Range Mgmt. & Agroforestry 40 (2) : 202-206, 2019
ISSN 0971-2070

Comparison of maximum entropy and logistic regression for distribution modeling of *Prangos pabularia* lindl. in southern rangelands of Ardabil province, Iran

J. Esfanjani[1], A. Ghorbani[*1], M. Moameri[1], M. A. ZareChahouki[2], A. Esmali Ouri[1] and A. Mirzaei Mossivand[3]
[1]Department of Range and Watershed Management, University of Mohaghegh Ardabili, Ardabil, Iran
[2]Department of Rehabilitation of Arid and Mountainous Regions, University of Tehran, Iran
[3]Faculty of Agricultural and Natural Resources, Lorestan University, Lorestan, Iran
*Corresponding author e-mail: a_ghorbani@uma.ac.ir
Received: 22nd February, 2019 Accepted: 20th November, 2019

Abstract
This study was aimed to model *Prangos pabularia* in the rangeland of southern Ardabil province, Iran using logistic regression (LR) and the Maximum Entropy Method (MaxEnt). The data for this study were soil factors, topographic factors and climatic factors. The slope and elevation maps (1: 25000 scale) were obtained from the DEM (digital elevation model) map. Six sites with the distribution of *P. pabularia* (presence and absence) were identified. Three 100 m transects were established. On each transect, ten plots (4 m^2) were located and the total canopy cover and density of plants were recorded. Overall, 180 plots were sampled in six sites. Soil samples were collected from a depth of 0-30 cm on each transects at the beginning and end of each transect. The LR model showed that rainfall/precipitation was the most effective factor on the habitat distribution of *P. pabularia*. The accuracy of the LR method for the prediction map was good (Kappa index = 0.51). The MaxEnt method indicated that variables such as elevation, rainfall, phosphorus (P) were the most effective factors on the distribution of the habitat of *P. pabularia*. The appraisal of the software performance and the accuracy of the model prediction were at an excellent level (area under the curve =0.94). The accuracy of the MaxEnt method was low (Kappa index = 0.15). Thus, the accuracy of the LR method was more reliable than that of MaxEnt method.

Keywords: Ardabil province, AUC, Logistic regression, MaxEnt, *Prangos pabularia*

Introduction

Modeling the habitat of plant species relates field observations to a set of environmental variables such as climate, topography, soil characteristics, geology, or land cover. Finally, these provide spatial prediction. These spatial predictions indicate the suitability of the area for species, communities and biodiversity (Hidalgo *et al.,* 2008). Logistic regression as one of the methods for forecasting the distribution of plant species, fits the probabilistic model between the presence of vegetation (as a dependent variable) and its effective factors (as an independent variable) using the maximum likelihood method and using the probabilistic related function with logistic regression, probabilities are obtained from zero to one. In other words, the zero value is the probability of not being present and the value of one is the 100% probability of being present (Homser and Lemeshows, 1989; Zare Chahouki and Esfanjani, 2015). One of the best methods to predict vegetation habitat modeling is the isolated entropy method. This method requires the presence of species data. This model examines the location of presence with environmental variables in that area. It uses principles of maximum entropy to produce a prediction of fitting habitat in areas not sampled across the study area (Yanga *et al.,* 2013; Piri Sahragard and Zare chahouki, 2016; Bagheri *et al.,* 2017; Esfanjani *et al.,* 2018). Zare Chahouki and Ahvazi (2012) predicted potential distributions of *Zygophyllum eurypterum* by three modeling techniques (ENFA, ANN and logistic) in North East of Semnan, Iran. The results of the ENFA method showed that 25200 hectares (34 per cent) of the study site were a potential habitat of *Z. eurypterum*. The study also revealed that maps generated using LR and ANN models for *Z. eurypterum* species were highly consistent with their corresponding actual maps of the area. This species is distributed in the rangeland with alkali-saline soil, high in lime per cent, silty-sandy texture and in 1000-2000 meters elevation. Esfanjani *et al.* (2017) reported similarity of the actual map to the predictive one at satisfactory level (Kappa coefficient =0.65) for habitat distribution of *Festuca ovina-Astragalus gossypinus* using the maximum entropy method. The habitat requirements of *Festuca ovina-Astragalus gossypinus* indicated that the highest probability of presence occurred in the soils with good N (0.14- 0.18%),

sand (35-38%), clay (20-24%) contents, but low in lime (11-13%) content. *Prangos pabularia,* a range plant, where the aerial parts are used as animal fodder with high nutritive value, while roots and fruits are used as medicine (Razavi, 2012). In some habitats, due to the excess grazing of livestock, their presence/availability is under threat. Thus, modeling the habitat of this plant is extremely important. The objective of this study was to compare logistic regression model and maximum entropy in predicting the habitat of the *Prangos pabularia* and to determine the most important environmental factors affecting its distribution.

Materials and Methods

Study area: The study was conducted in southern part of Ardabil province, Iran (37° 12' to 38° 07' N and 47° 51' to 48° 48' E). The annual precipitation varied from 200 to 500 mm, and the average temperature from 0.4 to 17 ^{0}C (Mossivand *et al.,* 2017).

Sampling method: A homogeneous unit map was prepared based on slope, elevation and satellite images. The selection of the sampling site in each unit was based on the species representing each vegetative type and presence of homogeneous vegetation. Three transects of 100 meters were placed in each homogeneous unit. Two transects were located along the most important environmental gradients (elevation, direction and slope) and another transect was perpendicular to other two transects. During each transect, 10 plots were placed at a distance of 50 meters (due to the great length of the range and the changing environmental conditions, this distance was considered). Thus, 30 plots were deployed in each homogeneous unit. The size of sampling plots was 4 square meters (according to the kind and distribution of plant species by a minimum surface area). In each plot, the kind and number of plant species and their coverage were recorded. For soil sampling, soil profiles were excavated at the beginning and end of each transect. Depending on the depth of the soil and the effective depth of the anisotropy, the species was selected at a depth of 0-30 cm. Then soil variables such as clay, sand, silt, organic matter, acidity, electrical conductivity, nitrogen (N), potassium (K) and phosphorus (P) were measured. Additionally, in each sampling unit, latitude, longitude, slope, aspect and elevation of sea level and climatic factors (rainfall and temperature) were recorded. To predict vegetation mapping, in addition to a number of environmental factors, mapping these factors were also needed. The environmental factors map was prepared by the Kriging interpolation method using the GIS 10.4.1 software. The slope and elevation map was also obtained.

MaxEnt method

To evaluate the environmental data in the maximum entropy method, the environmental maps were prepared. In this method, 70% of the attendance points were randomly assigned to the educational data, and 30% of the remaining data were used to evaluate the model results. Moreover, two options were used to construct response curves for environmental characteristics and the Jackknife test was used to determine the effective variables (Phillips *et al.,* 2006). The Jackknife test was also used to evaluate the significance of each variable in the model preparation. In the receiver operating characteristic (ROC) analysis, the accuracy of each model was expressed in terms of the surface area below the curve (0 to 100). ROC curves were a way of graphically displaying true positives versus false-positives across a range of cut-offs and of selecting the optimal cut-off for modeling of plant species to be selected (Florkowski, 2008). Curve analysis and receiver operating characteristic and area under the curve (AUC) were used to evaluate the overall quality of the model. The ROC curve showed the achievement of AUC in three different modes (Reed and Martens, 2008). Additionally, in the output of the MaxEnt software, the AUC was also presented as a method to evaluate prediction models. Eventually, the model output was the species habitat prediction map. The validation of prediction map was done by comparing to the plant ground map with the Kappa index.

Logistic regression method: In general, the logistic regression model predicts the presence and absence of plant species with the following equation (Warton *et al.,* 2012):

$$Y = [Exp (b_0+b_1x_1+\ldots+b_nx_n)] /[(1+Exp(b_0+b_1x_1+\ldots+b_nx_n)]$$

Where Y is the occurrence of species probability, b is regression model coefficients and x is predictive variables (environmental factors). In this method, plant species was considered as dependent variables and environmental variables as independent variables. Finally, the prediction map was obtained for the plant species. To test the obtained models, Homser and Lemeshows statistics was used (Homser and Lemeshows, 1989).

Accuracy of prediction maps: The accuracy of the pre-

-diction map with the actual maps was investigated calculating the kappa coefficient in the IDRISI Selva 17 software.

Results and Discussion

MaxEnt method

Analysis of omission/ commission for P. pabularia: The omission rate and the predicted area at different thresholds have been illustrated in graph (Fig 1). The shading surrounding the lines on the graph represented variability. Analysis of omission/commission depicted model performance/bias as a function of predicted occurrence values, which contributed to make decisions on an occurrence probability threshold to model 'habitat' *vs*. 'non-habitat'. Thus, good model increases more steeply than a straight line, and the area under the curve (AUC) was high (Philips *et al.,* 2006). In this case, our AUC value of 0.94 was good. Various criteria listed were used to choose logistic cutoff values in order to represent 'habitat' *vs*. 'non-habitat', along with tests of whether such cutoffs would be superior to a random selection of points (Table 1). The 'minimum training presence' criterion was the logistic threshold (0.197) resulting in inclusion of all training presence sessions.

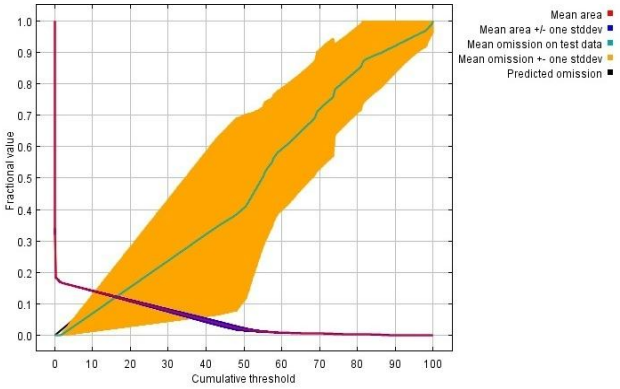

Fig 1. Average omission and predicted area for *P. pabularia*

ROC curves of sensitivity vs. specificity for P. pabularia: The AUC values show the performance of one model with another, and it is useful in evaluating multiple MaxEnt models. Philips *et al.,* (2006) reported that AUC value closer to 1.0 indicated better performance of the model. According to the AUC classification, the model predictive accuracy of the habitat *P. pabularia* was assessed (Fig 2) as an excellent level (AUC = 0.946).

Jackknife of regularized training gain for P. pabularia: Based on the jackknife operation results (Fig 3), elevation, rainfall (average rainfall varied from 227 to 410 mm) and P (amount of phosphorus from 1.42 to 4.53) were the most important variables in the distribution of the habitat *P. pabularia*. Results showed that the response curves of the species, *P. pabularia* was significantly affected by elevation, mean rainfall and phosphorus (P) variables. Elevation gradients have become important tools to assess the effects of temperature changes on vegetation properties, since these gradients enable temperature effects to be considered over larger spatial and temporal scales than what is possible through conventional experiments (Gale, 2004). The elevation affects plant physiology and ecology, therefore, all environmental factors should be taken into account.

Predication map: The predication map was obtained based on two levels of presence and absence of *P. pabularia*, and the map was compared to the actual plant map (Fig 4).

Table 1. Description, logistic threshold and cumulative threshold in analysis of omission/ commission for *P. pabularia*

Description	Logistic threshold	Cumulative threshold
Fixed cumulative value 1	0.122	1
Fixed cumulative value 5	0.407	5
Fixed cumulative value 10	0.407	10
Minimum training presence	0.197	1.172
10 percentile training presence	0.197	1.172
Equal training sensitivity and specificity	0.197	1.172
Maximum training sensitivity plus specificity	0.197	1.172
Equal test sensitivity and specificity	0.197	1.172
Maximum test sensitivity plus specificity	0.197	1.172
Balance training omission, predicted area and threshold value	0.042	0.301
Equate entropy of threshold and original distributions	0.407	48.128

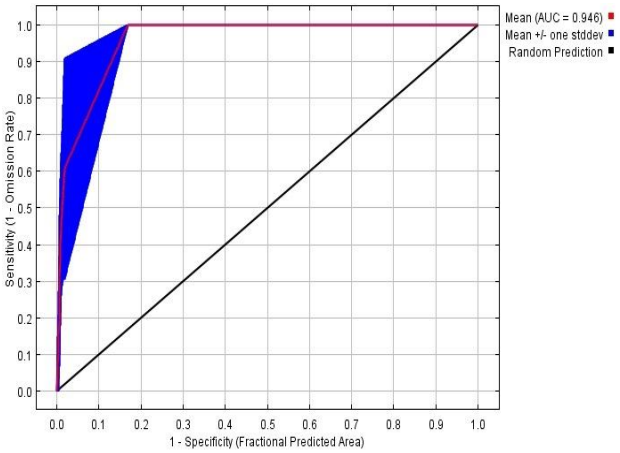

Fig 2. ROC curves of sensitivity vs. specificity for *P. pabularia*

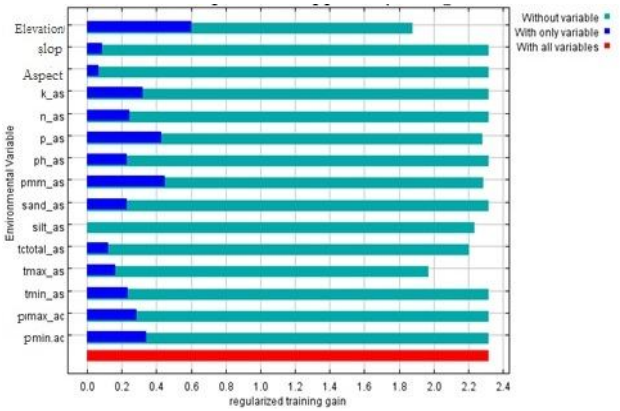

Fig 3. Jackknife results of important variables (elevation, Pmm, P) for *P. pabularia*; Y axis indicates environmental variables: elevation, slope, aspect, K (potassium), N (nitrogen), P (phosphorus), pH (acidity), Pmm: middle precipitation, sand, silt, Tc: total carbon, T minimum (ºC), T maximum (ºC), P min: minimum precipitation, P max: maximum precipitation

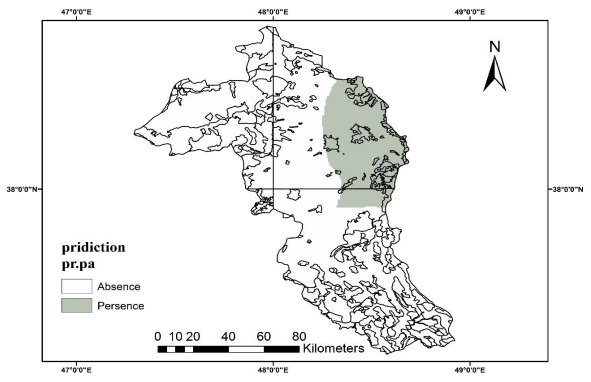

Fig 4. Predicted distribution for *P. pabularia* habitat from the MaxEnt model

Logistic regression method

An equation was derived between presence and absence of species habitat and environmental factors using the logistic regression model. The presence and absence of the habitat *P. pabularia* was related to rainfall. The relationship indicated the correlation between the presence of *P. pabularia* and the average rainfall. This showed that as the rainfall increases, the probability of the presence of *P. pabularia* increases.

P (*Pr.pa*) = Exp(-3.004 rainfall+1091.612)/1+Exp(-3.004 rainfall+1091.612)

Variable rainfall is well known to drive fluctuations in annual plant populations, yet the degree to which population response is driven by between-year variation in germination cueing, water limitation or competitive suppression is poorly understood (Levine *et al.,* 2008). If the amount of Hausmare and Lmshaw (HL) is high, it results in greater compliance. The significance of this model (Table 2) with the coefficients of diagnosis and HL statistic showed the significance of these relationships at 99 percent level.

Table 2. Logistic regression statistics to predict presence and absence of habitat

Habitat species	R^2	HL statistic
P. pabularia	0.75	0.97**

**Significant at 99 percent level

The rainfall map was prepared using the Kriging interpolation method and above equation was applied on the environmental map made (rainfall map) in part of the raster calculator in the GIS 10.4.1 software. The final prediction map based on two levels of presence (1) and absence (0) of plant species was compared with the actual plant map (Fig 5).

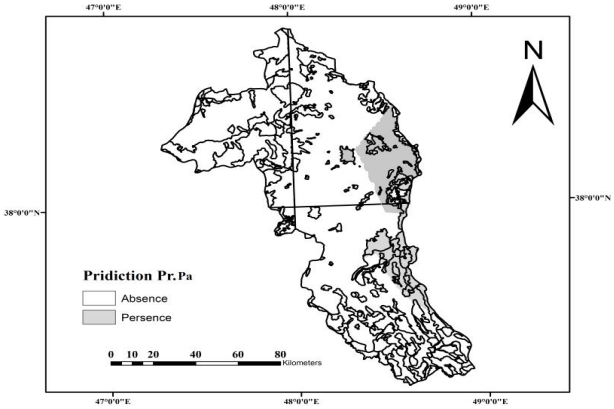

Fig 5. Predicted distribution for *P. pabularia* habitat from the LR model

The ROC curve was also used to analyze the environmental data (AUC = 0.94). The modeling accuracy was excellent. Again, the Kappa coefficient was used for the validation of the map of *P. pabularia* with its ground reality. The Kappa coefficient was 0.51 and 0.15 in the logistic regression and MaxEnt method, respectively. Thus, the habitat map of *P. pabularia* using the logistic regression method was predicted with good accuracy, whereas accuracy was low with MaxEnt.

Conclusion

In this investigation, the factors affecting the presence of *P. pabularia* were studied. In MaxEnt method, elevation, Pmm (middle precipitation) and P (phosphorus) factors had the most effect on the distribution of habitat *P. pabularia*, but in logistic regression method, the precipitation/rainfall had the most effect on the distribution of the habitat *P. pabularia*. Therefore, it was concluded that the environmental variables used to formulate the final model in the regression method, have a good ability to predict the habitat *P. pabularia* compared to the MaxEnt method. The prediction maps derived from the logistic regression model were more accurate than those derived from the MaxEnt method so that it can play a crucial role in proposing species consistent with different physiographic conditions in rangeland regeneration programmes.

References

Bagheri, H., A. Ghorbani., M. A. Zare chahouki., A. A. Jafari and K. Sefidi. 2017. Halophyte species distribution modeling with MaxEnt model in the surrounding rangelands of Meghan Playa in Iran. *Applied Ecology and Environmental Research* 15: 1473-1484.

Esfanjani. J., A. Ghorbani and M. A. Zare Chahouki. 2017. Modeling habitat distribution of *Festuca ovina-Astragalus gossypinus* by using maximum entropy method in the Chaharbagh rangelands of Iran. *Range Management and Agroforestry* 38:171-175.

Esfanjani. J., A. Ghorbani and M. A. ZareChahouki. 2018. MaxEnt modeling for predicting impacts of environmental factors on the potential distribution of *Artemisia aucheri* and *Bromus tomentellus-Festucaovina* in Iran. *Polish Journal of Environmental Studies* 27: 1041-1047.

Florkowski, C.M. 2008. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clinical Biochemist Reviews* 29: 83-87.

Gale, J. 2004. Plants and altitude- revisited. *Annals of Botany* 94:190-199.

Hidalgo, P. J., J. M. Marin., J. Quijada and J. M. Moreira. 2008. A spatial distribution model of cork oak (*Quercus sub*er) in southwestern Spain: a suitable tool for reforestation. *Forest Ecology and Management* 255: 25-34.

Homser, D.W and J. R. Lemeshows. 1989. *Applied Logistic Regression*. Wiley, New York. pp. 1-582.

Levine, J. M., A. K. McEachern and C. Cowan. 2008. Rainfall effects on rare annual plants. *Journal of Ecology* 96: 795-806.

Mossivand, A. M., A. Ghorbani and F. K. Behjou. 2017. Effects of some ecological factors on distribution of *Prangos uloptera* and *Prangos pabularia* in rangelands of Ardabil province, Iran. *Applied Ecology and Environmental Research* 15: 957-968.

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

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

Razavi, S.M. 2012. Breaking of seed dormancy in *Prangos pabularia* and *Prangos uloptera* growing in Iran. *Insight Botany* 2: 7-11.

Reed D.D. and B.K. Martens. 2008. Sensitivity and bias under conditions of equal and unequal academic task difficulty. *Journal of Applied Behavior Analysis* 41: 39-52.

Warton, D.I., S.T. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evaluation* 3: 89-101.

Yanga, L., S.P.S. Kushwahab., S. Saranb and P.S. JianchuXuc. 2013. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in lesser Himalayan foothills. *Journal of Ecological Engineering* 51: 83-87.

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.

ZareChahouki, M. A. and L. K. Ahvazi. 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.