

RESEARCH

Open Access



MaxEnt-based modeling of suitable habitat for rehabilitation of *Podocarpus* forest at landscape-scale

Birhane Gebrehiwot Tesfamariam^{1,2*} , Berhan Gessesse² and Farid Melgani³

Abstract

Modeling the current distribution and predicting suitable habitats of threatened species support proper planning processes for conservation and restoration. The aim of this study was thus to model the actual distribution and predict environmentally suitable habitats for *Podocarpus falcatus*, a locally threatened native tree species in Ethiopia. To realize this objective, species' presence samples, BIOCLIMATIC, and topographic predictors were combined to run a MaxEnt model. Finally, a model-generated habitat suitability map was produced with AUC accuracy of 0.783. Among the variables used for modeling, elevation range was found to be a key predictor of *Podocarpus* distribution, followed by precipitation of the driest quarter and isothermality. An extensive area (> 48%) of the studied landscape has been predicted to be environmentally suitable for the target species. However, only a small portion open-land area is practically available for rehabilitation since the area has been intensively cultivated to support the densely inhabited population. Therefore, potential areas for a small-scale plantation of *Podocarpus* trees remain to be pocket sites in religious places and around farmers' homesteads. So far, many farmers in this area have demonstrated a successful experience of growing this degraded native tree species. Thus, encouraging privately owned small-scale plantations could enhance rehabilitation and more sustainable conservation of the locally threatened native tree species.

Keywords: Ethiopia, MaxEnt, *Podocarpus falcatus*, Rehabilitation, Species' distribution models, Suitable habitat prediction

Introduction

Podocarpus falcatus is a native tree species to African countries including Ethiopia, while it is exotic to India (Orwa et al. 2009). This tree is known by the scientific and common names of *Podocarpus falcatus*, *Afrocarpus falcatus*, African yellow wood, and 'Zigba' in Amharic (Bekele 2007; Negash 2010; Doffana 2014). The general ecological range of *P. falcatus* was claimed to be 1550–3000 m above sea level (masl) altitude, 13–20 °C annual mean temperature, and 1200–1800 mm annual mean rainfall, over well-drained mainly humus-rich sandy soils

(Gilman and Watson 1994; Orwa et al. 2009). However, moderately different elevation ranges that limit the distribution of the species were reported in different geographic regions. While its ecological limit descends to low altitudes from subtropical/tropical moist lowlands to subtropical/tropical moist montane in South Africa ranging from 10 to 1700 masl (Farjon 2013), it was claimed to be dominantly found in Afromontane forest ecosystems with altitudinal ranges of 1550–2800 masl in Ethiopian (Tesfaye et al. 2002; Negash 2010; Teketay 2011).

This native tree is one of the most important tree species in Ethiopia, economically and ecologically, offering pervasive ecosystem functions, a source of food and shelter for wildlife, and yields high quality timber (Vivero et al. 2005; Bekele 2007; Teketay 2011). However, because of its multipurpose use and excellent softwood product,

*Correspondence: gebrehiwotb@yahoo.com; mikirodi24@gmail.com

¹ Department of Land Administration and Surveying, Dilla University, P. O. Box 419, Dilla, Ethiopia

Full list of author information is available at the end of the article

it has been under intensive exploitation for several decades, both legally and illegally from the natural forests of Ethiopia (Vivero et al. 2005; Teketay 2011). Some pieces of evidences indicate that this tree species was unsustainably exploited beginning in the 1920s from the dry Afromontane forests of Ethiopia (Teketay 1992). Following such increased overexploitation, it has continued to decline and is subsequently classified under the category of serious conservation concern within the country (Vivero et al. 2005; Teketay 2011).

After the species has already become rare or significantly declined, a banning proclamation was ratified by the Transitional Government of Ethiopia (TGE 1994) to prohibit further cutting of its remnants. However, it was reported that despite the ratification of such a legal prohibition, illegal felling of the tree has continued unrestricted, causing a more threatening decline (Teketay 2011). In addition, there are neither large-scale plantations nor known planned programs to rehabilitate this locally threatened native tree species (Teketay 2011). On the other hand, in spite of such a continuous shrinking in the extent of *Podocarps* forest cover, quantitative empirical information on the current distribution of its remnants and potentially suitable habitats is lacking. So that there is a need for a detailed survey of its remaining populations to estimate its present distribution and also predict its suitable habitat or potential distributions, which may help to establish the foundation for future rehabilitation and sustainable conservation plans.

While knowledge on the geographic distribution and suitable habitats of threatened species is crucial for conservation planning, detailed data on species' actual and potential distribution is usually lacking, as collecting such data is costly and labor-intensive. Conservationists thus largely rely on predictive models for identifying the location and patterns of species distribution for developing future conservation plans (Elith et al. 2011; Phillips et al. 2006). When reliable species' data are available, species distribution models (SDMs) enable to overcome the aforementioned constraints, as SDMs predict species' geographic distribution by establishing a relationship between species' presence sites and the environmental conditions prevailing at these locations (Phillips et al. 2006; Kumar and Stohlgren 2009; Elith et al. 2011; Eastman 2012).

Species distribution modelling (SDM) has stimulated the development of numerous statistical models having broad application potentials in biogeography, conservation biology, and ecological sciences (Elith et al. 2011; Castilho 2015). Several SDM algorithms are available currently, depending on the nature of species' data at hand (Eastman 2012). The relevant species' data are categorized as presence-absence, presence-only, and

abundance data. Presence-only data consists of species' occurrence samples, where the target species is known to inhabit; presence-absence data includes both samples of species' presence and absence locations; while abundance data indicates the numbers of species found at each site per unit area. Since presence-only records are the most readily available type of species' data, obtained from field-work or museum collections (example, from the Global Biodiversity Information Facility: www.gbif.org), modeling algorithms that require presence-only data are preferred more often (Phillips et al. 2006; Elith et al. 2011; Castilho 2015). As a result, these modeling approaches have been extensively studied and proven to be useful for modelling species' distributions that greatly enhance conservation planning programs. Among the presence-only modeling approaches, the maximum entropy (MaxEnt) algorithm has received increased attention, which has been widely used attributed to its comparatively better prediction performance (Phillips et al. 2004, 2006; Elith et al. 2006; Kumar and Stohlgren 2009; Girma et al. 2015; Abrha et al. 2018). MaxEnt has demonstrated better predictive performance than other presence-only species distribution models, specifically the Genetic Algorithm for Rule-Set Prediction (GARP) (Phillips et al. 2006; Elith et al. 2006). For that reason, this algorithm was chosen to characterize the actual and potential distribution of *Podocarpus falcatus* at a catchment level. This study offers valuable insights on the distribution and suitable habitats of *Podocarpus falcatus*, which could support future rehabilitation plans.

Materials and methods

Study area

This study was conducted in part of the Southeastern Escarpment of the Main Ethiopian Rift, occupying a geographic location between 6.52° and 6.94° north and 38.24° to 38.64° east. As depicted by the one-arc-second Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM Fig. 1), it is characterized by a wider altitudinal variation, ranging from 1397 to 3213 m above sea level. Since the catchment is dominantly occupied by highland areas, it exhibits a cool-tropical type of climate (Moat et al., 2017). This catchment is contained in the Gidabo river basin, characterized by average monthly temperature variations of 11.5 to 25 °C (Mechal et al., 2016). The area receives a bimodal rainfall pattern with peaks in April and May during the short rainy season, and in September and October during the main rainy season. Based on a bias-corrected CHIRPS rainfall, the climatological annual mean rainfall of this catchment calculated over 1981–2010 varies from 1078.85 mm in the downstream to 1430.71 mm in the upstream areas (Teskamariam et al.

2019). The marked altitudinal variations and associated physical features across the landscape have given rise to the spatial variation in rainfall distribution over the catchment.

Since this area comprises one of the known agro-ecological zones in Ethiopia, where *P. falcatus* grows naturally (Vivero et al. 2005; Teketay 2011), it was considered appropriate and chosen for modeling the present distribution and suitable habitats of *P. falcatus*. As such the watershed contains highly scattered remnant trees and a pocket *Podocarpus* forest, which were found useful for the species' distribution modelling and prediction of its suitable habitats in this area.

Species' occurrence data

Species' presence points instead of presence-absence data were collected, as the MaxEnt modeling algorithm applied in this study needs presence-only samples (Phillips et al. 2006). To properly estimate the geographic distribution along with the accompanying environmental conditions considered suitable for the study species, the occurrence data should represent a random sample of suitable conditions to the possible degree (Phillips et al. 2017). Because species' occurrence localities allow establishing the relationship between the species' geographic distribution and associated environmental conditions across the rest of the study area (Phillips et al. 2006; Elith et al. 2011). To that end, a field survey was conducted over the study catchment to identify and geographically record presence localities of the target species. Moving along transects throughout the landscape, geographic locations (longitude, latitude) and elevation of presence samples were recorded using a hand-held Global Positioning System (GPS receiver), where a total of 76 sample points were collected (Fig. 1). In the study watershed, the presence localities of the target species were found concentrated between 1625 and 2120 masl over the surveyed area. While no *Podocarpus* tree was found below 1625 masl in the watershed, sample collection over a mountain with an elevation higher than 2120 masl was constrained by physical inaccessibility (i.e., topographic barriers). Nevertheless, the size of this mountainous area is negligible compared to the extent of the surveyed area. On the other hand, except the protected pocket *Podocarpus* forest located at a hilly terrain that we surveyed and sampled it, the majority of the presence samples have been concentrated around homesteads of private holdings and religious places.

Selection of environmental predictors

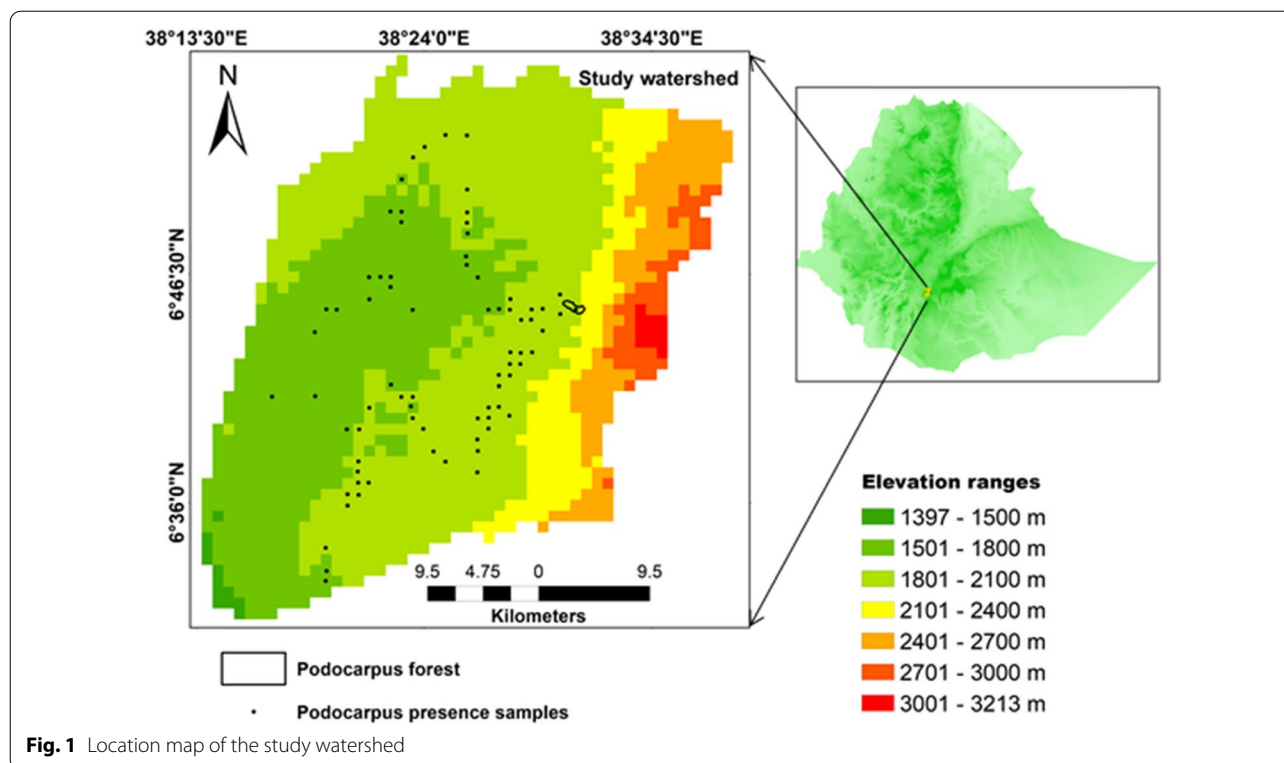
Environmental predictors impose constraints on the geographic distribution of the species and represent our ecological assumption that these features characterize the

major environmental factors that limit the spatial distribution of the species (Phillips et al. 2006). Thus, the ultimate goal of species distribution and habitat modeling is to estimate the inhabited and potentially suitable environmental conditions for the target species. This suggests the need for a careful selection of relevant environmental variables in order to generate reliable predictions (Phillips et al. 2017). Relevant variables representing various environmental dimensions were selected as potential predictors of the actual and potential distribution of the species. These comprise topographic variables (Elevation, slope, and aspect), soil texture, geological map, and the WorldClim bioclimatic variables which are among other potential applications designed for species distribution modelling (<https://www.worldclim.org/data>). These bioclimatic variables (Table 1) represent near-current climatic conditions averaged over 30 years from 1970 to 2000 (Hijmans et al. 2005; Fick and Hijmans 2017). In addition to their biological relevance to plant species distributions, these environmental variables (covariates) were chosen in accordance with related previous studies (Hijmans et al. 2005; Girma et al. 2015; Fick and Hijmans 2017; Abrha et al. 2018).

The raster-based bioclimatic variables were originally derived from monthly temperature and rainfall datasets that include the driest, wettest, coldest, and warmest extremes as well as intermediate and mean precipitation and temperature values representing biologically meaningful climate variables (Hijmans et al. 2005; Fick and Hijmans 2017). However, since the WorldClim precipitation data has been recognized to be less reliable over areas with sparse rain gauge stations, which characterizes the present study area, the precipitation variables used in this study were derived from a bias-corrected CHIRPS rainfall product, averaged over 1981–2010 (Tesfamariam et al. 2019). To bring all the environmental layers to the same resolution, the CHIRPS rainfall grid was down-scaled, while the USGS SRTM DEM was up-scaled to 1-km pixel size. The DEM was then used to derive three topographic variables (i.e., elevation, slope, and aspect).

Methods of reducing correlated and less important variables

A multicollinearity test using a variance inflation factor (VIF) and jackknife test of variable importance were employed as methods of feature reduction in order to filter out highly correlated and less important variables. Reduction or removal of (multi) collinear and less important variables, while retaining the less correlated and more important environmental predictors, helps to get more reliable results and improve the prediction accuracy of the model (Girma et al. 2015). Multicollinearity generally occurs when high correlations exist between



two or more predictor variables. When two independent variables contain redundant information largely, little is gained by using both in the model; rather it leads to unstable and unreliable (biased) predictions, as it tends to increase the variances of regression coefficients. Thus, the solution is to keep only one of the two highly correlated independent variables in the model (Yoo et al. 2014; Marco and Júnior 2018). Accordingly, after a multicollinearity test, those variables having $VIF > 5$ were removed because of their strong correlation with other variables. These include eight continuous variables consisting of BIO1, BIO5, BIO6, BIO8, BIO9, BIO10, BIO11, and BIO16.

Additionally, other non-collinear variables were reduced due to their little contribution (low gain) to the MaxEnt model based on a jackknife importance test, which consists of BIO4, BIO12, BIO13, BIO14, BIO15, aspect, geology, and soil texture. Since none of these variables contains useful information that is not contained in the other remaining variables, omitting each variable one by one did not decrease the training and test gains of the model. Finally, the MaxEnt model was run using 8 less correlated and more important variables derived from temperature (BIO2, BIO3, BIO7), precipitation (BIO17–BIO19), and topography (elevation, and slope). Measured in their relative contribution and jackknife importance test, these variables were found to be more relevant in

Table 1 Description of Bioclimatic variables preliminarily selected as predictors

Label	Variable	Unit
BIO1	Annual mean temperature	Degrees Celsius
BIO2	Mean diurnal range [mean of monthly (max-temp – min-temp)]	Degrees Celsius
BIO3	Isothermality (BIO2/BIO7) × 100	Percent
BIO4	Temperature seasonality (standard deviation) × 100	Percent
BIO5	Max temperature of warmest month	Degrees Celsius
BIO6	Min temperature of coldest month	Degrees Celsius
BIO7	Annual temperature range (BIO5–BIO6)	Degrees Celsius
BIO8	Mean temperature of wettest quarter	Degrees Celsius
BIO9	Mean temperature of driest quarter	Degrees Celsius
BIO10	Mean temperature of warmest quarter	Degrees Celsius
BIO11	Mean temperature of coldest quarter	Degrees Celsius
BIO12	Annual precipitation	Millimeters
BIO13	Precipitation of wettest month	Millimeters
BIO14	Precipitation of driest month	Millimeters
BIO15	Precipitation seasonality (coefficient of variation)	Millimeters
BIO16	Precipitation of wettest quarter	Millimeters
BIO17	Precipitation of driest quarter	Millimeters
BIO18	Precipitation of warmest quarter	Millimeters
BIO19	Precipitation of coldest quarter	Millimeters

Table 2 Environmental variables used as predictors (selected after multicollinearity and jackknife importance test)

Label	Variable	Unit	Description
BIO2	Mean diurnal range [mean of monthly (max-temp – min-temp)]	Degrees Celsius	Helps to assess species distribution in relation to diurnal temperature fluctuations
BIO3	Isothermality (BIO2/BIO7) × 100	Percent	Evaluates how large the day-to-night temperatures oscillates relative to annual temperature range
BIO7	Annual temperature range (Max temperature of warmest month minus Min temperature of coldest month)	Degrees Celsius	Enables to assess whether ranges of extreme temperature conditions affect the species' distribution
BIO17	Precipitation of driest quarter	mm	Describes how precipitation amount during the driest quarter of the year may affect distribution of the species
BIO18	Precipitation of warmest quarter	mm	Helps to understand how the amount of precipitation during the warmest quarter of the year may affect distribution of the species
BIO19	Precipitation of coldest quarter	mm	Enables to characterize how the amount of precipitation during the coldest quarter may affect distribution of the species
–	Elevation	Meters	Identifies the altitudinal range where the species could inhabit
–	Slope	Degree	Describes whether the species' distribution is affected by slope differences

Table 3 Relative contribution of the variables in predicting species' distribution

Labels	Variable	% contribution
Elevation	Elevation	37.5
BIO17	Precipitation of driest quarter	30.8
BIO3	Isothermality (mean diurnal range/annual temperature range) × 100	13.5
BIO19	Precipitation of coldest quarter	7.9
BIO7	Annual temperature range (Max T° of warmest month – Min T° of coldest month)	4.4
BIO2	Mean diurnal range [mean of monthly (max-temp – min-temp)]	2.9
Slope	Slope	2.5
BIO18	Precipitation of warmest quarter	0.6

T° temperature

characterizing the environmental conditions related to the species' distribution (Table 2).

Description of maximum entropy (MaxEnt) algorithm

Maximum entropy (MaxEnt) is an efficient modeling algorithm for making predictions from incomplete information that is particularly designed for applications demanding presence-only data on species' distribution (Phillips et al. 2006). It determines how the environmental conditions at the species' occurrence localities relate to the environmental conditions across the rest of the study area. The central idea in MaxEnt is to search for a probability distribution having a maximum entropy (most spread out), subject to the constraints imposed by the available information on species' presences and the associated environmental conditions across the study area (Phillips et al. 2006; Elith et al. 2011). MaxEnt uses a deterministic sequential-update algorithm that iteratively picks and adjusts weights of predictors, which is guaranteed to converge to the maximum entropy probability distribution (Phillips et al. 2004, 2006).

MaxEnt employs a probability distribution that belongs to the family of Gibbs distributions (exponential distributions), where these probability distributions are derived from a set of features f_1, \dots, f_n , parameterized by weights $\lambda_1, \dots, \lambda_n$, (Phillips and Dudík 2008). The MaxEnt probability distribution is guaranteed to converge to the best Gibbs distribution, as long as the occurrence sites are drawn independently at random (Phillips et al. 2006; Phillips and Dudík 2008).

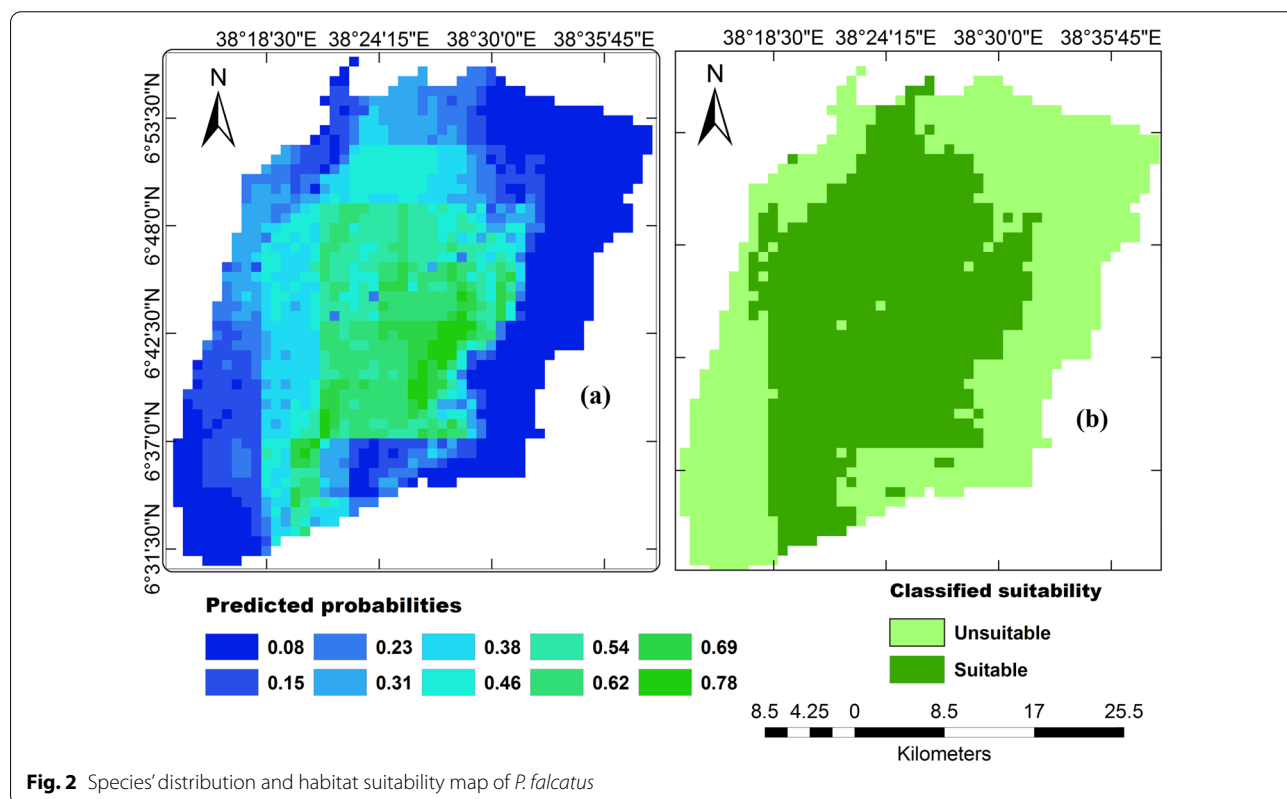
In practice, a regularized probability distribution is recommended to reduce overfitting to the training data. Regularization is a common modern approach in general and not specific to MaxEnt, which is a way of penalizing the coefficients (weights of features) to produce a simpler model that balances fit and complexity to enhance the generalization of independent test data. The probability assigned to each pixel in MaxEnt is typically very small, as the total values must sum to 1 over all the pixels in the study area. Thus, the probability distribution is displayed in terms of 'gain', computed as the log of the probability of the presence samples,

minus a constant that makes the uniform distribution has a zero gain (Phillips and Dudi'k 2008; Elith et al. 2011). MaxEnt generates a probability distribution over the pixels in the grid, starting from a uniform distribution (i.e., gain starts at zero) and repeatedly improves the fit to the data. At the end of the run, the gain indicates how closely the model is concentrated around the presence samples. For instance, if the gain is 2, it means that the average likelihood of the presence samples is $\exp(2) \approx 7.4$ times higher than that of a random background pixel (Phillips 2017). The gain increases iteration by iteration until the change from one iteration to the next falls below the convergence threshold, or until maximum iterations have been performed (Phillips et al. 2006; Phillips and Dudi'k 2008).

Unlike generalized linear models (GLM) and generalized additive (GAM) models, MaxEnt, requires species' presence records and its output involves a natural probabilistic interpretation with a smooth gradation ranging from most to least suitable environmental conditions (Phillips et al. 2004, 2006). Still, where a binary prediction is desired, MaxEnt is flexible in the choice of a threshold (Phillips et al. 2006). Conversely, in no sense are pixels without species' records interpreted as absences in the probability distribution of MaxEnt models. Because the current species' distribution and

its suitable habitat depend on the interplay of various factors, including knowledge on past disturbances, dispersal limitations, and biotic interactions; as well as the scale of analysis (Elith et al. 2011). Above all, previous comprehensive studies have substantiated the outperforming predictive power of MaxEnt compared to the commonly used presence-only models of Genetic Algorithm for Rule-Set Prediction (GARP) (Phillips et al. 2006; Elith et al. 2006).

Therefore, the latest version of MaxEnt software (version 3.4.1) was used in the present study to model the current distribution and suitable habitats for *P. falcatus* tree species. From the three types of replicate run options in the MaxEnt software (Subsample, Crossvalidate, and Bootstrap), the 'subsample' option was chosen for its convenience to randomly divide the species' presence sample into a user-defined proportion of training and test samples (Phillips 2017). The 76 points of total presence samples were thus partitioned to 75% and 25%, for model training and testing, respectively. Applying a default regularization multiplier (i.e., 1) to reduce overfitting, the MaxEnt logistic regression was run with 100 replicates and 500 iterations to allow the model has a sufficient time for convergence (Girma et al. 2015; Abrha et al. 2018). Finally, the species distribution and habitat suitability map was produced by averaging the 100 replicate runs.



Accuracy assessment of model performance

The receiver operating characteristics (ROC) area under the curve (AUC), a widely used and robust approach of model evaluation was employed to assess the accuracy of the model prediction (Pearson et al. 2004; Phillips et al. 2006; Aguirre-Gutiérrez et al. 2013; Girma et al. 2015). This was performed with independent test data, consisting of 25% (19) of the total samples (Girma et al. 2015; Abrha et al. 2018). AUC characterizes the performance of a model at all possible levels, independent of any particular threshold. It quantifies how a random positive instance and a random negative instance are correctly ordered by the classifier (with random ordering in the case of ties), in which a perfect classifier attains an AUC of 1, while the AUC value of 0.5 corresponds to a random prediction (Phillips et al. 2006; Girma et al. 2015).

For all possible thresholds, AUC is depicted by plotting sensitivity on the y-axis and 1-specificity on the x-axis (Phillips et al. 2006; Girma et al. 2015). Sensitivity shows the proportion of test localities correctly predicted present (1-extrinsic omission rate), while 1-specificity corresponds to the proportion of all pixels predicted to have suitable conditions for the species. In other words, sensitivity represents a true positive rate whereas 1-specificity represents a false positive rate (commission error). Contrastingly, when presence-only data is used (as the case in MaxEnt), ROC curve seems to be inapplicable. Because without absences, there will be no source of negative instances with which to measure specificity.

However, this can be solved by using a different classification scheme to distinguish presences from random (background), instead of presences from absences. Accordingly, to use the ROC curve with presence-only data, we have to label all the pixels with no occurrence localities as negative instances, even if they support good environmental conditions for the species. The maximum achievable AUC for models using presence-only data is, therefore, less than one (Wiley et al. 2003; Phillips et al. 2006), and is also smaller for species with a broader range of environmental conditions (Phillips et al. 2004). As a result, since the MaxEnt program uses presence-only data, the value 1-specificity represents the ‘fractional predicted area’ (fraction of the total study area predicted present), instead of the more standard commission rate (fraction of absences predicted present) (Phillips et al. 2006; Phillips 2017).

Results and discussion

Prediction of species’ distribution and habitat suitability

The MaxEnt model generated species’ distribution map of *P. falcatus* was produced at a 1-km spatial resolution, maintaining the pixel size of the input data. To classify the species’ distribution map into suitability classes, we have used the minimum predicted value assigned to any of the presence samples, following a related approach by Phillips et al. (2006). This minimum predicted probability value (i.e., >0.26) was obtained by averaging 100 replicate MaxEnt runs. Indeed, the areas labeled as suitable

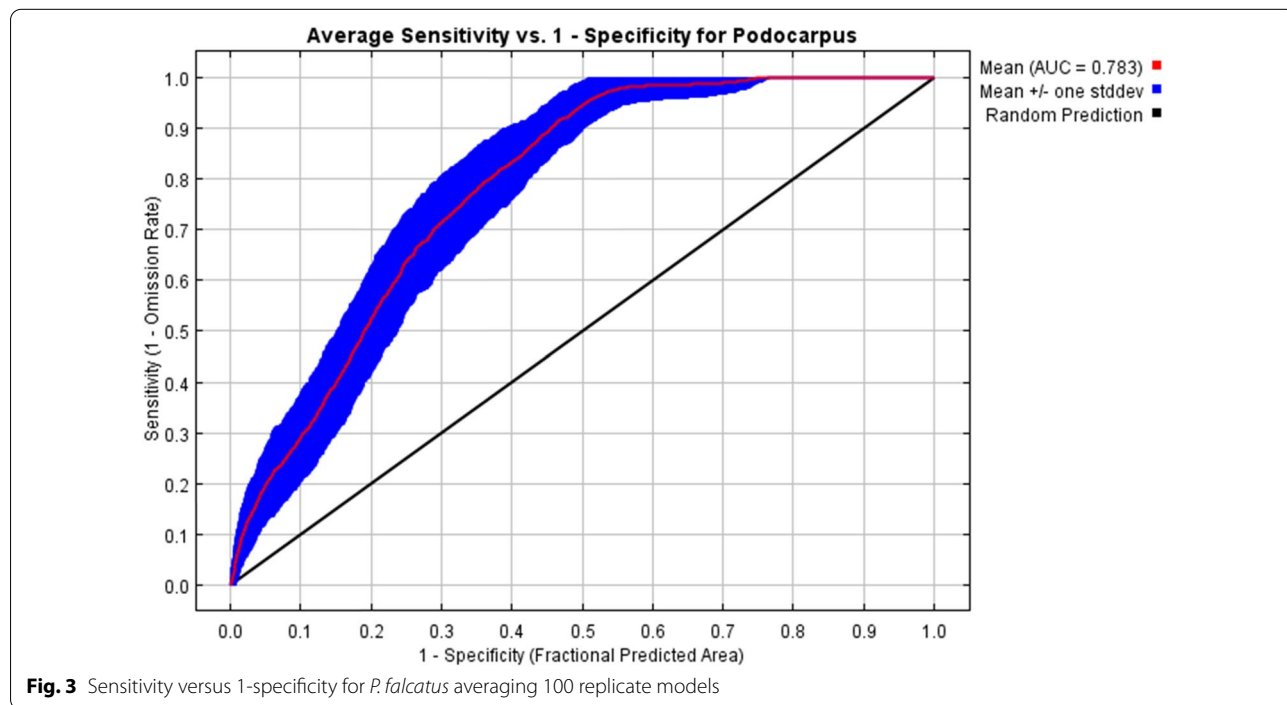


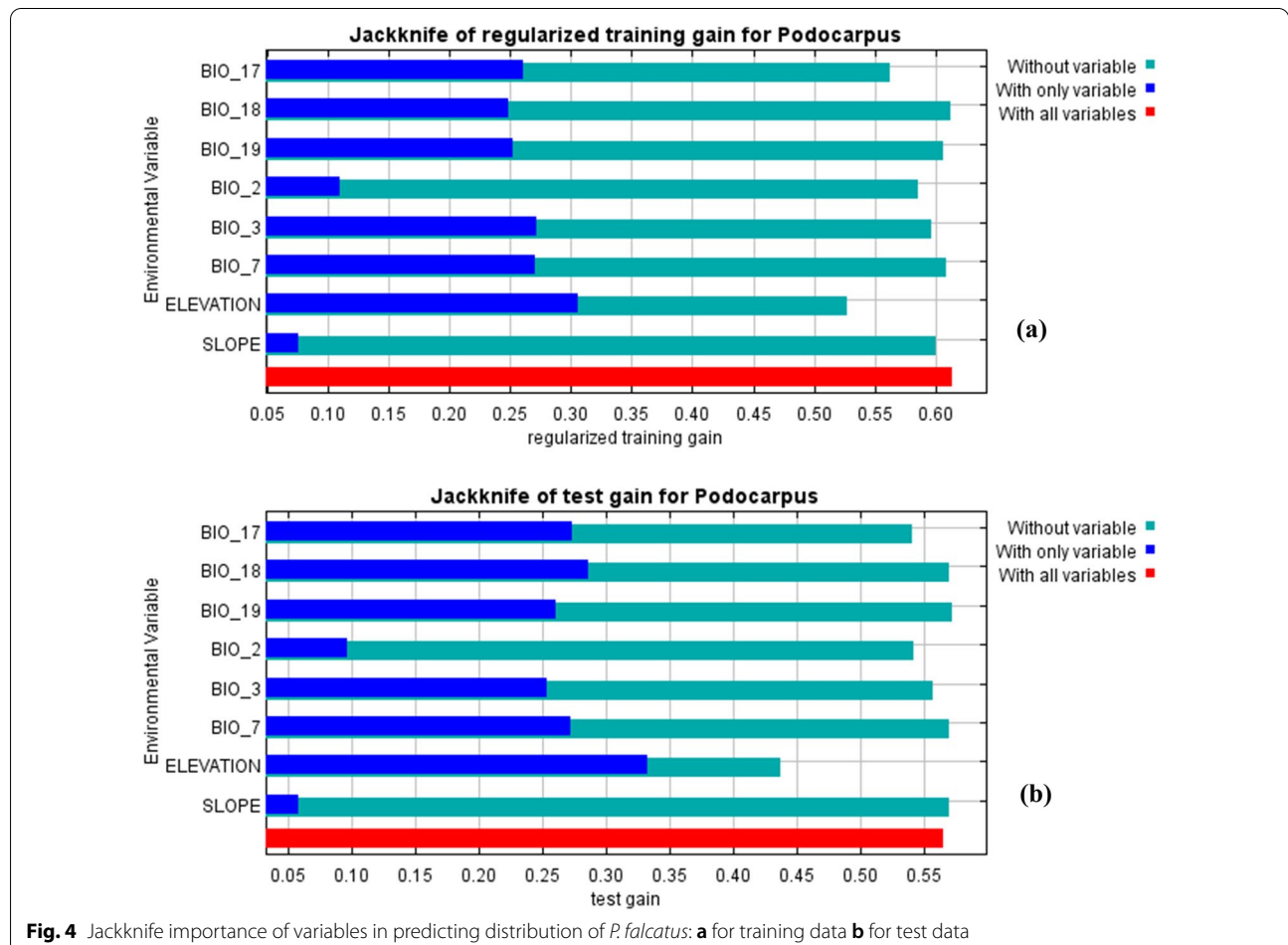
Fig. 3 Sensitivity versus 1-specificity for *P. falcatus* averaging 100 replicate models

in the species distribution and habitat suitability map (Fig. 2) fulfills, at least, the minimum environmental conditions currently inhabited by the species, as exhibited in the presence localities. Alternatively, the areas labeled as unsuitable in this map should be interpreted cautiously, as some locations might marginally support the growth of the species.

In modeling species distribution and habitat suitability, understanding the ecological concepts of ‘fundamental niche’ and ‘realized niche’ helps to properly interpret model predictions (Phillips et al. 2006). A fundamental niche comprises a set of all environmental conditions that satisfy the long-term survival of a particular species, while the environmental space that the species actually occupies constitutes a realized niche. In other words, a species’ fundamental niche represents its potential distribution, whereas the space it actually inhabits constitutes its realized distribution (Hutchinson 1957). Because of human influence, biotic interactions, or geographic barriers which might hinder further dispersal, the species’ realized niche may become smaller than its fundamental

niche with respect to the environmental conditions being modeled (Phillips et al. 2006). For this reason, the predicted area of species’ presence is usually larger than its realized distribution. Therefore, in this study, the geographic space classified to be suitable (Fig. 2) dominantly covers potentially suitable areas, with the species’ present distribution constituting a smaller portion.

On the other hand, despite the presence of an extensive environmentally suitable area that accounts for greater than 48% of the studied landscape, a smaller open space is actually available for the restoration of *P. falcatus* tree species. Because of high population density, the landscape has been intensively cultivated, dominantly covered by agroforestry crops extending up to steep-sloped terrains. As a result, except the pocket fragmented open spaces within the existing protected *Podocarpus* forest, potential open areas for a small-scale rehabilitation of this native tree species are mostly found at religious sites and around homesteads of individual farmers. In connection with this, as we realized in our field survey conducted in 2018/2019, many farmers in this area have



a successful experience of growing the tree around their homesteads as well as in the compounds of religious places. Moreover, the farmers who have implemented a small-scale plantation of this degraded native tree species have got government recognition. Thus, the local government office of forestry has been encouraging them by providing seedlings regularly during the rainy season for further plantation and rehabilitation of the degraded tree species.

Accuracy of model prediction

Accuracy of the MaxEnt model prediction was assessed using the receiver operating characteristics area under the curve (AUC). After 100 replicate runs, an average of training and test AUC of 0.783 with a standard deviation of 0.03 was attained (i.e., mean \pm study in Fig. 3). This indicates the probability that a randomly selected presence point is located in a raster cell with a probability value representing the species occurrence than a randomly generated point is 78.3%. Compared to those involving presence-absence data, the maximum achievable AUC for models using presence-only data, is normally lower (Wiley et al. 2003; Phillips et al. 2006), which is also smaller for species characterized by wider range of environmental conditions (Phillips et al. 2004). In this regard, while the accuracy is significantly better than a random prediction, this relatively smaller value of AUC could be attributed to the wide environmental range of the species (Tefsaye et al. 2002; Teketay 2011), in addition to the inherent characteristics of modeling with presence-only data (Wiley et al. 2003; Phillips et al. 2006). As shown below (Fig. 3), sensitivity corresponds to the proportion of test localities correctly predicted present (1–extrinsic omission rate), whereas 1–specificity (fractional predicted area) represents the proportion of all pixels predicted to have suitable conditions supporting the species (Phillips et al. 2006; Phillips 2017).

Major environmental predictors determining *Podocarpus* tree distribution

The principal application of species' distribution modeling is to empirically answer which variable(s) matter most in predicting the species being modeled (Phillips 2017). To that end, the MaxEnt's jackknife importance test (both for the training and test data) and relative percent contribution were applied. Comparatively, elevation range has scored the highest relative contribution (37.5%), followed by precipitation of the driest quarter

(30.8%), and isothermality (13.5%) (Table 3). By contrast, precipitation of the warmest quarter has contributed the least (i.e., 0.6%).

The jackknife importance score of the training and test data have revealed that elevation has contributed the highest gain to the model when used in isolation, demonstrating that it has the most useful information by itself. Elevation was also the variable that most decreased the model gain when it was omitted, and the most effective single variable in predicting the distribution of the independent test data. Therefore, based on its relative contribution and jackknife importance test, elevation appears to be the key predictor, as it contains the most useful information that is not present in the other predictor variables. This agrees with the recent SDM study conducted in northern Ethiopia for *Juniperus procera*, a tree species characterized by a similar environmental condition as *Podocarpus falcatus*. In this study, which has involved all the 19 bioclimatic variables and elevation as environmental predictors, Abrha et al. (2018) reported that elevation range was found to be one of the key predictors of the distribution and suitable habitats of *Juniperus procera*. In the present study, precipitation of the driest quarter (BIO17) and mean diurnal temperature range (BIO2) were also found to be useful to a certain extent following elevation gradient, as demonstrated by a decreased model gain when one of them was omitted (Fig. 3a, b).

Response curves of individual predictors also provide more specific and useful information that characterizes the species' distribution. These curves show how the predicted probability of the species' distribution changes with the variation of each predictor, keeping all other variables at their average sample value. The curves show the mean response of the 100 replicate MaxEnt runs (red) and the mean \pm one standard deviation (blue shades). Below are the response curves of the three predictors with the greatest contribution to the MaxEnt model. Based on the minimum predicted value assigned to any of the presence samples used for prediction, the response curve of elevation range shows that the maximum probability of species' occurrence exists from 1600 to 2000 masl (Fig. 4a). This indicates that suitability increases with elevation gradient from about 1600–1900 masl, while it gradually declines after it reaches a peak around 1900 masl. Similarly, the response curve of the driest quarter precipitation amount indicates that the species could exist in areas with 100 to about 150 mm quarterly rainfall during

(See figure on next page.)

Fig. 5 Individual response curves of major environmental predictors: **a** elevation, **b** precipitation of driest quarter **c** isothermally

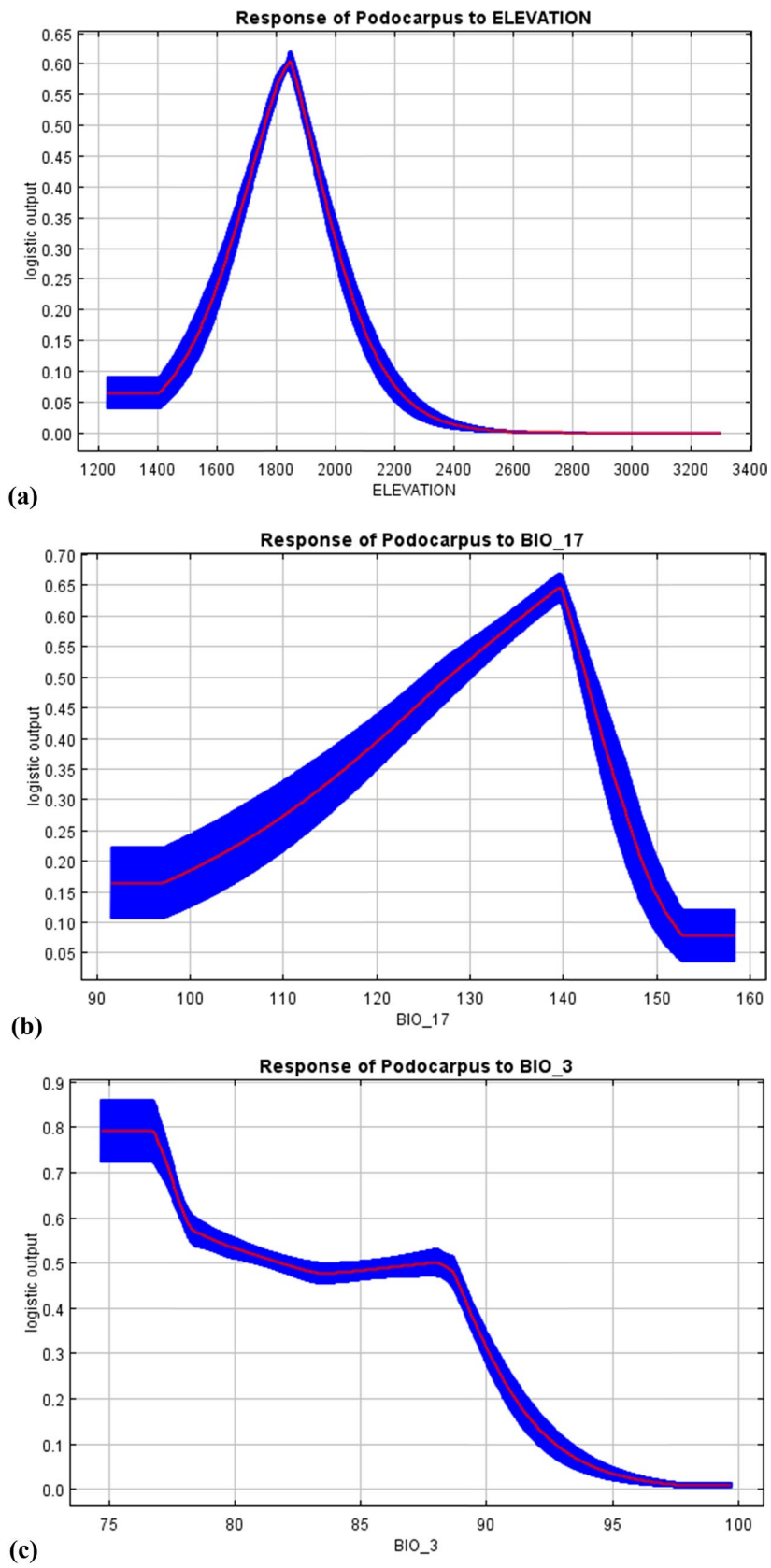


Fig. 5 (See legend on previous page.)

the driest period, with the maximum probability of presence being around 140 mm (Fig. 5b). The species could also tolerate from 75 to above 90% isothermality variations (the degree of diurnal to annual temperature oscillations), even though its probability of occurrence declines rapidly after about 88% (Fig. 5c).

Conclusions

Based on the MaxEnt model generated result, above 48% of the study catchment has been predicted to be potentially suitable for restoration of the threatened *P. falcatus* native tree species. From the eight variables used to run the MaxEnt model, elevation range was found to be a key environmental predictor of the present distribution and suitable habitats of *P. falcatus*, followed by precipitation of the driest quarter and mean diurnal temperature range ((Isothermality). Thus, the predicted distribution and potentially suitable habitats are found dominantly occupying a highland agro-ecological zone with elevation differences ranging from 1600 to 2200 masl and slope gradients varying from 0 to 25 degrees. This potentially suitable area is characterized by climate conditions of 16–21 °C annual mean temperature (ranging from 7 to 28 °C), and annual mean rainfall of 1070–1418 mm. These climatic and topographic conditions are consistent with the environmental characteristics prevailing at the protected natural *P. falcatus* forest site found at an altitude of 2120 m within the study catchment. These are also in line with the known general environmental conditions supporting the species distributions (Gilman and Watson 1994; Tesfaye et al. 2002; Orwa et al. 2009; Teketay 2011; Farjon 2013).

On the other hand, although an extensive portion of the study catchment was predicted to be environmentally suitable for rehabilitation of the target tree species, the largest portion of this area has been found to be intensively cultivated and unlikely to be practically available for tree planting. Since the study area is among the densely populated parts of the country, it has been intensively cultivated and dominantly covered with a traditional agroforestry system extending up to steeply sloped terrains. For this reason, large-scale restoration of this threatened tree species is unlikely to be practically implemented. Therefore, some pocket open spaces within the existing protected *P. falcatus* forest site, compounds of religious places, and pocket areas around homesteads of farmers have been identified as potential areas for small-scale rehabilitation of this seriously declined native tree species. In fact, the local government office of forestry has been encouraging farmers by providing seedlings regularly during the wet season to undertake and expand small-scale plantations of this target tree species.

Moreover, in addition to the successful experiences that many farmers have demonstrated so far in conserving this native tree, privately-owned plantations may enhance more sustainable management and conservation of this degraded tree species.

Acknowledgements

We are grateful to the CHIRPS and WorlClim data providers for availing these datasets at no charge. We also acknowledge the Ethiopian Space Science and Technology Institute, Entoto Observatory and Research Center (EORC) for the all support it provided throughout the course of this study.

Authors' contributions

BGT and BG have performed data collection, analysis and report writing, while FM has edited the manuscript. All authors have read and approved the final manuscript.

Funding

This work was carried out by a Student Research Grant funded by the Ethiopian Space Science and Technology Institute (ESSTI), Entoto Observatory and Research Center (EORC).

Availability of data and materials

The satellite-derived rainfall estimate, Climate Hazards group Infrared Precipitation with Stations version 2.0 (CHIRPS), was obtained from the CHIRPS website (<ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/>). Similarly, the BIOCLIMATIC variables were downloaded from the WorlClim webpage. In addition, we have used in-situ rainfall data acquired from the National Meteorological Agency of Ethiopia (NMA).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they do not have any competing interests.

Author details

¹Department of Land Administration and Surveying, Dilla University, P. O. Box 419, Dilla, Ethiopia. ²Department of Remote Sensing, Entoto Observatory and Research Center, Ethiopian Space Science and Technology Institute (ESSTI), Addis Ababa, Ethiopia. ³Department of Information Engineering and Computer Science, University of Trento, Trento, Italy.

Received: 9 December 2021 Accepted: 14 February 2022

Published online: 18 March 2022

References

- Abrha H, Bihane E, Hagos H, Manaye A (2018) Predicting suitable habitats of endangered *Juniperus procera* tree under climate change in Northern Ethiopia. *J Sustain For* 37(8):842–853
- Aguirre-Gutiérrez J et al (2013) Fit-for-purpose: species distribution model performance depends on evaluation criteria—Dutch Hoverflies as a Case Study. *PLoS ONE* 8(5):e63708. <https://doi.org/10.1371/journal.pone.0063708>
- Bekele A (2007) Useful trees and shrubs for Ethiopia: identification, propagation and management in 17 agroclimatic zones. RELMA in ICRAF Project, Nairobi, p 552
- Castilho R (2015) Species distribution modeling 101: basics hands-on experimental work notes. Johns Hopkins University, Baltimore

- De Marco P, Júnior CCN (2018) Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0202403>
- Doffana ZD (2014) 'Daguchoch (Podocarpus falcatus) is Abbo!' Wonsho Sacred Sites, Sidama, Ethiopia: origins, maintenance motives consequences and conservation threats. University of Kent, Canterbury
- Eastman J (2012) Idrisi selva tutorial, vol 45. Idrisi Production, Clark Labs, Clark University, Worcester, pp 51–63
- Elith J et al (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2):129–151
- Elith J et al (2011) A statistical explanation of MaxEnt for ecologists. *Divers Distrib* 17:43–57
- Farjon A (2013) *Afrocarpus falcatus*. The IUCN red list of threatened species 2013. Int Union Conserv Nat Nat Resour. <https://doi.org/10.2305/IUCN.UK.2013-1.RLTS.T42438A2980290.en>
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. <https://doi.org/10.1002/joc.5086>
- Gilman EF, Watson DG (1994) *Podocarpus falcatus*: fact sheet ST-492 October 1994. US Forest Service Department of Agriculture, US
- Girma A, de Bie CAJM, Skidmore AK, Valentijn V, Frans B (2015) Hyper-temporal SPOT-NDVI dataset parameterization captures species distributions. *Int J Geogr Inform Sci*. <https://doi.org/10.1080/13658816.2015.1082565>
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25(15):1965–1978
- Hutchinson GE (1957) Concluding remarks: Cold Spring Harbor Symposia on quantitative biology. In: Phillips SJ, Anderson RP, Schapire RE (eds) Maximum entropy modeling of species geographic distributions, vol 190. Ecological Modelling, Amsterdam, pp 231–259
- Kumar S, Stohlgren TJ (2009) MaxEnt modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. *J Ecol Nat Environ* 1(4):94–98
- Mechal A, Birk S, Winkler G, Wagner T, Mogessie A (2016) Characterizing regional groundwater flow in the Ethiopian Rift: a multimodel approach applied to Gidabo River Basin. *Austrian J Earth Sci*. <https://doi.org/10.17738/ajes.2016.0005>
- Moat J et al (2017) Coffee farming and climate change in Ethiopia: Impacts, forecasts, resilience and opportunities—Summary. The Strategic Climate Institutions Programme (SCIP). Royal Botanic Gardens, Kew, p 37
- Negash L (2010) A selection of Ethiopia's indigenous trees: biology, uses and propagation techniques. Addis Ababa University Press, Addis Ababa
- Orwa C, Mutua A, Kindt R, Jamnadass R, Anthony S (2009) Agroforestry Database: a tree reference and selection guide version 4.0. FAO, Rome
- Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* 27:285–298
- Phillips SJ, Dudík M (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography* 31:161175. <https://doi.org/10.1111/j.2007.0906-7590.05203.x>
- Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. Proceedings of the 21st International Conference on Machine Learning. ACM Press, New York, pp 655–662
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190:231–259
- Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of MaxEnt. *Ecography* 40(7):887–893
- Phillips SJ (2017) A brief tutorial on MaxEnt. http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 12 Mar 2020
- Teketay D (1992) Human impact on a natural Montane forest in southeastern Ethiopia. *Mt Res Dev* 16:393–400
- Teketay D (2011) Natural regeneration and management of *Podocarpus falcatus* (Thunb.) Mirb. in the Afromontane forests of Ethiopia. In: Gunter S et al (eds) Silviculture in the tropics. Tropical forestry 8. Springer, Berlin. https://doi.org/10.1007/978-3-642-19986-8_21
- Tesfamariam BG, Gessesse B, Melgani F (2019) Characterizing the spatiotemporal distribution of meteorological drought as a response to climate variability: the case of rift valley lakes basin of Ethiopia. *Weather Clim Extrem* 26:100237
- Tesfaye G, Teketay D, Fetene M (2002) Regeneration of fourteen tree species in Harenna forest, southeastern Ethiopia. *Flora Morphol Distrib Funct Ecol Plants* 197(6):461–474
- TGE (1994) Forestry conservation, development and utilization proclamation no. 94/1994. Negarit Gazeta of The Transitional Government of Ethiopia, Ethiopia
- Vivero JL, Kelbessa E, Demissew S (2005) The red list of endemic trees and shrubs of Ethiopia and Eritrea. Cambridge University, UK
- Wiley EO, McNyset KM, Peterson AT, Robins CR, Stewart AM (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* 16(3):120–127
- Yoo W et al (2014) A study of effects of multicollinearity in the multivariable analysis. *Int J Appl Sci Technol* 4(5):9

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)