Modeling effects of abiotic factors on the abundances of eight woody species in the Harana forest using artificial networks, random forest, and generalized linear models

Girma Ayele Bedane[1*] , Gudina Legese Feyisa[1] and Feyera Senbeta Wakjira[2]

Abstract

Background Abiotic factors exert different impacts on the abundance of individual tree species in the forest but little has been known about the impact of abiotic factors on the individual plant, particularly, in a tropical forest. This study identified the impact of abiotic factors on the abundances of *Podocarpus falcatus*, *Croton macrostachyus, Celtis africana, Syzygium guineense*, *Olea capensis*, *Diospyros abyssinica, Feliucium decipenses,* and *Coffea arabica.* A systematic sample design was used in the Harana forest, where 1122 plots were established to collect the abundance of species. Random forest (RF), artificial neural network (ANN), and generalized linear model (GLM) models were used to examine the impacts of topographic, climatic, and edaphic factors on the log abundances of woody species. The RF model was used to predict the spatial distribution maps of the log abundances of each species.

Results The RF model achieved a better prediction accuracy with $R^2 = 71\%$ and a mean squared error (MSE) of 0.28 for *Feliucium decipenses*. The RF model differentiated elevation, temperature, precipitation, clay, and potassium were the top variables that influenced the abundance of species. The ANN model showed that elevation induced a negative impact on the log abundances of all woody species. The GLM model reaffirmed the negative impact of elevation on all woody species except the log abundances of *Syzygium guineense* and *Olea capensis*. The ANN model indicated that soil organic matter (SOM) could positively affect the log abundances of all woody species. The GLM showed a similar positive impact of SOM, except for a negative impact on the log abundance of *Celtis africana* at $p < 0.05$. The spatial distributions of the log abundances of *Coffee arabica*, *Filicium decipenses,* and *Celtis africana* were confined to the eastern parts, while the log abundance of *Ol*ea *capensis* was limited to the western parts.

Conclusions The impacts of abiotic factors on the abundance of woody species may vary with species. This ecological understanding could guide the restoration activity of individual species. The prediction maps in this study provide spatially explicit information which can enhance the successful implementation of species conservation.

Keywords Species distribution model, Random forest, Artificial neural network, Generalized linear model, Species abundance, Woody species, Environmental variable

*Correspondence:
Girma Ayele Bedane
girmaayele35@gmail.com
[1] Center for Environmental Science, College of Computational and Natural Science, Addis Ababa University, P. O. Box 1176, Addis Ababa, Ethiopia
[2] Centre for Environment and Development, College of Development Studies, Addis Ababa University, Addis Ababa, Ethiopia

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 2 of 14

Introduction

Abundances of woody plant species in a forest ecosystem are influenced by various abiotic factors (Feng et al. 2021; Rahman et al. 2021; Waldock et al. 2022). The abiotic factors that commonly affected the spatial distributions of woody plant species comprise of climatic factors, topographic features, and soil properties (Nguyen et al. 2015; Feng et al. 2021; Rahman et al. 2021; Waldock et al. 2022). Climatic factors which include variability of precipitation and temperature influence the abundance of plant species at different geographical scales (Yang et al. 2006; Qian 2013). Topographical features indirectly affect woody plant species by regulating microclimate and soil conditions (Nguyen et al. 2015; Feng et al. 2021; Ahmed et al. 2022). Soil physical and chemical properties play essential roles at a locale scale in influencing the growth, abundance, and distribution of woody plants species (Condit et al. 2013; Nguyen et al. 2015; Rahman et al. 2021; Yaseen et al. 2022).

Understanding the impacts of abiotic factors on plant species could have an important implication in sustainable forest management. Primarily, it helps to provide scientific explanations of how plant species could interact with abiotic factors, which might be positive or negative interactions (Scarnati et al. 2009; Seo et al. 2021). Second, it uses to produce prediction maps that provide spatially explicit information to easily differentiate species conservation priority areas (Park et al. 2003; Clément et al. 2014; Seo et al. 2021). The prediction maps broadly enable an understanding of how plant species are distributed in relation to the interaction between plants and abiotic factors (Barker et al. 2014; Clément et al. 2014). With this recognition, a research study about the impacts of abiotic factors on plant species has remained an essential topic in forest ecology (Guisan and Zimmermann 2000; Park et al. 2003; Nguyen et al. 2015; Rahman et al. 2021).

Particularly, it is essential in tropical forests, where the impact of abiotic factors on the abundance of plant species has not been sufficiently studied (Newbold 2009; Condit et al. 2013). For instance, Ethiopia is the fifth largest flora-rich country in tropical Africa, having more than 6500 species (Husen et al. 2012). However, previous studies in the country have focused mainly on floristic composition with less attention to the impacts of abiotic factors on the abundance of woody plant species in different forest areas in the country (Asefa et al. 2020).

The Harana Afromontane forest is one of the priority forest areas in Ethiopia that has experienced a similar limitation, regarding the effects of abiotic factors on individual woody species (Lulekal et al. 2008). Apart from this, Harana forest is a center of carbon storage with an average above-ground carbon stock of 131.505 tons ha^{-1} (Bedane et al. 2022). In addition, Harana forest provides a buffering service to the Bale Mountains National Park which is one of 34 global biodiversity hotspot areas (Nelson 2012).

Therefore, this study targeted the Harana forest to investigate the impacts of abiotic factors on the abundance of eight woody plant species. Eight woody species were considered in this study with the understanding that the analysis of abiotic factors on every individual species in the forest could not be manageable (Gowda 2011). The selection of the first seven woody species was made taking into account the ecological contribution of those species in accumulating 45% of the above-ground carbon stock in the Harana forest (Bedane et al. 2021). These seven species include *Podocarpus falcatus, Croton macrostachyus, Celtis africana, Syzygium guineense, Olea capensis, Diospyros abyssinica* and *Feliucium decipenses.* Moreover, *Coffea arabica* was included considering its largest abundance in the forest (Kewessa et al. 2019).

This study was to bridge the identified knowledge gaps by addressing two important objectives. The first objective was to analyze the impacts of climate, topographic, and soil factors on the log abundances of eight woody plant species using generalized linear models (GLMs), random forest (RF), and artificial neural networks (ANNs). These models have been commonly used in the study of ecology and reported to show good performances (Guisan et al. 2002; Aksu et al. 2019; Zhang et al. 2019; Yudaputra et al. 2019). Meanwhile, the use of more than one model could provide an opportunity to look into the reliability and consistency of the impacts of abiotic factors on the abundance of woody species under applications of different models (Antúnez 2022). The second objective was to predict the spatial distributions of the log abundances of eight woody species. This information is essential to establish a benchmark to monitor the potential spatiotemporal dynamics that might occur in the future in connection with anthropogenic pressures (Barker et al. 2014).

Materials and methods

Description of the study area

Harana forest is an Afromontane forest type that covers a total area of 107,298 ha and is located between latitude 6° 14′ 40″ to 6° 38′30″ N and longitude 39° 22′10″ to 39° 27′50″ E (Fig. 1). The Harana forest lies in the northern parts of Delo Mena and Harana Buluk Districts in the Oromia National Regional State (Kewesa et al. 2019; Bedane et al. 2022). The Northern and North-Eastern parts of the Harana forest lie inside the Bale Mountains National Park (Nelson 2012).

The elevation of the Harana forest ranges from 1270 to 3030 m asl with increasing trends from the south to north, and from east to west directions. The highest
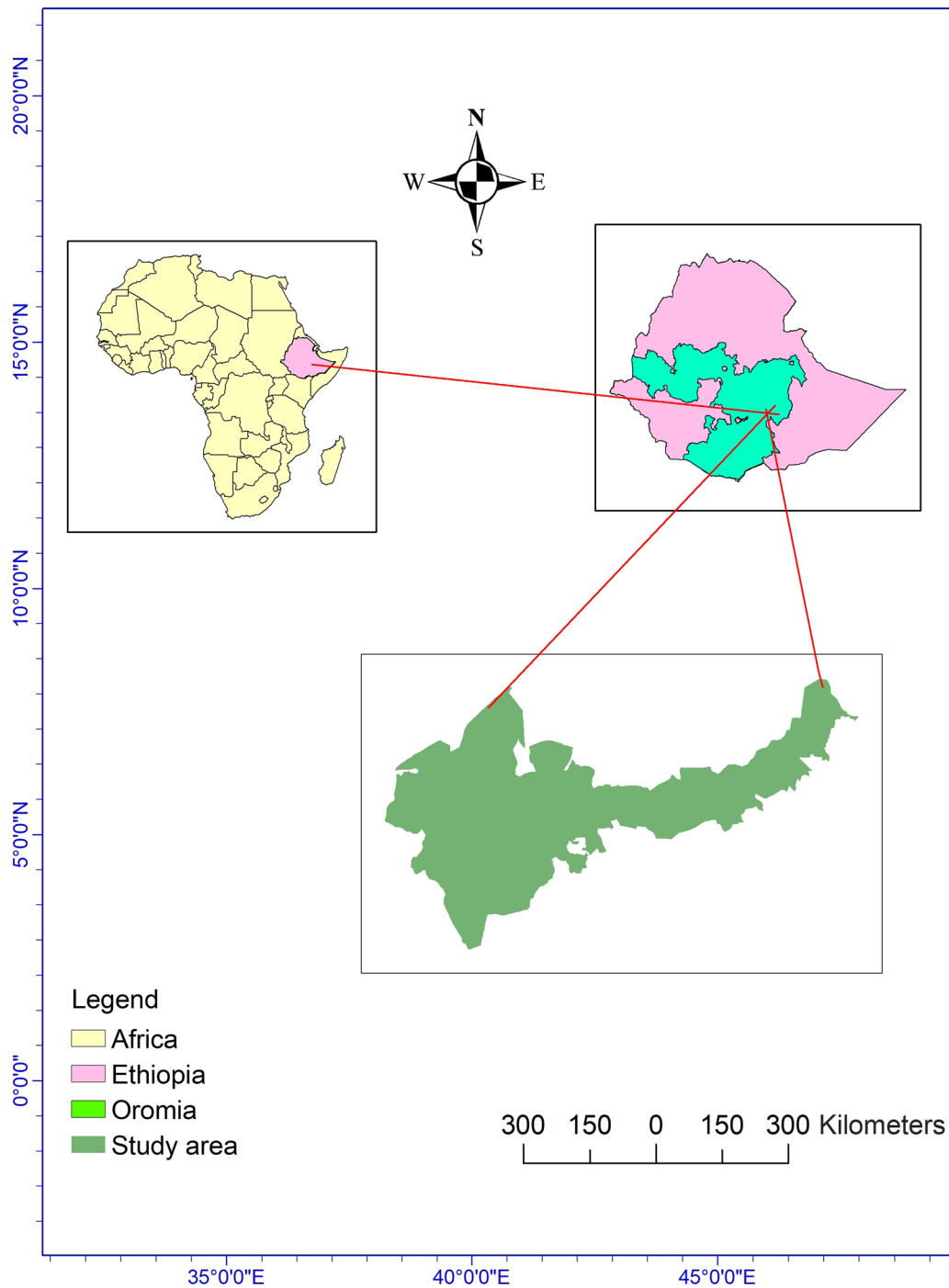
Fig. 1 Map of the study area in relation to the bigger geographical locations

peaks are located in the northwest and western parts of the forest, while the central and southern parts are characterized by relatively uniform elevation (Ayele et al. 2019; Bedane et al. 2022). The study area is characterized by a bimodal pattern of rainfall. The first rainy season starts in April and extends to June. The second short rainy season begins in the middle of September and lasts in early November. The annual rainfall ranges from 765 to 1110 mm, while the annual mean temperature ranges between 14 °C and 22 °C (Ayele et al. 2019).

Harana forest is mainly occupied by dominant woody species, which include *Podocurpus falcatus*, *Warburgia*

ugandensis, Celtis africana, Syzygium guineense, Olea capensis, Diospyros abyssinica, and *Filicium decipiens* (Kewesa et al. 2019; Bedane et al. 2022). The composition of soil types is characterized by nitosols, chromic, pellic vertisols, Ortic acrisols, and Eutric and Calcic cambisols (Lulekal et al. 2008).

Data collection and analysis
Sampling methods, species data collection and analysis
The study area was divided into a grid square of 1 km by 1 km which helped to establish a total of 1122 plots at the point, where two grid lines intersect each other. This was conducted using ArcGIS version 10.4. The coordinate points of each plot were fed into the GPS receivers to identify each location during the field data collection. Three nested circular plots with radii of 15, 5, and 2 m were established from the center of a plot using measuring tape and ribbons. The count of each tree species with ≥ 5 cm diameter at breast height (DBH) was conducted within the circular plot of a 15 m radius. Saplings of woody species with DBH of 1–5 cm were counted within the circular plot of a 5 m radius, while seedlings of less than 1.0 cm root collar diameter were counted in the smallest circle. The circular plot was selected for its advantage of being less susceptible to the inclusion and omission of trees around the boundary of a circular plot (UNFCCC 2015).

The abundance of species within different plot sizes was standardized to the area of a 15 m radius (0.07065 ha) to facilitate the statistical analysis with a uniform unit. The Statistical Package for the Social Sciences (SPSS) version 20.0 was used to compute the relative frequency and relative mean abundance of woody species. The relative frequency of each species was computed as a percentage of several plots with a specific species divided by a total of plots. The relative mean abundance of each woody species was also calculated as a percentage of the mean abundance of each woody species divided by the mean abundance of all woody species.

Selection of predictor variables
We conducted the selection of the predictor variables based on an in-depth review of the literature. The variables included elevation, annual precipitation, annual mean temperature, and edaphic factors. The digital elevation model (DEM) with 30 m resolution was downloaded from the USGS data portal, whereas an elevation of the study area was clipped from the DEM using the appropriate tools in ArcGIS v10.4. The annual mean temperature and annual precipitation were downloaded from the Bioclim data portal (https://www.Bioclim.org). The soil data were accessed from the GIS database of Farm Africa. These soil data included soil pH, organic matter

(OM), total nitrogen (N), carbon to nitrogen ratio (C/N), available phosphorus (P), available potassium (K), magnesium, calcium, and clay, silt, and sand contents in percentage. Each layer was used to extract values of each soil property that corresponded to the location of 1122 plots from where data on species abundance were collected. The extraction of soil values was conducted using multiple-point extraction techniques in ArcGIS v10.4. The purpose of extracting soil data was to establish the relationships between species abundance and soil variables to determine the effects of soil on the abundance of different woody species at the specified location.

Data exploration and transformation
Statistical procedures assume variable data are normally distributed (Osborne 2002). However, ecological data in nature do not often follow a normal distribution, where there is skewness (Feng et al. 2014). The abundances of woody species in this study did not follow the normal distributions and were transformed using the log base 10 (abundance $+ 1$) and this expression is referred to as the log abundance of species through this paper. We did the log transformation because the ANN model would require the transformed value of a response variable to facilitate the computation of complex ecological interactions (Sola and Sevilla 1997; Puheim and Madarász 2014). Log transformation has been widely used to convert skewed ecological data (Feng et al. 2014), and a constant number is usually added to the original data before data transformation. This was done to avoid undefined values for the log numbers less than one (Feng et al. 2014; Osborne 2002).

The ANN model also requires the normalized value of input variables to a similar scale, otherwise, predictors with different measurement scales produce unrealistic outputs (Puheim and Madarász 2014). Predictor variables were normalized using Eq. 1 cited in (Olden and Jackson 2002):

$$N = (X_i - Min(X)/(Max(X) - Min(X)) \qquad (1)$$

whereas N refers to a normalized value for observation X_i, X_i is a list of observations.

Modelling of the abundance of woody species
Random forest modelling
The random forest regression model in the R software ver. 4.1.0 was used to examine the relationship between the log abundance of each species and various abiotic factors. A total of 1122 samples were classified into 80% training and 20% testing. This was done as the RF model would not require any independent data set to conduct a model validation (Kapwata and Gebreslasie 2016). The

RF model was run for each species using three steps that have been commonly described in the literature (Kapwata and Gebreslasie 2016). First, we set ntree=500 which was the number of trees to grow in a model depending on bootstrap sampling. Second, we selected mtry=5 as several predictor variables to perform splitting of data at each node to grow un-pruned regression trees with identified independent variables. Third, estimated values of regression trees were added and averaged to predict outputs of the log abundance of each woody species.

The RF model was further used to produce the map of the log abundance of each woody species using raster layers of 16 predictor variables and the shapefile of the log abundance of each woody species. The RF model was selected for this purpose because it achieved smaller prediction errors for many species as compared to the ANN and GLM models. The accuracy of the RF model could be assessed using mean squared error (MSE) and coefficient of determination (R^2) (Scarnati et al. 2009). Meanwhile, the RF model identified the variable importance using a percent increase in the mean squared errors (%IncMSE). A large %IncMSE value shows a high degree of variable importance in predicting a given response variable (Scarnati et al. 2009; Kapwata and Gebreslasie 2016).

Artificial neural network modeling

We constructed the ANNs with one input layer, one hidden layer, and one output layer using the neural net package in the R Software ver. 4.1.0. A total of 1122 sample locations was split into 80% training and 20% testing data to run the model for each of the eight woody species. The number of hidden layers was determined to be one as increasing the number of hidden layers could lead to overfitting of the model (Chen et al. 2022). A supervised backpropagation algorithm (BPN) was used to train the model using forward and backward processes. During the forwarding learning process, 16 input variables were propagated from the input layer to the hidden layer, where weights and bias values were assigned to the neuron connection in the ANN model (Chen et al. 2022). Weights were multiplied by the corresponding input variable to determine the contribution of input variables in the model (Kukreja et al. 2016; Aksu et al. 2019; Chen et al. 2022). The sigmoid activation function was used to decide whether the sum weight would be transferred into the output layer or not (Chen et al. 2022) as the sigmoid activation function is advantageous of handling data that are characterized by non-linear features (Kukreja et al. 2016; Aksu et al. 2019). The back-propagation learning was conducted to adjust weights and biases to minimize prediction errors between actual and prediction values (Ahmadzadeh et al. 2017; Aksu et al. 2019; Chen et al. 2022).

The accuracy of the ANN model was determined using a coefficient of determination (R^2) and mean squared error (MSE) (Chen et al. 2022). The Olden algorithm was further used to identify variable importance based on a quantified magnitude and direction of the effects of input variables on a response variable (Olden and Jackson 2002). The Olden variable importance in this study was standardized using Eq. 2 proposed by Olden and Jackson (2002). This standardization process was required to display the variable importance graphically to provide a clear illustration for better visualization:

$$Z_n = (X_n - \overline{x})/\delta_X \tag{2}$$

where Z_n is the standardized value of observation n, X_n is the original value of observation n, $\overline{x}$ is the mean, while δ_X is the standard deviation of the variable X.

Generalized linear modeling

The generalized linear model in R software version 4.1.0 was used to quantify the impact of predictor variables on the log abundance of each woody species. The total sample data was divided into 80% training and 20 testing data. The GLM model was run by specifying the identity link function to compute the linear relationship between predictor variables and response variables (Lopatin et al. 2016; Goldburd et al. 2020). The identity link uses to create a linear relationship between response and predictor variables (Dunn and Smyth 2018; Goldburd et al. 2020). Effects of abiotic factors on the log abundances of woody species were interpreted using parameter estimates at the given significance levels. The performance of the GLM model was assessed using the coefficient of determination (R^2) and the Akaike information criterion (AIC) (Rion 2010; Sakate and Kashid 2016). A higher R^2 indicates the selected environmental predictors could better explain variation in the response variable (Kapwata and Gebreslasie 2016; Chen et al. 2022).

Results

Descriptive statistics

Descriptive statistics of tree abundances and frequencies

The mean abundance of eight woody species was estimated at 2404 plants per plot of 0.07065 ha. The mean abundance of *C. arabica* was 1395 individuals per plot which accounted for 58% of the total abundance of all woody plants in the Harana forest. The mean abundance of *F. decipenses* was 376 individuals per plot with a relative abundance of 16%. The mean abundance of *P. falcatus* was 68 individuals per plot with a relative abundance of 3%, while its relative frequency was 39%. *O. capensis* was identified as the smallest woody species both in terms of a relative abundance of 1% and a relative frequency of 22% (Table 1).

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 6 of 14

Descriptive statistics of tree stem density within different DBH classes

Small trees with < 10 cm DBH accounted for 16% of the total stem density for each of *P. falcatus* and *D. abyssinica*, while each of *C. macrostachyus* and *F. decipenses* accounted for 11% (Table 2). *S. guineense, O. capensis,* and *C. africana* were characterized by a small proportion of the stem density for DBH < 10 cm. A large proportion of stem density per ha in each species consisted in the DBH class of 20–60 cm, while a proportion of stem density was limited for the DBH class that exceeded 60 cm.

Prediction outputs of different models
Random forest model

The RF model in this study identified the elevation, mean annual temperature, annual precipitation, clay content, and available potassium as the top important predictors that could influence the log abundances of the majority of

woody species (Fig. 2). Specifically, the distribution of the log abundances of *C. arabica* was mainly affected by the annual precipitation, elevation, silt content, mean annual temperature, and clay content. The spatial distribution of *P. falcatus* appeared to be influenced by the mean annual temperature, calcium, available potassium, clay content, and carbon-to-nitrogen ratio. Continually, elevation, soil cation exchange capacity (CEC), mean annual temperature, available potassium, and annual precipitation were the important predictors that shaped the spatial distribution of the log abundances of *F. decipenses* (Fig. 2).
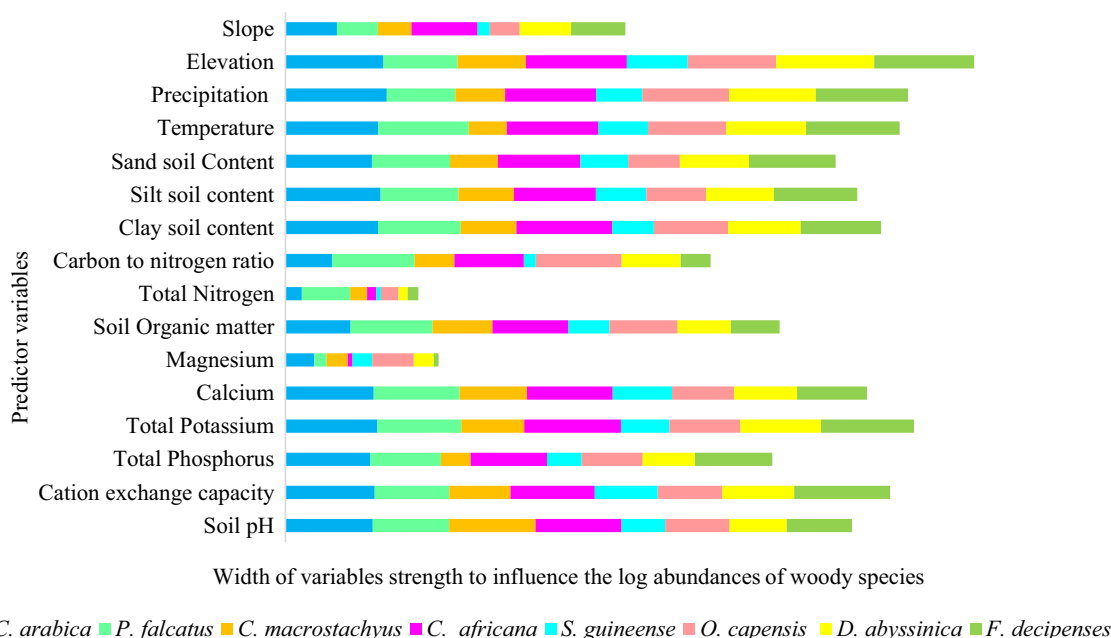
Artificial neural networks

Elevation appeared to exert a negative effect on the log abundances of all species showing a more negative impact on *F. decipenses* and *C. macrostachyus* (Table 3). The ANNs indicated that SOM was positively related to the log abundances of all species, while a strong positive relationship was detected with the log abundances of

Table 1 Descriptive statistics of eight woody species in the Harana forest

Species	Frequency	Percent (%)	Mean abundance	Relative abundance (%)	Std. Error ($+$, $-$)
Coffea arabica	449	40	1395	58	99.57
Podocarpus falcatus	434	39	68	3	5.80
Croton macrostachyus	570	51	60	2	7.24
Celtis africana	491	44	236	10	20.80
Syzygium guineense	384	34	70	3	11.01
Olea capensis	245	22	19	1	5.42
Diospyros abyssinica	318	28	178	7	18.34
Feliucium decipenses	350	31	376	16	57.82
Overall mean stem density			2402	100	150

Table 2 Stem density of each tree species per ha and its proportion in the Harana forest

DBH classes (cm)	Proportion of woody species (%)						
	S. guineense	*P. falcatus*	*D. abyssinica*	*C. macrostachyus*	*O. capensis*	*F. decipenses*	*C. african*
< 10	5	16	16	11	5	11	8
10–20	13	20	17	19	6	25	13
20–30	16	12	19	19	11	25	15
30–40	16	9	20	13	11	14	16
40–50	12	9	9	11	12	10	18
50–60	11	9	8	10	15	7	11
60–70	8	7	3	5	9	4	7
70–80	6	5	3	4	11	2	5
80–90	4	5	2	4	9	2	4
90–100	7	5	2	3	6	0	4
>100	2	2	0	2	4	2	0
Tree density/ha	86	91	58	73	54	90	67

Fig. 2 Strength of the impacts of variables in predicting the log abundances of woody species based on %incMSE values using the RF model

D. abyssinica, and *C. africana.* Total nitrogen exhibited a negative impact on the majority of the log abundances of woody species with a strong negative impact on the log abundances of *C. macrostachyus* and *F. decipenses.* A carbon-to-nitrogen ratio exhibited a positive effect with all species except that a positive impact was detected on the log abundances of *C. arabica* and *F. decipenses,* and *C. macrostachyus.* Available phosphorus tended to negatively affect the log abundances of all species except for a positive impact on the log abundances of *C. arabica* and *F. decipenses.* Available potassium displayed a negative impact on the log abundances of all species apart from the log abundances of *C. arabica, C. macrostachyus,* and *S. guineense.* Precipitation exerted a strong negative impact on the log abundances of *D. abyssinica,* while temperature appeared o show a positive effect on *F. decipenses* (Table 3).

Table 3 Impacts of environmental variables on the log abundances of woody species using ANNs

Variables	Standardized Olden value on log abundances of woody species							
	C. arabica	*P. falcatus*	*C. macrostachyus*	*C. africana*	*S. guineense*	*O. capensis*	*D. abyssinica*	*F. decipenses*
Soil pH	0.2	− 0.5	2.7	− 0.2	− 0.1	− 0.1	− 0.2	0.3
CEC	0.7	− 0.4	0.5	0.0	0.1	− 0.3	− 0.3	1.0
Phosphorus	0.1	− 0.6	− 0.8	− 0.3	− 1.5	− 0.4	− 0.5	0.5
Potassium	0.0	− 0.6	0.4	− 0.2	0.1	− 0.3	− 0.3	− 0.6
Calcium	0.3	− 0.5	0.8	− 0.2	− 0.4	− 0.4	− 0.6	0.3
Magnesium	− 3.8	3.0	0.0	− 2.1	0.0	− 0.5	− 0.2	0.1
SOM	0.9	2.1	0.4	3.4	2.2	2.6	3.7	1.1
Nitrogen	− 0.2	− 0.5	− 2.3	0.6	− 1.1	− 1.1	0.2	− 2.6
C/N	− 0.2	1.0	− 0.2	0.5	2.7	2.7	1.0	− 0.1
Clay content	0.7	− 0.5	0.4	0.0	− 0.3	− 0.2	− 0.3	0.6
Silt content	0.1	− 0.5	0.0	− 0.3	− 0.2	− 0.2	− 0.2	0.5
Sand content	0.7	− 0.7	0.0	− 0.3	− 1.0	− 0.8	− 0.1	-0.6
Temperature	0.3	− 0.2	0.0	0.0	0.2	− 0.3	− 0.3	0.9
Precipitation	0.0	− 0.2	− 0.1	− 0.2	0.0	0.1	− 0.4	0.3
Elevation	0.0	− 0.2	− 1.5	− 0.3	− 0.4	− 0.3	− 0.6	− 2.2
Slope	0.1	− 0.5	− 0.2	− 0.2	− 0.3	− 0.2	− 0.3	− 0.1

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 8 of 14

Generalized linear model (GLM)

The GLM parameter estimates indicated that elevation showed a significant negative relationship with the log abundances of *C. arabica*, *C. macrostachyus*, *C. africana*, and *D. abyssinica*, while a positive effect of elevation showed only with the log abundance of *P. falcatus* at $p < 0.05$ (Table 3). SOM appeared to show a positive effect on the log abundances of all species, but a significant effect was detected only with the log abundance of *C. africana* at $p < 0.05$. C/N ratio exhibited a positive impact on the log abundances of *P. falcatus*, *C. africana*, *O. capensis*, and *F. decipenses*, while a significant negative effect was detected with the log abundance of *C. arabica* at $p < 0.05$. Available phosphorus and available potassium exhibited significant negative impacts on the log abundances of *C. arabica*, *P. falcatus*, *C. africana*, and *F. decipenses*. Annual precipitation showed a significant negative impact on the log abundances of *C. arabica*, *C. macrostachyus*, and *C. africana*. The mean annual temperature posed a significant negative impact on the log abundances of *C. macrostachyus* but a positive effect on the log abundance of *S. guineense* at $p < 0.05$ (Table 4).

Prediction accuracy of ANN, RF, and GLM models

The mean squared error in Table 5 shows that the RF produced better prediction accuracy as compared with the ANN model. The RF model predicted the spatial distribution of log abundances of *F. decipenses*, *S. guineense*, and *P. falcatus* with small prediction errors (MSE) of 0.28, 0.38, and 0.39, respectively. Consistently, the GLM predicted the log abundances of *F. decipenses*

and *S. guineense* with smaller AIC values (Table 4). The RF, GLM, and ANN models predicted the log abundance of *F. decipenses* achieving the R^2 values of 71%, 64%, and 59%, respectively. The ANN model produced larger R^2 values for the log abundances of *P. falcatus*, *C. macrostachyus*, *S. guineense*, and *O. capensis*.

Spatial distribution map

The random forest model predicted the spatial distribution of the log of abundances of eight woody species showing more species abundance in the eastern parts (Fig. 3). The geographical distribution of the log abundance of *C. arabica* was mainly concentrated in the eastern parts of the study area (Fig. 3a). The distribution of the log abundances of *P. falcatus* appeared to have a wider ecological range, while its large abundance was detected around a border in the south direction, including a few locations in the central parts of the study area (Fig. 3b). The spatial distribution of the log abundances of *F. decipenses* and *C. africana* appeared to be confined to the eastern parts (Fig. 3c, and d). The log abundance of *D. abyssinica* was found denser in the southeastern parts of the study area (Fig. 3e). The distribution of log abundances of *C. macrostachyus*, and *S. guineense* were sparsely distributed (Fig. 3f–h).

Discussion

Impacts of multivariate variables on the abundance of woody species

Effects of elevation on the abundances of species

The elevation is one of the abiotic factors which controls the spatial distribution of plants at various spatial

Table 4 Parameter estimates of predictor variables on the log abundances of eight woody species

The GLM parameter estimates and significant level under each woody species								
Variables	*C. arabica*	*P. falcatus*	*C. macrostachyus*	*C. africana*	*S. guineense*	*O. capensis*	*D. abyssinica*	*F. decipenses*
pH	0.41	0.03	0.5*	− 0.11	0.03	0.30	0.05	− 0.16
CEC	4.45*	− 0.46	1.58	1.52	0.16	− 0.61	0.13	0.39
P	− 0.93*	− 1.35*	− 0.4	− 1.7*	− 0.49	0.78	− 0.74	− 1.61**
K	− 1.09**	− 0.9**	0.22	− 0.6*	0.12	− 0.06	0.26	− 1.19**
Ca	1.89*	− 0.28	0.83	1.35*	− 0.16	− 1.17*	0.45	1.14*
Mg	− 242**	82.87	− 24.9	− 105*	− 3.37	− 38.15	− 50.77	− 34.67
SOM	76.64	0.82	31.6	147.8*	11.18	93.87	96.92	5.84
N	− 12.58	1.90	− 7.47	21.7*	− 7.07	− 6.39	14.73**	− 4.61
C/N	− 17.18*	21.8**	− 0.36	14.1*	8.42	32.71**	17.06*	− 1.8
Clay	3.2*	− 0.4	0.93	0.69	− 0.17	− 0.41	0.18	− 1.00
Silt	7.78*	− 2	− 2.73	− 3.68	− 0.26	− 0.59	0.69	3.75
Sand	12.36*	− 1.97	0.68	− 3.19	− 2.29	− 4.2	0.74	0.84
Temperature	− 0.6	1.2	− 1.37*	0.71	1.38*	0.07	− 0.43	0.87
Precipitation	− 4.78**	1.18*	− 1.6*	− 3.1*	0.85	2.58**	− 3.73**	− 1.94**
Elevation	− 3.05**	0.99*	− 1.5**	− 1.9*	− 0.25	0.25	− 2.34**	− 1.93*
Slope	− 0.53	− 0.19	− 0.04	− 0.2	0.01	− 0.03	0.52*	− 0.17

Significant at ** < 0.01; Significant * at < 0.05

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 9 of 14

Table 5 Prediction accuracies of the ANN, RF, and GLM models using MSE, AICs and R^2

Species	Prediction models with associated prediction accuracy estimators					
	ANN	RF	GLM	ANN	RF	GLM
	MSE			AICs	R^2 %	R^2 %	R^2 %
C. arabica	1.19	0.83	3279	41	58	47
P. falcatus	0.49	0.39	2336	42	32	23
C. macrostachyus	0.52	0.52	2471	8	5	7
C. africana	0.65	0.56	2750	48	48	40
S. guineense	0.38	0.38	2094	7	6	1
O. capensis	0.46	0.49	2448	29	23	21
D. abyssinica	0.54	0.6	2543	29	29	29
F. decipenses	0.44	0.28	2067	59	71	64
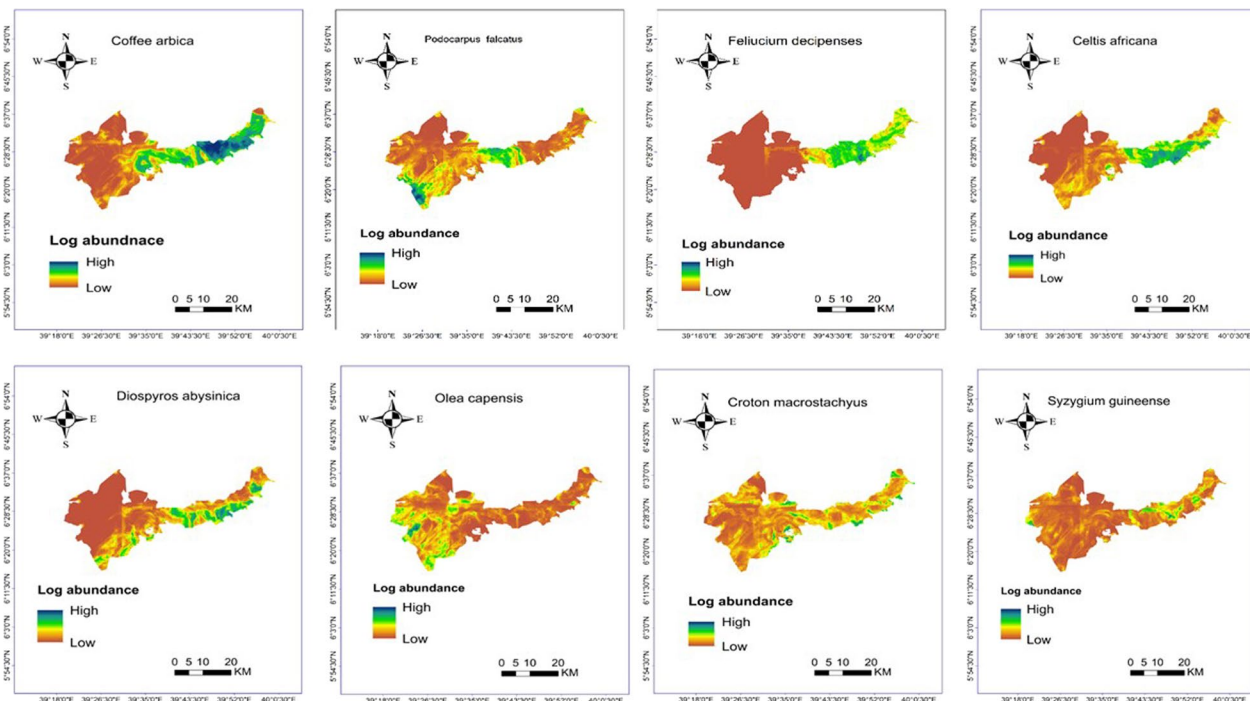
Fig. 3 Spatial distribution maps produced by the RF model of the log abundances of the eight woody species

locations (Nguyen et al. 2015; Ahmed et al. 2022). The ANN model in this study indicated elevation could exert a negative impact on the log abundances of all species. The GLM revealed a similar observation except a positive impact on the log abundance of *P. falcatus* (Table 4). The negative impact of elevation on the log abundance of plant species mainly relates to the declined temperature with an increase in elevation that creates unfavorable conditions for plant growth (Rahman et al. 2021; Ahmed et al. 2022). This is evident that elevation in this study area showed a strong negative correlation with temperature (Annex 1). The positive effect of elevation on the log abundance of *P. falcatus* can suggest this species might be less susceptible to the effect of elevation (Lulekal et al. 2008). Eventually, the impact of elevation on plant species could not be a linear relationship as the effects of elevation can be moderated by slope, and edaphic features (Wakjira 2006; Nguyen et al. 2015; Ahmed et al. 2022). For instance, at a medium elevation range, a steep slope could develop poor soil fertility which leads to less abundance of plant species (Nguyen et al. 2015).

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 10 of 14

Effects of soil organic matter, total nitrogen, and carbon-to-nitrogen ratio

The ANN model appears to provide a positive impact of the SOM on the log abundances of all species (Table 3). The observation with the ANN model looks to be consistent with the understanding that soil organic matter is an important variable to support the vigorous growth of plant species (Booth et al. 2005; Zhou et al. 2019; Yaseen et al. 2022).

Nitrogen is an essential nutrient required in large amounts to construct a plant organ, facilitating metabolism activities, and fruit development (Chapin et al. 1987; Xu et al. 2020). Contrastingly, the ANN model revealed that total nitrogen could exert a negative impact on the log abundances of all woody species except a positive effect on the log abundances of *C. africana* and *D. abyssinica* (Table 3). This observed negative impact is probably related to the bondage of nitrogen in the organic materials which could not be easily available for plant use (Matkala et al. 2020). This may create that nitrogen might be a limiting factor, where the decomposition rate of organic matter is lower (Matkala et al. 2020; Rahman et al. 2021; Devi 2022; Gerke 2022).

The ANN and GLM models appear to display a similar observation in depicting a positive impact of carbon to nitrogen ratio on the log abundances of all woody species, except a negative impact on the log abundances of *C. arabica, F. decipenses,* and *C. macrostachyus*. This positive observation seems to be reasonable as the carbon to nitrogen ratio is an indicator of soil fertility (Zhou et al. (2019).

Effects of magnesium, available phosphorus, and available potassium

The ANN model indicated a positive effect of magnesium on the log abundance of *P. falcatus* (Table 3). This positive impact could be a plausible observation as magnesium is an essential element for plant physiological and biochemical processes (Gransee and Führs 2013; Ishfaq et al. 2022). Contrastingly, the ANN and GLM models showed that magnesium exerts a negative impact on the log abundances of *C. arabica,* and *C. africana.* This negative impact might be explained from two different perspectives. First, magnesium tends to form gypsum and magnesium carbonate compounds which create an unavailable form of magnesium for plants (Gransee and Führs 2013). Second, the availability of magnesium could be reduced by the acidic soil property, particularly for

a soil pH < 6 (Gransee and Führs 2013; Chaudhry et al. 2021).

The variable importance analysis using the RF model indicated that available phosphorus appears to be a strong variable to influence the spatial distribution of the log abundances of *C. arabica*, *F. decipenses*, *P. falcatus*, *C. africana,* and *O. capensis.* The ANN and GLM results in Tables 3 and 4, respectively, showed that phosphorus appears to exhibit negative impacts on the majority of the log abundances of woody species. This seems to deviate from the understanding that phosphorous plays an important role in seed germination and plant growth (Rahman et al. 2021; Nguyen et al. 2015; Mathew et al. 2016). However, the observed negative impacts of phosphorus might have occurred as phosphorous exists in unavailable form by creating dihydrogen phosphate and iron compounds in tropical soil (Vitousek et al 2010; Yaseen et al. 2022), where our study area is also located in the tropics. It has been reported that in the absence of phosphorus species can adapt to low-phosphorus and can grow fast in tropical areas (STRI 2018; Yaseen et al. 2022).

On the other hand, the ANN model in this study shows a positive effect of phosphorus on the log abundances of *C. arabica*, and *F. decipenses,* while the GLM indicates the opposite (Table 4). These opposing observations support Antúnez (2022) who has reported that the use of more than one model is important to check the reliability and consistency of the impacts of abiotic factors on the abundance of woody species under different models.

Available potassium is identified as one of the top variables to influence the spatial distribution of the log abundances of *C. arabica*, *F. decipenses, C. africana*, and *P. falcatus* (Fig. 3)*.* The ANN model identified that available potassium could exert a positive effect on the log abundances of *C. arabica, C. macrostachyu*s, and *S. guineense.* This positive impact could be directly related to the roles of potassium which helps to enhance the growth of plants (Nguyen et al. 2015; Xu et al. 2020). Phosphorus is presumed to show a positive effect on plants as it is very essential for plant growth and the transfer of energy in plant cells (Long et al 2018; Yaseen et al. 2022). Differently, The ANN and GLM models showed that available potassium exerted a negative impact on the log abundances of *P. falcatus, F. decipenses,* and *C. africana.* This negative impact might be attributed to the formation of poorly soluble dihydrogen phosphate concentrations in the topsoil of tropical areas (Matkala et al. 2020; Yaseen et al. 2022). Besides, the impact of potassium might

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 11 of 14

be either positive (Nguyen et al. 2015) or negative, depending on the types of plant species (Mathew et al. 2016).

Effects of climatic factors on the abundances of species

The ANN model in this study showed a positive effect of precipitation on the log abundances of *C. arabica*, and *F. decipenses*; but the GLM displayed a negative impact. The GLM finding looks to be logical as increasing precipitation in the natural forest tends to develop a cooling effect which creates unfavourable conditions for the growth of *C. arabica*, while this species prefers warmer temperatures (Wakjira 2006). The ANN and GLM models showed that precipitation could induce a negative impact on the log abundances of many species. This negative impact might be linked to various reasons. First, extreme precipitation may create water saturation in the soil which may affect nutrient availability to plant roots (Stan and Sanchez-Azofeifa 2019). Second, high precipitation can cause soil erosion which may result in poor soil fertility that limits the growth of woody species (Wakjira 2006).

Concerning the impact of temperature on individual plants, the ANN and GLM models look to fail in detecting a similar impact of temperature on the log abundance of many species, except that both models showed a positive impact of the temperature on the log abundance of *S. guineense* and a negative impact on the log abundance of *C. macrostachyus*. The temperature has been widely reported to show a positive impact on plant growth and distribution (Amissah et al. 2014; Stan and Sanchez-Azofeifa 2019), despite that increase in temperature could negatively affect the physiological processes of plant species (Clark 2004; Yang et al. 2006).

Spatial distribution of the abundance of woody species

The spatial distribution pattern of the log abundance of *C. arabica* looks to be denser in the eastern parts, where the elevation range is moderate as compared to a higher elevation in the western parts. This may imply *C. arabica* could be sensitive to the effect of elevation as increasing in elevation directly drops the temperature to the extent, where it would be less suitable for the growth of this species. Consistent with this, *C. arabica* in Ethiopia normally grows within elevation ranges of 940–2400 m asl (Wakjira 2006; Lulekal et al. 2008; Schmitt et al. 2009), and a suitable mean annual temperature range from 15 to 20 °C (Wakjira 2006).

The spatial distribution of the log abundance of *P. falcatus* appears to distribute widely in different parts of the study area (Fig. 3b) and the GLM revealed that the log abundance of *P. falcatus* does not show significant interactions with a majority of abiotic factors (Table 4). These findings support us to argue the abundance of woody species with a wider geographical distribution may not be significantly influenced by different environmental factors. The ecological implication of this observation depicts that *P. falcatus* might not be sensitive to the variability of abiotic factors and could grow under different environmental gradients.

Spatial distributions of the log abundances of *F. decipenses*, *C. africana*, and *D. abyssinica* show dense patterns in the eastern parts of the study area (Fig. 3c–e), where the area is characterized by a hot temperature as compared to the western parts. This observation implies that the ecological requirement of these three species is limited to a hotter location. Exceptionally, the spatial distribution of the log abundance of *O. capensis* appears to be denser in the southwestern parts of the study area (Fig. 3f). This may suggest *O. capensis* requires a higher elevation which could be characterized by cooler temperatures and higher precipitation.

Conclusions

Elevation, precipitation, temperature, clay content, and available potassium were identified as the top abiotic factors to influence the abundance of woody species in this study area. The impacts of these factors appeared either positive or negative with species. This indicates different woody species exhibit various ecological responses to the effects of different abiotic factors. The spatial distribution pattern of the log abundance of *P. falcatus* in this study seems to indicate species with a wider ecological range might be less sensitive to the impacts of abiotic factors. The log abundances of *C. arabica*, *F. decipenses*, *C. africana*, and *D. abyssinica* could appear more sensitive to the cold temperature, whereas *O. capensis* seems less sensitive to the cold temperatures and it grows at a high elevation area in western parts of the study area. This empirical finding could be useful to identify ecological requirements for those identified woody species to enhance restoration activities. The spatial distribution maps of these species could further help to differentiate the locations with high or low species richness to guide appropriate conservation measures.

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 12 of 14

Annex 1: Pearson correlation coefficients between various abiotic factors

Variables	pH	CEC	P	K	Ca	Mg	OM	N	C/N	clay	Sand	Silt	Temperature	Precp	Elevation	Slope
pH	1	− 0.138**	− 0.003	− 0.061*	− 0.140**	0.053	0.503**	0.072*	0.338**	0.143**	0.104**	− 0.106**	0.039	− 0.007	− 0.016	0.052
CEC	− 0.138**	1	0.079**	0.136**	0.858**	0.202**	− 0.01	− 0.104**	− 0.037	− 0.867**	− 0.919**	0.921**	− 0.059*	− 0.131**	0.039	0.080**
P	− 0.003	0.079**	1	0.016	0.129**	0.107**	− 0.088**	− 0.038	− 0.041	− 0.113**	− 0.085**	0.090**	− 0.011	− 0.01	0.072*	0.082**
K	− 0.061*	0.136**	0.016	1	0.297**	0.049	− 0.252**	− 0.064*	− 0.180**	− 0.320**	− 0.101**	0.110**	− 0.745**	0.639**	0.717**	0.326**
Ca	− 0.140**	0.858**	0.129**	0.297**	1	0.286**	− 0.079**	− 0.108**	− 0.089**	− 0.986**	− 0.811**	0.816**	− 0.216**	− 0.03	0.216**	0.116**
Mg	0.053	0.202**	0.107**	0.049	0.286**	1	0.056	− 0.014	0.018	− 0.555**	− 0.521**	0.522**	− 0.031	0.024	0.058	0.124**
OC	0.513**	− 0.009	− 0.087**	− 0.252**	− 0.079**	0.056	0.999**	0.349**	0.821**	0.079**	− 0.008	0.005	0.221**	− 0.174**	− 0.206**	− 0.006
OM	0.503**	− 0.011	− 0.088**	− 0.252**	− 0.079**	0.056	1	0.348**	0.823**	0.080**	− 0.008	0.004	0.220**	− 0.172**	− 0.206**	− 0.006
N	0.072*	− 0.104**	− 0.038	− 0.064*	− 0.108**	− 0.014	0.348**	1	0.199**	0.105**	0.089**	− 0.094**	0.02	0.018	− 0.035	− 0.058
C/N	0.338**	− 0.037	− 0.041	− 0.180**	− 0.089**	0.018	0.823**	0.199**	1	0.089**	0.034	− 0.039	0.161**	− 0.108**	− 0.147**	0.022
clay	0.143**	− 0.867**	− 0.113**	− 0.320**	− 0.986**	− 0.555**	0.080**	0.105**	0.089**	1	0.858**	− 0.864**	0.234**	− 0.001	− 0.221**	− 0.123**
Sand	0.104**	− 0.919**	− 0.085**	− 0.101**	− 0.811**	− 0.521**	− 0.01	0.089**	0.034	0.858**	1	− 0.994**	0.032	0.126**	− 0.018	− 0.085**
Silt	− 0.106**	0.921**	0.090**	0.110**	0.816**	0.522**	0.004	− 0.094**	− 0.039	− 0.864**	− 0.994**	1	− 0.037	− 0.126**	0.023	0.080**
Temp	0.039	− 0.059*	− 0.011	− 0.745**	− 0.216**	− 0.031	0.220**	0.02	0.161**	0.234**	0.032	− 0.037	1	− 0.857**	− 0.755**	− 0.328**
Precp	− 0.007	− 0.131**	− 0.01	0.639**	− 0.03	0.024	− 0.172**	0.018	− 0.108**	− 0.001	0.126**	− 0.126**	− 0.857**	1	0.702**	0.337**
Elv	− 0.016	0.039	0.072*	0.717**	0.216**	0.058	− 0.206**	− 0.035	− 0.147**	− 0.221**	− 0.018	0.023	− 0.755**	0.702**	1	0.404**
Slope	0.052	0.080**	0.082**	0.326**	0.116**	0.124**	− 0.01	− 0.058	0.022	− 0.123**	− 0.085**	0.080**	− 0.328**	0.337**	0.404**	1

Abbreviations
ANN   Artificial neural networks
DBH   Diameter at breast height
GLM   Generalized linear models
RF   Random forest
SDMs   Species distribution models

Author contributions
The first author was entirely involved in field data collection, data entry, and data analysis. The second author provided a strong analytical analysis component. The third author mainly provided critical review and feedback during the manuscript preparation. All authors read and approved the final manuscript.

Availability of data and materials
All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
We the authors declare that we have no competing interests.

References
Ahmadzadeh E, Lee J, Moon I (2017) Optimized neural network weights and biases using particle swarm optimization algorithm for prediction applications. J Korea Multimed Soc 20(8):1406–1420. https://doi.org/10.9717/kmms.2017.20.8.1406

Ahmed S, Lemessa D, Seyum A (2022) Woody Species Composition, Plant Communities, and Environmental Determinants in Gennemar Dry Afromontane Forest, Southern Ethiopia. Hindawi Scientific Article. https://doi.org/10.1155/2022/797043

Aksu G, Güzeller CO, Eser MT (2019) The effect of the normalization method used in different sample sizes on the success of artificial neural network model. Int J Assess Tool Educ 6(2):170–192. https://doi.org/10.21449/ijate.479404

Amissah L, Mohren GM, Bongers F, Hawthorne WD, Poorter L (2014) Rainfall and temperature affect tree species distribution in Ghana. J Tropic Ecol 30:435–446. https://doi.org/10.1017/S026646741400025X

Antúnez P (2022) Main environmental variables influencing the abundance of plant species under risk category. J For Res 33:1209–1217. https://doi.org/10.1007/s11676-021-01425-6

Asefa M, Ca M, He Y, Mekonnen E, Song X, Yang J (2020) Ethiopian vegetation types, climate and topography. Plant Divers 42:302–311. https://doi.org/10.1016/j.pld.2020.04.004

Ayele G, Hussein H, Mersha A (2019) Land use land cover change detection and deforestation modelling: in Delomena District of Bale Zone, Ethiopia. J Environ Prot 10(4):532–561. https://doi.org/10.4236/jep.2019.104031

Barker NK, Slattery SM, Darveau M, Cumming SG (2014) Modeling distribution and abundance of multiple species: different pooling strategies produce similar results. Ecosphere 5(12):158. https://doi.org/10.1890/ES14-00256

Bedane *et al. Ecological Processes*   (2023) 12:10

Page 13 of 14

Bedane GA, Feyisa GF, Senbeta F (2022) Spatial distribution of above-ground carbon density in Harana Forest, Ethiopia. Ecol Process 11:4. https://doi.org/10.1186/s13717-021-00345-x

Booth CA, Fullen MA, Jankauskas B, Jankauskiene G (2005) The role of soil organic matter content in soil conservation and carbon sequestration studies: case studies from Lithuania and the UK. Sustainable Development and Planning II, Vols. 1 and 2. WIT Trans Ecol Enviro 84:463–473

Cardona NU, Blair ME, Londono MC, Loyola R, Velásquez-Tibatá J, Morales-Devia H (2019) Species distribution modeling in latin america: a 25-year retrospective review. Trop Conserv Sci 12:1–19. https://doi.org/10.1177/1940082919854058

Chapin FS, Bloom AJ, Field CB, Waring RH (1987) Plant responses to multiple environmental factors, physiological ecology provides tools for studying how interacting environmental resources to control plant growth. Bioscience 37(1):49–57

Chaudhry AH, Nayab S, Hussain SB, Ali M, Pan Z (2021) Current understandings on magnesium deficiency and future outlooks for sustainable agriculture. Int J Mol Sci 22:1819. https://doi.org/10.3390/ijms22041819

Chen Y, Yang X, Yang Q, Li D, Long W et al (2014) Factors affecting the distribution pattern of wild plants with extremely small populations in Hainan Island, China. PLoS ONE 9(5):e97751. https://doi.org/10.1371/journal.pone.0097751

Chen WY, Chan YJ, Lim JW, Liew CS, Mohamad MH, Usman A, Lisak G, Hara H, Tan WN (2022) Artificial neural network (ANN) modelling for biogas production in pre-commercialized integrated anaerobic-aerobic bioreactors (IAAB). Water 14:1410. https://doi.org/10.3390/w14091410

Clark DA (2004) Sources or sinks? The responses of tropical forests to current and future climate and atmospheric composition. Philos Trans R Soc B-Biol Sci 359:477–491. https://doi.org/10.1098/rstb.2003.1426

Clément L, Catzeflis F, Richard-Hansen C, Barrioz S, Thoisy B (2014) Conservation interests of applying spatial distribution modeling to large vagile Neotropical mammals. Trop Conserv Sci 7(2):192–213. https://doi.org/10.1177/194008291400700203

Condit R, Engelbrecht MJ, Pinob D, Péreza R, Turnera BL (2013) Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. Natl Acad Sci USA 110(13):5064–5068. https://doi.org/10.1073/pnas.1218042110

Devi AS (2022) Influence of trees and associated variables on soil organic carbon: a review. J Ecol Environ 45:40–53. https://doi.org/10.1186/s41610-021-00180-3

Dunn PK, Smyth GK (2018) Generalized Linear Models with Examples in R. Springer Texts in Statistics, Springer Science Business Media, LLC. https://doi.org/10.1007/978-1-4419-0118-7_5

Feng C, Wang H, Lu N, Chen T, He H, Lu Y, Tu XM (2014) Log-transformation and its implications for data analysis. Shanghai Arch Psychiatry 26(2):105–109. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Feng G, Huang J, Xu Y, Li J, Zang R (2021) Disentangling environmental effects on the tree species abundance distribution and richness in a subtropical forest. Front Plant Sci 12:622043. https://doi.org/10.3389/fpls.2021.622043

Gerke J (2022) The central role of soil organic matter in soil fertility and carbon storage. Soil Syst 6:33. https://doi.org/10.3390/soilsystems6020033

Goldburd M, Khare A, Tevet D, Guller D (2020) Generalized linear models for insurance rating. CAS Monograph Series number 5, 2nd Edition, the Casualty Actuarial Society

Gowda DM (2011) Probability models to study the spatial pattern, abundance and diversity of tree species. Conference on Applied Statistics in Agriculture. https://doi.org/10.4148/2475-7772.1048

Gransee A, Fuhrs H (2013) Magnesium mobility in soils as a challenge for soil and plant analysis, magnesium fertilization and root uptake under adverse growth conditions. Plant Soil 368:5–21. https://doi.org/10.1007/s11104-012-1567-y

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157(2–3):89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Husen A, Mishra VK, Semwal K, Kumar D (2012) Biodiversity status in Ethiopia and challenges. In: Bharati KP, Chauhan A, Kumar P (eds) Environmental pollution and biodiversity, vol 1. Discovery Publishing House Pvt Ltd., New Delhi, pp 31–79

Ishfaq M, Wang Y, Yan M, Wang Z, Wu L, Li C, Li X (2022) Physiological essence of magnesium in plants and its widespread deficiency in the farming

system of China. Front Plant Sci 13:802274. https://doi.org/10.3389/fpls.2022.802274

Kapwata T, Gebreslasie MT (2016) Random forest variable selection in spatial malaria transmission modeling in Mpumalanga Province, South Africa. Geospat Health 11:251–262. https://doi.org/10.4081/gh.2016.434

Kewessa G, Tiki L, Nigatu D, Datiko D (2019) Effect of forest coffee management practices on woody species diversity and composition in bale eco-region, Southeastern Ethiopia. Open J Forest 9(4):265–282. https://doi.org/10.4236/ojf.2019.94015

Kukreja H, Bharath N, Siddesh CS, Kuldeep S (2016) AN Introduction to an artificial neural network. Int J Adv Res Innov Ideas Educ 1:27–30

Long C, Yang X, Long W, Li D, Zhou W, Zhang H (2018) Soil nutrients influence plant community assembly in two tropical coastal secondary forests. Trop Conserv Sci 11:1–9. https://doi.org/10.1177/1940082918817956

Lopatin J, Dolos K, Hernández HJ, Galleguillos M, Fassnacht FE (2016) Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in Central Chile. Remote Sens Environ 173:200–210. https://doi.org/10.1016/j.rse.2015.11.029

Lulekal E, Kelbessa E, Bekele T, Yineger H (2008) Plant species composition and structure of the Mana Angetu Moist montane Forest, South-Eastern Ethiopia. J East Afr Nat History 97(2):165–185. https://doi.org/10.2982/0012-8317-97.2.165

Matkala L, Salemaa M, Bäck J (2020) Soil total phosphorus and nitrogen explain vegetation community composition in a northern forest ecosystem near a phosphate massif. Biogeosciences 17:1535–1556. https://doi.org/10.5194/bg-17

Nelson A (2012) General Management Planning for the Bale Mountains National Park, Walia-Special Edition on the Bale Mountains

Newbold T (2009) The value of species distribution models as a tool for conservation and ecology in Egypt and Britain. Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Nguyen TV, Mitlöhner R, Bich NV, Do TV (2015) Environmental factors affecting the abundance and presence of tree species in a tropical lowland limestone and non-limestone forest in Ben En National Park, Vietnam. J Forest Environ Sci 31(3):177–191. https://doi.org/10.7747/JFES.2015.31.3.177

Olden JD, Jackson DA (2002) Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol Model 154:135–150. https://doi.org/10.1016/S0304-3800(02)00064-9

Osborne JW (2002) Notes on the use of data transformations. Pract Assess Res Eval 8:6. https://doi.org/10.7275/4vng-5608

Puheim M, Madarász L (2014) Normalization of Inputs and Outputs of Neural Network Based Robotic Arm Controller in Role of Inverse Kinematic Model, a conference paper

Qian H (2013) Environmental determinants of woody plant diversity at a regional scale in China. PLoS ONE 8:e75832. https://doi.org/10.1371/journal.pone.0075832

Rahman A, Khan SM, Ahmad Z, Alamri S et al (2021) Impact of multiple environmental factors on species abundance in various forest layers using an integrative modeling approach. Glob Ecol Conserv 29:e01712. https://doi.org/10.1016/j.gecco.2021.e01712

Rion V (2010) Modeling the plant species richness: a comparison between two approaches. Master thesis of Science in Behavior, Evolution and Conservation, Lausanne University

Sakate DM, Kashid DN (2016) A new robust model selection method in GLM with application to ecological data. Environ Syst Res 5:9. https://doi.org/10.1186/s40068-016-0060-7

Scarnati L, Attorre F, Farcomeni A, Francesconi F, Sanctis MD (2009) Modelling the spatial distribution of tree species with fragmented populations from abundance data. Comm Ecol 10(2):215–224. https://doi.org/10.1556/comec.10.2009.2.12

Schmitt CB, Denich M, Boehmer HJ (2009) Plant diversity and conservation of Afromontane forest with *Coffea arabica* in the Bonga region (SW Ethiopia). In: van der Burgt X, van der Maesen J, Onana JM (eds) Systematics and conservation of African plants. Royal Botanic Gardens, Kew, pp 679–690

Seo E, Hutchinson RA, Fu X et al (2021) StatEcoNet: Statistical Ecology Neural Networks for Species Distribution Modeling. The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), Association for the Advancement of Artificial Intelligence (www.aaai.org)

Sola J, Sevilla J (1997) Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Trans Nucl Sci 44(3):1464–1468. https://doi.org/10.1109/23.589532

STRI (Smithsonian Tropical Research Institute) (2018) Diverse Tropical Forests Grow Fast Despite Widespread Phosphorus Limitation, ScienceDaily

UNFCCC (2015) Measurements for estimation of carbon stocks in afforestation and reforestation project activities under the clean development mechanism, a field manual. Retrieved 01 July 2021.

Vincenzia S, Zucchetta M, Franzoi P, Pellizzato M, Pranovi F, Deleo GA, Torricelli P (2011) Application of a Random Forest algorithm to predict the spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. Ecol Model 222(8):1471–1478. https://doi.org/10.1016/j.ecolmodel.2011.02.007

Vitousek PM, Porder S, Houlton BZ, Chadwick OA (2010) Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. Ecol Appl 20:5–15. https://doi.org/10.1890/08-0127.1

Wakjira FS (2006) Biodiversity and ecology of Afromontane rainforests with wild *Coffea arabica* L. populations in Ethiopia. Ecology and Development Series No. 38

Waldock C, Stuart-Smith RD, Albouy C et al (2022) A quantitative review of abundance-based species distribution models. Ecography. https://doi.org/10.1111/ecog.05694

Xu X, Du X, Wang F, Sha J, Chen Q, Tian G, Zhu Z, Ge S, Jiang Y (2020) Effects of potassium levels on plant growth, accumulation and distribution of carbon, and nitrate metabolism in apple dwarf rootstock seedlings. Front Plant Sci 11:904. https://doi.org/10.3389/fpls.2020.00904

Yang Y, Watanabe MF, Lii F, Zhang J, Zhang W, Zhai J (2006) Factors affecting forest growth and possible effects of climate change in the Taihang Mountains, northern China. Forestry 79(1):135–147. https://doi.org/10.1093/forestry/cpi062

Yaseen M, Fan G, Zhou X, Long W, Feng G (2022) Plant diversity and soil nutrients in a tropical coastal secondary forest: association ordination and sampling year differences. Forests 13:376. https://doi.org/10.3390/f13030376

Yudaputra A, Robiansyah I, Sumeru RD (2019) The Implementation of Artificial Neural Network and Random Forest in Ecological Research: Species Distribution Modelling with Presence and Absence Dataset. Research Center for Plant Conservation and Botanic Gardens, Indonesian Institute of Sciences, Jl. Ir. H. Juanda No. 13, Bogor, Indonesia

Zhang L, Huettmann F, Zhang X, Liu S, Sun P, Yu Z, Mi C (2019) The use of classification and regression algorithms using the random forests method with presence-only data to model species distribution. MethodsX 6:2281–2292. https://doi.org/10.1016/j.mex.2019.09.035

Zhou W, Han G, Liu M, Li X (2019) Effects of soil pH and texture on soil carbon and nitrogen in soil profiles under different land uses in Mun River Basin, Northeast Thailand. PeerJ 7:e7880. https://doi.org/10.7717/peerj.7880

Publisher's Note