

ORIGINAL RESEARCH ARTICLE

Modeling and mapping the distribution and priority seed zone for the conservation of the vulnerable *Vitellaria Paradoxa* (Shea nut tree) in the Guinea savanna ecosystem

Gabriel Salako^{1,2†*}, Oluwasogo Olalubi^{3†}, Abdulrasheed Adio⁴,
Henry Sawyerr⁵, Moses Adetumbi⁴, Abel Adebayo⁶, and Andrey Zaitsev²

¹Department of Environmental Management and Toxicology, Faculty of Pure and Applied Sciences,
Kwara State University, Malete, Kwara, Nigeria

²Department of Soil Zoology, Senckenberg Museum of Natural History Görlitz, Senckenberg Society for
Nature Research, Görlitz, Saxony, Germany

³Department of Public Health, Faculty of Pure and Applied Sciences, Kwara State University, Malete, Kwara, Nigeria

⁴Department of Plant and Environmental Biology, Faculty of Pure and Applied Sciences,
Kwara State University, Malete, Kwara, Nigeria

⁵Department of Environmental Health Science, Faculty of Pure and Applied Sciences,
Kwara State University, Malete, Kwara, Nigeria

⁶Department of Geography, Faculty of Environmental Sciences, Modibbo Adama University, Yola, Adamawa, Nigeria

*Corresponding author: Gabriel Salako (gabriel.salako@senckenberg.de)

[†]These authors contributed equally to this work.

Received: May 19, 2025; Revised: June 15, 2025; Accepted: June 19, 2025; Published online: July 10, 2025

Abstract: *Vitellaria paradoxa* C.F. Gaertn is a native and valuable economic tree species found in the Guinea savanna ecosystem of West Africa. The majority of rural populations, especially women, depend on it for food, domestic energy (fuelwood), and as a source of employment and income. Unfortunately, the extensive felling of this tree species for charcoal production over the past decades poses a grave threat to both the environment and the livelihoods of people; therefore, efforts to restore and conserve the tree species are urgently required. The salient question here is where to obtain viable seeds for its propagation and restoration. This study applied remote sensing technology to extract vegetation-related phenotypic data from satellite images (Landsat 8 – 9), combined with climate data, using a machine learning-based species distribution model. This approach aimed to identify environmentally suitable habitats for *V. paradoxa* and locate areas likely to contain healthy and viable seed sources. These areas were identified through the spatial combination of thresholded habitat suitability maps and vegetation indices – an approach herein referred to as the seed zone priority location index (SZPI). The SZPI is an area that is not only climatically suitable for *V. paradoxa* distribution and survival but also where healthy and viable tree populations can be found. The SZPI is expected to provide vital information necessary to guide the location and collection of suitable and viable seeds required for the restoration and conservation of *V. paradoxa*.

Keywords: Conservation; Tree species; Seed zones; Vegetation index

1. Introduction

Vitellaria paradoxa C.F. Gaertn (shea nut tree) is primarily found in the Guinea savanna ecosystem of West Africa and the Nigeria states of Kwara, Niger, and Oyo North.¹ It holds immense socioeconomic, industrial, health, and ecological values.² Thousands of rural citizens, especially women, depend on this tree species for food, condiments, income, and household needs, such as cream, soap, and fuelwood.^{3,4} However, the excessive cutting and burning of these trees for domestic fuelwood and charcoal production over the past 30 years poses a dangerous trend not only to the survival of the species but also to the economic and ecological services it provides. Over 5 million metric tons of charcoal have been produced from this tree species in West Africa, primarily in Ghana and Nigeria.^{5,6} Notably, *V. paradoxa* is currently classified as a vulnerable species on the International Union for Conservation of Nature's Red List.⁷

Although this tree species primarily grows in the wild, its propagation and cultivation remain major challenges in the tropics. Seed germination and maturation take a long time, and the absence of reliable and viable seeds for vegetative propagation further complicates the process.⁴

Given the socioeconomic benefits of this tree species, especially to rural females, as previously mentioned, and its ecological importance in carbon sequestration (with an average of 36 *V. paradoxa* trees capable of sequestering approximately 2.49 Mg/ha),⁸ the present rate of exploitation for charcoal production is unsustainable. Therefore, it is imperative to implement conservation measures and to identify and select viable, healthy seeds for propagation and restoration. *V. paradoxa* is a deciduous tree species that sheds its leaves during the dry season,⁹ underscoring the importance of monitoring its vegetation phenology – particularly its growth patterns, leaf development, and greenness – to effectively secure viable seeds for its propagation and restoration.

A preliminary investigation showed that, despite existing research efforts to improve seed propagation through vegetative techniques,^{2,4} no vegetation index assessment has been conducted on *V. paradoxa* to identify reliable seed sources for vegetative propagation. Therefore, measuring the species' leaf spectral reflectance to assess its health status or to quantify the extent of its viability, particularly distinguishing between diseased and healthy tissues, could be a promising tool for tackling this challenge.^{10,11}

To achieve this, a careful selection of relevant vegetation phenotypic data from the tree species is needed to infer chlorophyll content (greenness) and, invariably, plant health, biomass, and yield – key factors in selecting viable seeds from tree populations. The combination of geospatial technology and species distribution modeling (SDM) is one of the feasible approaches for this task, as traditional field surveys are often unfeasible at larger spatial scales due to resource limitations. SDM works by correlating species field observations, primarily species occurrence data, with environmental predictor variables that reflect the ecological requirements of the species to predict habitat suitability.¹²

Remote sensing platforms, such as spaceborne/satellite (optical and radar) and unmanned aerial vehicle mounted with different sensors, have become valuable instruments to rapidly and non-destructively assess plant and vegetation status. A number of useful vegetation phenotype parameters or vegetation indices can be derived from these remote sensing platforms to infer plant productivity (e.g., chlorophyll contents or index), plant health status (e.g., green index), plant biomass and yield (e.g., normalized difference vegetation index [NDVI], plant water stress (normalized difference water index [NDWI])). These indices have been well described in the literature.¹³⁻¹⁷ Vegetation phenological parameters or vegetation indices in remote sensing are usually estimated using arithmetic combinations of spectral reflectance values, mainly from the red and near-infrared regions of the electromagnetic spectrum.

Similarly, a strong understanding of tree population seed zones can provide valuable guidance for tree restoration and conservation activities. A seed zone or seed transfer zone is defined as a geographically delineated area having similar climatic characteristics from which seeds could be transferred with a minor risk of maladaptation.^{18,19} The delineation of seed zones has mostly relied on climatic variables, such as precipitation and temperature.^{18,19} Generalized provisional seed zones can be constructed using geographic information system (GIS)-based models of spatial layers combination, for example, the most influential geospatial climate/environmental variables,^{18,20,21} or through multivariate clustering of environmental variables.^{22,23} In temperate regions, the temperature of the coldest month has been identified as one of the most important climatic variables affecting plant adaptation and distribution. This has informed the use of winter minimum temperatures in constructing seed zone maps (SZM) in the United States²⁰ and the plant hardiness zone map by the United

States Department of Agriculture.²⁴ However, in tropical regions, total annual rainfall and its distribution play a more significant role in determining plant distribution and survival, as temperature remains relatively uniform with little or no seasonal variation.²⁵ This is true for the study area in the present project, where the average annual temperature rarely exceeds 30°C and shows little or no seasonal variation.

This study seeks to answer the following questions: (i) Where is the suitable habitat for *V. paradoxa*? (ii) What are the essential vegetation phenotype parameters that can guide the selection of viable/healthy seeds for *V. paradoxa* propagation? and (iii) Can the locations where viable seeds can be sourced and the associated seed zone in the study area be identified? Accordingly, this study applied remote sensing imaging technologies to extract vegetation-related phenotypic and climatic data for habitat suitability or SDM, with the goal of identifying suitable sites and potential sources of viable and healthy seeds for *V. paradoxa*.

2. Materials and methods

2.1. Study area

The study area covered the Guinea savanna region of Nigeria, including Niger, Kwara, the Federal Capital Territory, and parts of Kebbi and Kogi states (Figure 1). The vegetation consists primarily of deciduous trees and grasses. Rainfall is seasonal, occurring mostly from June to September, with total annual rainfall of 1,200 mm in the southern region and 700 – 900 mm in

the north. The mean annual temperature is around 26°C, with little or no seasonal variation.¹

2.2. Shea nut tree (*V. paradoxa*) occurrence data

The occurrence data of *V. paradoxa* were obtained from field surveys and the Global Biodiversity Information Facility (GBIF). Field data were collected from March to August 2021 and from October to November 2024, while GBIF data were accessed on November 26, 2024 (occurrence data and citations in Appendix A1 and Table S1). Special attention was paid to the identification of *V. paradoxa* in the field with the aid of local guides and experts. The data collected from both sources included species name, geographic coordinates (in decimal degrees), sources of information (for data collected from GBIF), and the name of the closest settlement (only in the field survey). Additional information collected included the date and time of each observation.

For GBIF data, all duplicates and entries with questionable geographic coordinates were removed. The remaining points were spatially thinned at a 1 km resolution using the “spThin” package in R²⁶ to minimize spatial or geographical bias commonly associated with large online databases. This resulted in a total of 100 unique occurrence datasets. Given that *V. paradoxa* data collected in this study represent presence-only records, pseudo-absence data (locations where the species is not known to be present) were required for most machine learning algorithms in spatial modeling. To generate this, between 500 and 2,000 random background points were sampled, and 1,000 points were selected

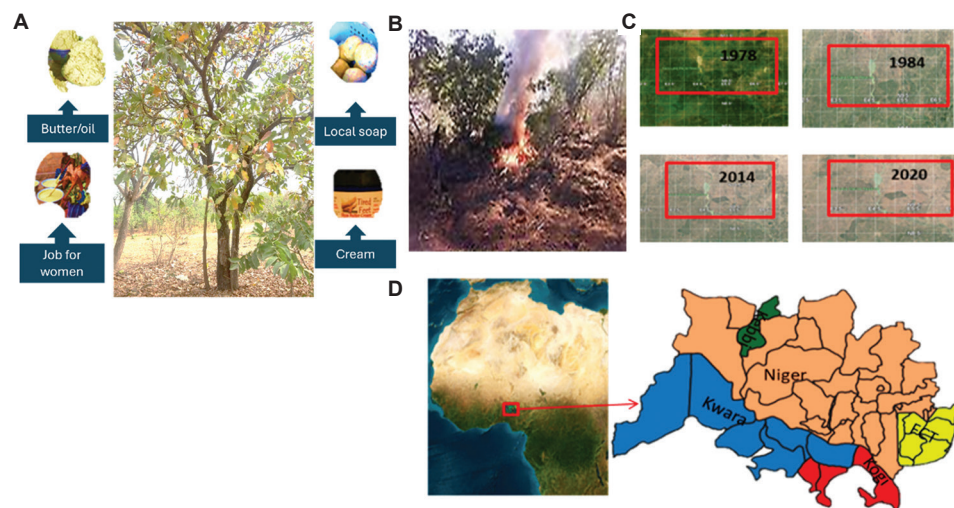


Figure 1. Illustration of the use of *Vitellaria paradoxa* and the study locations. (A) Benefits of the species. (B) Charcoal burning sites. (C) Land cover changes. (D) Study area showing local and state administration boundaries.

as optimal based on a cross-validation score >0.7 . These random background points were used as pseudo-absence data to characterize the modeling environment in the study area.²⁷ The combined data (presence and background data/pseudo-absence records) were then used to calibrate the model.

2.3. Remote sensing data collection and processing

Satellite images of Landsat 8 – 9 surface reflectance at Collection 2, Level 1 (radiometrically corrected) for the year 2024 were acquired from the United States Geological Survey EarthExplorer platform (<https://earthexplorer.usgs.gov/>). Landsat 8 and Landsat 9 were launched on February 11, 2013, and September 27, 2021, respectively, with both providing moderate-resolution imagery across 11 spectral bands, ranging from 15 to 100 m of global land surface.

Landsat 8 and Landsat 9 are equipped with two sensors: The Operational Land Imager and Thermal Infrared Sensor, both of which offer improved signal-to-noise ratio radiometric performance, enhancing land-cover characterization capabilities (Table 1). All data were provided in GeoTIFF format. Looking ahead, the launch of Landsat Next is planned for late 2030 or early 2031. This next-generation mission will feature a 26-band “super spectral” sensor with improved spatial and temporal resolution (<https://landsat.gsfc.nasa.gov/satellites/landsat-next/>). In addition, high-resolution GeoEye-1 imagery at 70 cm spatial resolution was acquired for areas within Kainji Lake National Park, located in the study area. This imagery was used to

assess the geographic coordinates of questionable occurrence data obtained from GBIF.

A total of six image scenes were downloaded from the USGS EarthExplorer, covering the spatial extent of the study area along Landsat paths 189 – 191 and rows 053 – 054. From each scene, four spectral bands were selected: the green band (Band 3), red band (Band 4), near-infrared band (Band 5), and short-wave infrared 2 (Band 7). These images, collected between July and September 2024, were 90% cloud-free and coincided with the planting and growing season or the peak of plant greenness in the study area. These selected bands have been used in several studies to extract data and information on vegetation health.¹⁷ All image scenes and their respective spectral bands (a total of 24) were mosaicked for further analysis (Figure 2).

2.4. Estimation of vegetation phenology parameters/indices

To estimate relevant vegetation phenology/indices, several mathematical operations were performed on the selected spectral bands. For example, the NDVI is the most frequently used and simplest vegetation index to quantitatively and qualitatively evaluate vegetation cover.^{15,28} It measures the difference between near-infrared and red reflectance, producing a numerical value for each pixel that ranges from -1 to $+1$. A value of 0 indicates no vegetation, negative values signify water bodies, and values approaching $+1$ indicate dense, green vegetation. A summary of the mathematical operations

Table 1. Basic band spectral information on the Landsat 8 – 9 mission

S. No.	Spectral band	Description	Spectral range (μm)	Resolution (m)
OLI spectral bands				
1	Band 1	Coastal/aerosol	0.43 – 0.45	30
2	Band 2	Blue	0.45 – 0.53	30
3	Band 3	Green	0.53 – 0.59	30
4	Band 4	Red	0.64 – 0.67	30
5	Band 5	Near-infrared	0.85 – 0.88	30
6	Band 6	Short-wave infrared 1	1.57 – 1.65	30
7	Band 7	Short-wave infrared 2	2.11 – 2.29	30
8	Band 8	Panchromatic	0.5 – 0.68	15
9	Band 9	Cirrus	1.36 – 1.38	30
TIRS spectral band				
10	Band 10	Long-wave infrared 1	10.60 – 11.19	100
11	Band 11	Long-wave infrared 2	11.5 – 12.51	100

Abbreviations: OLI: Operational land imager; TIRS: Thermal infrared sensor.

on all four spectral bands and the application of these indices are presented in Table 2.

All computations and visualizations were performed using the raster and terra packages in R programming.

To address potential multicollinearity among variables, the variance inflation factor step function was applied.

The ratio vegetation index showed a markedly high correlation (>9.0) with all vegetation indices. While the

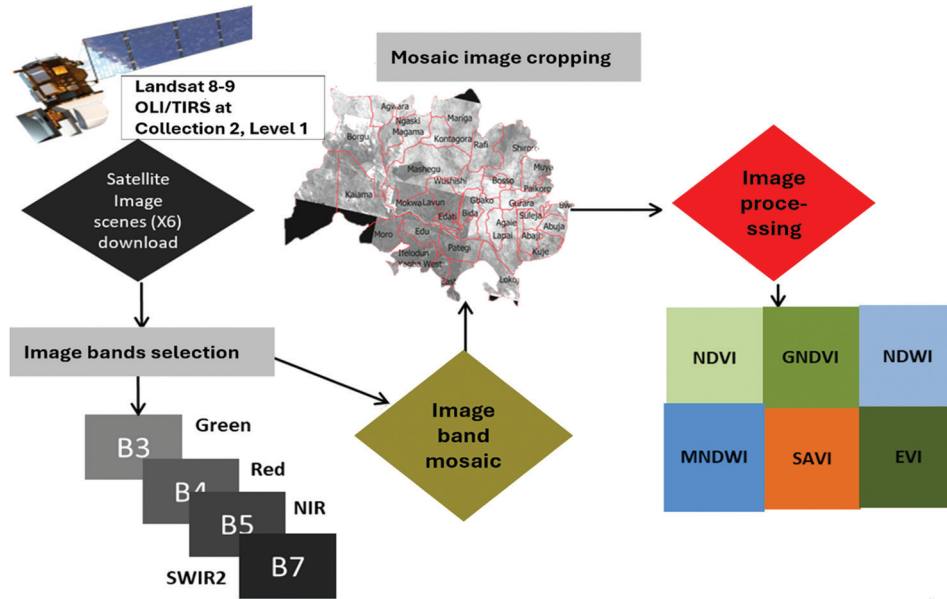


Figure 2. Image processing workflow into vegetation indices (dark area = no data). A total of 24 images were obtained from six scenes of four spectral bands each (6 × 4).

Abbreviations: EVI: Enhanced vegetation index; GNDVI: Green normalized difference vegetation index; MNDWI: Modified normalized difference water index; NDVI: Normalized difference vegetation index; NDWI: Normalized difference water index; OLI: Operational Land Imager; SAVI: Soil-adjusted vegetation index; TIRS: Thermal Infrared Sensor.

Table 2. Vegetation indices’ description and arithmetic computation

Index	Description	Source	Bands	Formula	Application/uses
NDVI	Normalized difference vegetation index	29	Red, NIR	$(NIR - red) / (NIR + red)$	Plant growth, yield
RVI	Ratio vegetation index		Red, NIR	Red / NIR	Biomass
GNDVI	Green normalized difference vegetation index	13	Green, NIR	$(NIR - green) / (NIR + green)$	Chlorophyll, plant health, nitrogen
EVI	Two-band enhanced vegetation index	17	Red, NIR	$G \times (NIR - red) / (NIR + 2.4 \times red + 1)$	Biomass, yield
SAVI	Soil-adjusted vegetation index	16	Red, NIR	$(NIR - red) \times (1 + L) / (NIR + red + L); L = 0.50$	Monitoring seed and leaf development
NDWI	Normalized difference water index	30	Green, NIR	$(Green - NIR) / (Green + NIR)$	Water bodies detection
MNDWI	Modified normalized difference water index	31	Green, SWIR2	$(Green - SWIR2) / (Green + SWIR2)$	Improved water stress, drought detection
SLAVI	Specific leaf area vegetation index	32	Red, NIR, SWIR2	$NIR / (Red + SWIR2)$	Biomass, canopy cover

Abbreviations: NIR: Near infrared; SWIR2: Short-wave infrared 2.

specific leaf area vegetation index was excluded, the green NDVI (GNDVI), NDVI, and enhanced vegetation index (EVI) were retained to capture subtle yet important information peculiar to each vegetation index (Figure 3).

Climate data (Table 3) at a 1 km spatial resolution were obtained from the climatologies at high resolution for the Earth's land surface areas (CHELSA) dataset.³³ CHELSA provides interpolated average climate data for the period 1981 – 2010. Four climate data points considered ecologically and physiologically relevant to the growth and distribution of *V. paradoxa* were preselected: mean annual temperature (Bio 1), total

annual precipitation (Bio 12), precipitation amount of the wettest month (Bio 13), and precipitation amount of the driest month (Bio 14). Bio 1 was subsequently excluded from the modeling, as it showed minimal variation across the study area. In addition, both the satellite imagery and climate data in GeoTIFF format were resampled to 30 m spatial resolution.

2.5. Species distribution model calibration and assessment

The SDMs were fitted using three machine learning algorithms (random forest [RF], boosted regression

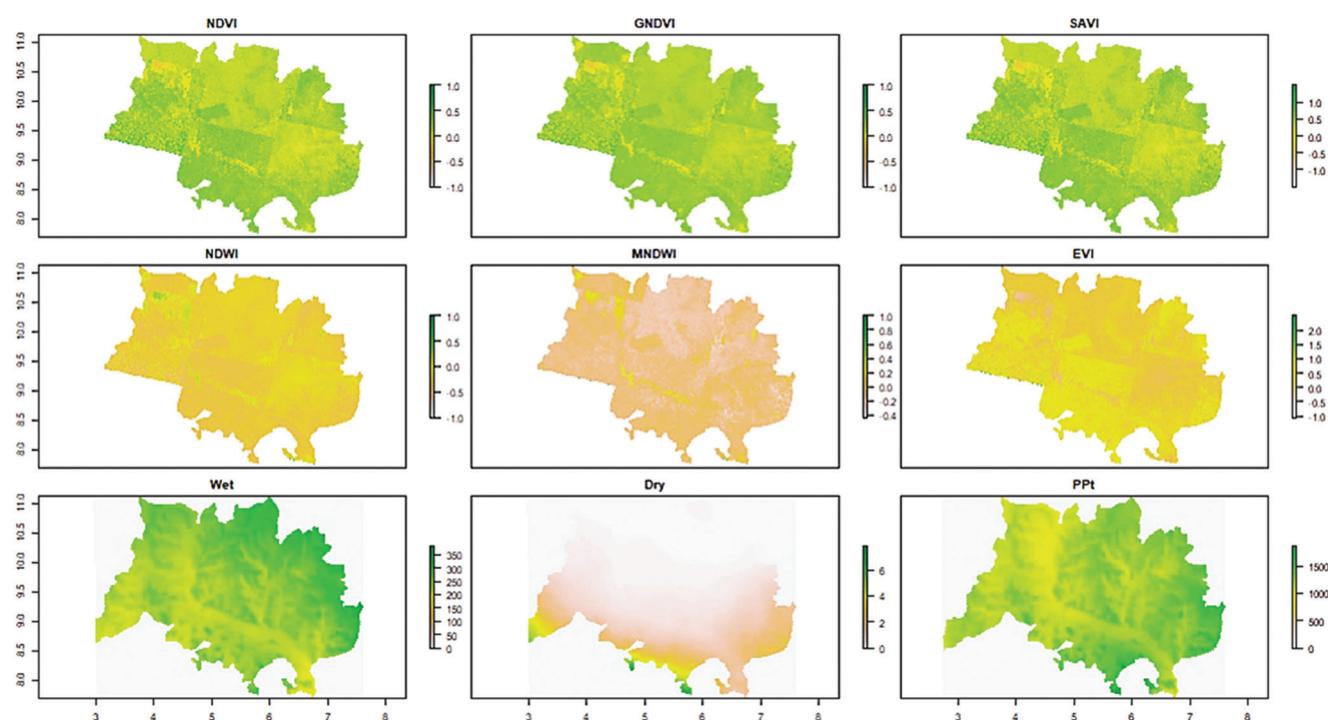


Figure 3. Remote sensing-derived vegetation indices (NDVI, GNDVI, SAVI, NDWI, MNDWI, and EVI) and climate variables (total annual precipitation [PPt], precipitation of the wettest month [Wet], and precipitation of the driest month [Dry])

Abbreviations: EVI: Enhanced vegetation index; GNDVI: Green normalized difference vegetation index; MNDWI: Modified normalized difference water index; NDVI: Normalized difference vegetation index; NDWI: Normalized difference water index; OLI: Operational Land Imager; SAVI: Soil-adjusted vegetation index; TIRS: Thermal Infrared Sensor.

Table 3. Climate variables downloaded from the climatologies at high resolution for the Earth's land surface areas dataset

S. No.	Variable	Description	Abbreviation	Unit
1	Bio 1	Mean annual air temperature	Mat	°C
2	Bio 12	Annual precipitation amount	PPt	mm/year
3	Bio 13	Precipitation of the wettest month	Wet	mm/month
4	Bio 14	Precipitation of the driest month	Dry	mm/month

trees [BRT], and artificial neural network [ANN]). An RF is an ensemble method that incorporates multiple classification (decision) trees and produces predictions either by averaging (e.g., regression) or voting (e.g., classification).^{12,34} It demonstrates high predictive performance among machine learning algorithms.^{35,36}

A BRT, also known as generalized boosted regression modeling, uses a boosting algorithm that iteratively applies regression-tree algorithms in a forward, stage-wise procedure to construct a combination of trees. The model focuses on the weakest parts of the present model by constructing a new tree based on the residual of the previously fitted tree.³⁷

An ANN is a complex interconnected system of several nodes, mimicking the structure and functioning of neurons in the human brain. It consists of three major components: an input/nodes (comprising the predictors/environmental variables), a hidden layer (receiving information with several connections to the input through interconnected nodes), and an output layer (the prediction). ANNs can produce outputs similar to those of BRT and RF, in the form of regression/probability predictions or classification for binary inputs, though the latter is the most widely used form.³⁸

The models were fitted using the following packages in R: “randomForest” for RF,³⁹ “DISMO” for BRT,⁴⁰ and “neuralnet” for ANN.⁴¹ The binary occurrence data for *V. paradoxa* (presence and background) served as the response variable, while selected climate and vegetation indices were used as predictors or explanatory variables. For the RF model optimization, a tuning method was adopted⁴² based on the number of trees (ntree) and node size. Meanwhile, the BRT parameters were tuned by changing the number of trees/tree complexity (tc) in the range of 1,000 – 5,000 for each model until the optimum configuration was achieved, while maintaining the learning rate (lr) at 0.001. Diagnostic plots (showing error trees and residuals) and cross-validation correlations were used to visually assess and select the optimal model. For the ANN model, input data were transformed into a common scale (normalization) to ensure accurate comparison between predicted and actual values. The ANN architecture was configured with hidden layers set at 4 and 2, activation set at rectified linear unit, and linear output as false (regression) or true (classification). SDM probability predictions were generated across the study area. Models with an area under the curve (AUC) >0.7 were included in a weighted ensemble or consensus to obtain the final SDM prediction.

To evaluate the predictive performances of the three models’ algorithms selected in this study, a simple holdout, also known as split-sample cross validation method, was used by splitting the data into an 80:20 ratio, where 80% data was used to train the model while the remaining 20% was set aside to assess the model predictive performance.^{12,40} The models were also replicated 10 times to obtain the best model fit based on their cross-validation correlation. Both the threshold-independent statistic, “the area under the receiver operating characteristic curve (AUC_{ROC}),”⁴³ and the threshold-dependent, the confusion matrix statistics, were employed to assess the predictive accuracy, precision, and specificity of the models. Another model evaluation method used was the Kappa coefficient statistic,⁴⁴ which measures the level of agreement between two raters based on binary classification.

The AUC_{ROC} class ranges from 0 to 1. An AUC value of 0.5 or lower indicates performance no better than random prediction, values between 0.7 to 0.8 are considered tolerable, 0.8 – 0.9 are regarded as very good or excellent, and values above 0.9 are classified as outstanding.^{42,45} Meanwhile, the Kappa scores range from –1 to 1, where 0 indicates no agreement and random, 1 represents perfect agreement, and the rare occasions of negative values signify less agreement than would be expected by chance. The Kappa index is determined by Equation I:

$$-K = (p_0 - p_e)/(1 - p_e) \quad (I)$$

Where K is the Kappa index, p_0 is the observed agreement, and p_e is the expected agreement by chance.

The relative contribution of each predictor/environmental variable to the model predictions was also assessed, and their ecological plausibility was evaluated through extracting response curves, also known as partial dependence plots.

2.6. Climate-based provisional seed zone and seed zone priority location index (SZPI)

Based on the *V. paradoxa*’s predicted suitability and the relative contributions of the top climate variables, SZMs were constructed by spatially intersecting the values in their respective cells in a GIS environment (QGIS 3.40.6 and ArcMap 10.8). A quantile classification (ordering) method was then applied to group the data into five broad, homogeneous classes, arranged according to their relative wetness or dryness.

The SZPI was defined as an area that is not only climatically suitable for the distribution and survival of

V. paradoxa but also supports healthy and viable tree populations, as identified through could be found based on the thresholded vegetation indices. To determine the SZPI, the following steps were taken:

- (i) The predicted SDM probability raster was reclassified into two classes⁴⁶ by setting a suitability threshold at $\geq 60\%$ probability: suitable (1) and not suitable (0; less than the threshold).^{19,22}
- (ii) Vegetation indices were reclassified based on optimal point (greenness peak) thresholds, derived from a response-curve mode. This is the peak point along the vegetation index gradient (e.g., NDVI, GNDVI, modified NDWI [MNDWI]) and, therefore, is considered indicative of viable seed locations (1). Values below this threshold were classified as poor/non-viable seed locations (0).
- (iii) The SZPI maps were generated by spatially combining the thresholded *V. paradoxa* suitability map with the thresholded vegetation indices—GNDVI, MNDWI, and NDVI. The resulting map contained two classes: high-priority seed location (2) and poor/non-priority seed location (1).
- (iv) The SZPI maps were intersected with climate-based SZMs to identify areas of overlap or intersection with seed zones and associated land uses, such as forest, parks, and reserves. It should be noted that the vegetation index thresholds used were assumed to be species-specific (*V. paradoxa*), as different species may respond differently to NDVI and other vegetation indices, potentially leading to varied results.

3. Results

3.1. Models' predictive performances

Figure 4 presents the predictive performances of the three models' algorithms across the evaluation metrics (AUC, accuracy, and Kappa index). Both RF and BRT performed excellently in all three metrics, with RF exhibiting a marginally higher performance. In contrast, ANN performed poorly in all the metrics; its AUC score indicates the prediction was not better than random.

3.2. Estimation of the variable importance and responses of *V. paradoxa* to climate and vegetation indices/phenology parameters

Figure 5A presents the partial dependence plots or response curves of *V. paradoxa* distribution along the environmental gradients (climate and vegetation indices), while Figure 5B and Table 4 demonstrate the optimal threshold (used for the reclassification into

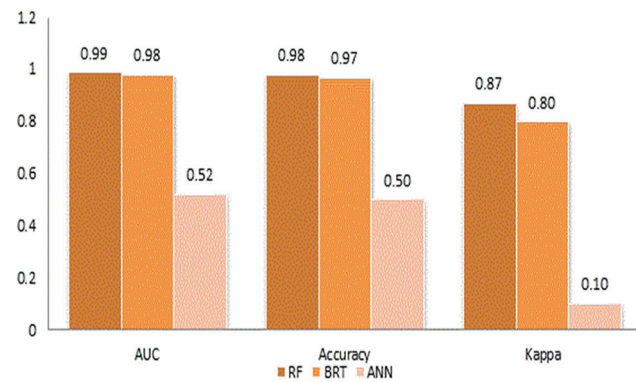


Figure 4. Models' performances of random forest (RF), boosted regression trees (BRT), and artificial neural network (ANN). Class ranges from 0 to 1, with 0 indicating very poor model performance, 0.5 not better than a random prediction, 0.7 – 0.8 acceptable, and >0.9 outstanding.

Abbreviation: AUC: Area under the curve.

viable and non-viable seed location) and threshold ranges, respectively. The predicted optimal threshold for total annual precipitation and the wettest months were at 1,000 mm and 250 mm, respectively, while that of the driest month should not be < 0.4 mm of rainfall. In terms of water stress, it was predicted to respond positively as the value increased from 0.01, reaching a peak response at 0.2. Overall, *V. paradoxa* was predicted to exhibit positive responses to most of the vegetation indices (NDVI, EVI, soil-adjusted vegetation index [SAVI], and NDWI).

Figure 5C presents the relative contributions of each predictor variable to the distribution of *V. paradoxa* as estimated by the BRT and RF models, along with the model consensus shown in Figure 5D. The ANN model was excluded from further analysis due to its poor performance. There was an agreement between the RF and BRT models regarding the importance of most variables. Water-related variables, specifically annual precipitation, precipitation of the wettest month, and MNDWI, emerged as the most important predictors. Among the vegetation indices, both models ranked the GNDVI as the most important vegetation phenology parameter. However, NDVI was ranked higher by the BRT model compared to the RF model, while both SAVI and EVI were consistently ranked as the least important by both algorithms.

3.3. Predicted spatial distribution of *V. paradoxa* and the mapping of seed-zone priority index

Figure 6A-D demonstrates the predicted distribution of *V. paradoxa* as modeled by RF, BRT, and ANN, and the

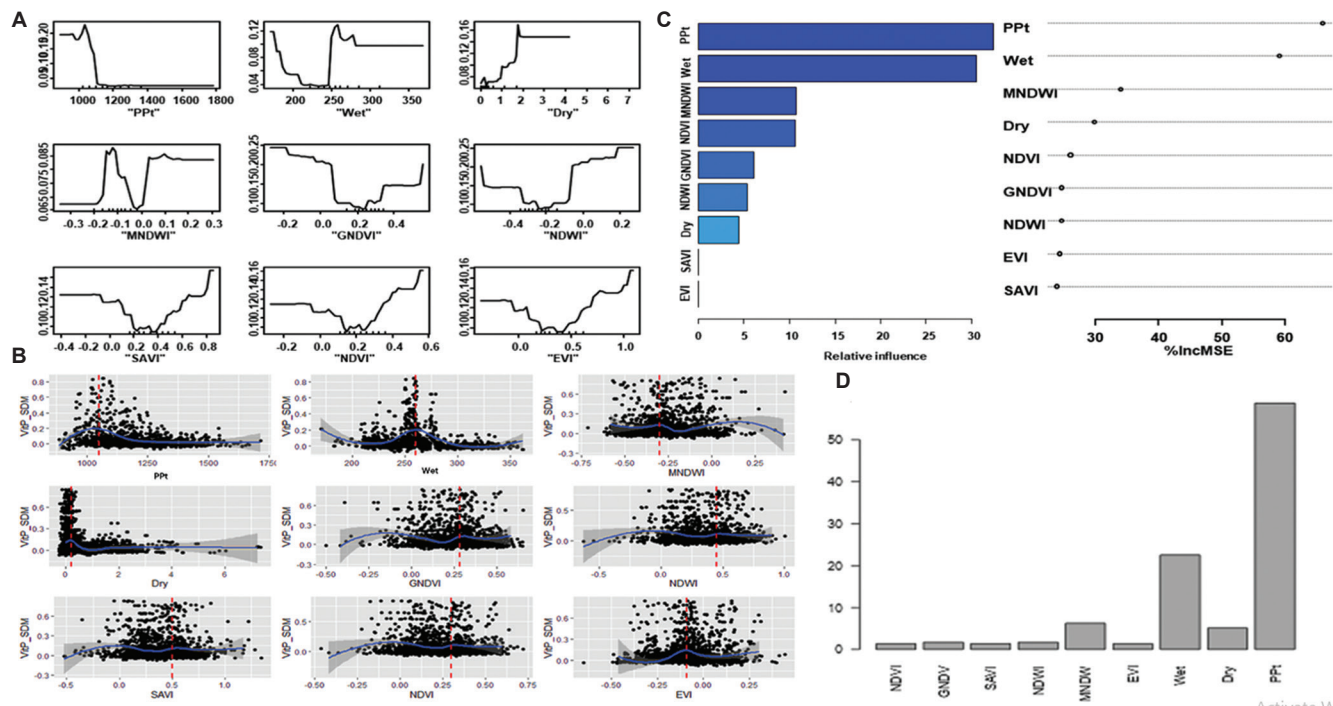


Figure 5. Model response curves, optimal threshold, and variable contributions. (A) Model response curves of the variables along the environmental gradient. (B) The optimal/modal thresholds (red vertical dotted line). The thresholds are used to reclassify the selected vegetation indices (GNDVI, MNDWI, and NDVI) into high and low priority seed locations. (C) Variable contributions of BRT and RF. (D) Consensus of BRT and RF. The NDWI and MNDWI usually exhibit positive values. Zero or negative values signify vegetation.

Abbreviations: Dry: Precipitation of the driest month; EVI: Enhanced vegetation index; GNDVI: Green normalized difference vegetation index; MNDWI: Modified normalized difference water index; NDVI: Normalized difference vegetation index; NDWI: Normalized difference water index; PPT: Total annual precipitation; SAVI: Soil-adjusted vegetation index; Wet: Precipitation of the wettest month.

Table 4. Selected essential environmental thresholds for seed source selection

S. No.	Variables	Optimal threshold range ^a	Predicted seed source locations
1	Total annual precipitation (PPT)	1,000 – 1,200 mm	Mashegu, Lavun, Borgu
2	Precipitation of the wettest month (Wet)	250 – 300 mm	Mashegu, Lavun, Borgu
3	Precipitation of the driest month (Dry)	0.20 – 0.30 mm	-
4	Modified normalized difference water index	0.0 – 0.2	Borgu Agwara, Mokwa
5	Normalized difference vegetation index	0.3 – 0.6	Mashegu, Lavun, Agwara, Borgu
6	Green normalized difference vegetation index	0.3 – 0.6	Mashegu, Lavun, Agwara, Borgu
7	Soil-adjusted vegetation index	0.3 – 0.8	-
8	Enhanced vegetation index	0.4 – 0.7	-

Note: ^aThe optimal range was derived from a partial dependence plot or response curve.

mean prediction derived from the RF and BRT models. While both BRT and RF successfully predicted suitable habitats for the tree species, the ANN model failed to generate reliable predictions. High suitability areas for *V. paradoxa* were predicted around the northern area of Kainji Lake National Park, forest regions around Eban,

Bambafu, and Fellagi in Niger state, and at the border of Borgu, Kaiama, and Baruten local government area of Niger and Kwara states. Additional clusters of high suitability (>60%) were predicted in central Mashegu, northern Lavun, and Borgu local government areas.

Figure 6F highlights that areas predicted as suitable (in green) were not necessarily predicted as high-priority seed source locations (in red). This discrepancy results from the integration of both habitat suitability thresholds and vegetation index thresholds, where only locations meeting both criteria were classified as high-priority seed zones. With respect to the relationship

between priority seed source, the associated seed zone, and land use, an overlay of the predicted SZPI map onto the five major seed zones across the study areas (Figure 7A-C) showed that *V. paradoxa* populations were predicted to be climatically suitable and viable in three seed zones. These zones are characterized by total annual rainfall between 1,000 and 1,200 mm and

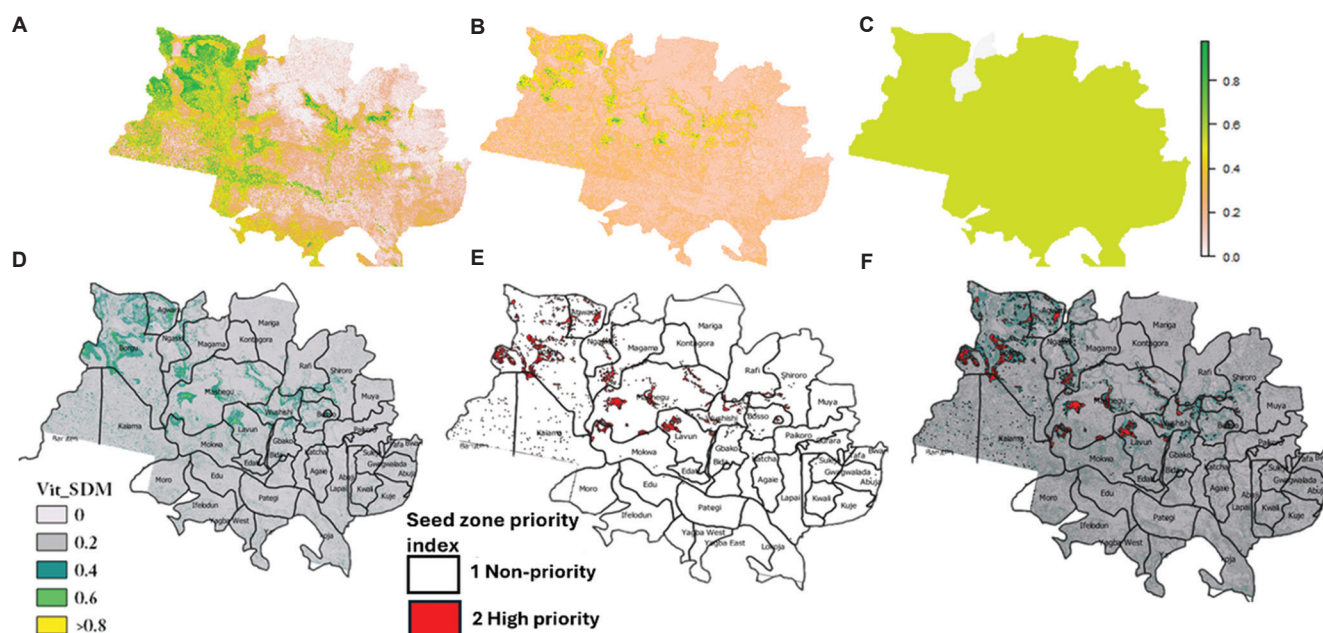


Figure 6. Predictions on *Vitellaria paradoxa* parameters. Species distribution modeling (SDM) prediction using (A) random forest (RF), (B) boosted regression tree (BRT), and (C) artificial neural network. (D) Weighted mean prediction using RF and BRT. (E) Seed zone priority location index (SZPI). (F) Overlay of SDM on SZPI. A red dot indicates high-priority seed source locations, while a green dot indicates suitable but non-priority seed source locations.

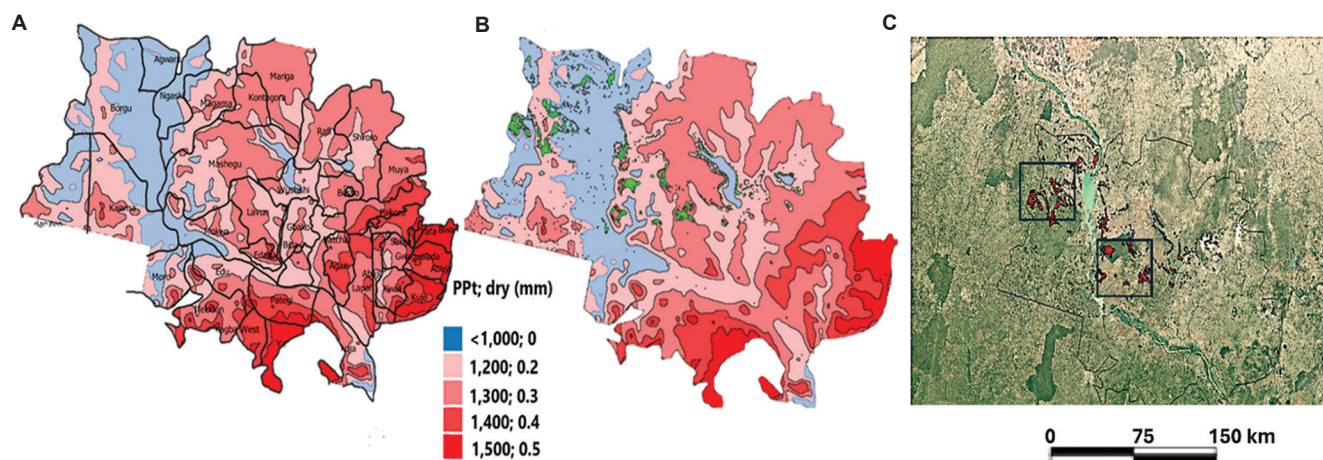


Figure 7. Relationship between provisional seed zone and predicted seed zone priority index. (A) Provisional five seed zones within the local administrative boundary. (B) Overlay of the predicted seed zone priority location index (green polygon). (C) Overlap of the seed zones with forest reserves and parks. Abbreviations: Dry: Precipitation of the driest month; PPT: Total annual precipitation.

the rainfall of the driest month ranging from 0.20 to 0.30 mm.

4. Discussion

Both BRTs and RFs are the two best-performing machine learning algorithms in this study. They have been described as top-performing machine learning algorithms³⁸ and applied in several predictive modeling in ecology and biogeography^{35,36} for conservation purposes. In contrast, the ANN model failed to produce meaningful predictions in this study. It has also been reported to be the least-performing model in comparable SDM studies.³⁸ One potential explanation for this poor performance is the relatively small sample size used ($n = 100$), as ANN performance has been shown to improve with larger sample size⁴⁷ and is more suitable for classification tasks than for regression-based predictions.³⁸ It is therefore recommended that ANN be used primarily for classification purposes with a relatively larger sample size.

A weighted average of the two best-performing algorithms, BRT and RF, was applied as a consensus approach for the final SDM prediction.⁴⁰ This ensemble method helps to reduce uncertainty or bias associated with relying on a single algorithm.

Several plant phenotyping studies in ecology and agriculture have utilized vegetation indices, such as NDVI, GNDVI, NDWI, and SAVI to characterize phenological changes in plants and crops.^{48,49} Understanding these essential vegetation indices provides insights into plant health, nutrient status, and the stress level, which could be a major step toward tree conservation strategies and rational choices guiding the selection of seed sources for propagation. Earlier studies have applied remote sensing technologies to assess health status, such as nitrogen and chlorophyll contents of various plants and crops,⁵⁰⁻⁵² which have proven valuable in assessing plant productivity, yield, and vulnerability to environmental stress.

While most habitat suitability/SDM studies have focused on the use of climate variables as predictors,^{53,54} and occasionally land use or habitat types,^{36,42} this study integrates both climate variables and remote sensing technology. By applying machine learning algorithms, this approach not only predicts suitable habitats for *V. paradoxa* but also identifies priority locations for viable seed sources (SZPI) within seed zones that could be useful for tree propagation and restoration.

Although *V. paradoxa* is mostly found in West Africa, its distribution has often been erroneously mapped to

cover the entire subregion.^{55,56} For example, a previous study¹ has mapped and identified savanna woodland ecosystems as the primary habitat for *V. paradoxa* in Kwara State in north-central Nigeria. In contrast, another study⁵⁷ described its distribution around the savanna national parklands/reserves in Africa, where naturally occurring tree stands are protected by locals or government authorities. Besides the protection of the tree species in a few national game reserves, no extensive programs have been made toward the conservation of the tree species in the study area. These earlier studies aligned with the model predictions from this research, which indicate high habitat suitability and the presence of viable seed sources for *V. paradoxa* in forest clusters and protected areas within the Guinea savanna zone. The high suitability areas, mostly located in forest clusters, parks, and game reserves, could be indicative of tree exploitations and habitat degradation in unprotected areas over the years, leading to a significant decline in *V. paradoxa* populations across much of its natural range.

Climate variables, particularly precipitation and its associated variables, contributed more than 25% to the SDM predictions for *V. paradoxa*. This finding is consistent with previous research⁵⁸ that identified rainfall and soil fertility as key determinants of phenotypic traits, such as leaf size, fruit weight, and nut size in *V. paradoxa*. In regions where rainfall is abundant and the soil is fertile, the trees tend to produce larger nuts with higher oil content. Although soil variables were not included in the present model, vegetation indices representing plant greenness may serve as effective proxies for assessing soil nutrient availability.

The inclusion of vegetation phenology parameters, such as NDWI, GNDVI, and NDVI, not only enhanced model performance but also facilitated the identification of *V. paradoxa* populations with viable/healthy seeds. This has significant implications for assessing tree growth, nutrient status, water stress levels, productivity, and restoration efforts.

The positive responses of *V. paradoxa* to most vegetation indices (e.g., NDWI/MNDWI, EVI, and NDVI) signify the importance of these variables to the growth and health status of the species, both of which directly influence seed quality. An optimal prediction was reached at approximately 1,050 mm for total annual precipitation and around 250 mm during the wettest month. Previous studies have also shown that the leaf and fruit quality of *V. paradoxa*, as well as the trees' circumference, are positively correlated with rainfall.⁹

Water stress is another important environmental variable that not only affects the distribution of

V. paradoxa but also its health.⁵⁹ Negative indices signify water stress, which could be injurious to seed germination. This study's findings highlight NDWI/MNDWI as crucial vegetation indices for detecting water stress in the species. The model predicted a positive response begins at a 0.2 water index gradient. Although this study did not project species distribution and seed priority index under future scenarios, the threshold information on precipitation and MNDWI (Table 4) can guide conservation strategies and serve as baseline values of the species' water requirements under changing climatic conditions.

Despite occurrence data were obtained from both an online database (GBIF) and field surveys, data gaps remain in certain parts of the study areas, especially in the Federal Capital Territory and northern Niger state (e.g., Mariga and Kontagora). Therefore, the model's predictions of low habitat suitability and low seed location priority index in these areas could not be fully validated. Nonetheless, an examination of the spatial distribution of both the GNDVI and MNDWI showed low greenness and potential water stress in these regions, which may explain the model's low suitability predictions. It should be emphasized, however, that the vegetation indices and climate variables used in this study represent conditions at the landscape level and are not exclusive to *V. paradoxa*; they may also reflect the presence and influence of other plant species in the same locations.⁶⁰ In addition, seed viability in this study was inferred solely from vegetation phenology parameters, and therefore, the findings should be interpreted within these limitations.

5. Conclusion

This study employed remotely sensed vegetation and climate data in combination with machine learning algorithms to identify not only suitable habitats for *V. paradoxa*, but also to locate potential viable and productive sites for seed sources – referred to as the SZPI. Out of five seed zones, three were found to be climatically suitable and viable for seed propagation. The integration of vegetation index thresholds with habitat suitability mapping allowed for the classification of areas into poor/non-priority and high-priority seed source locations. This approach provides a valuable framework for guiding the collection of healthy and viable seeds, while also offering insights into key climate variables across seed zones that influence the species' propagation and conservation – an effort aligned with the objectives of the Sustainable Development Goals.

The restoration of this threatened yet important tree species demands urgent attention from all stakeholders, including the government, indigenous people, and scientists. Future efforts will involve collaboration with experts in agroforestry, genetics, and relevant governmental agencies to initiate experimental trials using seeds collected from the high-priority locations identified in this study. Such trials will not only help to validate this work and reveal areas for improvement, but could also significantly support long-term conservation measures aimed at preserving and restoring this tree species.

Acknowledgments

We express our gratitude to the Senckenberg Museum of Natural History Gortitz for providing the opportunity for us to use their office facilities and support for staff members working on an externally funded project.

Funding

This work was funded by the Nigerian Federal Government TETFUND (Grant ID: Kwasu I B R/C RIT/270921 t/0 L 1 ff ETF2020/00 09).

Conflict of interest

The authors declare no conflict of interest in any form.

Author contributions

Conceptualization: Gabriel Salako, Oluwasogo Olalubi, Henry Sawyerr

Formal analysis: Gabriel Salako, Henry Sawyerr, Oluwasogo Olalubi

Investigation: Abdulrasheed Adio, Moses Adetumbi

Methodology: Gabriel Salako, Oluwasogo Olalubi, Abdulrasheed Adio

Writing – original draft: Gabriel Salako, Oluwasogo Olalubi

Writing – review & editing: Gabriel Salako, Oluwasogo Olalubi, Abel Adebayo, Andrey Zaitsev

Availability of data

R code and data are available at: <https://github.com/gabsalako/Vitellaria-Paradoxa>

Online maps available at: https://rpubs.com/gabsalak/Shea_nut_tree

References

- Salako G, Sawyerr H, Bashir A, Adebayo A, Adio A. Deductive and multi-criteria approach to ecosystem modelling and habitat mapping of shea butter trees (*Vitellaria paradoxa*) in the tropical savanna. *Int J Environ Agric Biotechnol (IJEAB)*. 2017;2(6):3078-3088. doi: 10.22161/ijeab/2.6.38
- Goreja WG. *Shea Butter: The Nourishing Properties of Africa's Best-Kept Natural Beauty Secret*. Delhi: Amazing Herbs Press Paperback, TNC International Inc; 2004.
- Natural Resources Commission NRC. *Lost Crops of Africa*. Vol. 2. Washington, D.C. The National Academies Press; 2006. p. 379.
- Amissah N, Akapo B, Yeboah J, Blay E. Asexual propagation of sheanut tree (*vitellaria paradoxa* c.f. Gaertn.) using a container layering technique. *Am J Plant Sci*. 2013;4:1758-1764. doi: 10.4236/ajps.2013.49216
- FAOSAT. *Wood Charcoal Production in Africa GRID-Arendal (Grida.No)*. Rome: FAOSAT; 2014.
- FAOSTAT. *Energy Statistics Database, United Nations Statistics Division*. Rome: FAOSTAT; 2019.
- IUCN; 2013. Available from: <https://www.iucnredlist.org/species/37083/10029534> [Last accessed on 2024 Oct 14].
- Shu-Aib JS, Collins PB, Abbas SK. Prospects of shea trees (*Vitellaria paradoxa*) in climate change mitigation - a case study in cheyohi in the Kumbungu district of ghana. *Ghana J Sci Technol Dev*. 2023;9(1):141-159. doi: 10.47881/209.967x
- Choungou Nguenkeng PB, Hendre P, Tchoundjeu Z, et al. The current state of knowledge of shea nut tree (*Vitellaria paradoxa* C.F.Gaertner.) for nutritional value and tree improvement in west and central Africa. *Forests*. 2021;12(12):1740. doi: 10.3390/f12121740
- Simko I, Jimenez-Berni JA, Sirault XRR. Phenomic approaches and tools for phytopathologists. *Phytopathology*. 2017;107:6-17. doi: 10.1094/PHYTO-02-16-0082-RVW
- Yang G, Jiangang L, Chunjiang Z, et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front Plant Sci*. 2017;30:1111. doi: 10.3389/fpls.2017.01111
- Guisan A, Thuiller W, Zimmermann NE. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge: Cambridge University Press; 2017. p. 478.
- Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens Environ*. 1996;58(3):289-298. doi: 10.1016/S0034-4257(96)00072-7
- Gao BC. NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ*. 1996;58(3):257-266. doi: 10.1016/S0034-4257(96)00067-3
- Reed BC, Schwartz MD, Xiao X. Remote sensing phenology: Status and the way forward. In: Noormets A, editor. *Phenology of Ecosystem Processes*. New York, New York, NY: Springer; 2009. p. 231-246.
- Huete AR, Didan K, Shimabukuro YE, et al. Amazon rainforests green-up with sunlight in dry season. *Geophys Res Lett*. 2006;33(6):L06405, doi: 10.1029/2005GL025583
- Jiang Z, Huete AR, Didan K, Miura T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens Environ*. 2008;112(10):3833-3845. doi: 10.1016/j.rse.2008.06.006
- Youtie B, Shaw N, Fisk M, Jensen SA. Strategy for maximizing native plant material diversity for ecological restoration, germplasm conservation and genecology research. In: *Proceedings of the 8th European Conference on Ecological Restoration*. České Budějovice: Czech Republic; 2012. p. 9-14.
- Kim C, Kim W, Song W, Cho J, Choi J. Prediction of native seed habitat distribution according to SSP scenario and seed transfer zones: A focus on acer pictum subsp. Mono and *Quercus acuta*. *Forests*. 2023;14(1):87. doi: 10.3390/f14010087
- Bower AD, St Clair JB, Erickson V. Generalized provisional seed zones for native plants. *Ecol Appl*. 2014;24(5):913-919. doi: 10.1890/13-0285.1
- Pike C, Potter KM, Berrang P, et al. New seed-collection zones for the eastern United States: The eastern seed zone forum. *J Forestry*. 2020;9(2):271. doi: 10.1093/jofore/fvaa013
- Crow TM, Albeke SE, Buerkle CA, Hufford KM. Provisional methods to guide species-specific seed transfer in ecological restoration. *Ecosphere*. 2018;9:e02059. doi: 10.1002/ecs2.2059
- Shryock DF, Defalco LA, Esque TC. Spatial decision-support tools to guide restoration and seed-sourcing in the desert Southwest. *Ecosphere*. 2018;9(10):e02453. doi: 10.1002/ecs2.2453
- Widrlechner MP, Daly C, Keller M, Kaplan K. Horticultural applications of a newly revised USDA plant hardiness zone map. *Hort Technol*. 2012;22:6-19.
- Richter M. Temperatures in the tropics. In: Pancel L, Köhl M, editors. *Tropical Forestry Handbook*. Berlin, Heidelberg: Springer; 2016.
- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. SpThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*. 2015;38:541-545. doi: 10.1111/ecog.01132
- Phillips SJ, Dudik M, Elith J, et al. Sample selection

- bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol Appl*. 2009;19:181-197. doi: 10.1890/07-2153.1
28. Ghorbani A, Amir MM, Esmali Ouri A. Utility of the NDVI for land/canopy cover mapping in Khalkhal County (Iran). *Ann Biolo Res*. 2012;3:5494-5503.
 29. Rouse JW, Haas RH, Scheel JA, Deering DW. Monitoring vegetation systems in the great plains with ERTS. *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium*. 1974;1:48-62.
 30. McFeeters SK. The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int J Remote Sens*. 1996;17:1425-1432. doi: 10.1080/0143116960894871
 31. Xu H. Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int J Remote Sens*. 2006;27:3025-3033. doi: 10.1080/01431160600589179
 32. Lymburner L, Beggs PJ, Jacobson CR. Estimation of canopy-average surface-specific leaf area using landsat tm data. *Photogrammetr Eng Remote Sens* 2000;66(2):183-191.
 33. Karger DN, Conrad O, Böhrner J, et al. Climatologies at high resolution for the earth's land surface areas. *Sci Data*. 2017;4:170122. doi: 10.1038/sdata.2017.122
 34. Breiman L, Cutler A. *Description: Classification and Regression Based on a Forest of Trees Using Random Inputs, Based on Breiman. Package: RandomForest*; 2001. doi: 10.1023/A:1010933404324
 35. Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J. Predictive performance of presence only species distribution models: A benchmark study with reproducible code. *Ecolo Monograp*. 2022;92(1):e0148. doi: 10.1002/ecm.1486
 36. Salako G, Russell DJ, Stucke A, Einar E. Assessment of multiple model algorithms to predict earthworm geographic distribution range and biodiversity in Germany: Implications for soil-monitoring and species-conservation needs. *Biodivers Conserv*. 2023;32:2365-2394. doi:10.1007/s10531-023-02608-9
 37. Elith J, Graham CH, Anderson RP, et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*. 2006;29:129-151.
 38. Li X, Wang YL. Applying various algorithms for species distribution modelling. *Integr Zool*. 2013;8:124-135. doi: 10.1111/1749-4877.12000
 39. Liaw A, Wiener M. Classification and regression by random forest. *R News*. 2002;2:18-22.
 40. Hijmans RJ, Elith J. *Spatial Distribution Models, Spatial Data Science With R*; 2019. Available from: <https://rspsatial.org/sdm/sdm.pdf> [Last accessed on 2024 Jun 10].
 41. Fritsch S, Guenther F, Wright M. *Neuralnet: Training of Neural Networks R Package e Version*; 2019. Available from: <https://cran.r-project.org/package=neuralnet> [Last accessed on 2024 Apr 12].
 42. Salako G, Zaitsev A, Betancur-Corredor B, Russell DJ. Modelling and spatial prediction of earthworms ecological-categories distribution reveal their habitat and environmental preferences. *Ecol Indic*. 2024;169:1128. doi: 10.1016/j.ecolind.2024.112832
 43. Jiménez-Valverde A. Insights into the area under the receiver operating characteristic curve (AUC) as discrimination measure in species distribution modelling. *Glob Ecol Biogr*. 2011;21(4):498-507. doi: 10.1111/j.1466-8238.2011.00683.x
 44. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure*. 1960;20:37-46. doi: 10.1177/001316446002000104
 45. Manel S, Ceri WH, Ormerod SJ. Evaluating presence-absence models in ecology: The need to account for prevalence. *J Appl Ecol*. 2001;38:921-931. doi: 10.1046/j.1365-2664.2001.00647.x
 46. Silva MC, Moonlight P, Oliveira RS, Rowland L, Pennington TR. COSST: A tool to facilitate seed provenancing for climate-smart ecosystem restoration. *J Appl Ecol*. 2024;62:677-688. doi: 10.1111/1365-2664.14854
 47. Varvatkar A. *Improve Neural Network Performance*; 2023. Available from: <https://www.kaggle.com/code/avadhutvarvatkar/improve-neural-network-performance> [Last accessed on 2024 May 15].
 48. Kipp S, Mistele B, Schmidhalter U. Identification of stay-green and early senescence phenotypes in high-yielding winter wheat, and their relationship to grain yield and grain protein concentration using high-throughput phenotyping techniques. *Funct Plant Biol*. 2013;41:227-235. doi: 10.1071/FP13221
 49. Jansen M, Pinto F, Nagel KA, et al. Non-invasive phenotyping methodologies enable the accurate characterization of growth and performance of shoots and roots. In: Tuberosa R, editor. *Genomics of Plant Genetic Resources*. Dordrecht: Springer Science Business Media; 2014.
 50. Li C, Zhu X, Wei Y, et al. Estimating apple tree canopy chlorophyll content based on sentinel-2A remote sensing imaging. *Sci Rep*. 2018;8:3756. doi: 101038/s41598-018-21963-0
 51. He L, Ren X, Wang Y, et al. Comparing methods for estimating leaf area index by multi-angular remote sensing in winter wheat. *Sci Rep*. 2020;10:13943. doi: 101038/s41598-020-70951-w
 52. Bao J, Yu M, Li J, Wang G, Tang Y, Zhi J. Determination of leaf nitrogen content in apple and jujube by near infrared spectroscopy. *Sci Rep*. 2024;14:20884. doi: 10.1038/s41598-024-71590-1
 53. Oyebanji OO, Salako G, Nneji LM, et al. Impact of

- climate change on the spatial distribution of endemic legume species of the Guineo-Congolian forest, Africa. *Ecol Indic.* 2020;122:107282.
doi: 10.1016/j.ecolind.2020.107282
54. Salako G, Oyebanji OO, Olagunju TE, Howe GT. Potential impact of climate change on the distribution of some selected legumes in Cameroon and adjoining Nigeria border. *Afr J Ecol.* 2021;59:959-975.
doi: 10.1111/aje.1291
 55. Atalla JA. *A Review of Institutional Frameworks for Conservation of the Shea Nut Tree in Uganda. United Nations University Land Restoration Training Programme*; 2015. Available from: <https://www.grocentre.is/static/gro/publication/427/document/atalla2015.pdf> [Last accessed on 2025 Jan 15].
 56. Wundengba D. *Key Facts About the Shea Tree Northernghana*; 2020. Available from: <https://northernghana.net/facts-about-the-shea-tree/> 2020 [Last accessed on 2024 Oct 23].
 57. Okao M, Odoi JB, And Okia CA. Growth performance of shea nut tree (*Vitellaria Paradoxa*) collections in an *ex-situ* trial plot in lira district, Uganda. *Ijrdo J Agric Res.* 2019;5:2455-2476.
doi: 10.53555/ar.v5i4.2776
 58. Bondé L, Ouédraogo O, Traoré S, Thiombiano A, Boussim JI. Impact of environmental conditions on fruit production patterns of shea tree (*Vitellaria paradoxa* C.F.Gaertn) in West Africa. *Afr J Ecol.* 2019;57:353-362.
doi: 10.1111/aje.12621
 59. Bayala J, Sanon Z, Bazié P, *et al.* Relationships between climate at origin and seedling traits in eight pan African. Provenances of *Vitellaria paradoxa* C.F. Gaertn. Under imposed drought stress. *Agrofor Syst.* 2018;92:1455-1467.
doi: 10.1007/s10457-017-0091-8
 60. Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ, Stenseth NC. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol Evol.* 2005;20(9):503-510.
doi: 10.1016/j.tree.2005.05.011

Appendix

A1. Occurrence data and citations

[[1]]

<<rgbif citation>>

Citation: iNaturalist contributors, iNaturalist (2024). iNaturalist Research-grade Observations. iNaturalist.org. Occurrence dataset <https://doi.org/10.15468/ab3s5x> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by-nc/4.0/legalcode>

[[2]]

<<rgbif citation>>

Citation: AFFOUARD A, JOLY A, LOMBARDO J, CHAMP J, GOEAU H, CHOUET M, GRESSE H, BOTELLA C, BONNET P (2023). Pl@ntNet automatically identified occurrences. Version 1.8. Pl@ntNet. Occurrence dataset <https://doi.org/10.15468/mma2ec> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[3]]

<<rgbif citation>>

Citation: Onuminya T, Ogundipe O, Igbari A, Kemabonta K (2022). Higher Plant Collections of the Lagos University Herbarium (LUH): Families A-Z. Version 1.5. University of Lagos. Occurrence dataset <https://doi.org/10.15468/ty22qw> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[4]]

<<rgbif citation>>

Citation: TOKO IMOROU I, SOGBOSSI E, BIAOU S, GOBI MOHAMED A (2018). Diversité floristique des espèces menacées prioritaires du Bénin (Afrique de l'Ouest). Version 1.1. Laboratory of Forest Sciences (University of Abomey-Calavi). Occurrence dataset <https://doi.org/10.15468/m0cqdu> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[5]]

<<rgbif citation>>

Citation: GBAGUIDI R (2023). Plant-specific diversity in forest reserves and riparian areas of Benin. Version 1.3. SOS Biodiversity. Sampling event dataset <https://doi.org/10.15468/uq5xe6> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[6]]

<<rgbif citation>>

Citation: TOKO IMOROU I, SOGBOSSI E, BIAOU S (2017). Diversité des espèces menacées des zones classées et terroirs villageois du Bénin (Afrique de l'Ouest). Laboratory of Forest Sciences (University of Abomey-Calavi). Occurrence dataset <https://doi.org/10.15468/g8pbz5> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[7]]

<<rgbif citation>>

Citation: Chukwuma E, Agbo-Adediran A, Chukwuma D, Ugbogu O (2022). Sampling Event Dataset from Oba Hills, Osun State. Version 1.5. Forestry Research Institute of Nigeria. Sampling event dataset <https://doi.org/10.15468/zkkrum> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by-nc/4.0/legalcode>

[[8]]

<<rgbif citation>>

Citation: KAKPO S B (2017). Census of agroforestry plant species of Benin (West Africa) by SOS Biodiversity. Laboratory of Forest Sciences (University of Abomey-Calavi). Occurrence dataset <https://doi.org/10.15468/9fvpm9> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[9]]

<<rgbif citation>>

Citation: KPOSSI H (2020). Diversité des plantes des parcours naturels et des forêts au Bénin. Laboratoire d'Ecologie Appliquée/Université d'Abomey-Calavi (LEA/UAC). Occurrence dataset <https://doi.org/10.15468/xxnkja> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[10]]

<<rgbif citation>>

Citation: APELETE E (2020). Census of fodder species through central and north of Benin. Direction Générale des Eaux, Forêts et Chasse. Occurrence dataset <https://doi.org/10.15468/qnqvt5> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[11]]

<<rgbif citation>>

Citation: GBAGUIDI R, DOSSA GBO C F, DEGBOE E (2023). Study of specific plant diversity in forest reserves and riparian areas in Benin. Version

1.6. Direction Générale des Eaux, Forêts et Chasse. Sampling event dataset <https://doi.org/10.15468/yx8ay6> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[12]]

<<rgbif citation>>

Citation: Nnabude P, Egboka T, Ugbogu O, Osiyemi O (2019). Plants Of Nnamdi Azikiwe University Herbarium (NAUH) Awka. Version 1.5. Forestry Research Institute of Nigeria. Occurrence dataset <https://doi.org/10.15468/qqzpf> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by-nc/4.0/legalcode>

[[13]]

<<rgbif citation>>

Citation: MAGLOE G A, Tohinnou C (2020). Ethnobotanique et diversité floristique des écosystèmes temporairement inondés et fragiles du Bénin. Laboratoire d'Ecologie Appliquée/Université d'Abomey-Calavi (LEA/UAC). Occurrence dataset <https://doi.org/10.15468/zckesy> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[14]]

<<rgbif citation>>

Citation: Yêhouénu Tessi D R (2017). Census of medicinal and agroforestry plants of Benin. Laboratory of Forest Sciences (University of Abomey-Calavi). Occurrence dataset <https://doi.org/10.15468/y8jm57> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[15]]

<<rgbif citation>>

Citation: DOSSA GBO C F, VITOEKPON I, GBO F C (2020). Census of plant species in Benin as part of several research projects. ONG Rêve développement. Occurrence dataset <https://doi.org/10.15468/9w3pr5> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[16]]

<<rgbif citation>>

Citation: AGBAHOUNGBA F, DOSSA GBO C F (2016). Elaboration du tarif de cubage multispécifique. Version 1.1. GBIF Benin. Occurrence dataset <https://doi.org/10.15468/lzrybq> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by-nc/4.0/legalcode>

[[17]]

<<rgbif citation>>

Citation: Gbetoho A J, Gangnibo C (2019). Agroforestry plant species found in protected areas and riparian lands in Sudano-guinean and Sudanian agroecological zones. Data mobilized in the framework of BID National project BID-AF2015-0065-NAC and funded by EU. Version 1.3. GBIF Benin. Occurrence dataset <https://doi.org/10.15468/lcbrcr> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[18]]

<<rgbif citation>>

Citation: KOURA K (2017). Agroforestry species in *Parkia biglobosa*'s parklands in Northern Benin. Laboratory of Forest Sciences (University of Abomey-Calavi). Occurrence dataset <https://doi.org/10.15468/iokuu0> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[19]]

<<rgbif citation>>

Citation: TOTIN H, OGOU C, MOROU N (2020). Floral inventory of the natural pastures of the municipalities bordering the W Transboundary Biosphere Reserve in the Alibori department and the municipalities of Dassa_Zoume and Glazoué in the Collines department (Benin). Jeunesse et Emplois Verts pour une Economie Verte (JEVEV-ONG). Occurrence dataset <https://doi.org/10.15468/onygso> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[20]]

<<rgbif citation>>

Citation: Essou F (2020). Inventaire des plantes alimentaires au Bénin. Laboratoire d'Ecologie Appliquée/Université d'Abomey-Calavi (LEA/UAC). Occurrence dataset <https://doi.org/10.15468/bgidnz> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[21]]

<<rgbif citation>>

Citation: S. Gaston A, Direction D G F E D R N (2016). The census of National Forest of Benin. Version 1.2. GBIF Benin. Occurrence dataset <https://doi.org/10.15468/5cdfgv> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: Public Domain (CC0 1.0)

[[22]]

<<rgbif citation>>

Citation: Kiki P, Ganglo J (2017). Census of the threatened species of Benin. Version 1.5. GBIF Benin. Occurrence dataset

<https://doi.org/10.15468/fbbbfl> accessed via GBIF.org on 2024-11-26.
Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on
2024-11-26
Rights: <http://creativecommons.org/publicdomain/zero/1.0/legalcode>

[[23]]

<<rgbif citation>>

Citation: Nkemehule F, Onuminya T, Ogundipe O (2023). Herbarium specimens held at the College of Medicine, University of Lagos, Lagos, Nigeria. Version 1.3. University of Lagos. Occurrence dataset <https://doi.org/10.15468/5ngcgq> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[24]]

<<rgbif citation>>

Citation: FAVI F, WABI F, AIKPON G (2017). Distribution de quelques espèces de plantes agroforestières et médicinales au Bénin. Faculté des Sciences et Techniques (FAST). Occurrence dataset <https://doi.org/10.15468/kpwam2> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[25]]

<<rgbif citation>>

Citation: Onuminya T, Ogundipe O, Igbari A (2022). A survey of medicinal plants sold in Lagos herbal markets in Nigeria. University of Lagos. Occurrence dataset <https://doi.org/10.15468/cmb5cw> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[26]]

<<rgbif citation>>

Citation: Taugourdeau S, Daget P (2019). FLOTROP, a massive contribution to plant diversity data for open ecosystems in Tropical Africa. Version 1.5. CIRAD SELMET. Occurrence dataset <https://doi.org/10.15468/oxunfl> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[27]]

<<rgbif citation>>

Citation: Onuminya T, Ogundipe O, Igbari A (2022). A record of Plants used In Treating Dental Diseases in Lagos Nigeria. Version 1.1. University Of Lagos. Occurrence dataset <https://doi.org/10.15468/bygfuz> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[28]]

<<rgbif citation>>

Citation: KAKPO S, KOURA K, GANGLO J C (2019). Census of plant species of Benin (West Africa): Database4. SOS Biodiversity. Occurrence dataset <https://doi.org/10.15468/3zjlhv> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[29]]

<<rgbif citation>>

Citation: Houndonougbo A, WIDEDJI J B, KAKPO S B, KOURA K, GANGLO J C (2019). Inventory of plant species through Benin (West Africa). ONG Rêve développement. Occurrence dataset <https://doi.org/10.15468/kiwmh2> Accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[30]]

<<rgbif citation>>

Citation: Crop Wild Relatives Occurrence data consortia (2024). A global Database for the distributions of crop wild relatives. Version 1.13. Centro Internacional de Agricultura Tropical - CIAT. Occurrence dataset <https://doi.org/10.15468/jyrthk> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[31]]

<<rgbif citation>>

Citation: Bijmoer R, Arias Guerrero S, Creuwels J (2024). Naturalis Biodiversity Center (NL) - Botany. Naturalis Biodiversity Center. Occurrence dataset <https://doi.org/10.15468/ib5ypt> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/publicdomain/zero/1.0/legalcode>

[[32]]

<<rgbif citation>>

Citation: Royal Botanic Gardens, Kew (2024). Royal Botanic Gardens, Kew - Herbarium Specimens. Occurrence dataset <https://doi.org/10.15468/ly60bx> Accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[33]]

<<rgbif citation>>

Citation: HOUNSA M, Gbetoho A (2019). Biodiversité végétale de quelques écosystèmes au Bénin. Laboratory of Forest Sciences (University of Abomey-Calavi). Occurrence dataset <https://doi.org/10.15468/a1cdnr> Accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[34]]

<<rgbif citation>>

Citation: Gangnibo C N, Gbètoho A J (2016). Census of species recorded During phytosociological surveys in Benin. Version 1.1. GBIF Benin. Occurrence dataset <https://doi.org/10.15468/6habyb> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: Public Domain (CC0 1.0)

[[35]]

<<rgbif citation>>

Citation: Meise Botanic Garden (2024). Meise Botanic Garden Herbarium (BR). Version 1.33. Meise Botanic Garden. Occurrence dataset <https://doi.org/10.15468/wrthhx> accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[36]]

<<rgbif citation>>

Citation: Teisher J, Stimmel H (2024). Tropicos MO Specimen Data. Missouri Botanical Garden. Occurrence dataset <https://doi.org/10.15468/hja69f> Accessed via GBIF.org on 2024-11-26. Accessed from R via rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26
Rights: <http://creativecommons.org/licenses/by/4.0/legalcode>

[[37]]

<<rgbif citation>>

Citation: Ugbogu O, Osiyemi O, Chukwuma E, Adeniji K, Odewo A, Okanlawon T, Soyewo T, Oyebola T, Aina S (2019). Some Angiosperm Data of the Forest Herbarium Ibadan (FHI) Nigeria. Version 1.8. Forestry Research Institute of Nigeria. Occurrence dataset <https://doi.org/10.15468/uhnd5n> accessed through GBIF.org on 2024-11-26. Accessed from R through rgbif (<https://github.com/ropensci/rgbif>) on 2024-11-26