of novel variables that, independently or in synergy with other variables, can help explain the spatial variability of tree cover characteristics and enhance the reliability of models.

Lastly, to capture a greater diversity of environmental conditions across the GME and enhance the reliability of predictions, future efforts should incorporate data from areas outside MMNP and protected zones. By using data solely from MMNP in the current study, we lacked data from sites where human activities like agriculture or cattle herding occur – practices prohibited in MMNP but

widespread outside the GME – that are known to impact tree cover characteristics (Chidumayo, 2013; Macave et al., 2022; Buramuge et al., 2023).

## 5 | CONCLUSIONS

This study provides a crucial baseline for the development of spatial models and tree maps in the heterogeneous miombo landscape of the GME, addressing a previously lacking area of research. The challenges encountered underscore the complexity of modeling trees in this landscape. Our results emphasize the significance of the existing PA network where high composite values align with unique species assemblages and conservation efforts. The strong positive correlations found between tree cover characteristics suggest a potential functional relationship between tree diversity and biomass, which may have implications for broader biodiversity conservation efforts. This finding warrants further investigation as it may yield insights into biodiversity and biomass patterns across taxa that can inform broad monitoring and conservation strategies. We are cautious in interpreting some results due to methodological limitations and encourage future research to incorporate additional explanatory factors and data from areas outside protected zones to enhance the reliability of estimates. Given the environmental impacts we can expect to occur as a result of climate change and human population growth (Warren et al., 2018; Elisa et al., 2024), there is an urgent need to expand and refine spatial products and tools that can help monitor any changes and guide the effective management of the GME.

### ACKNOWLEDGEMENTS

The successful completion of our research project would not have been possible without the support and partnership of several organizations. We extend our gratitude to the Greater Mahale Ecosystem Research and Conservation (GMERC) and the Frankfurt Zoological Society. Additionally, we acknowledge the UCSD/Salk Institute Center for Academic Research and Training in Anthropogony (CARTA) for supporting GMERC and the permanent research station at Issa Valley. We are also thankful to the Tanzania Wildlife Research Institute (TAWIRI), Tanzania Commission for Science and Technology (COSTECH), and Tanzania National Parks (TANAPA) for granting us permission to conduct this research and for their support during the fieldwork. Lastly, we would like to express our appreciation to Godfrey Stephano, Makasa Mlekwa, Shabani Kabangula, Mustafa Kizimba, and the many Tanzania field assistants and TANAPA rangers for their invaluable expertise in the field.

### FUNDING INFORMATION

Alex K. Piel and Fiona A. Stewart received funding from the Arcus Foundation and the UCSD/Salk Institute Centre for Academic Research and Training in Anthropogony (CARTA).

## DATA AVAILABILITY STATEMENT

The data and code that support the findings of this study are openly available in figshare (<https://figshare.com/>) at <https://doi.org/10.6084/m9.figshare.c.6874501>.

## ORCID

Adrienne B. Chitayat  <https://orcid.org/0000-0003-2651-4010>

Matthew Lewis  <https://orcid.org/0000-0001-5475-3524>

Fiona A. Stewart  <https://orcid.org/0000-0002-4929-4711>

Serge A. Wich  <https://orcid.org/0000-0003-3954-5174>

Alex K. Piel  <https://orcid.org/0000-0002-4674-537X>

## REFERENCES

- Amatulli, G., Domisch, S., Tuanmu, M.N., Parmentier, B., Ranipeta, A., Malczyk, J. et al. (2018) A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, 5(1), 180040. Available from: <https://doi.org/10.1038/sdata.2018.40>
- Ampoorter, E., Barbaro, L., Jactel, H., Baeten, L., Boberg, J., Carnol, M. et al. (2020) Tree diversity is key for promoting the diversity and abundance of forest-associated taxa in Europe. *Oikos*, 129(2), 133–146. Available from: <https://doi.org/10.1111/oik.06290>
- Banda, T., Schwartz, M.W. & Caro, T. (2006) Woody vegetation structure and composition along a protection gradient in a miombo ecosystem of western Tanzania. *Forest Ecology and Management*, 230(1–3), 179–185. Available from: <https://doi.org/10.1016/j.foreco.2006.04.032>
- Barati, S., Rayegani, B., Saati, M., Sharifi, A. & Nasri, M. (2011) Comparison the accuracies of different spectral indices for estimation of vegetation cover fraction in sparse vegetated areas. *Egyptian Journal of Remote Sensing and Space Science*, 14(1), 49–56. Available from: <https://doi.org/10.1016/j.ejrs.2011.06.001>
- Barreras, A., Alanis de la Rosa, J.A., Mayorga, R., Cuenca, R., Moreno-G, C., Godínez, C. et al. (2023) Spatial predictions of tree density and tree height across Mexico forests using ensemble learning and forest inventory data. *Ecology and Evolution*, 13(5), e10090. Available from: <https://doi.org/10.1002/ece3.10090>
- Bastin, J.F., Berrahmouni, N., Grainger, A., Maniatis, D., Mollicone, D., Moore, R. et al. (2017) The extent of forest in dryland biomes. *Science*, 356(6338), 635–638. Available from: <https://doi.org/10.1126/science.aam6527>
- Bietsch, K., Naik, R., Patterson, K.P., Honzak, C. & Kudrati, M. (2016) *Analyzing the building blocks of resilience: findings from a baseline survey of the Tuungane population, health, and environment project in western Tanzania*. Washington, DC: USAID.
- Buramuge, V.A., Ribeiro, N.S., Olsson, L., Bandeira, R.R. & Lisboa, S.N. (2023) Tree species composition and diversity in fire-affected areas of Miombo woodlands, Central Mozambique. *Firehouse*, 6(1), 26. Available from: <https://doi.org/10.3390/fire6010026>
- Bütikofer, L., Anderson, K., Bebbler, D.P., Bennie, J.J., Early, R.I. & Maclean, I.M.D. (2020) The problem of scale in predicting biological responses to climate. *Global Change Biology*, 26(12), 6657–6666. Available from: <https://doi.org/10.1111/gcb.15358>
- Cabacinha, C.D. & de Castro, S.S. (2009) Relationships between floristic diversity and vegetation indices, forest structure and landscape metrics of fragments in Brazilian Cerrado. *Forest Ecology and Management*, 257(10), 2157–2165. Available from: <https://doi.org/10.1016/j.foreco.2009.02.030>
- Carvalho, J.S., Stewart, F.A., Marques, T.A., Bonnini, N., Pintea, L., Chitayat, A. et al. (2022) Spatio-temporal changes in chimpanzee density and abundance in the greater Mahale ecosystem, Tanzania. *Ecological Applications*, 32(8), e2715. Available from: <https://doi.org/10.1002/eap.2715>
- Cavada, N., Ciolli, M., Rocchini, D., Barelli, C., Marshall, A.R. & Rovero, F. (2017) Integrating field and satellite data for spatially explicit inference on the density of threatened arboreal primates. *Ecological Applications*, 27(1), 235–243. Available from: <https://doi.org/10.1002/eap.1438>
- Chidumayo, E.N. (2013) Forest degradation and recovery in a miombo woodland landscape in Zambia: 22 years of observations on permanent sample plots. *Forest Ecology and Management*, 291, 154–161. Available from: <https://doi.org/10.1016/j.foreco.2012.11.031>
- Chitayat, A.B., Wich, S.A., Lewis, M., Stewart, F.A. & Piel, A.K. (2021) Ecological correlates of chimpanzee (pan troglodytes schweinfurthii) density in Mahale Mountains National Park, Tanzania. *PLoS One*, 16, e0246628. Available from: <https://doi.org/10.1371/journal.pone.0246628>
- David, R.M., Rosser, N.J. & Donoghue, D.N.M. (2022) Improving above ground biomass estimates of southern Africa dryland forests by combining Sentinel-1 SAR and Sentinel-2 multispectral imagery. *Remote Sensing of Environment*, 282, 113232. Available from: <https://doi.org/10.1016/j.rse.2022.113232>
- de Castilho, C.V., Magnusson, W.E., de Araújo, R.N.O., Luizão, R.C.C., Luizão, F.J., Lima, A.P. et al. (2006) Variation in aboveground tree live biomass in a central Amazonian Forest: effects of soil and topography. *Forest Ecology and Management*, 234(1–3), 85–96. Available from: <https://doi.org/10.1016/j.foreco.2006.06.024>
- Dickson, R., Baker, M., Bonnini, N., Shoch, D., Rifkin, B., Stewart, F.A. et al. (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(11), 1–14. Available from: <https://doi.org/10.3390/f11111195>
- Divíšek, J. & Chytrý, M. (2018) High-resolution and large-extent mapping of plant species richness using vegetation-plot databases. *Ecological Indicators*, 89, 840–851. Available from: <https://doi.org/10.1016/j.ecolind.2017.11.005>
- Dong, S.X., Davies, S.J., Ashton, P.S., Bunyavejchewin, S., Supardi, M.N.N., Kassim, A.R. et al. (2012) Variability in solar radiation and temperature explains observed patterns and trends in tree growth rates across four tropical forests. *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), 3923–3931. Available from: <https://doi.org/10.1098/RSPB.2012.1124>
- Elisa, M., Caro, T., Yon, L., Hardy, I.C.W., Roberts, S. & Symeonakis, E. (2024) Wildlife corridor degradation and human-wildlife conflict: a case study from Tanzania. *African Journal of Ecology*, 62(2), e13264. Available from: <https://doi.org/10.1111/aje.13264>
- Engelbrecht, B.M.J., Comita, L.S., Condit, R., Kursar, T.A., Tyree, M.T., Turner, B.L. et al. (2007) Drought sensitivity shapes species distribution patterns in tropical forests. *Nature*, 447(7140), 80–82. Available from: <https://doi.org/10.1038/nature05747>
- FAO. (2019) *Trees, forests and land use in drylands: the first global assessment: full report*. Rome: FAO.
- Fisher, J. B., Whittaker, R. J., & Malhi, Y. (2010). ET come home: potential evapotranspiration in geographical ecology. *Global Ecology and Biogeography*, 20(1), 1–18. Portico. <https://doi.org/10.1111/j.1466-8238.2010.00578.x>
- Fricke, G.A., Wolf, J.A., Saatchi, S.S. & Gillespie, T.W. (2015) Predicting spatial variations of tree species richness in tropical forests from high-resolution remote sensing. *Ecological Applications*, 25(7), 1776–1789. Available from: <https://doi.org/10.1890/14-1593.1>
- Frost, P. (1996) The ecology of Miombo woodlands. In: *The Miombo in transition: woodlands and welfare in Africa*. Bogor: Centre for International Forestry Research, pp. 11–57.
- Graham, J.H. & Duda, J.J. (2011) The humpbacked species richness-curve: a contingent rule for community ecology. *International Journal of Ecology*, 2011, 1–15. Available from: <https://doi.org/10.1155/2011/868426>

- Grossman, J.J., Vanhellemont, M., Barsoum, N., Bauhus, J., Bruelheide, H., Castagneryrol, B. et al. (2018) Synthesis and future research directions linking tree diversity to growth, survival, and damage in a global network of tree diversity experiments. *Environmental and Experimental Botany*, 152, 68–89. Available from: <https://doi.org/10.1016/j.envexpbot.2017.12.015>
- Harvey, C.A., Medina, A., Sánchez, D.M., Vilchez, S., Hernández, B., Saenz, J.C. et al. (2006) Patterns of animal diversity in different forms of tree cover in agricultural landscapes. *Ecological Applications*, 16, 1986–1999. Available from: [https://doi.org/10.1890/1051-0761\(2006\)016\[1986:POADID\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2006)016[1986:POADID]2.0.CO;2)
- Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.F., Kaufman, D.M. et al. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84(12), 3105–3117. Available from: <https://doi.org/10.1890/03-8006>
- Homeier, J., Breckle, S.W., Günter, S., Rollenbeck, R.T. & Leuschner, C. (2010) Tree diversity, forest structure and productivity along altitudinal and topographical gradients in a species-rich Ecuadorian montane rain forest. *Biotropica*, 42(2), 140–148. Available from: <https://doi.org/10.1111/j.1744-7429.2009.00547.x>
- Huete, A.R. (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. Available from: [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Itoh, N., Nakamura, M., Ihobe, H., Uehara, S., Zamma, K., Pintea, L. et al. (2012) Long-term changes in the social and natural environments surrounding the chimpanzees of the mahale mountains national park. In: Plumptre, A.J. (Ed.) *The ecological impact of long-term changes in Africa's Rift Valley*. New York, NY: NOVA Science, pp. 249–277.
- Ivory, S.J. & Russell, J. (2016) Climate, herbivory, and fire controls on tropical African forest for the last 60ka. *Quaternary Science Reviews*, 148, 101–114. Available from: <https://doi.org/10.1016/j.quascirev.2016.07.015>
- Jantz, S., Pintea, L., Nackoney, J. & Hansen, M. (2016) Landsat ETM+ and SRTM data provide near real-time monitoring of chimpanzee (*Pan troglodytes*) habitats in Africa. *Remote Sensing*, 8(5), 427. Available from: <https://doi.org/10.3390/rs8050427>
- Jinga, P. & Palagi, J. (2020) Dry and wet miombo woodlands of south-central Africa respond differently to climate change. *Environmental Monitoring and Assessment*, 192(6), 1–16. Available from: <https://doi.org/10.1007/s10661-020-08342-x>
- Kalaba, F.K., Quinn, C.H., Dougill, A.J. & Vinya, R. (2013) Floristic composition, species diversity and carbon storage in charcoal and agriculture fallows and management implications in Miombo woodlands of Zambia. *Forest Ecology and Management*, 304, 99–109. Available from: <https://doi.org/10.1016/j.foreco.2013.04.024>
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W. et al. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 122. Available from: <https://doi.org/10.1038/sdata.2017.122>
- Kerr, J.T. & Ostrovsky, M. (2003) From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, 18, 299–305.
- Kikula, I.S. (1986) The influence of fire on the composition of Miombo woodland of SW Tanzania. *Oikos*, 46(3), 317. Available from: <https://doi.org/10.2307/3565829>
- Knapp, N., Fischer, R., Cazcarra-Bes, V. & Huth, A. (2020) Structure metrics to generalize biomass estimation from lidar across forest types from different continents. *Remote Sensing of Environment*, 237, 111597. Available from: <https://doi.org/10.1016/j.rse.2019.111597>
- Leathwick, J.R., Elith, J. & Hastie, T. (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, 199(2), 188–196. Available from: <https://doi.org/10.1016/j.ecolmodel.2006.05.022>
- Lindsey, P., Allan, J., Brehony, P., Dickman, A., Robson, A., Begg, C. et al. (2020) Conserving Africa's wildlife and wildlands through the COVID-19 crisis and beyond. *Nature Ecology and Evolution*. *Nature Research*, 4, 1300–1310. Available from: <https://doi.org/10.1038/s41559-020-1275-6>
- Macave, O. A., Ribeiro, N. S., Ribeiro, A. I., Chaúque, A., Bandeira, R., Branquinho, C., & Washington-Allen, R. (2022). Modelling Aboveground Biomass of Miombo Woodlands in Niassa Special Reserve, Northern Mozambique. *Forests*, 13(2), 311. <https://doi.org/10.3390/f13020311>
- Mapaure, I. (2013) A preliminary simulation model of individual and synergistic impacts of elephants and fire on the structure of semi-arid Miombo woodlands in northwestern Zimbabwe. *Journal of Ecology and the Natural Environment*, 5(10), 285–302. Available from: <https://doi.org/10.5897/jene12.065>
- Mattivi, P., Franci, F., Lambertini, A. & Bitelli, G. (2019) TWI computation: a comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, 4(1), 1–12.
- Mayes, M., Mustard, J., Melillo, J., Neill, C. & Nyadzi, G. (2017) Going beyond the green: senesced vegetation material predicts basal area and biomass in remote sensing of tree cover conditions in an African tropical dry forest (Miombo woodland) landscape. *Environmental Research Letters*, 12(8), 85004. Available from: <https://doi.org/10.1088/1748-9326/aa7242>
- McNellie, M.J., Oliver, I., Ferrier, S., Newell, G., Manion, G., Griffioen, P. et al. (2021) Extending vegetation site data and ensemble models to predict patterns of foliage cover and species richness for plant functional groups. *Landscape Ecology*, 36(5), 1391–1407. Available from: <https://doi.org/10.1007/s10980-021-01221-x>
- Meng, R., Wu, J., Schwager, K.L., Zhao, F., Dennison, P.E., Cook, B.D. et al. (2017) Using high spatial resolution satellite imagery to map forest burn severity across spatial scales in a pine barrens ecosystem. *Remote Sensing of Environment*, 191, 95–109. Available from: <https://doi.org/10.1016/j.rse.2017.01.016>
- Mittermeier, R.A., Mittermeier, C.G., Brooks, T.M., Pilgrim, J.D., Konstant, W.R., da Fonseca, G.A.B. et al. (2003) Wilderness and biodiversity conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18), 10309–10313.
- Moudrý, V., Lecours, V., Malavasi, M., Misiuk, B., Gábor, L., Gdulová, K. et al. (2019) Potential pitfalls in rescaling digital terrain model-derived attributes for ecological studies. *Ecological Informatics*, 54, 100987. Available from: <https://doi.org/10.1016/j.ecoinf.2019.100987>
- Mutowo, G. & Murwira, A. (2012) Relationship between remotely sensed variables and tree species diversity in savanna woodlands of southern Africa. *International Journal of Remote Sensing*, 33(20), 6378–6402. Available from: <https://doi.org/10.1080/01431161.2012.687472>
- Naftal, B., Kija, H.K., William, C., Noe, C., Anderson, D., Stewart, F.A. et al. (2022) Spatial and temporal pattern of wildfires in the Masito Ugalla ecosystem (2008–2019), Tanzania. *IOSR Journal of Environmental Science*, 16, 12–19. Available from: <https://doi.org/10.9790/2402-1607021219>
- Nakamura, M. (2012) Long-term field studies of chimpanzees at Mahale Mountains national park, Tanzania. In: Watts, D.P. & Kappeler, P.M. (Eds.) *Long-term field studies of primates*. Berlin: Springer, pp. 339–356.
- Ogawa, H., Idani, G., Moore, J., Pintea, L. & Hernandez-Aguilar, A. (2007) Sleeping parties and nest distribution of chimpanzees in the savanna woodland, Ugalla, Tanzania. *International Journal of Primatology*, 28(6), 1397–1412. Available from: <https://doi.org/10.1007/s10764-007-9210-0>
- Ouyang, S., Xiang, W., Wang, X., Zeng, Y., Lei, P., Deng, X. et al. (2016) Significant effects of biodiversity on forest biomass during the succession of subtropical forest in south China. *Forest Ecology and*

- Management, 372, 291–302. Available from: <https://doi.org/10.1016/J.FORECO.2016.04.020>
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C. et al. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33(10), 1704–1711. Available from: <https://doi.org/10.1111/j.1365-2699.2006.01460.x>
- Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J. & Stenseth, N.C. (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510. Available from: <https://doi.org/10.1016/j.tree.2005.05.011>
- Piel, A.K., Bonnin, N., Ramirez-Amaya, S., Wondra, E. & Stewart, F.A. (2019) Chimpanzees and their mammalian sympatriates in the Issa Valley, Tanzania. *African Journal of Ecology*, 57(1), 31–40. Available from: <https://doi.org/10.1111/aje.12570>
- Pintea, L. (2012) *Modeling potential conflict between agricultural expansion and biodiversity in the greater Mahale ecosystem, Tanzania*. Report to Africa Biodiversity Collaborative Group, The Jane Goodall Institute. [http://www.abcg.org/action/document/show?document\\_id=357](http://www.abcg.org/action/document/show?document_id=357). Accessed 15 December 2020.
- Plumptre, A.J., Davenport, T.R.B., Behangana, M., Kityo, R., Eilu, G., Ssegawa, P. et al. (2007) The biodiversity of the Albertine rift. *Biological Conservation*, 134(2), 178–194. Available from: <https://doi.org/10.1016/j.biocon.2006.08.021>
- Polley, E.C. & Van Der Laan, M.J. (2010) *Super learner in prediction*. Berkeley, CA: University of California, Berkeley Division of Biostatistics.
- Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A. et al. (2021) Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253, 112165. Available from: <https://doi.org/10.1016/j.rse.2020.112165>
- QGIS Development Team. (2020) *QGIS geographic information system*. Open Source Geospatial Foundation Project.
- R Core Team. (2022) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ribeiro, N.S., Silva de Miranda, P.L. & Timberlake, J. (2020) Biogeography and ecology of Miombo woodlands. In: *Miombo woodlands in a changing environment: securing the resilience and sustainability of people and woodlands*. Cham: Springer, pp. 9–53. [https://doi.org/10.1007/978-3-030-50104-4\\_2](https://doi.org/10.1007/978-3-030-50104-4_2)
- Ribeiro, N.S., Syampungani, S., Nangoma, D. & Ribeiro, A.I. (2015) Miombo woodlands research towards the sustainable use of ecosystem Services in Southern Africa. In: *Biodiversity in ecosystems – linking structure and function*. IntechOpen: Rijeka, pp. 144–153. <https://doi.org/10.5772/59288>
- Rodrigues, A.C., Villa, P.M., Ali, A., Ferreira-Júnior, W. & Neri, A.V. (2020) Fine-scale habitat differentiation shapes the composition, structure and aboveground biomass but not species richness of a tropical Atlantic forest. *Journal of Forestry Research*, 31(5), 1599–1611. Available from: <https://doi.org/10.1007/S11676-019-00994-X/FIGURES/5>
- Ryan, C.M. & Williams, M. (2011) How does fire intensity and frequency affect miombo woodland tree populations and biomass? *Ecological Applications*, 21(1), 48–60. Available from: <https://doi.org/10.1890/09-1489.1>
- Ryan, C.M., Pritchard, R., McNicol, I., Owen, M., Fisher, J.A. & Lehmann, C. (2016) Ecosystem services from southern African woodlands and their future under global change. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 371, 312. Available from: <https://doi.org/10.1098/rstb.2015.0312>
- Ryan, M.G. & Yoder, B.J. (1997) Hydraulic limits to tree height and tree growth: what keeps trees from growing beyond a certain height? *Bioscience*, 47(4), 235–242. Available from: <https://doi.org/10.2307/1313077>
- Saito, M., Luysaert, S., Poulter, B., Williams, M., Ciais, P., Bellassen, V. et al. (2014) Fire regimes and variability in aboveground woody biomass in miombo woodland. *Journal of Geophysical Research – Biogeosciences*, 119(5), 1014–1029. Available from: <https://doi.org/10.1002/2013JG002505>
- Schröder, J.M., Ávila Rodríguez, L.P. & Günter, S. (2021) Research trends: tropical dry forests: the neglected research agenda? *Forest Policy and Economics*, 122, 102333. Available from: <https://doi.org/10.1016/J.FORPOL.2020.102333>
- Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. Available from: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shirima, D.D., Totland, Ø., Munishi, P.K.T. & Moe, S.R. (2015) Relationships between tree species richness, evenness and aboveground carbon storage in montane forests and miombo woodlands of Tanzania. *Basic and Applied Ecology*, 16(3), 239–249. Available from: <https://doi.org/10.1016/j.baae.2014.11.008>
- Siyum, Z.G. (2020) Tropical dry forest dynamics in the context of climate change: syntheses of drivers, gaps, and management perspectives. *Ecological Processes*, 9(1), 1–16. Available from: <https://doi.org/10.1186/s13717-020-00229-6>
- Slik, J.W.F., Aiba, S.I., Brearley, F.Q., Cannon, C.H., Forshed, O., Kitayama, K. et al. (2010) Environmental correlates of tree biomass, basal area, wood specific gravity and stem density gradients in Borneo's tropical forests. *Global Ecology and Biogeography*, 19(1), 50–60. Available from: <https://doi.org/10.1111/j.1466-8238.2009.00489.x>
- Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P. & Macomber, S.A. (2001) Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote Sensing of Environment*, 75(2), 230–244. Available from: [https://doi.org/10.1016/S0034-4257\(00\)00169-3](https://doi.org/10.1016/S0034-4257(00)00169-3)
- Sunderland, T., Apgaua, D., Baldauf, C., Blackie, R., Colfer, C., Cunningham, A.B. et al. (2015) Global dry forests: a prologue. *International Forestry Review*, 17(2), 1–9. Available from: <https://doi.org/10.1505/146554815815834813>
- Sweke, E.A., Assam, J.M., Matsuishi, T. & Chande, A.I. (2013) Fish diversity and abundance of Lake Tanganyika: comparison between protected area (Mahale Mountains National Park) and unprotected areas. *International Journal of Biodiversity*, 2013, 1–10. Available from: <https://doi.org/10.1155/2013/269141>
- Syampungani, S., Chirwa, P.W., Akinnifesi, F.K., Sileshi, G. & Ajayi, O.C. (2009) The Miombo woodlands at the cross roads: potential threats, sustainable livelihoods, policy gaps and challenges. *Natural Resources Forum*, 33(2), 150–159. Available from: <https://doi.org/10.1111/j.1477-8947.2009.01218.x>
- Tarimo, B., Dick, Ø.B., Gobakken, T. & Totland, Ø. (2015) Spatial distribution of temporal dynamics in anthropogenic fires in miombo savanna woodlands of Tanzania. *Carbon Balance and Management*, 10(1), 1–15. Available from: <https://doi.org/10.1186/s13021-015-0029-2>
- Thomsen, S., Loos, J., Stewart, F.A. & Piel, A.K. (2023) Wildlife habitat association over a twelve-year period (2008–2020) in the greater Mahale ecosystem, western Tanzania. *Journal for Nature Conservation*, 75, 126464. Available from: <https://doi.org/10.1016/J.JNC.2023.126464>
- Thuiller, W., F. Midgley, G., Rougeti, M. & M. Cowling, R. (2006) Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, 29(5), 733–744. Available from: <https://doi.org/10.1111/j.0906-7590.2006.04674.x>
- Timberlake, J. & Chidumayo, E. (2011) Miombo ecoregion vision report. *Occasional Publications in Biodiversity*, 3, 25. Available from: <https://doi.org/10.1109/mcs.1983.1104758>
- Timberlake, J., Chidumayo, E. & Sawadogo, L. (2010) Distribution and characteristics of african dry forests and woodlands. In: Gumbo, D.J. & Chidumayo, E.N. (Eds.) *The dry forests and woodlands of Africa: managing for products and services*. Routledge: London, pp. 11–41. <https://doi.org/10.4324/9781849776547-10>

- Torres, A.B. & Lovett, J.C. (2013) Using basal area to estimate aboveground carbon stocks in forests: La primavera Biosphere's reserve, Mexico. *Forestry: An International Journal of Forest Research*, 86(2), 267–281. Available from: <https://doi.org/10.1093/FORESTRY/CPS084>
- Van Der Laan, M.J., Polley, E.C. & Hubbard, A.E. (2007) Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 1218.
- Verhegghen, A., Kuzelova, K., Syrris, V., Eva, H. & Achard, F. (2022) Mapping canopy cover in African dry forests from the combined use of Sentinel-1 and Sentinel-2 data: application to Tanzania for the year 2018. *Remote Sensing*, 14(6), 1522. Available from: <https://doi.org/10.3390/rs14061522>
- Warren, R., Price, J., VanDerWal, J., Cornelius, S. & Sohl, H. (2018) The implications of the United Nations Paris agreement on climate change for globally significant biodiversity areas. *Climatic Change*, 147(3–4), 395–409. Available from: <https://doi.org/10.1007/s10584-018-2158-6>
- White, F. (1983) *The vegetation of Africa: a descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa, natural resources research*. Paris: United Nations.
- White, J.D. & Swint, P. (2014) Fire effects in the northern Chihuahuan Desert derived from Landsat-5 thematic mapper spectral indices. *Journal of Applied Remote Sensing*, 8(1), 83667. Available from: <https://doi.org/10.1117/1.jrs.8.083667>
- Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, 21, 213–251. Available from: <https://doi.org/10.2307/1218190>
- Williams, M., Ryan, C.M., Rees, R.M., Sambane, E., Fernando, J. & Grace, J. (2008) Carbon sequestration and biodiversity of regrowing miombo woodlands in Mozambique. *Forest Ecology and Management*, 254(2), 145–155. Available from: <https://doi.org/10.1016/j.foreco.2007.07.033>
- Wilson, M.L. (2012) Long-term studies of the chimpanzees of Gombe National Park, Tanzania. In: Watts, D.P. & Kappeler, P.M. (Eds.) *Long-term field studies of primates*. Berlin: Springer, pp. 357–384.
- Wolf, J.A., Fricker, G.A., Meyer, V., Hubbell, S.P., Gillespie, T.W. & Saatchi, S.S. (2012) Plant species richness is associated with canopy height and topography in a Neotropical Forest. *Remote Sensing*, 4(12), 4010–4021. Available from: <https://doi.org/10.3390/RS4124010>
- Yates, L.A., Aandahl, Z., Richards, S.A. & Brook, B.W. (2023) Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*, 93, 1557. Available from: <https://doi.org/10.1002/ECM.1557>
- Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., ... Arino, O. (2021). ESA WorldCover 10 m 2020 v100 (Version v100) [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.5571935>
- Zellweger, F., Baltensweiler, A., Ginzler, C., Roth, T., Braunisch, V., Bugmann, H. et al. (2016) Environmental predictors of species richness in forest landscapes: abiotic factors versus vegetation structure. *Journal of Biogeography*, 43(6), 1080–1090. Available from: <https://doi.org/10.1111/jbi.12696>
- Zellweger, F., Braunisch, V., Morsdorf, F., Baltensweiler, A., Abegg, M., Roth, T. et al. (2015) Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests. *Forest Ecology and Management*, 349, 36–44. Available from: <https://doi.org/10.1016/j.foreco.2015.04.008>
- Zomer, R.J., Trabucco, A., van Straaten, O. & Bossio, D.A. (2006) Carbon, land and water: a global analysis of the hydrologic dimensions of climate change mitigation through afforestation/reforestation. IWMI Report.
- Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. Available from: <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Appendix S1.** Maps showing the calibration range of predictor variables used in ensemble models of tree cover characteristics.

**Appendix S2.** Grid search space for hyperparameter tuning of ensemble models of tree cover characteristics.

**How to cite this article:** Chitayat, A.B., Lewis, M., Anyelwisy, M., Laizer, G.S., Stewart, F.A., Wich, S.A. et al. (2024) Development of spatial models and maps for tree species diversity and biomass in a miombo ecosystem, western Tanzania. *Applied Vegetation Science*, 27, e70002. Available from: <https://doi.org/10.1111/avsc.70002>