INTERNATIONAL JOURNAL OF
INNOVATIVE RESEARCH AND KNOWLEDGE
ISSN-2213-1356 www.ijirk.com

MaxEnt Modeling for Predicting a Suitable Habitat for a Threatened and Endangered Medicinal Plant Species *Strychnos henningsii* in Kenya

***KURIA MW**
Kabarak University

NGUMI VW (DVC, APD)
NJENGA PK, PhD
Jomo Kenyatta University of Agriculture and Technology

WANGAI LN
Kirinyaga University College

***Corresponding author**

Abstract

Strychnos henningsii is a threatened and endangered medicinal plant species in Kenya. It is highly over-exploited for medicinal purposes, which has resulted in its decline in the natural habitat. For instance, it has been reported as disappearing in Ukambani areas in Kenya. A study on ethnobotany of this species reported that the species was obtained with difficulties in the study areas. Recently, Neem Foundation in Kenya reported that this species is important for malaria treatment. Therefore, this species is a serious case for conservation. The main objective of this study is to model the potential suitable habitats for growth and reintroduction of this species which can be targeted for its conservation.

The use of species distribution models has gained popularity in biological sciences. These tools find use in predicting species distribution across a given area of study. The maximum entropy (or MaxEnt) distribution model was used for predicting potential a suitable habitat for S. henningsii. Therefore, field-collected co-ordinates of

270 presence data locations of this species were used. Additionally, environmental data downloaded from the WorldClim data portal were also used. MaxEnt was then run using default settings with 75% of the locations being used for the training and the remaining 25% for the testing of the model. It generated a map of predicted suitable habitat of a species on the scale from 0 to 1. The lowest suitable areas represented by 0, and 1 representing the highest suitable areas.

The results indicated that the area under the curve (AUC) for the receiver operating characteristic (ROC) analyses measured at all possible threshold values for training and test data were 0.986 and 0.983 respectfully. These values were close to 1 thereby showing that the model performed better than random and therefore points to the accuracy of the model in prediction. The areas identified for growth and reintroduction of these species were Kilifi, Kwale, Taita Taveta Mt. Kilimanjaro Game Reserve (coastal regions), Maasai Mara Game Reserve, Marigat, Maralal, Baragoi, Marsabit and Huri Hills (dry land areas in Kenya). These areas can be used to develop S. henningsii demonstration plots. People living within these niches can also be encouraged to adopt the species for agroforestry in order to conserve this important plant species.

Keywords: *MaxEnt; species distribution models; threatened; Strychnos henningsii; conservation*

INTRODUCTION

Strychnos henningsii (Gilg.), Loganiaceae common names: Muteta (Kikuyu/Kamba), Muchimbi (Meru) (Maundu and Tengäs, 2005), is an indigenous medicinal and threatened plant species in Kenya. The wide-scale use of this species resulted in its over-exploitation by the traditional medicinal practitioners, hoteliers and restaurant owners, and the local people. It is used in the preparation of milk soups and fatty-meat among Kikuyu, Maasai and Kamba people (Chapman *et al*., 1997; Palgrave 1988; Beentje, 1994; Maundu *et al*., 1999; Gachathi, 2007). In African traditional medicine, *S. henningsii* has been used for treatment of various diseases including rheumatism, syphilis, gastrointestinal disorders (purgative) and snakebites (Hutchings, 1989; Watt and Breyer, 1962; Pujol, 1993; Hutchings, 1996; Oyedemi *et al.,* 2009). Compounding over-exploitation, the species has a slow growth rate, seed production is erratic and seed germination is poor (Maundu and Tengäs, 2005). It has been reported as disappearing in Mwingi areas in Kenya (Musila *et al*., 2004; Schmeltzer, 2008). A study by Kuria, Njenga and Ngumi, 2012) revealed that *S. henningsii* is a threatened species in the areas of study. Most respondents of this study obtained the plant with difficulties and among the key difficulties encountered were scarcity and long distance walking.

The demand for this species is continually increasing. According to Neem Foundation of Kenya (2015), *S. henningsii* was discovered to have the ability to cure malaria. Kenya being a tropical country, the fight against malaria is never ending. Therefore, the present study aims at the modeling of *S. henningsii* distribution in Kenya to delineate the potential distribution areas for growing this species using MaxEnt software. Such areas could be used for reintroduction and conservation of these species.

One approach to assessing the potential distribution of a species is through the use of Species Distribution Models (SDMs). Such models correlate known occurrences of a species with environmental variables of the location(s) from where it has been recorded (Elith and Leathwick, 2009) and predict the potential distributions (Peterson *et al*., 2011; Yang *et al*., 2013). They have been used for prioritizing field surveys (Graham *et al*., 2008). SDMs are based on the classical concept of niche in ecology and model potential or realized distribution of a species based on the modeling algorithm used (Kumar *et al*., 2014).

Distribution Models are broadly classified into two groups: "correlative models" and "process-based or mechanistic models" (Dormann *et al.,* 2012). Correlative models associate species occurrence data with spatial environmental layers of the area of study and produce maps of probability presence or relative suitability of a species (Kumar *et al*., 2014). The process based or mechanistic model uses species' functional traits and physiological tolerances for model fitting (Kearney *et al*., 2010). This approach requires detailed experimental data that may not be available for the target species (Dormann *et al.,* 2012). Both methods have been used in quantifying and mapping the potential distribution of species in areas outside their current distributional range (Lozier and Mill, 2011; Ni *et al.,* 2012). SDMs are becoming increasingly popular in ecology and are being widely used in many ecological applications (Elith *et al*., 2006; Peterson *et al*., 2003).

Predicting and mapping of potential suitable habitat for threatened and endangered species is critical for the monitoring and restoration of their declining native populations in the natural habitats, artificial introductions or selecting conservation sites, and conservation and management of native habitats (Gaston 1996). However, distribution data of threatened and endangered species are often sparse and clustered, making it difficult to model their suitable distribution using commonly used modeling approaches (Kumar and Stohlgren, 2009). Some of the popular and commonly used SDMs include: Bioclimatic Variable (BIOCLIM), Domain (DOMAIN), Generalized Linear Model (GLM), Multivariate Adaptive Regression Splines (MARS), Genetic Algorithm for Rule-set Production (GARP), Maximum Entropy Modeling (MaxEnt) and Boosted Regression Trees (BRT). The use of these models/algorithms is guided by the ultimate objective of the study (Jaryan *et al*., 2013).

BIOCLIM is a model that identifies locations where the climatic indices fall within the range that has been determined based on ground observations. The model treats each climatic axis independently, thus leading to unsound predictions (Carpenter *et al.,* 1993). The model does not perform well with respect to precipitation. DOMAIN is a multivariate distance-based model. Though it uses presence-only data, performance of the model is limited and sometimes additional information on absence locations is required (Graham *et al*., 2008). GLM is a linear regression model that considers response variables that have other than normal distribution (Guisan *et al*., 2002). Compared to traditional models, such as DOMAIN, GLM performs well (Elith *et al*., 2006). MARS provides an alternative regression-based method for fitting nonlinear responses using piece-wise linear fits (Elith *et al*., 2006). It is faster than GLM, but highly sensitive to sample size (Wisz *et al*., 2008). BRT constructs a combination of trees and is quite effective. It is difficult to identify the significant predictor variable in BRT and it is relatively time-consuming (Graham *et al*., 2008). MaxEnt is a comparatively popular model for accurately predicting species distribution and, therefore, it has been recommended (Wisz *et al*., 2008; Meyer *et al.*, 2006; Berry *et al*., 2002; Hijmans *et al*., 1999).

MaxEnt is a maximum entropy-based program in which relative entropy is minimized between the two probabilities of presence data and landscape (Elith *et al*., 2011). It focuses on relating the environmental conditions of the area where a species is present to the environmental conditions across the area of study (Phillips *et al*., 2006; Phillips and Dudik, 2008). MaxEnt has been found to perform best among many different modeling methods (Elith *et al*., 2006; Ortega-Huera and Peterson, 2008) and remain effective despite small sample sizes (Hernandez *et al*., 2006; Peterson *et al*., 2007; Papes and Gaubert, 2007; Wisz *et al*., 2008; Benito *et al*., 2009). This method has been used in modeling potential distribution areas for many plant species among which *Canacomyrica monticola* in New Calendonia (Kumar and Stohlgren 2009), *Basella alba* L (Reddy *et al*., 2015), *Bradypus variegatus* and *Microryzomys minutus* (Philips *et al*., 2006), *Sapium sebiferum* (Jaryan *et al.,* 2013); baobab trees (Sanchez *et al.,* 2010), dipterocarps species (Amaludin; 2012) and also in wildlife research (Baldwin,

2009).

In this study, the correlative model MaxEnt was used to predict the potential distribution of *S. henningsii* in Kenya. The specific guiding objectives were to determine the environmental conditions characterizing the geographical areas of *S. henningsii,* and map the potential distribution areas of this species in Kenya based on the determined environmental conditions. These areas can then be used to develop *S henningsii* demonstration plots, or orchids. People living within the niche could be encouraged to adopt the plant for agro-forestry in order to conserve this species.

MATERIALS AND METHODS

A field study was conducted in areas where *S. henningsii* naturally grew. Nine populations were included in the study. The locational data is required as a comma-separated values (CSV) file. It had information on attributes like, ID, species name, county name, longitude, latitude and altitude. Therefore, this information was converted into CSV format and later used as one of the inputs to MaxEnt. The locational data requires three mandatory fields, which are ID, latitude and longitude.

In addition to species occurrence data, environmental data are a key input in MaxEnt. Therefore, information on bioclimatic parameters was downloaded from the WorldClim data portal (www.worldclim.org). Data downloaded from the WorldClim website was in four sets of variables: Maximum temperature, minimum temperature, precipitation and altitude. The data downloaded lacked spatial projections but contained latitudinal and longitudinal information with a datum of WGS84 (World Geodetic System 1984). On extraction of the ZIP files of the variables, the minimum temperature had 12 layers. This file represented data in months of the year with 1 (January) and 12 (December). This case was found similar for maximum temperature and precipitation. Altitude had only one layer. In total the 37 layers were downloaded. To add a spatial projection each layer was processed one at a time. A layer was added in QGIS (Quantum Geographical information) using the “add” raster layer button and assigned a projection, EPSG4326 (European Petroleum Survey Group 4326). A wall-to-wall vector map of Kenya was added onto the layer and assigned the same projection.

Considering that the area of study is Kenya, the bioclimatic pixels that covered it were clipped from the WorldClim data and used for the study. However, re-sampling of the KenyaClim data is necessary to ensure that the pixel sizes of the data are similar in all layers and throughout the image. This was done through the use of the Wrap (Warwick Research Archive portal) (Re-project) tool. The re-sampling method was cubic. The exact pixel size was given as 0.008 by 0.008. This is done by typing –tr 0.008 0.008 after the GTiff command at the edit option in the Wrap (Re-project) dialogue box. The re-sampled layer was saved as a GTiff (GeoTag Image File Format) file. The native format of this file was “BIL” (Band Interleaved by Line). In order to be compatible with MaxEnt, these files were converted to ASCII (American Standard Code for Information Interchange) using ArcGIS (Aeronautical Reconnaissance Coverage Geographic Information System) 9.3 version.

The CSV file that contains *S. henningsii* presence data was added on the samples upload option and the ASCII file containing the environmental layers were used as MaxEnt inputs. The Environmental layers’ option can only run 19 variables at a time. The altitude ASCII file was added. The remaining 36 files had to be sampled. Six layers representing six different months in maximum temperature, minimum temperature and precipitation were selected. The selection was based on seasons, in such a way that the months selected were each from a different season. Therefore, January, March, May, July, October, and December were the representative layers for the three environmental variables. The 18 layers were added.

MaxEnt (version 3.3.3k) downloaded from the portal (www.cs.princeton.edu) (Phillips *et al.,* 2006) was used in this study. 75% of the location data points were used for training and the remaining 25% for testing the model. The model was then run using the default settings. The outputs generated by MaxEnt predicted suitability of a habitat of a species on the scale from 0 to 1, with the lowest suitability areas represented by 0 and 1 representing the high suitability areas.

The evaluation of model accuracy is an essential step as it indicates the level of accuracy of the estimations. The concept of model validation (Bair, 1994; Oreskes, 1998) is generally accepted and interpreted in terms of suitability for a particular purpose (Rykiel Jr., 1996; Sargent, 2001). Several methodologies have been used for model accuracy assessment in species distribution modeling. The Receiver Operating Characteristic (ROC) and defined thresholds are important methodologies used for the evaluation of MaxEnt model quality (Reddy *et al*., 2015). Table 1 provides the codes and details of the bioclimatic variables used for MaxEnt modeling.

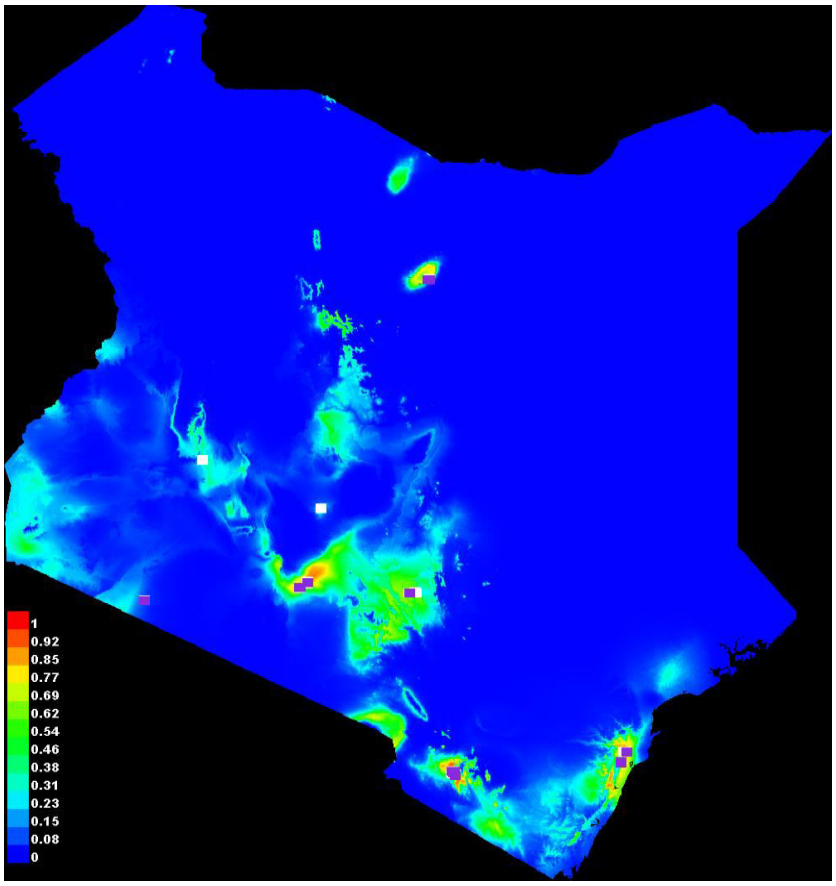
Table 1: Codes and Details of the Bioclimatic Variables Used for MaxEnt Modeling

Codes	Parameters
prec_1_res	Precipitation for the month of January
pre_10_res	Precipitation for the month of October
prec_7_res	Precipitation for the month July
tmax_10_res	Maximum temperature for the month of October
tmax_1_res	Maximum temperature for the month of January
alt_res	Altitude of the geographical locations
prec_12_res	Precipitation for the month December
prec_5_res	Precipitation for the month May
tmin_12_res	Minimum temperature for the month of December
tmin_5_res	Minimum temperature for the month of May
prec_3_res	Precipitation for the month March
tmin_1_res	Minimum temperature for the month of January
tmax_7_res	Maximum temperature for the month of July
tmin_10_res	Minimum temperature for the month of October
tmax_12_res	Maximum temperature for the month of December
tmax_5_res	Maximum temperature for the month of May
tmax_3_res	Maximum temperature for the month of March
tmin_7_res	Minimum temperature for the month of July
tmin_3_res	Minimum temperature for the month of March

RESULTS

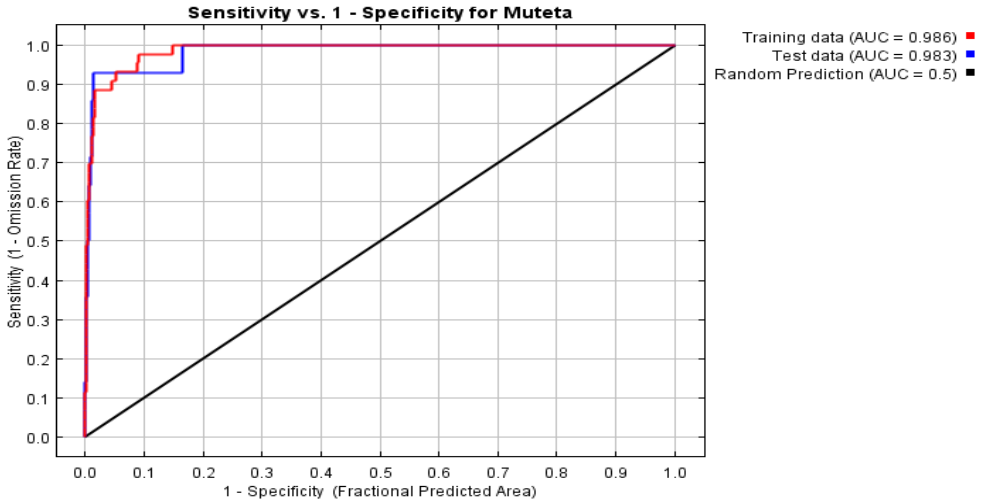
In this study, environmental conditions of the current locations of *S. henningsii* were utilized to predict the probability of suitable conditions existing for the same species in other locations in Kenya using the MaxEnt model. In Figure 1, the image uses colors to indicate the predicted probability that conditions are suitable, with red indicating high probability of suitable conditions for *S. henningsii*, green indicating conditions typical of those areas where the species is found, lighter green shades indicating low predicted probability of suitable conditions, and blue representing areas of unsuitable conditions.

Figure 1: MaxEnt Model Output Showing the Presence Locations used for Training (in white) and the Test Locations (in violet) and Potential Distribution Areas for *S. henningsii*



The relevance and predictive accuracy of the MaxEnt output model was evaluated using two approaches: Receiver Operating Characteristic (ROC)-plot and a defined threshold. ROC is a threshold-independent approach analyzed for both training and test data. In it, the performance is measured on the basis of area under curve (AUC) (Hanley and McNeil, 1982). The ROC curve is a plot between sensitivity (true positive fraction), i.e. absence of omission error, and the proportion of incorrectly predicted observed absences (1-specificity) or false positive fraction, i.e. commission error. The specificity is defined using predicted area, rather than true commission. The area below the ROC curve, i.e. the AUC value indicates the predictive accuracy of the model. The value of AUC in the case of training data was 0.986 and that of the test data was 0.983, both close to 1, which shows that the model performed better than random and therefore points to the accuracy of the model (Figure 2).

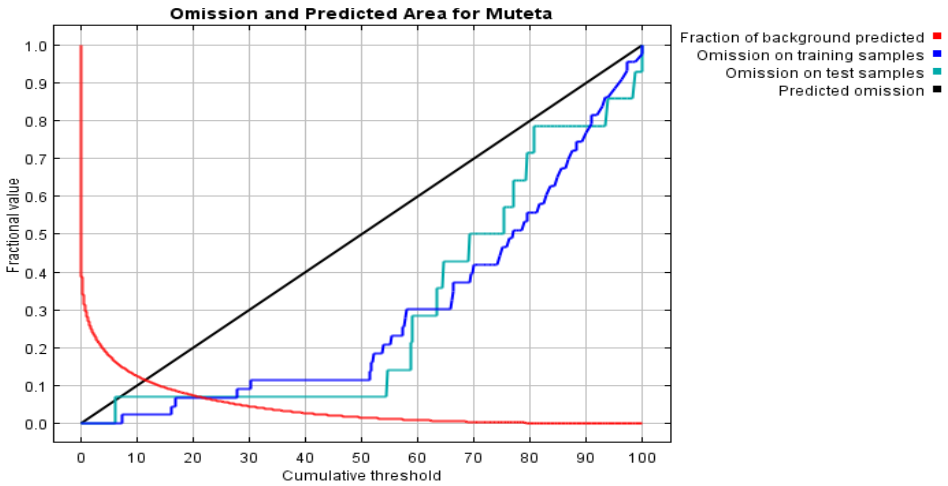
Figure 2: Receiver Operating Characteristic Curve with Area Under Curve (AUC)



The same data for training and testing were used and thus the red and blue lines were identical. The red (training) line shows the "fit" of the model to the training data. The blue (testing) line indicates the fit of the model to the testing data, and is the real test of the models predictive power (Phillips *et al*., 2006). The black line shows the line that would be expected if the model was no better than random. If the blue line (the test line) falls below the black line, then this indicates that the model performs worse than a random model would. The further towards the top left of the graph that the blue line is, the better the model is at predicting the presences contained in the test sample of the data (Phillips *et al*., 2006). It is important to note that AUC-values tend to be higher for species with narrow ranges, as in the case of *S. henningsii* relative to the study area described by the environmental data.

AUC's are developed from ROC plots for assessing differences in species suitability for developed models compared to a random distribution. A binomial test of omission (known areas of presence/ predicted absence) can then be used to test whether or not this difference is significant (Philips *et al.,* 2006). This test is a threshold-dependent method based on omission and predicted area to test the suitability of the model in prediction (Phillips *et al*., 2006; Phillips and Dudik, 2008) (see Figure 3).

***Figure 3: Omission vs. Predicted Area for** S. henningsii.*



Omission rate is the fraction of the test localities that fall into the pixels that are not predicted as suitable for the species, and predicted area is the fraction of all the pixels that are predicted as suitable for the species (Phillips *et al*., 2006). The red line indicates fraction of background predicted, black line indicates predicted omission, blue line indicates omission on training samples and light blue line indicates omission on test samples. The omission rate is calculated both on the training presence records and on the test records (Fielding and Bell, 2007; Anderson *et al*., 2003). The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold. Figure 3 shows how testing and training omission and predicted area vary with the choice of cumulative threshold.

The other approach used for evaluation of the MaxEnt model in this study was the defined threshold. This approach involves selecting thresholds to establish the sites that are considered suitable or unsuitable for the species of interest. Once a threshold has been identified, the locations can be classified as suitable or unsuitable for the species of interest. These thresholds are established to maximize sensitivity while minimizing specificity (Fielding and Bell, 1997; Phillips *et al*., 2006). Threshold values differ for each model and are selected to provide a desired balance between omission and commission (Fielding and Bell, 1997; Hernandez *et al.,* 2006). Where this threshold is applied is selected at the discretion of the modeler, for example, when dealing with endangered species the modeler may decide to maintain zero omission error while identifying the minimum predicted areas. However, if the modeler is interested in identifying any possible area that a species may utilize, the approach might be to minimize commission error (Pearson *et al*., 2007).

In this study some common thresholds and corresponding omission rates for the evaluation of MaxEnt accuracy are as shown in (Table 2). Since the test data are available and the number of test samples is at most 25, binomial probabilities are calculated using a normal approximation to the binomial. These are 1- sided-p values for the null hypothesis that the test points are predicted no better than by a random prediction with the same fractional predicted area. The "balance" threshold minimizes 6* training omission rate, +0.04* cumulative threshold, and +1.6* fractional predicted area. Table 2 describes the common thresholds and the corresponding omission rates for the threshold-dependent binomial tests of omission.

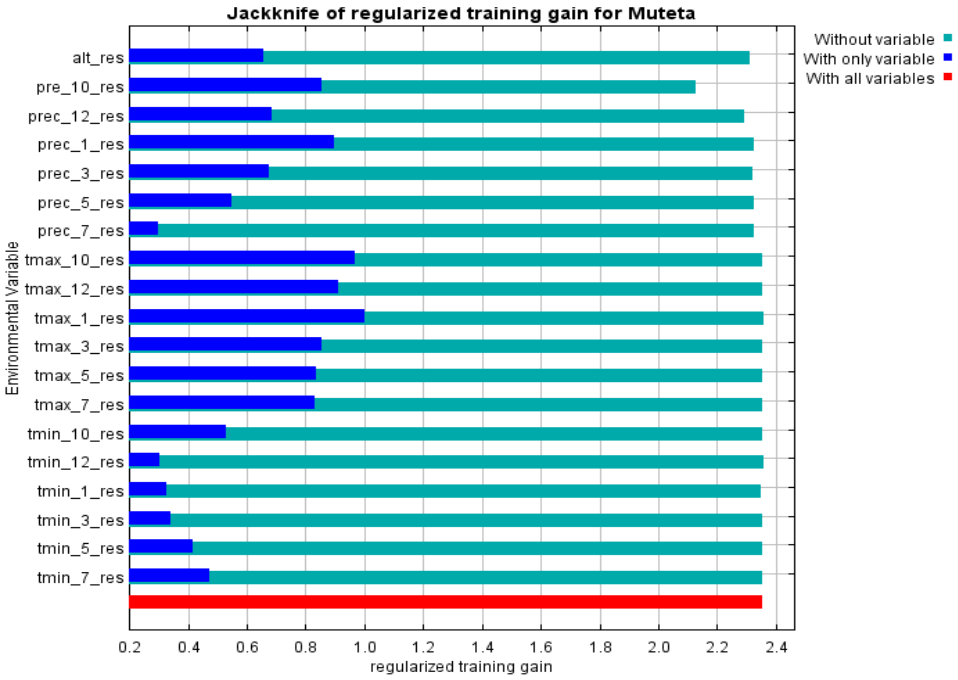
Table 2: Common Thresholds and Corresponding Omission Rates for the Threshold- Dependent Binomial Tests of Omission

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.016	Fixed cumulative value 1	0.286	0.000	0.000	2.51E-8
5.000	0.058	Fixed cumulative value 5	0.180	0.000	0.000	3.618E-11
10.000	0.109	Fixed cumulative value 10	0.125	0.023	0.071	2.363E-11
7.408	0.081	Minimum training presence	0.149	0.000	0.071	2.147E-10
30.210	0.299	10 percentile training presence	0.046	0.093	0.071	4.954E-17
21.451	0.222	Equal training sensitivity and	0.070	0.070	0.071	1.224E-14

		specificity				
16.143	0.178	Maximum training sensitivity plus specificity	0.090	0.023	0.071	3.225E-13
20.975	0.219	Equal test sensitivity and specificity	0.071	0.070	0.071	1.638E-14
54.476	0.575	Maximum test sensitivity plus specificity	0.013	0.209	0.071	5.335E-24
2.722	0.037	Balance training omission, predicted area and threshold value	0.224	0.000	0.000	8.037E-10
15.059	0.165	Equate entropy of thresholded and original distributions	0.095	0.023	0.071	6.554E-13

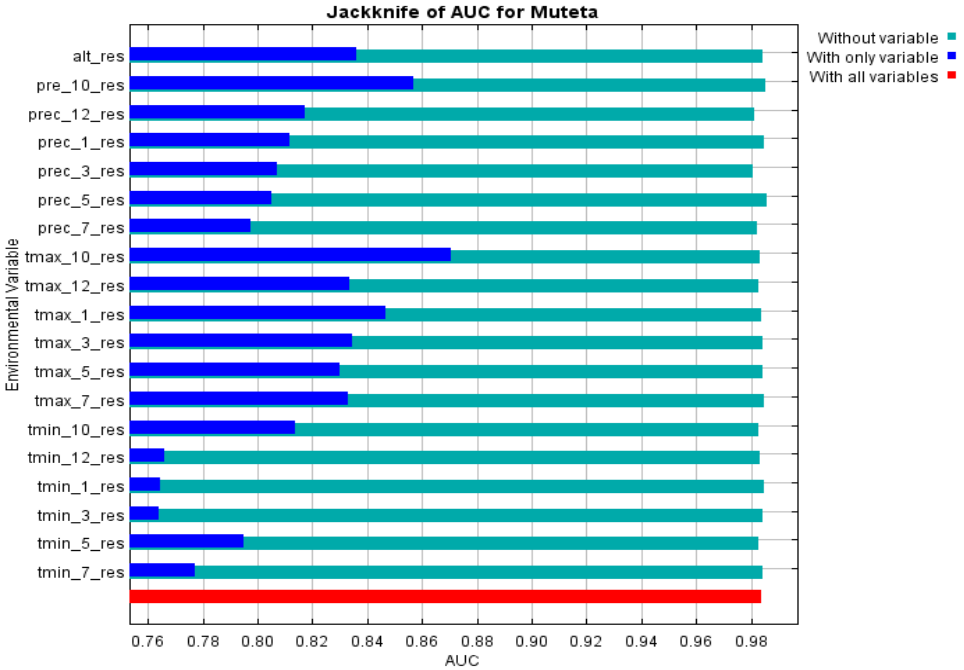
MaxEnt also performed the jack-knife test to evaluate the relative influence of different environmental variables in the model prediction of *S. henningsii* distribution. Therefore, the environmental variables with highest gain when used in isolation are tmax_1_res (Maximum temperature for January) in Figure 4(a), tmax_10_res (Maximum temperature for October) in Figures 4 (b) and (c), which therefore appears to have the most useful information by themselves. The environmental variable that decreases the gain the most when it is omitted is pre_10_res [Figure 4 (a)], which therefore appears to have the most information that is not present in the other variables. MaxEnt also generated species' response curves that showed relationships between predicted probabilities of presence for *S. henningsii* and the different variables [Figures 5(a)-(d)].

Figure 4(a): Relative Importance of Environmental Variables Based on Jackknife test: (a) Training Gain as a Measure of Model's Predictive Ability



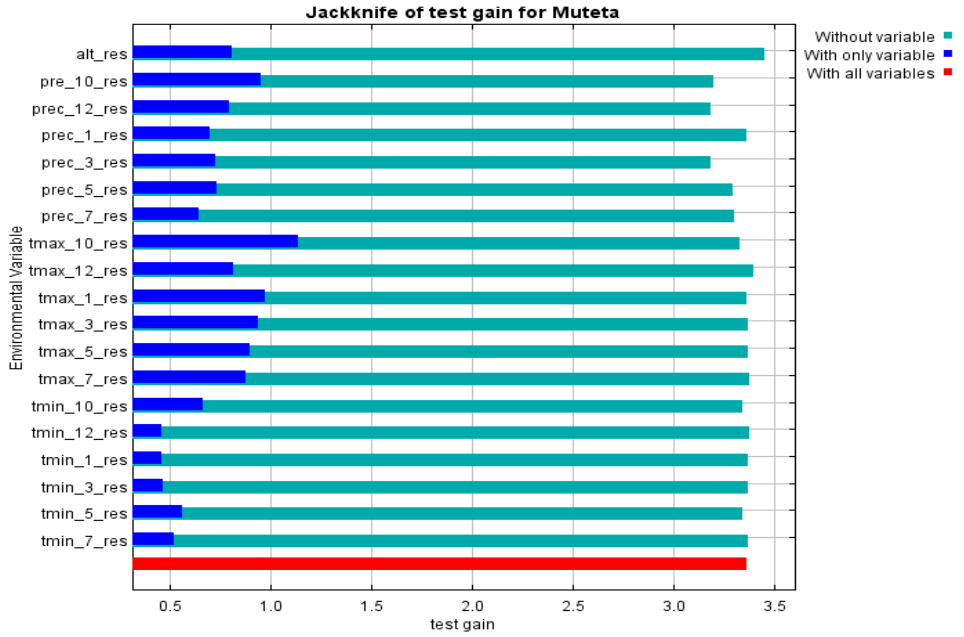
(a)

Figure 4(b): Relative Importance of Environmental Variables Based on Jackknife test: Area Under Curve (AUC) as a Measure of Model's Predictive Ability



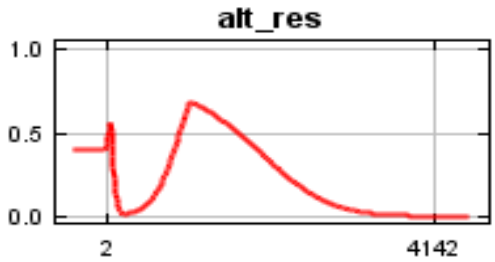
(b)

Figure 4(c): Relative Importance of Environmental Variables Based on Jackknife test: Testing Gain as a Measure of Model's Predictive Ability

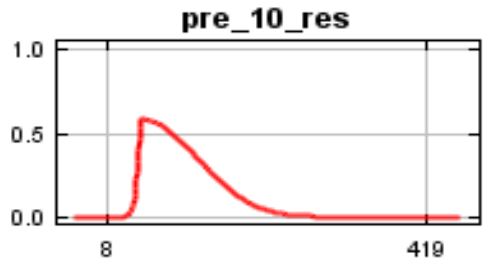


(c)

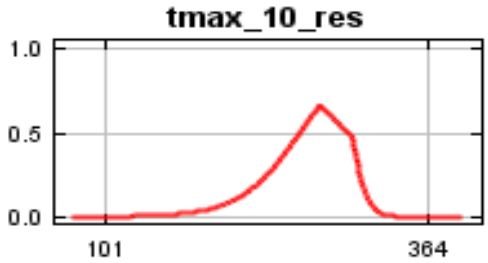
Figure 5: Relationship between the strongest environmental variables and S. henningsii probability of presence: (a) Altitude of different ecological zones; (b) Precipitation for the month of October; (c) Maximum temperature for the month of October; (d) Maximum temperature for the month of January. Each of these curves is based on different MaxEnt models created using only the corresponding variable



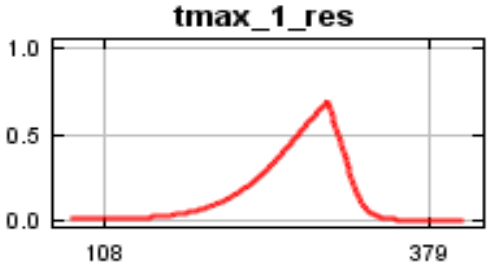
(a)



(b)



(c)



(d)

DISCUSSION

In this study, we evaluated the performance of the correlative niche model MaxEnt in predicting the potential distribution areas of *S. henningsii* in Kenya. Two approaches were used in achieving this goal: The ROC analysis and the omission/commission rate. The location data was split into two portions: Training (75%) and test (25 %) data.

The potential habitats distribution for *S. henningsii* predicted by MaxEnt showed that the areas suitable for growth and reintroduction of this species are along the coastal regions (Taita Hills, Shimba Hills, Kilifi, Mt Kilimanjaro game reserve), Marsabit, Huri Hills, Marigat, Baragoi, Malaral (dryland areas in Kenya), Ngong and Karura forests (see Figure 1). Some of these areas also represent the high potential areas for the growth of *S. henningsii*. The predicted probability of suitable condition was generated on a scale from 0 to 1. Lowest suitability areas are represented by 0 while 1 represents the areas with highest suitability. The probability <0.23 are areas of unsuitability represented by the blue shade. The probabilities 0.23 to 0.38 are low suitable area represented by lighter green color, 0.38 to 0.69 are moderate or areas typical of those where *S. henningsii* is growing represented by green color, while areas of higher suitability >0.77 are represented by the red color (Figure 1).

The area under the curve (AUC) for the Receiver Operating Characteristic (ROC) is a widely used statistic to evaluate the performance of species distribution models (Jemênez-Valverde, 2012). It was developed for radar signal detection and ROC graphs were adapted in clinical medicine (Hanley and Meniel 1982, 1983; Zweig and Campbell, 1993; Pepe, 2000). The main advantage with ROC analysis is that it provides a single measure of model performance independent of any particular choice of threshold (Phillip *et al*., 2006). ROC has been applied in a variety of discriminatory problems in machine learning (Provost and Fawcett, 1997). In the last decade the AUC index has been accepted as a standard measure for assessing the accuracy of species distribution models (Elith, 2002; Fielding and Bell; 1997).

ROC is a graph obtained by plotting sensitivity on the y-axis and [1- specificity] on the x- axis for all possible thresholds. Sensitivity refers to the fraction of all pixels that fall in the areas suitable for *S. henningsii,* i.e. absence of omission error (true positive rate) while [1- specificity] is the fraction of the pixels that falls in the areas unsuitable for *S. henningsii* but predicted as suitable or (false positive fraction), i.e. commission error across all possible threshold between 0 and 1 (Jimênez-Valverde 2012). A model will be considered to perform better than chance if the curve lies above the diagonal of no discrimination that is if the AUC is higher than 0.5. (Jimênez-Valverde 2012)

ROC analysis was carried out both on the training data and test data records for *S. henningsii.* The AUC achieved on the training and the test data were 0.986 and 0.983 respectively. AUC has values that range from 0.5 –1.0. Values close to 0.50 indicate that the model is close to random and is a poor indicator, whereas a value of 1 indicates best run (Engler *et al*., 2004; Swets, 1988). AUC is a ranked approach for assessing model fit that determines the probability that a presence location will be ranked higher than a random background location, that is: AUC > 0.9 = very good; AUC 0.7-0.9 = good, and AUC < 0.7 = uninformative (Swets, 1988; Phillips *et al*., 2006). The results revealed that MaxEnt model performance was very good in predicting the potential suitable areas for *S. henningsii* in Kenya since the values for AUC for training and test records were greater than 0.9 (Figure 2).

Omission vs. predicted area analysis is a threshold-dependent binomial test of omission (known areas of presence predicted as absent). These thresholds are set up to maximize sensitivity and minimize specificity. Thus, the proportion of sites that are correctly classified as suitable can be compared to the proportion of the unsuitable sites

to determine the accuracy of the model (Baldwin, 2009). In this study, the '25' entered for random test percentage commanded the program to set aside 25% of the sample (presence records) for testing. This allows the program to conduct simple statistical analysis; much of which uses a threshold to make a binary prediction with suitable conditions predicted above and unsuitable below (Reddy *et al*., 2015). Figure 3 shows the omission rate and predicted areas as a function of the cumulative threshold. The omission rate was calculated on both training and test records (75% and 25% of the presence records, respectfully). The omission rate should be close to the predicted omission because of the definition of the cumulative threshold. Figure 3 shows how the training and test omission and predicted areas vary with the choice of cumulative threshold. A threshold of 0.5 that is a threshold, above which a species is more likely to be present, was used (Jimênez-Varverde and Lobo, 2007).

The omission on test records (sky blue line) is a good match to the predicted omission (black line), the omission rate for test data drawn from the MaxEnt distribution itself. The predicted omission is a straight line by definition of the cumulative output format. In some situations, the test omission line lies well below the predicted omission line while in some other situations the test omission line lies well above the predicted omission line. A common reason is that the test and training data are not independent (Phillips *et a*l., 2006). This showed that the MaxEnt model was significantly better than random in binomial test of omission and predicted area curve.

The 'jack-knife test' approach is a method that excludes one variable at a time when running the model. This provides information on the performance of each variable in the model in terms of how each variable is important in explaining the species distribution and how much unique information each variable provides (Phillips, 2009; Yost *et al*., 2008). The results also reveal that maximum temperature, precipitation and altitude contributed a lot in developing the model for the three jack-nife tests as compared to the minimum temperature [Figures 4(a)-(c)]. This indicates that *S. henningsii* prefers areas with high temperature and moist conditions. It also conforms with the potential areas predicted on the map by MaxEnt. These areas are found in the dry land areas of Kenya with high temperature and also experience short term rainy seasons [Figures 4(b)-(d)]. The majority of the potential suitability areas for *S. henningsii* are found in low altitude areas and the probability of presence declines with increasing altitude [Fig 4(a)]. MaxEnt response curves also show that the variables that strongly contributed to the model prediction were maximum temperature, altitude and precipitation [Figures 5(a)-(d)].

CONCLUSIONS

This study presents the first predicted potential habitat distribution map for *S. henningsii*. The sensitivity versus [1-specificity] graph indicated that the MaxEnt model had 'good' predictive accuracy (AUC=0.986 for training data, and 0.983 for test data) in terms of *Strychnos henningsi*i potential distribution areas in Kenya. The potential habitat map can help in planning land use management around its existing populations and set priorities to restore its natural habitat for more effective conservation. This approach can be used to quantify and map the potential distribution habitats for other threatened and endangered species to guide their conservation and restoration efforts.

ACKNOWLEDGEMENTS

We wish to thank the Institute of Research Production and Extension in Jomo Kenyatta University of Agriculture and Technology for funding this research work. We are also grateful to Mr. Kiama and Ms. Cynthia Kanyingi from GIS (Geographical Information System) Department in Kenya Forestry Research Institution for their assistance in data analysis and interpretation in species modeling. The corresponding author (Mrs. Kuria) also thanks her supervisors for their continued support in this research work.

REFERENCES

Amaludin Nazahatul Anis (2012). Modeling tropical tree species for improved conservation assessments of dipterocarps in Malaysia and Philippines using maxent Biotropica **44** (5):649-657.

Anderson RP, Lew D, Peterson AT, (2003). Evaluating predictive models of species' distributions: Criteria for selecting optimal models, Ecological Modeling **162:**211–32.

Bair ES (1994). Model (in) validation - A view from courtroom, Ground Water **32**: 530-1.

Baldwin RA (2009). Use of maximum Entropy Modeling in wildlife research Entropy **11**: 854-866 doi; 10.3390/e11040854

Beentje HJ (1994). Kenya Trees, Shrub and Lianas, National Museums of Kenya, Nairobi, Pg 465.

Benito BM, Martınez-Ortega MM, Munoz LM, Lorite J and Penas J (2009). Assessing extinction-risk of endangered plants using species distribution models: A case study of habitat depletion caused by the spread of greenhouses, Biodiversity Conservation doi: 10.1007/s10531-0099604.

Berry PM, Dawson TP, Harrison PA and Pearson RG (2002). Modeling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland, Global Ecological Biogeography **11**: 453–62.

Carpenter G, Gillison AN and Winter J (1993). DOMAIN: A flexible modeling procedure for mapping potential distributions of plants and animals, Biodiversity Conservation **2**: 667–80.

Chapman L, Johns T and Mahunnah RLA (1997). Sapon-like in vitro characteristic of extracts from selected non-nutrient wild plant food additive used by the Maasai in meat and milk based soups, Ecology of food and nutrition **36** (1): 1-22.

Dormann CF, Schymanski, SJ, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Romermann C, Schroder B and Singer A (2012). Correlation and process in species distribution models: Bridging a Dichotomy, Journal of Biogeography **39**: 2119–31.

Elith, J. (2002). Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. In: Quantitative Methods for Conservation Biology (Ferson S., Burgman M., Eds.). New York, Springer, pp. 39-58

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Takanawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS and Zimmermann NE (2006). Novel methods improve prediction of species' distributions from occurrence data, Ecography **29:** 129-51.

Elith J and Leathwick JR (2009). Species distribution models: Ecological explanation and prediction across space and time, Annual Review: Ecological Evolution System **40:** 677–97.

Elith J, Phillips SJ, Hastle T, Dudik M, Chee YE and Yates CJ (2011). A statistical explanation of Maxent for ecologists, Diversity Distribution **17**: 43–57.

Engler R, Guisan A and Rechsteiner L (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data, Journal of Applied Ecology **41**, 263–74.

Fielding AH and Bell JF (2007). A review of methods for assessment of prediction errors in conservation presence/absence models, Environmental Conservation **24:**38–49.

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

Gachathi M (2007). Kikuyu Botanical dictionary. A guide to plant names uses and culture values, 2nd Ed Tropical Botany, Pgs 143, 212.

Gaston KJ (1996). Species richness: measure and measurement. Biodiversity: A biology of numbers and difference (ed. K.J. Gaston), Blackwell Science, Oxford pp. 77–113.

Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loiselle BA and the NCEAS (2008). Predicting Species Distributions Working Group, the influence of spatial errors in species occurrence data used in distribution models, Journal of Applied Ecology **45:** 239–47.

Guisan A, Edwards TC and Hastie T (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene, Ecological Modeling **157:** 89–100.

Hanley JA and McNeil BJ (1982) The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve, Radiology **143:** 29-36.

Hanley JA, McNeil, BJ (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology **148:** 839–843.

-Hernandez PA, Graham CH, Master LL and Albert L (2006). The effect of sample size and species characteristics on performance of different species distribution modeling method, Ecography **29**: 773–85.

Hijmans RJ, Schreuder M, de la Cruz, J and Guarino L (1999). Using GIS to check co-ordinates of germplasm accessions, Genetic Resource. Crop Evolution **46:** 291–6.

Hutchings A (1989). *A survey and analysis of traditional medicinal plants as used by the Zulu, Xhosa and Sotho,* Bothalia **19:** 111–23.

Hutchings A (1996). Zulu medicinal plants, Natural University Press. Pietermaritzburgs.

Jaryan V, Uniyal SK, Gupta RC and Singh RD (2013). Extent of occurrence and area of occupancy of tallow tree (Sapium sebiferum) using the red list criteria for documenting invasive species expanse, Natl. Academic. Science Letter **36:** 85–91.

Jimênez-Valverde A and Lobo JM (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence, Acta Oecologica-International Journal of Ecology **31:** 361-9.

Jimênez-Valverde A (2012). Insight into the area under the Receiver Operating Characteristic Curve (AUC) as a discrimination measure in species distribution modeling. Global Ecology and Biogeography 21: 498-507.

Kearney MR, Wintle BA and Porter WP (2010). Correlative and mechanistic models of species distribution provide

congruent forecasts under climate change, Conservation Letters **3:** 203–13.

Kenyaneem.com (2015). Muteta Soup, *a* Powerful Fever Reliever, Neem Foundation Kenya. Retrieved from http://www.kenyaneem.com/?q=muteta- soup-powerful-fever- reliever.

Kumar S, Stohlgren TJ (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. Journal of ecology and Natural Environment **1** (4):94-98.

Kumar S, Noven LG and Tee WL (2014). Evaluating correlative and mechanistic niche models for assessing the risk of pest establishment, Ecosphere **5**(7):86.

Kuria MW, Njenga PK and Ngumi VW (2012). Ethnobotanical studies of *Strychnos henningsii* in five (Gilg.) natural habitats in Kenya, International Journal of Medicinal Plant Research ISSN: 2169-303X **1**(6):63-74, November, 2012. Available online at: www.internationalscholarsjournals.org

Lozier JD and Mills. NJ (2011). Predicting the potential invasive range of light brown apple moth (Epiphyas postvittana) using biologically informed and correlative species distribution models, Biological Invasions **13**:2409–21.

-Maundu P and Tengäs T (2005). Useful Trees and Shrubs of Kenya, Technical handbook No. 35. Nairobi, Kenya: World Agroforestry Centre-Eastern and Central Africa Regional Program (ICRAF-ECA). Pp 400.

Maundu PM, Ngugi GW and Kabuye CHS (1999). Traditional food plants of Kenya. Kenya Resource Centre for Indigenous Knowledge *(KENRIK),* National Museums of Kenya, Nairobi, Pg 216.

Meyer EM, Peterson AT, Servin JI and Kliff LF (2006). Ecological niche modeling and prioritizing areas for species reintroductions, Oryx **40**(4):411–8.

Musila W, Kisangau DP and Muema J (2004). Conservation status and use of medicinal plants by traditional medicinal practitioners in Machakos district, Kenya - Traditional aspects of health and wellness, Proceedings of Inter-Institutional Consortium for indigenous knowledge - Transforming the academy, Pennsylvania pg. 27-9.

Ni WL, Li ZH, Chen HJ. Wan FH, Qu WW, Zhang Z and Kriticos DJ (2012). Including climate change in pest risk assessment: The peach fruit fly, *Bactrocera zonata* (Diptera: Tephritidae), Bulletin of Entomological Research **102**:173–83.

Oreskes N (1998). Evaluation (not validation) of quantitative models, Environmental Health Perspective **106:**1453-60.

Ortega-Huerta MA and Peterson AT (2008). Modeling ecological niches and predicting geographic distributions: A test of six presence-only methods, Revista Mexicana De Biodiversidad **79:** 205- 16.

Oyedemi SO, Bradley G and Afolayan AJ (2009). Ethnobotanical survey of medicinal plants used for the management of Diabetes mellitus in the Nkonkobe municipality of South *Africa,* Journal of Medicinal Plants Research **3**(12):1040-4.

Palgrave KC (1988). Trees of South Africa, 5th Ed, Struik Publishers, Cape Town Pg 765-766.

Papes M and Gaubert P (2007). Modeling ecological niches from low numbers of occurrences; Assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents, Diversity Distribution **13:** 890-902.

Pearson RG, Raxworthy CJ, Nakamura M and Peterson AT (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar, Journal of Biogeography **34:**102-17.

Pepe MS (2000). Receiver operating characteristic methodology, Journal of the American Statistical Association **95:**308–11.

Peterson AT, Papes M and Kluza DA (2003). Predicting the potential invasive distributions of four alien plant species in North America, Weed Science **51:** 863–8.

Peterson AT; Papeş M; Eaton M (2007) Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. Ecography 30: 550–560.

Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Meyer E, Nakamura M and Araujo MB (2011). Ecological niches and geographic distributions, Princeton University Press, Princeton, New Jersey, USA.

Phillips SJ (2009). A brief tutorial on Maxent, versions: 3.3.1. Available online: http://www.cs.princeton.edu/~schapire/maxent/

Phillips SJ, Anderson RP and Schapire RE (2006). Maximum entropy modeling of species geographic distributions, Ecological Modeling **190:** 231–59.

Philips SJ and Dudik M (2008). Modeling of species distributions with maxEnt: New extensions and a comprehensive evaluation, Ecography **31**: 16175.

Provost FJ, Fawcett T (1997). Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. Knowledge Discovery and Data Mining. ACM Press, New York, 43–48.

Pujol J (1993). Naturafrica-the herbalists' handbook, Jean Pujol Natural Healer; Foundation, Durban

Reddy MT, Begun H, Sunil N, Pandravada SR, Sirvaraj N and Kumar S (2015). Mapping the climate suitability using Maxent approach for Ceylon spinach *(Basella alba L.*) cultivation in India, Journal of Agricultural Science **10** (2): 87 – 97.

Rykiel Jr. EJ (1996). Testing ecological models: The meaning of validation, Ecological Modeling **90:** 229-44.

Sanchez AC, Osborne PE, Haq N (2010). Identifying the global potential for baobab trees cultivation using Ecological niche modeling. Agroforestry systems 80: 90 doi:10. 1007/S10457-010-9282-2.

Sargent RG (2001). Verification, validation and accreditation of simulation models, in: Peters BA, Smith JS, Medeiros DJ, Rohrer MW (Eds.), Proceedings of the 2001 Winter Simulation Conference, December 10-13, Arlington, VA, USA, 106-14.

Schmeltzer GH (2008). Plant resource of Tropical Africa (11): Medicinal Plant 1, PROTA Foundation Backhuys, CTA Wageningen, Netherlands pg. 570-1.

Swets JA (1988). Measuring the accuracy of diagnostic systems, Science **240:**1285–93.

Watt MP and Breyer NG (1962). Medicinal and poisonous plants of Southern and Eastern Africa, Living stone, Edinburgh

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A and NCEAS (2008). Effects of sample size on the performance of species distribution models, Diversity Distribution **14:**763–73.

Yang X, Kushwaha Q, Saran SPS, Xu SJ and Roy PS (2013). MaxEnt modeling for predicting the potential distribution of medicinal plant Justicia adhatoda L. in lesser Himalayan foothills, Ecological Engineering **51:**83–7.

Yost AC, Petersen SL, Gregg M and Miller R (2008). Predictive modeling and mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using Maximum Entropy and a long-term dataset from Southern Oregon, Ecological Informatics **3:**375–86.

Zweig MH and Campbell G (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine, Clinical Chemistry **39:**561–77.

ABBREVIATIONS

Arc GIS: Aeronautical Reconnaissance Coverage Geographic Information System
ASCII: American Standard Code for Information Interchange
AUC: Area Under the Curve
BIL: Band Interleaved by line file format
BioClim: BioClimatic variable
BRT: Boosted Regression Tree
CSV: Comma-Separated Values
DOMAIN: Domain
EPSG4326: European petroleum Survey group 4326
GARP: Genetic Algorithm for Rule-set Production
GLM: Generalized Linearized Model
Gtiff: GeoTag Image File format
MARS: Multivariate Adaptive Regression Splines
MaxEnt: Maximum Entropy
QGIS: Quantum Geographical Information System
ROC: Receiver Operating Characteristic
SDM: Species Distributed Model
WGS684: World Geodetic System 1984
WorldClim: World Climatology
Wrap: Warwick Research Archive Portal

ABOUT THE AUTHORS

KURIA MW is a Lecturer at Kabarak University and a PhD Student at Jomo Kenyatta University of Agriculture and Technology in Nairobi, Kenya. Her professional interests are in molecular biology and conservation of indigenous plant species.

NGUMI VW (DVC, APD) is a Professor at Jomo Kenyatta University of Agriculture and Technology. Her professional interests are to build capacity through teaching and research in her areas of interest which are plant physiology, plant tissue culture and conservation botany.

NJENGA PK, PhD is a Senior Lecturer, Botany Department, Jomo Kenyatta University of Agriculture and Technology. His professional interests include plant biotechnology, and sustainable utilization and conservation of plant genetic resources.

WANGAI LN is a Professor and Head of the School of Health Sciences, Kirinyaga University College. Her professional interests are in molecular biology and conservation of genetic resources.