Range Mgmt. & Agroforestry 33 (2) : 123-128, 2012
ISSN 0971-2070

Predicting potential distributions of *Zygophyllum eurypterum* by three modeling techniques (ENFA, ANN and logistic) in North East of Semnan, Iran

Mohammad Ali Zare Chahouki* and Lyla Khalasi Ahvazi
Department of Rehabilitation of Arid and Mountainous Regions, University of Tehran, Iran, P.O. Box: 31585-4314
*Corresponding author email: mazare@ut.ac.ir
Received 22nd February 2012 Accepted 7th August, 2012

Abstract
The paper investigates the use of 'Ecological niche factor analysis' (ENFA) method for modeling *Zygophylum eurypterum* species geographic distributions with presence-only data and 'Artificial Neural Network' (ANN), 'Logistic Regression' (LR) methods for investigating of *Zygophyllum eurypterum* species distribution with presence-absence data in North East of Semnan province. Plant density and cover, soil texture, available moisture, pH, electrical conductivity (EC), organic matter, lime, gravel and gypsum contents and topography (elevation, slope and aspect) are the variables sampled using the randomized systematic method. Within each vegetation type, the samples were collected using 15 quadrates placed at an interval of 50 m along three 750 m transects. To map soil characteristics, geostatistical method was used. The back propagation Neural Network in MATLAB software was used to generate the ANN model with one input, one hidden and one output layer and the Logistic Regression analysis was done in SPSS software and based on obtained models (ANN and LR methods) predicted maps in Arc Map software were created. The accuracy of the predicted maps (prepared by ENFA, ANN and LR) were tested with actual vegetation maps. Kappa coefficients prepared by these methods show good accordance with actual vegetation map prepared for the study area. The results of ENFA method show that 25200 hectares or 34 percent of study site is potential habitat of *Z. eurypterum*. The results also revealed that maps generated using the LR and ANN models for *Z. eurypterum* species have a high accordance with their corresponding actual maps of the study area. This species is distributed in rangeland with alkali-saline soil, high of lime percent, silty-sandy texture and in 1000-2000 meters elevation.

Key words: Artificial Neural Network, Biomapper, Ecological niche factor analysis, Kappa coefficients, Logistic Regression, North East of Semnan, *Zygophyllum eurypterum*

Introduction

Zygophyllum eurypterum is one of the most important range plants often seen as associated species and rarely as the dominant species in rangelands, very critical for soil conservation and grazing pressure in Iran.

Arid and semi arid ecosystems of Iran are exposed to anthropogenic and conservation programmes that require the study of vegetation population and environmental factors for rehabilitation. If we have access to presence/absence (used-vs.-unused) resource units, logistic regression can be used to generate models for vegetation distribution. Here we assign a 0 to sites where the *Z. eurypterum* is absent and a 1 to Sites where the *Z. eurypterum* is present.

Ecological-niche factor analysis (ENFA) is a multivariate approach to study geographic distribution of species on a large scale with only 'presence' data. ENFA compares the ecogeographical predictor distribution for a presence data set consisting of locations where the species has been detected with the predictor distribution of the whole area. Like the Principal Component Analysis, ENFA summarizes all predictors into a few uncorrelated factors retaining most of the information. But in this case, the factors have an ecological meaning: the first factor is the 'marginality', and reflects the direction in which the species niche mostly differs from the available conditions in the global area. Subsequent factors represent the 'specialization'. They are extracted successively by computing the direction that maximizes the ratio of the variance of the global distribution to that of the species distribution. A large part of the information is accounted for by a few of the first factors. The species distribution on these factors is used to compute a habitat suitability index for any set of descriptor values (Hirzel *et al.,* 2001).

The Artificial Neural Networks (ANN) is a promising area of predictive modeling of plant species distribution (Fukuda, 2011, Watts *et al*., 2011). ANNs have been used for a wide variety of applications where statistical methods are traditionally employed.

Logistic Regression (LR) is a kind of Generalized Linear Models (GLM) suitable to analyze a binary response variable (Miller and Franklin, 2002). The dependent variable (presence/absence of the species) is explained by a sum of weighted ecogeographical predictors. The weights are tuned in order to generate the best fit between the model and the calibration data set (Nicholls, 1989).

Different studies have shown that although competition influences the growth and distribution of the plants, but soil characteristics are of high importance in distribution of salt lands plants.

Tan *et al.* (2006) suggest that GLM and ANN models are the most suitable and robust models for studying ecosystems with time-dependent dynamics and periodicities whose frequency are possibly less than the time scale of the data considered. ENFA analyses seem to meet most species modeling requirements (Yee and Mackenzie, 2002).

The main purpose of this research was to investigate the relationship between soil and physiographic characteristics with *Z. eurypterum* species to determine the most important factors affecting the separation of this plant species and then preparing the prediction map three models including ENFA, ANN and Logistic in predicting plant species distribution. Investigating the new method for predicting plants in areas by different characteristics (in soil, topography and climatic) is very important in finding ways for conservation and management practices of *Z. eurypterum* to rehabilitate rangelands.

Materials and Methods

Study area: The study area (Fig.1) was semi arid rangelands located in northeast of the Semnan province, central Iran (35º 53´ N, 54º 24´ E to 35º50´ N, 53º43´ E) with elevation ranging from 1129 to 2260 m a.s.l. Average annual precipitation ranged between 128 mm in the saline lowlands to 275 mm in the mountains. Minimum temperature occurs in December (around -6ºC) while the highest temperature reaches 45ºC in June.

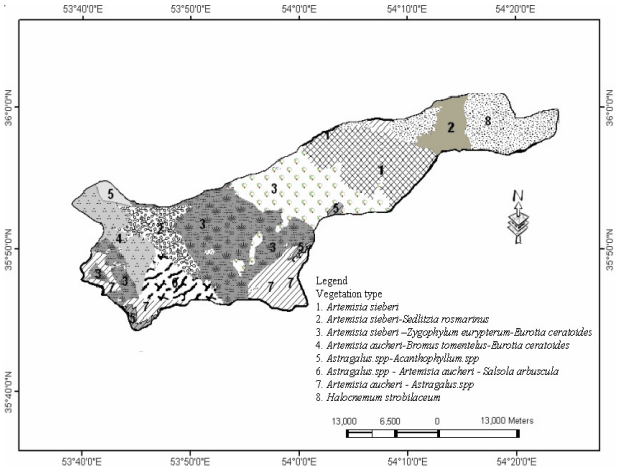

Fig. 1. Location of study area and distribution of vegetation types

Data collection: Based on field surveys dominant vegetation types were determined. Sampling was performed in the key area of each vegetation type using randomized-systematic method. Quadrate size was determined for each vegetation type using minimal area method (Asri, 1995). Soil samples were taken from 0-20 cm and 20-80 cm in starting and ending points of each transect. Measured soil factors included texture, available moisture, soil organic matter (Black, 1979), pH in the saturation extract, electrical conductivity (EC), and lime. The elevation and slope and slope direction were determined at the location of each quadrate.

Data analysis: For plant predictive mapping it is necessary to prepare the maps of all effective factors used in the models (Fig. 2). Topographic data (elevation, slope and aspect) were derived from Digital Elevation Models (DEM) with a resolution of 10 m. characteristics, Geostatistical methods were used to map soil characteristics. Block Kriging method has used by GS$^+$ (comprehensive geostatistics software) and GIS software (ESRI, 2004) to predict soil factor.

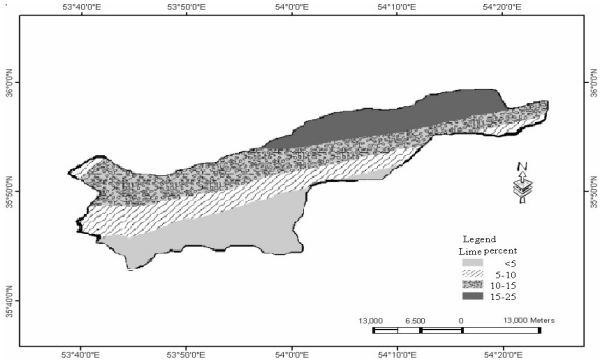

Fig 2. Spatial distribution of lime content in 0-20 cm depth

Three habitat models (ENFA, ANN and LR) with a binomial probability distribution and a logit link were fitted for *Zygophylum eurypterum*, based on 24 topography and soil predictor variables.

ENFA was performed with the BIOMAPPER software (Hirzel *et al.,* 2001). Data layers format exchanged to Raster layers in IDRISI software to entrancing the Biomapper software. Then the predictors were first normalised by the Box–Cox algorithm (Sokal and Rohlf, 1981). Ecological niche factors were then computed on these normalized predictors.

An overall suitability index of the focal cell was computed from a combination of its scores on each factor. In order to account for the differential ecological importance of the factors, we attributed equal weight to marginality and specialization, but, while the entire marginality component goes to the first factor, the specialization component is apportioned among all factors proportionally to their eigenvalue. Repeating this procedure for each cell allows producing a habitat-suitability map, where suitability values range from 0 to 1. To convert this quantitative (or semi-quantitative) map into a presence/absence one, a threshold value may be chosen, above which the cell will be considered as suitable.

Marginality $$M = \frac{|m_G - m_S|}{1.96\sigma_G} \quad (1)$$

Specialization $$S = \frac{\sigma_G}{\sigma_S} \quad (2)$$

By importing the information layers in appropriate model and using necessary statistical analysis in Biomapper software, the map of its potential habitat has been created. All vegetation types were included in the formulation of the ANN model and all 24 environmental variables were included in the procedure. The MLP network is trained using one of supervised learning algorithms, which uses the data to adjust the networks weights and thresholds so as to minimize the error in its predictions on the training set. The back propagation neural network (Cairns, 2001) was used to generate the ANN model with one input, one hidden and one output layer. The mean accuracy based on two validation data sets for each ANN model (7-10 nodes) was compared to determine the ANN model which best classified the data.

Logistic regression (LR) techniques were implemented for plant species predictive modeling. The probability of occurrence of each plant species according to Linear Regression was calculated with respect to the combined effect of site conditions using the following equation:

$$Y = \frac{\exp(LP)}{(1+\exp(LP))} = \frac{\exp(b_0 + b_1x_1 + ... + b_nx_n)}{1+\exp(b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n)}$$

Where b_0 is the constant and exp is an exponential function and where b_1, b_2,..., and bn are the logic coefficients of X_1, X_2,..., and X_n variables respectively**,** in which presence/absence of an object is transformed into a continuous probability ranging from 0 to 1.

The models were calculated with individual selected variables and their combination using SPSS, 15.0. The best model is selected using two criteria: (1) approximate variance explained (Nagelkerka R square) and (2) goodness of fit (Hosmer and Lemeshow test statistics). The accuracy of predicted maps and adequacy of vegetation types mapping were evaluated using the Kappa statistic.

Results and Discussion

ENFA based modeling: Marginality coefficients (Table 1) showed that most important variables are essentially linked to high Lime and PH. The next factors account for some more specialization, mostly regarding Lime frequency and altitude (second factor) as well as Gypsum (third factor), showing some sensitivity to shifts away from their optimal values on these variables.

Suitability map was built from these four factors for the whole of North East of Semnan (Fig 4). The results show that 25200 hectares of study site is potential habitat of *Z. eurypterum* which is 34 percent of study site. To evaluate the verity of this model, Boyce index has been used and model rectitude in this test was determined as 87.2 percent. The mean and the standard deviation of the accuracy assessment were calculated for modal validation.

ANN based modeling: The accuracy of the ANN models was variable and depended on the number of nodes existed in the hidden layer of the model. The ANN models with 7 and 10 nodes had the highest accuracy (i.e. 0.56). Since both the 7- and 10-node models produced the same accuracy (using the average of the two validation data sets), we chose the 7-node model as the best one because it performed better on the training data set than did the 10 node model. The average accuracy for the 7-node ANN model was 0.56 and ê = 0.51. Fig 5 shows the predicted map of *Z. eurypterum* generated using the ANN model.

Table 1. Variance explained by the first four (out of 24) ecological factors, and coefficient values for most important initial variables.

Variables	Depth	Marginality	Specialization1	Specialization2	Specialization3
Gravel	0-20	0.162	0.251	0.1302	-0.231
	20-80	0.25	0.0266	0.0604	0.248
Clay	0-20	-0.0073	0.1502	0.0448	0.179
	20-80	-0.42	-0.2168	0.259	0.1104
Silt	0-20	0.235	-0.191	0.0278	0.257
	20-80	0.132	-0.0028	0.0624	0.269
Sand	0-20	0.231	0.0828	-0.1383	-0.216
	20-80	0.124	0.181	-0.1862	-0.255
Lime	0-20	0.354	0.1644	-0.342	0.0918
	20-80	0.312	0.0101	-0.343	0.163
Organic matter	0-20	-0.0671	-0.0388	0.3744	-0.0263
	20-80	0.1304	0.2962	0.348	-0.0768
Available moisture	0-20	0.2038	-0.241	0.1148	0.244
	20-80	0.1725	0.2399	0.1206	0.287
Gypsum	0-20	-0.0845	0.0925	0.068	0.366
	20-80	-0.0831	0.0925	-0.068	0.243
Electrical conductivity	0-20	-0.0521	0.0957	-0.0693	0.231
	20-80	-0.0672	0.101	-0.0529	0.235
pH	0-20	0.264	-0.679	-0.113	-0.016
	20-80	0.323	0.2747	-0.133	-0.291
Elevation	-	0.235	0.0252	-0.2594	-0.126
Slope	-	-0.379	0.283	0.2219	-0.123

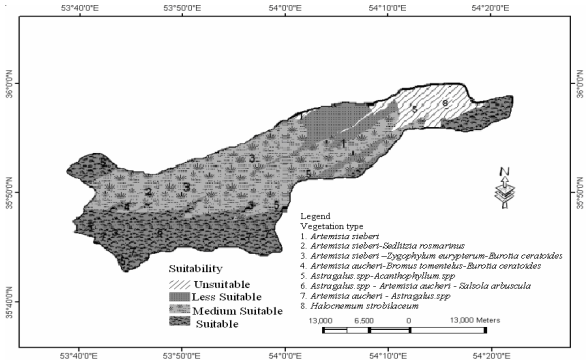

Fig 4. Habitat Suitability map of *Z. eurypterum* using the ENFA model

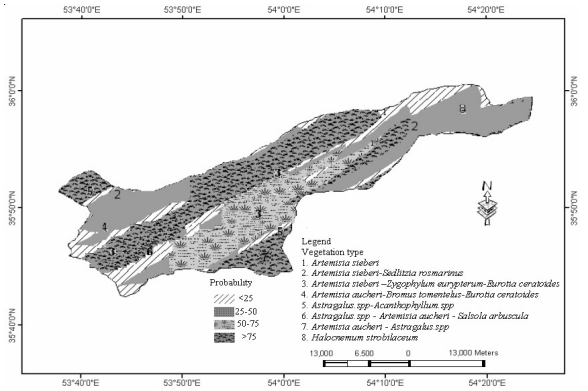

Fig 5. Predicted map of *Z. eurypterum* using ANN model

Logistic Regression based modeling: The predicted occurrence probability of *Z. eurypterum* is expressed *via* the equations 2 in below. Regarding equation (1), the occurrence of *Z. eurypterum* is dependent to the content of clay and gypsum in the soil depth of 0-20. Based on the predictive model obtained using the LR method, predictive vegetation maps were generated in the GIS environment. Fig 6 shows the predicted map of *Z. eurypterum* using the logistic regression model.

$$P\ (\textit{Zygophyllum eurypterum}) = \frac{\text{Exp}\ (-0.854\ \text{clay1} - 18.287\ \text{gypsum1} + 16.97)}{1 + \text{Exp}\ (-0.854\ \text{clay1} - 18.287\ \text{gypsum1} + 16.97)}$$

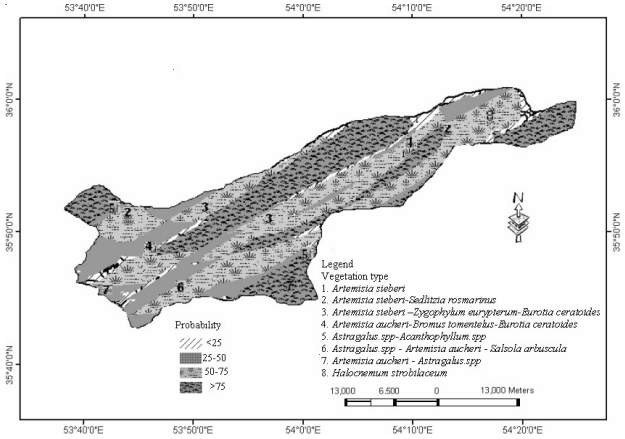

Fig 6. Predicted map of *Z. eurypterum* using LR model

Evaluation: The best measure of agreement between observed and predicted presence-absence is Kappa (ê) statistic. The accuracy of predicted maps and adequacy of vegetation types mapping were evaluated using the Kappa statistic. Kappa coefficients were prepared by these methods, successively include: 0.62, 0.61 and 0.58 which indicate that they have good accordance with actual vegetation map prepared for the study area.

The main factors predicting the occurrence of *Z. eurypterum* were the lime, gypsum, available moisture, texture and elevation percent obtained from analysis of the three models.

Available moisture is the most important factor limiting plant growth and distribution of plants. Jacobson (1997) showed that mycorrhizal colonization of roots was primarily affected by moisture availability. Because of Soil texture effects of plants available moisture and elements and root distribution depth, the soil texture has an important role in the distribution of *Z. eurypterum*. The result showed that *Z. eurypterum* habitat was positioned in the direction of coarse sand and gravely dry sites.

Elevation gradients create varied climates, along with resultant soil differentiation; promote the diversification of plant species (Brown, 2001; Lomolino, 2001). Many studies have investigated species distribution along elevation gradient across habits and taxa (Rahbek, 1997; Austrheim, 2002), as part of efforts to understand ecosystem effects on distribution of vegetation (Tilman and Downing, 1994). Research showed that the better elevation of studied species is 1000-2000m and *Z. eurypterum* habitat showed direct relationship with lime percent.

The research concluded that there was good agreement between each of three models (ENFA, ANN and LR) and actual vegetation map (0.62, 0.61 and 0.58). In this study, the relationship between *Z. eurypterum* habitat selection and niche factors were analyzed on a large scale by using ENFA and then habitat suitability was mapped based on this relationship. The marginality coefficients of ENFA interpret this relation, and the absolute values of coefficients reflect the extent of the preferences for niche factors.

ANNs are non-linear methods that could be preferred when quick and accurate predictions are more important than statistical and/or ecological interpretations. LR and ANN have clear advantages in predicting plant habitat such as *Z. eurypterum*, since they provide the clearest indications of possible causal effects on distribution. Finally, it was proved from analysis of prepared prediction maps in this research that ENFA model is easy, practical and economical in comparative of another models and this model have been used in every ecological studies.

This article has focused on applications in plant ecology. In the future, it appears certain that the need for accurate, powerful, data driven, and highly interpretable methods which enable visualization of large data sets in high-dimensional space will continue (Termansen *et al.,* 2006). We used soil and topography variables to model the spatial distributions and test to what extent the proposed approach captures the spatial pattern of *Z. eurypterum* species in the study area.

Acknowledgements

The authors would like to acknowledge the financial support of university of Tehran for this research under grant number, 7314844/1/3.

References

Asri, Y., 1995. *Phytosociology*. Institute of Forests and Rangelands Press. 198p.

Austrheim, G. 2002. Plant diversity patterns in semi-natural grasslands along an elevational gradient in Southern Norway. *Plant Ecology* 161:193-205.

Black, C. A., 1979. Methods of soil analysis. *American Society of Agronomy* 2: 771-1572

Brown, J., 2001. Mammals on mountainsides: elevational patterns of diversity. *Global Ecology and Biogeography* 10:101-109.

Cairns, D. M., 2001- A comparison of methods for predicting vegetation type. *Plant Ecol.* 156: 3-18.

ESRI. 2004. ArcGIS 9.0. ESRI, A review of methods for the assessment of prediction errors in conservation presence- absence models. *Environmental Conservation*, 24: 38-49.

Fukuda, S. 2011. Assessing the applicability of fuzzy neural networks for habitat preference evaluation of Japanese medaka (*Oryzias latipes*) - *Ecol. Inform.* 6: 286-295.

Hirzel, A. H., V. Helfer and F. Metral, 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145: 111-121.

Jacobson, M. K. 1997. Moisture and substrate stability determine VA-mycorrhizal fungal community distribution and structure in arid grassland. *Journal of Arid Environments.* 35: 59–75.

Lomolino, M. V. 2001. Elevation gradients of species - density: historical and prospective views. *Global Ecology and Biogeography* 10:3-13.

Miller, J. and J. Franklin, 2002. Modeling the distribution of four vegetation alliances using Generalized Linear Models and classification trees with spatial dependence- *Ecol. Model*. 157: 227-247.

Nicholls, A. O., 1989. How to make biological surveys go further with Generalised Linear Models. *Biol. Conserv.* 50: 51-75.

Rahbek, C. 1997. The relationship among area, elevation and regional species richness in neotropical birds. *American Naturalist.* 149:875-902.

Sokal, R. R. and F. J. Rohlf. 1981. *Biometry. The principles and practice of statistics in biological research,* 2nd edn. W. H. Freeman & Company, New York, NY, 550 pp.

Tan, C. O., U. Özesmi, M. Beklioglu, E. Per and B. Kurt, 2006. Predictive models in ecology: Comparison of performances and assessment of applicability- *Ecol. Inform*. 1: 195-211.

Termansen, M., C. J. McClean and C. D. Preston, 2006. The use of genetic algorithms and Bayesian classification to model species distributions. *Ecological Modelling.* 192: 410-424.

Tilman, D. and J. A. Downing. 1994. Biodiversity and stability in grasslands. *Nature* 367: 363-365.

Watts, M. J., Y. Li, B. D. Russell, C. Mellin, S. D. Connell and D. A. Fordham, 2011. A novel method for mapping reefs and subtidal rocky habitats using artificial neural networks. *Ecological Modelling*. 222: 2606-2614.

Yee, T. W. and M. Mackenzie, 2002. Vector generalized additive models plant ecology. *Ecological Modelling.* 157:141-156.