Range Mgmt. & Agroforestry 37 (2) : 142-147, 2016
ISSN 0971-2070

Evaluation of MaxEnt method for habitat distribution modeling of three plant species in Garizat rangelands of Yazd province, Iran

Mohammad Ali ZareChahouki[1*] and Hossein PiriSahragard[2]
[1] Department of Rehabilitation of Arid and Mountainous Regions, University of Tehran, Iran
[2] Range and Watershed Department, University of Zabol, Iran
*Corresponding author e-mail: mazare@ut.ac.ir
Received: 25th January, 2016 Accepted: 6th October, 2016

Abstract
This study aimed to predict geographical distribution of *Tamarix ramosissima*, *Seidlitzia rosmarinus* and *Cornulaca monocantha* in Poshtkouh rangelands and to find the influential variables in the distribution of these species in desert rangelands of central Iran. Eleven environmental factors used to explore the effective environmental variables on given species distribution. Maps of the environmental variables were generated using GIS and Geostatistics facilities. Predictive maps of distribution were produced with maximum entropy method (MaxEnt). Accuracy of model output was assessed by using area under the curve (AUC) and withholding 25 per cent of the data. The agreement of predictive map with actual map was checked by calculating Kappa coefficient. The results indicated that vegetation distribution pattern was mainly related to soil characteristics such as EC, available moisture (AW), lime, organic matter (OM) and elevation. AUC values indicated the high power of MaxEnt to create habitat distribution maps of plant species except *C. monocantha* (*S. Rosmarinus* = 0.98, *T. ramosissima* = 0.99, and *C. monocantha* = 0.78). Correspondence of actual map with predictive map for *S. rosmarinus*, *C. monocantha* and *T. ramosissima* was assessed at very satisfactory (Kappa=0.76), good (Kappa= 0.61) and poor (Kappa= 0.31) level, respectively.

Keywords: AUC, Environmental variables, Geographical distribution, Geostatistics, Kappa

Abbreviations: AUC: Area under the curve; **GIS**: Geographic information system; **MaxEnt:** Maximum entropy; **PVM**: Predictive vegetation modeling

Introduction
The distribution pattern of natural vegetation is associated with four types of environmental factors including climatic, physiographic, edaphic and biotic factors and combination of these factors can affect the establishment of plant species. Data of ecological parameters such as climate, soil, and biotic factor are generally difficult or expensive to measure, soil data are even more difficult to derive, and they tend to be less accurate than pure topographic characteristics (Guisan and Zimmermann, 2000). On the other hand, indirect parameters (*e.g.* topographic variables, elevation, slope, aspect) are most easily measured by remote sensing and are often used because of their good correlation with observed species patterns (Su, 1987; Guisan and Zimmermann, 2000).

Knowledge about the geographical distribution of species is critical for the development, monitoring and restoration of plant species in their natural habitat, selecting conservation sites, and conservation and management of their native habitat (Gaston, 1996; Pearson *et al*., 2007; Cayuela *et al*., 2009; PiriSahragard and ZareChahouki, 2015). In that context, spatially explicit models of species distributions can be useful tools to identify sites where conservation of species concerned are likely to be present. Species distribution modeling (SDM) tools are becoming increasingly popular in ecology and are being widely used in many ecological applications (Elith *et al*., 2006; Peterson, 2007; ZareChahouki and Esfanjani, 2015).

Now a days, a variety of statistical methods are used to predict the geographical distributions of species, such as BIOCLIM, maximum entropy (MaxEnt), DOMAIN, genetic algorithm for rule-set prediction (GARP), generalized linear models (GLM), generalized additive model (GAM) and discriminant analysis (DA) (Elith *et al*., 2006; Guisan *et al*., 2007; Peterson, 2007; Wisz *et al*., 2008; ZareChahouki *et al*., 2012). Different modeling approaches have the potential to yield substantially different predictions, so the choice of the right statistical method in a specific modeling context is an important issue (Segurado and Araújo, 2004; PiriSahragard and ZareChahouki, 2015). Tarkesh and Jetschke (2012) compared the performance of BIOCLIM, GARP and MaxEnt with multivariate adaptive regression spline (MARS), nonparametric multiplicative regression (NPMR) and logistic regression tree (LRT). They demonstrated that MaxEnt and MARS techniques achieved the best.

MaxEnt is a general-purpose method for making predictions or inferences from incomplete information (Pearson *et al*., 2007). The MaxEnt probability distribution has many advantages such as clear mathematical definition working with positive-only examples and usability of both continuous and categorical data, so it can incorporate interactions between different variables and is therefore suitable to analysis (Phillips *et al*., 2006). This study was conducted to assess the capability of Maximum Entropy model to predict the spatial distribution of plant species habitats and to analyse the ecological relationship between range plant species distribution and their environment in Garizat rangelands of Yazd province, central Iran.

Materials and Methods

***Study area*:** This research was conducted in Garizat rangelands (94130 ha), located in the southern of Garizat region of the Yazd province in center of Iran (31º 04´ 53´´N, 53º 40´ 04´´E to 31º 21´ 26´´N, 54º 14´ 58´´E) (Fig. 1). The maximum elevation is 2100 m and the minimum elevation is 1400 m. Average annual precipitation ranges from 200 mm to 45 mm (ZareChahouki *et al.,* 2010).

***Data collection and preprocessing of data*:** Homogeneous units were first delineated using basic maps of the study area (digital elevation, aspect, slope and geology maps in the scale of 1:25000). Vegetation sampling (plant species list, canopy cover per cent) was carried out using randomized-systematic method. Four transects with 200-1000 m in length in each unit was conducted for vegetation sampling (PiriSahragard and ZareChahouki, 2015). Quadrat size was determined with minimal area method which varied from 2 to 25 m^2 depending on the plant species. The sample size was calculated with statistical method in each unit. Vegetation was sampled in 60 quadrats with respect to vegetation cover variation. In order to soil sampling, 65 soil samples were collected in depths of 0-30 cm and 30-80 cm. Soil characteristics consisting of gravel per cent, texture, saturation moisture, available water, lime, gypsum, organic matter, acidity (pH), electrical conductivity (EC) and soluble solute (Na^+, Ca^{2+}, Mg^{2+}, K^+, Cl^-, Co_3^{2-}, HCo_3^- and So_4^{2-}) were measured using standard methods. Due to high precision of the recorded data, correlation within variables was assessed, and those with high correlation (>0.80) were removed. For geostatistical analysis and creating the maps, ArcGIS 10 and GS^+ 5.1.1 were used. In order to reduce the number of input for MaxEnt model, principal component analysis (PCA) was conducted on vegetation and environmental variables matrix using the program PC-ORD (Hosseini *et al*., 2013). Based on the result of PCA, the different environmental variables selected as input for MaxEnt model were elevation, aspect, slope, gypsum (gyps), lime, available moisture (AW), electrical conductivity (EC), clay, gravel, organic matter (OM) and acidity (pH). All environmental factors converted to ASCII raster grids and species occurrence coordinates were converted to decimal degrees in ArcGIS 9.3.

***Model building*:** Modeling was performed after preparation of the environment variable maps and their entry in the maximum entropy software. MaxEnt software generates an estimate of probability of presence of the species ranging from 0 to 1, where 0 stands for the lowest and 1 for the highest probability. Because of the continuous output of MaxEnt, it is necessary to determine an optimal threshold for determining the presence or absence of the target species (Phillips *et al*., 2006). After determining the optimal threshold using equal sensitivity and specificity method, presence or absence species maps were obtained and their compliance with actual maps was investigated through calculation of the kappa coefficient in the Idrisi 32 software.

***Model validation*:** Validation was made by dividing a dataset into two subsets, the first one (training data) that was used to build a model, typically comprising 75 per cent of all data and the other (test data) that was used to test the model comprising 25 per cent of all data (Death and Fabricius, 2000). The area under the curve (AUC) of receiver operating characteristic (ROC) function was used for evaluation of the discrimination ability (Fielding and Bell, 1997). The AUC as threshold- independent measure ranges from 0.5 for an uninformative model to 1 for perfect discrimination. In order to evaluate the importance of each environmental predictor variable, the Jackknife operation was used.

Results and Discussion

Model accuracy and predictive maps of plant species habitat: After PCA analysis, eleven environmental variables were selected for MaxEnt modeling. Model comparison by AUC revealed that based on AUC value, performance of modelling techniques for *T. ramosissima* and *S. rosmarinus* was bigger than *C. monocantha*. In other words, the accuracy of *T. ramosissima* model (0.99) was the best, followed by *S. Rosmarinus* (0.98) and *C. monocantha* (0.78). The AUC of ROC plots for the three

Table 1. AUC value and maps agreement between predictive and actual maps of habitat in the study area

No	Vegetation type	AUC	Kappa value	Level of agreement
1	*T. ramosissima*	0.99	0.31	Poor
2	*S. rosmarinus*	0.98	0.76	Very good
3	*C. monocantha*	0.78	0.61	Good

species are shown in Table 1. Hence, *C. monocantha* habitat could not be separated with high accuracy by the MaxEnt model. Based on results, AUC value for *T. ramosissima* and *S. rosmarinus* was bigger than *C. monocantha* species. Result showed AUC values tended to be lower for species that had widespread distribution, such as *C. monocantha*. This was in accordance with results of Ardestani *et al*. (2015). Furthermore, Phillips *et al.* (2006) compared MaxEnt predictions with the genetic algorithm for rule-set prediction (GARP) and found that MaxEnt almost always possessed higher AUC, indicating better discrimination of suitable versus unsuitable area for the species. MaxEnt also performed better than several other models when presented with low sample sizes (Hoffman *et al.,* 2008).

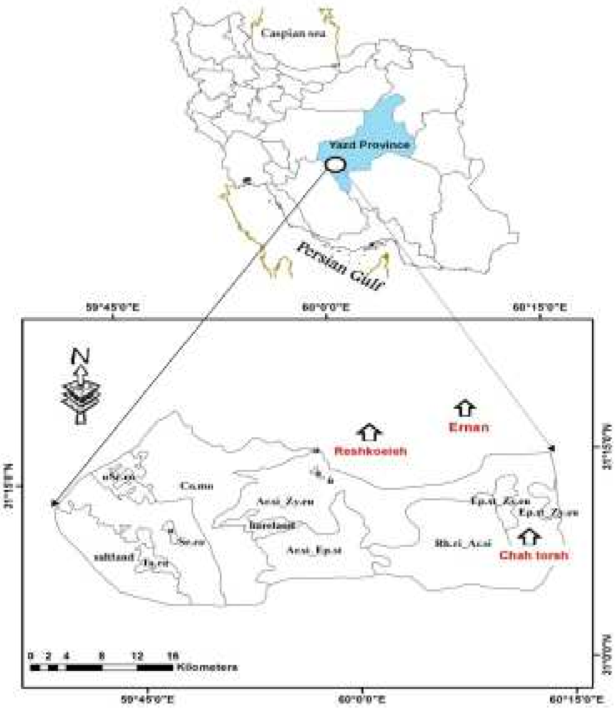

Fig 1. General location and vegetation type map of the study area

According to the results, level of agreement of predictive maps built upon layers of environmental variables at each site using MaxEnt approach showed a very good correspondence for actual and predictive maps of *S.*

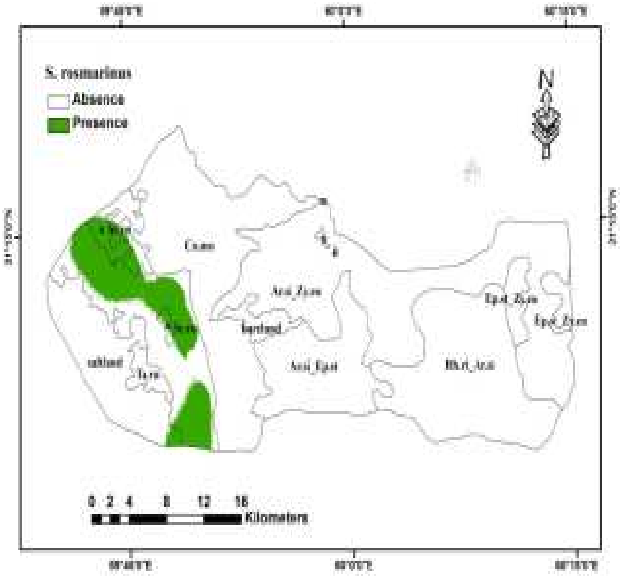

Fig 2. Actual and predicted distribution maps *of S. rosmarinus* species.

rosmarinus (Kappa coefficient=0.76). Moreover, it showed that predictive maps of *C. monocantha* had good correspondence with the actual map (Kappa coefficient=0.61), while predictive map of *T. ramosissima* had poor correspondence with actual map (Kappa coefficient=0.31) (Table1). Exemplary habitat distribution maps for *S. rosmarinus* species is presented in the figure 2. Thus, it can be said that MaxEnt model successfully predicted occurrence of *S. rosmarinus* and *C. monocantha* species in comparison with *T. ramosissima.* On the other hand, variability in model performances were affected by ecological niche of plant species as well as difference in their response shapes. PiriSahragard and ZareChahouki (2015) reported that MaxEnt models were more appropriate for species that had widespread ecological niche such as *C. monocantha*. Whereas, Yang *et al.* (2013) stated that the predicted potential distribution areas through MaxEnt almost always appeared as over predicted in some area compared to the realized niche of the species. Furthermore, PiriSahragard and ZareChahouki (2016) reported that that there was a strong relationship between model performance and the kinds of species distributions being modeled. Some methods performed generally better, but no method was superior in all circumstances. Because MaxEnt considers only niche-based presence data, it estimates the species fundamental niche (different from occupied niche) rather than realized niche (Kumar and Stohlgren, 2009; Yang *et al*., 2013).

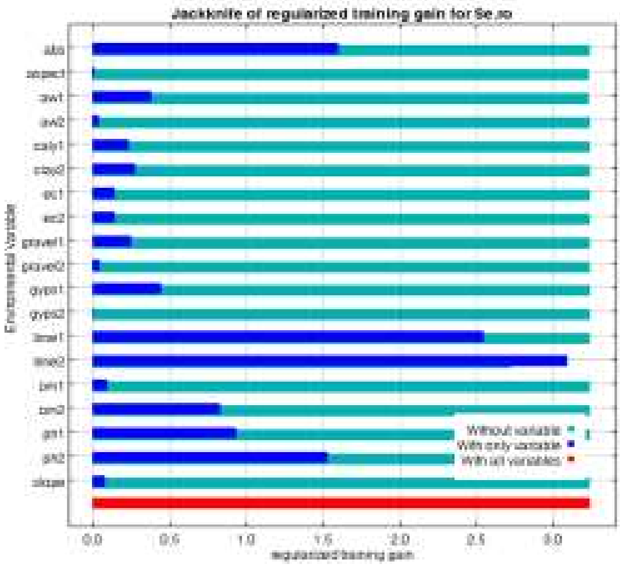

Fig 3. Jackknife evaluations of relative importance of predictor variables for *S. rosmarinus*

Predictor variable importance: According to the results, all models had a mix of soil and topographic variables (lime, Abs, OM, pH, EC and AW), suggesting species distribution is related to environmental niche. Overall, soil properties had the greatest impact on the predictive models. Based on the Jackknife operation results, EC1 and AW1 were the main factor influencing the spatial distribution of *T. ramosissima.* In other words, this species had high occurrence probability in the areas with EC more than 50 d S/m and AW more than 8 per cent (approximately 8 to 10%). Barnes and Harrison (1982) reported that soil available moisture had the great impact on occurrence of vegetation types. Furthermore habitat distribution of *S. rosmarinus* meaningfully was influenced by lime1 and lime2 and occurrence of *S. rosmarinus* had high probability in area with high lime amount (25% < lime 2 < 35%) (Fig. 3). Soil lime content effect on plant growth through its effect on soil pH and reduction in the availability of micronutrients such as Zn and Mn had already been emphasized in many studies (ZareChahouki *et al.,* 2010, 2012; Hosseini *et al.,* 2013, PiriSaharagard and ZareChahouki, 2015). In distribution of *C. monocantha* habitat, OM2 and pH were the most important variables, indicating that the habitat with low levels of OM2 (0-0.05%) and high pH values (8-8.2) can provide suitable condition for *C. monocantha* species. Additionally results indicated that elevation was one of the common predictors for all models. These findings were in accordance with the results of Azarnivand *et al.* (2002), ZareChahouki *et al.* (2010) and Hosseini *et al.* (2013).

Overall, in most models soil properties were one of the most important environmental variables affecting vegetation communities in the region and these were shown to be a controlling factor for many species. Hosseini *et al*. (2013) reported some soil parameters such as lime, gravel, organic matter and soil available moisture which affected vegetation communities in arid lands. Since the study area is located mostly at the southern slopes of Shirkouh, using only aspect variable, the MaxEnt model cannot achieve any gain (Fig. 3).

In MaxEnt, by performing iterations and changing coefficients for any single characteristic, per cent contribution of each variable was determined. This feature allowed users to identify variables which had more influence on the occurrence of different plant species and subsequently, researchers focused only on the important variables which saved cost and time of investigations while increasing the accuracy of prediction models (PiriSaharagard and ZareChahouki, 2015; PiriSahragard and ZareChahouki, 2016). In comparing the three species, it was observed that *C. monocantha* had widest distribution while both *T. ramosissima* and *S. rosmarinus* were concentrated in the west southern of Poshtkoh rangelands. Since MaxEnt was mapping the fundamental niche (different from occupied niche) of the species using bioclimatic variables, the suitable habitat for *C. monocantha* might be over predicted in some areas.

Conclusion

In this study, we showed that due to unite eco-physiological characteristics, the three-species studied could endure strict environmental condition and the habitat distribution patterns for *S. rosmarinus* and *C. monocantha* could be modeled using a small number of occurrence records and environmental variables using MaxEnt. In general, as MaxEnt only requires presence data of a plant species to map distribution of its habitat; it can be widely used compared with other standard methods. As findings from the present study indicated that MaxEnt is a generative method, results of MaxEnt modeling can provide key information about the environmental tolerances of the species that can be used for protecting susceptible habitats from future invasion and impacts of climate change, and its output can be easily understood by field practitioners, conservation planners and range lands managers of Iran as base information for grazing management and rangelands rehabilitation. Additionally application of statistical modeling techniques such as MaxEnt can be used to produce distribution models of sufficient quality for use in conservation planning. In general, when user-friendliness is more important, MaxEnt is preferred to other alternatives.

References

Azarnivand, H., M. Jafari, M. R. Moghaddam, A. Jalili and M. A. ZareChahouki. 2006. Investigation on environmental variables affecting habitat distribution of *A. sieberi* and *A. aucheri* in rangelands of Vardavard, Garmsar and Semnan. *Iranian Journal of Natural Resources* 56: 93-99.

Barnes, P. W. and A. T. Harrison. 1982. Species distribution and community organization in a Nebraska Sandhills mixed prairie as influenced by plant/soil water relationships. *Oecologia* 52: 192-201.

Cayuela, L., D. Golicher, A. Newton, H. Kolb, F. S. de Alburquerque, M. M., Arets, J. R. M. Alkemade and A. M. Pérez. 2009. Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Tropical Conservation Science* 2: 319-352.

Death, G. and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for the analysis of complex ecological data. *Ecology* 81: 3178-3192.

Ardestani, E. G., M. Tarkesh, M. Bassiri and M. R. Vahabi. 2015. Potential habitat modeling for reintroduction of three native plant species in central Iran. *Journal of Arid Land* 7: 381-390.

Elith, J., C. H. Graham and R. P. Anderson. 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29: 129-151.

Fielding, A. H. and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence methods. *Environmental Conservation* 24: 38-49.

Gaston, K. J. 1996. Species richness: measure and measurement. In: K.J. Gaston (ed). *Biodiversity: A Biology of Numbers and Difference*. Blackwell Science, Oxford. pp. 77-113.

Guisan, A. and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.

Guisan, A., C. H. Graham., J. Elith and F. Huettmann. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13: 332-340.

Hoffman, J. D., S. Narumalani and D. R. Mishra. 2008. Predicting potential occurrence and spread of invasive plant species along the North Platte River, Nebraska. *Invasive Plant Science and Management* 1: 359-367.

Hosseini, S. Z., M. Kappas and M. A. ZareChahouki. 2013. Modelling potential habitats for *Artemisia sieberi* and *Artemisia aucheri* in Poshtkouh area, central Iran using the maximum entropy model and geostatistics. *Ecological Informatics* 18: 61-68.

Kumar, S. and T. J. Stohlgren. 2009. MaxEnt modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola in* New Caledonia. *Ecology and Natural Environment* 1: 94-98.

Pearson, R. G., C. J. Raxworthy, M. Nakamura and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102-117.

Peterson, A. T. 2007. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3: 59-72.

Phillips, S. J., R. P. Anderson and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

PiriSahragard, H., M. A. ZareChahouki and H. Gholami. 2015. Predictive distribution models for determination of optimal threshold of plant species in central Iran. *Range Management and Agroforestry* 36:146-150.

PiriSahragard, H. and M. A. ZareChahouki. 2015. An evaluation of predictive habitat models performance of plant species in Hozesoltan rangelands of Qom province. *Ecological Modelling* 309-310: 64-71.

PiriSahragard, H. and M. A. ZareChahouki. 2016. Comparison of logistic regression and machine learning techniques in prediction of habitat distribution of plant species. *Range Management and Agroforestry* 37: 21-26.

Segurado, P. and M. Araújo. 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31: 1555-1569.

Su, H. J. 1987. Forest habitat factors and their quantitative assessment. *Quarterly Journal of Chinese Forestry* 20: 1-28.

Tarkesh, M. and G. Jetschke. 2012. Comparison of six correlative models in predictive vegetation mapping on a local scale. *Environmental and Ecological Statistics* 19: 437-457.

Wisz, M. S., R. J. Hijmans, J. Li., A. T. Peterson, C. H. Graham and A. Guisan. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763-773.

Yang, X. Q., S. P. S. Kushwaha and S. Saran. 2013. MaxEnt modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in lesser Himalayan foothills. *Ecological Engineering* 51: 83-87.

ZareChahouki, M. A., H. Azarnivand, M. Jafari and A. Tavili. 2010. Multivariate statistical methods as a tool for model-based prediction of vegetation types. *Russian Journal of Ecology* 41: 84-94.

ZareChahouki, M. A., L. KhalasiAhvazi and H. Azarnivand. 2012. Comparison of three modelling approaches for predicting plant species distribution in mountaneous scrub vegetation (Semnan rangelands, Iran). *Polish Journal of Ecology* 60: 277-289.

ZareChahouki, M. A. and L. KhalasiAhvazi. 2012. Predicting potential distributions of *Zygophyllum eurypterum* by three modeling techniques (ENFA, ANN and logistic) in North East of Semnan, Iran. *Range Management and Agroforestry* 33: 123-128.

ZareChahouki, M. A. and J. Esfanjani. 2015. Predicting potential distribution of plant species by modeling techniques in southern rangelands of Golestan, Iran. *Range Management and Agroforestry* 36: 66-71.