Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2018) 42: 28-37
© TÜBİTAK
doi:10.3906/bot-1612-5

A comparison of logistic regression and maximum entropy for distribution modeling of range plant species (a case study in rangelands of western Taftan, southeastern Iran)

Hossein PIRI SAHRAGARD*, Majid AJORLO
Department of Range and Watershed Management, Faculty of Water and Soil, University of Zabol, Zabol, Iran

Received: 02.12.2016 • **Accepted/Published Online:** 08.08.2017 • **Final Version:** 11.01.2018

Abstract: This study aimed to compare the efficiency of logistic regression and maximum entropy models for distribution modelling of plant species habitats in the rangelands of western Taftan, southeastern Iran. Vegetation cover was sampled using a systematic-randomized method. Soils were sampled at 0–30 and 30–60 cm depths through digging of eight soil profiles. The agreement between predictive maps generated by models with documented maps of habitats indicated that logistic regression was able to predict the distribution of *Artemisia aucheri* and *Artemisia sieberi* habitats at excellent (kappa value = 0.95) and weak (kappa value = 0.39) levels, respectively. On the other hand, the agreement between predicted maps generated by maximum entropy with documented maps was very good for *Amygdalus scoparia* and *Artemisia aucheri* habitats (kappa value = 0.82 and 0.76, respectively), and weak for *Artemisi aucheri* (kappa value = 0.55). This study indicates that logistic regression and maximum entropy methods had the same efficiency in distribution modelling of plant species with a limited ecological niche. However, the maximum entropy model can receive priority in distribution prediction of plant species with a limited ecological niche because it uses only presence data of plants and a small dataset.

Key words: Predictive modelling, logistic regression, maximum entropy, ecological niche, plant habitats

1. Introduction

Availability of spatial distribution of plant species is a major requirement for conservation of plant natural habitats. Identification of rangeland with high suitability for establishment of a specific plant species is also important. The logistic regression and maximum entropy methods predict potential distribution of species rather than their real distribution. Therefore, the probability of predicted presence of a species by these two models is equal to the suitability of a habitat for a specific species (Pearce and Ferrier, 2000; Keating and Cherry, 2004). It is clear that establishment of plant species in suitable habitats can be limited by some factors such as human intervention and biological interactions (Phillips et al., 2006). On the other hand, a field survey for data collection is always time consuming and costly. In this situation, predictive models for plant species distribution can be an important alternative (Hernandez et al., 2008). Predictive models are cause and effect tools that construct a relationship between real distribution patterns of plants and environmental variables for prediction of plant species distribution (Elith and Graham, 2009). In addition, predictive models are widely used in wildlife management, determination of suitability of a habitat for a specific species, and rangeland improvement projects (Araujo and Guisan, 2006; Zare Chahouki and Khalasi Ahvazi, 2012; Hosseini et al., 2013; Ardestani et al., 2015; Piri Sahragard and Zare Chahouki, 2016b).

The logistic regression model is one of the regression methods that can be used when a response variable is binary, and the predictor variable is continuous or categorical. Logistic regression uses logit function for description of the relationship between response variables and predictor variables (Miller and Franklin, 2002). In the general linear model (GLM), logistic regression is widely used for distribution modelling of plant species (Rushton et al., 2004). In logistic regression, variable input to the model is done on the basis of significance of maximum likelihood statistic, and variable output is done based on probability of this statistic and maximum likelihood estimation (Zare Chahouki, 2010).

The maximum entropy model as a machine learning method is also a common method in distribution modelling of plant species that uses only the presence data of species as response variable (Baldwin, 2009). Input variables can be continuous or categorical in this method (Phillips and Dudik, 2008). This model's need for a small dataset for precise model development, its low sensitivity

* Correspondence: hopiry@uoz.ac.ir

to spatial uncertainty of data, and its generation of maps that show probability of species presence in a specific area are some of its advantages (Baldwin, 2009). In other words, this method predicts presence probability of each plant species in a specific area using presence data and environmental variables data layer (Phillips et al., 2006). Continuous output of the predictive model, possibility of threshold determination based on the objectives of the study, and being user friendly are other characteristics of this method (Piri Sahragard and Zare Chahouki, 2015).

Kumar and Stohlgren (2009) used the maximum entropy model for distribution modelling of suitable habitats for *Canacomyrica monticola* and reported that the distribution pattern of this endangered species can be modeled with acceptable accuracy using this model. Tarkesh and Jetscheke (2012) compared prediction efficiency of the maximum entropy model with BIOCLIM[1] and GARP[2] methods, and showed that prediction efficiency of maximum entropy was better than by two other methods. Furthermore, Piri Sahragard and Zare Chahouki (2015) compared the prediction efficiency of logistic regression, maximum entropy, and artificial neural network in plant species potential habitats, and observed that the maximum entropy method was more appropriate for distribution modelling of species with vast ecological niches.

It seems that comparison of predictive efficiency of the two models, i.e. logistic regression and maximum entropy, is necessary in distribution modelling of range plant species because the two models vary in their input data type and modelling procedure. Moreover, by taking into consideration the different capability of the two models, limitation of resources, and budget, it is necessary to characterize the model (a model that uses presence and absence data, or presence data only for modelling) that is more reasonable on the large scale. The objectives of this study were to identify favorable environmental conditions for plant species establishment and to compare the prediction efficiency of the logistic regression and maximum entropy in distribution modelling of range plant species in the rangelands of western Taftan, southeast Iran.

2. Materials and methods

2.1. Study area

The study site, with a total area of 64,000 ha, is located between 60°39′36″ and 60°58′19″ E longitude and 28°20′35″ and 28°42′39″ N latitude on a hill slope of Taftan Mountain, Sistan and Baluchestan province, southeastern Iran (Figure 1). The highest and lowest altitude of the area was 2800 and 140 above msl, respectively. Geomorphologically, the study area was plain-mountainous. Annual precipitation ranges from 160 to 832 mm, snowing and frosting occur in winter (Ebrahimi et al., 2015). *Haloxylon persicum*, *Artemisia sieberi*, *Amygdallus scoparia*, and *Zygophyllum eurypterum* are dominant plant species; *Hamada salicornica*, *Cousinia stocksi*, and *Artemisia santolina* are co-dominant species.

2.2. Data collection

Environmental variables were quantified by using the digital elevation model (DEM) and geology map (scale 1:25,000), field survey and laboratory analysis for understanding the effective variables in distribution of plant species, and development of a model for distribution prediction. Sampling units were prepared through integration of landform and geology maps and separation of habitat was conducted on the basis of field survey and observations. Vegetation of each habitat was sampled by a randomized-systematic method along four sampling lines 150–200 m in length. The length of a sampling line was determined on the basis of plant density and variation of vegetation. Sample size was calculated by Cochran's Q test with regard to variations of vegetation and soil. Quadrat size (2–25 m^2) was determined according to vegetation type, plant density, and parameters to be measured (Table 1).

Species type, species richness, and vegetation cover percent were recorded in each quadrat during vegetation sampling. The soils of each habitat were sampled at 0–30 and 30–60 cm depths through the digging of eight soil profiles. In total, 40 soil samples were collected. Physical and chemical properties including gravel content, lime content, pH, EC, available moisture content, organic matter, gypsum, sand, silt, clay, and saturated moisture were measured for soil samples. In addition, data related to habitat border, slope, altitude, and geological structure were recorded for each habitat.

2.3. Data analysis

The distribution of plant species in each habitat was modelled using logistic regression and maximum entropy models after quantification of some variables and preprocessing of the dataset. Multicollinearity between independent variables was checked through the calculation of variance inflation factor (VIF) in the logistic regression method. The VIF value was lower than 10, indicating that there was no multicollinearity between independent variables. Predictive models of logistic regression were generated in SPSS 18 for each habitat. Layers of input variables of predictive models were prepared by using geostatistics and GIS facilities. Coefficients of these layers were assigned to each layer in the ARC GIS 9.3 environment and finally predicted maps of plant habitats were generated.

[1]Bioclimatic Prediction and Modeling System
[2]Genetic Algorithm for Rule Set Production

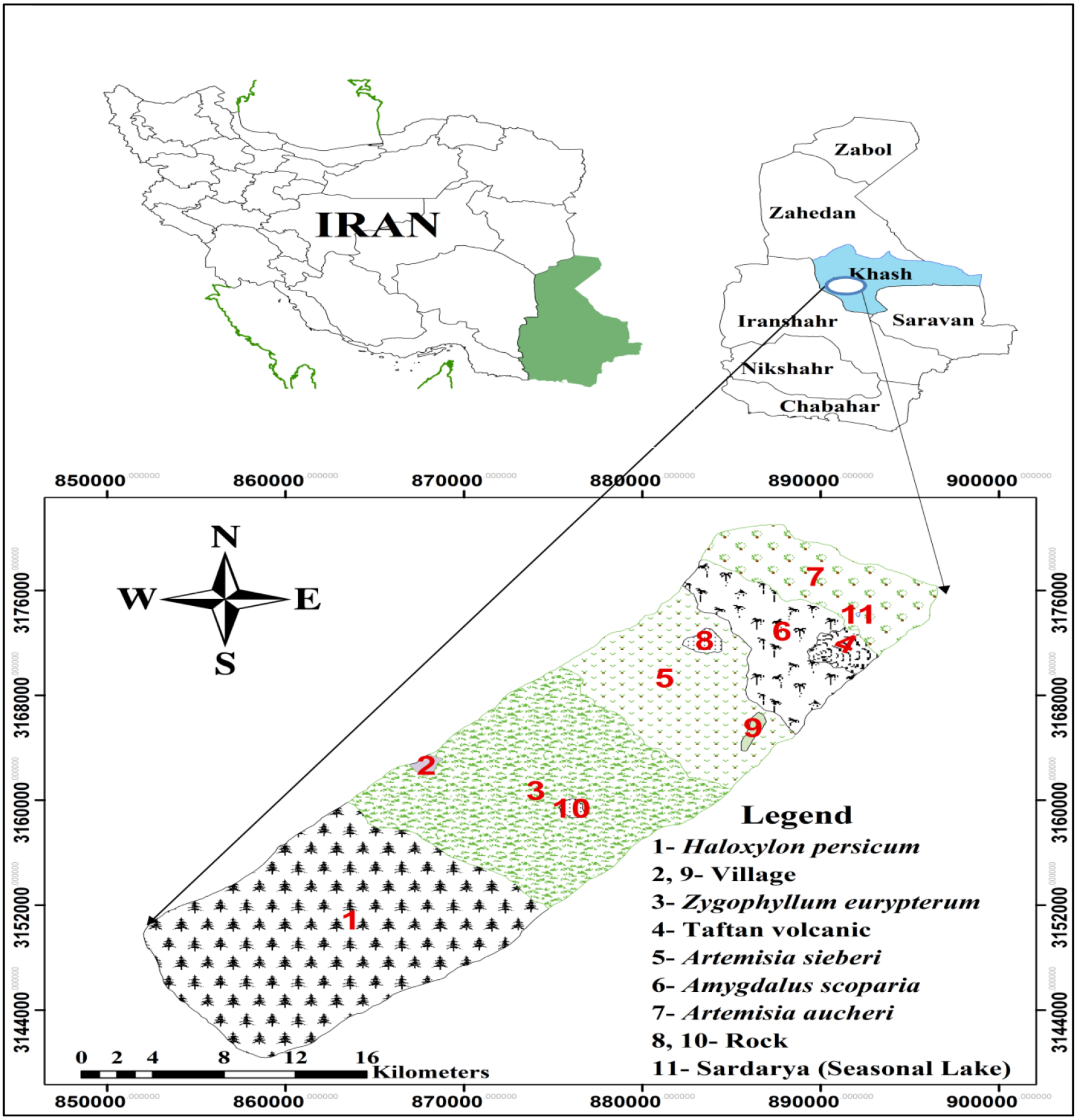

Figure 1. Major features of the study area and its location in southeastern Iran.

For maximum entropy, maps of environmental variables were generated in ASCII format and MaxEnt 3.3e was used for modelling of plant habitat distribution. It should be noted that 25% of data were used for model examination and the rest were used for training. The iteration of model generation was considered 1000. The jackknife test was employed for determination of the importance of environmental variables (Piri Sahragard and Ajorlo, 2016).

2.4. Assessment of the accuracy of predicted models and maps

The Hosmer–Lemeshow (HL) test was used for the assessment of logistic regression models. The HL statistic was used for the assessment of agreement between predicted and observed maps (Hosmer and Lemeshow, 2000). A high value of this statistic indicates greater agreement between them. In addition, the area under curve (AUC) statistic was used for the assessment of

Table 1. Length of sampling line, distance between quadrats, number of quadrats, and quadrat size in the studied plant habitats of western Taftan.

Plant habitat	Symbol on the map	Sampling line length (m)	Distance between quadrats (m)	Number of quadrats	Quadrat size (m^2)
Haloxylon persicum	Ha. pe	200	20	40	25
Zygophyllum eurypterum	Zy. eu	200	20	40	4
Artemisia sieberi	Ar. si	150	10	60	2
Amygdalus scoparia	Am. sc	150	10	60	25
Artemisia aucheri	Ar. au	150	10	60	2

generated models by maximum entropy method (Sweet, 1988). The value of this statistic varies from 0.5 (in cases where there is no difference between the points of two groups, i.e. correct presence and correct absence) to 1.0 (in cases where there is no overlap between the points of two groups and difference is excellent). The AUC statistic shows the power of the model in distinction between presence and absence. If the value of the statistic is close to 1.0, it indicates better agreement of the model with the real environment (Piri Sahragard and Zare Chahouki, 2015). After the generation of the predicted map, it is necessary to determine the optimal threshold for determination of presence or absence of desired species (Phillips et al., 2006). In this study, after the determination of optimal threshold using the equal sensitivity and specificity method, continuous predicted maps were converted into presence and absence maps (Piri Sahragard and Zare Chahouki, 2016b). The agreement between predicted and documented maps was calculated with kappa index in IDRISI release 32 (Zare Chahouki et al., 2010).

3. Results

3.1. Assessment of efficiency of predicted models

Logistic regression models for the studied habitats were significant at $\alpha = 0.01$ according to the HL test. The accuracy level of the models generated for *Haloxylon persicum*, *Amygdalus scoparia*, and *Artemisia aucheri* habitats was good, but it was acceptable for *Zygophyllum eurypterum* and *Artemisia sieberi* habitats (Table 2). Models were run with various numbers and arrangements of variables in the maximum entropy method. Models with the lowest number of variables and the highest AUC were selected as optimal models.

3.2. Assessment of agreement between predicted and documented maps

Assessment of efficiency of the models used in this study for determination of plant species ecological niche revealed that the accuracy level of generated models varies in prediction of species presence and absence. This leads to the emergence of significant differences in the

Table 2. Statistics of accuracy assessment of the models and the accuracy level of predictive models in the studied plant habitats of western Taftan.

Plant habitat	Logistic regression		Maximum entropy	
	R^2	HL	AUC	Accuracy level
Haloxylon persicum	0.86	0.99	0.93	Good
Zygophyllum eurypterum	0.70	1.00	0.86	Acceptable
Artemisia sieberi	0.87	1.00	0.84	Acceptable
Amygdalus scoparia	0.77	0.99	0.95	Good
Artemisia aucheri	0.89	1.00	0.96	Good

HL: Hosmer–Lemeshow test; AUC: area under curve efficiency of the models. Consequently, the agreement between predicted and observed maps of species in different plant habitats varied. In this study, the agreement between predicted and observed maps was assessed by kappa index. The highest and lowest kappa values related to the agreement between predicted maps generated by logistic regression and documented maps were 0.95 and 0.39 in *Artemisia aucheri* and *Artemisia sieberi* habitats, respectively. The agreement between predicted maps generated by the maximum entropy method and observed maps was very good for *Haloxylon persicum*, *Amygdalus scoparia*, and *Artemisia aucheri* habitats, but it was good and fair for *Zygophyllum eurypterum* and *Artemisia sieberi* habitats, respectively (Table 3). Predicted and documented maps with the highest and lowest agreement generated by logistic regression and maximum entropy are shown in Figure 2.

3.3. Relationship between presence of plant species and environmental variables using logistic regression models
Input variables of predictive models generated by logistic regression along with their coefficients are shown in Table 4. Percentages of sand and organic matter at 0–30 cm soil depth were the most effective variables in the distribution of *Haloxylon persicum* habitat. In other words, the presence of this species was directly related to soil light texture and organic matter content. Presence of *Zygophyllum eurypterum* was boosted by increase of gypsum percent at 30–60 cm and silt at 0–30 cm soil depths, indicating that the presence of this species was directly related to these variables (Table 4). However, the presence of *Artemisia sieberi* was mainly affected by soil lime content and pH at 0–30 cm depth. In the highland of the study area, physical characteristics of the habitats including altitude, slope, and geological structure also affect the distribution of plant habitats. For example, besides the percentage of sand at 0–30 cm soil depth, slope and geological structure were the most effective variables in *Amygdalus scoparia* habitat. Furthermore, altitude and lime percent at 0–30 cm soil depth had the greatest effect on the distribution of *Artemisia aucheri* habitat (Table 4).

3.4. Relative importance of variables
The relative importance of variables was determined by the jackknife test. The results showed that adding new variables to the predictive models of maximum entropy did not increase the accuracy of the model . Therefore, these variables were the most effective ones in the predictive models. For example, available soil moisture content and sand percent at 0–30 cm soil depth along with slope degree were the most effective variables in the predictive model of *Haloxylon persicum* (Table 5). Thus, a predictive model that includes these three variables was selected as the best one on the basis of AUC statistic value. In addition, the simplest and most accurate model of *Zygophyllum eurypterum* showed that soil gypsum and lime content at 30–60 cm depth and gravel percent at 0–30 cm depth played the greatest roles in the predictive model (Table 5). Exclusion of these variables can adversely affect the accuracy of the model. Moreover, soil moisture content at 0–30 cm depth and gypsum content at 30–60 cm depth were the most effective variables in the predictive model of *Artemisia sieberi* (Table 5). The effective variables in the distribution of *Amygdalus scoparia* were soil gravel percent (0–30 cm depth), lime content (30–60 cm depth), and geologic structure. The most precise model for *Artemisia aucheri* was the model that included altitude, lime, and silt at 30–60 cm soil depth. In the jackknife test, the response curves show a relationship between environmental variables and presence probability of plant species. Interpretation of these curves can introduce favorable environmental conditions for the occurrence of plant species. The output of the jackknife test for *Artemisia aucheri* is shown in Figure 3.

Table 3. Kappa index values and the level of agreement between predicted and documented maps in studied plant habitats of western Taftan.

Plant habitat	Model	Threshold	Kappa index	Level of agreement
Haloxylon persicum	Logistic regression Maximum entropy	0.5 0.4	0.62 0.75	Good Very good
Zygophyllum eurypterum	Logistic regression Maximum entropy	0.3 0.6	0.58 0.69	Good Good
Artemisia sieberi	Logistic regression Maximum entropy	0.3 0.7	0.39 0.55	Weak Fair
Amygdalus scoparia	Logistic regression Maximum entropy	0.3 0.1	0.86 0.82	Excellent Very good
Artemisia aucheri	Logistic regression Maximum entropy	0.2 0.4	0.95 0.76	Excellent Very good

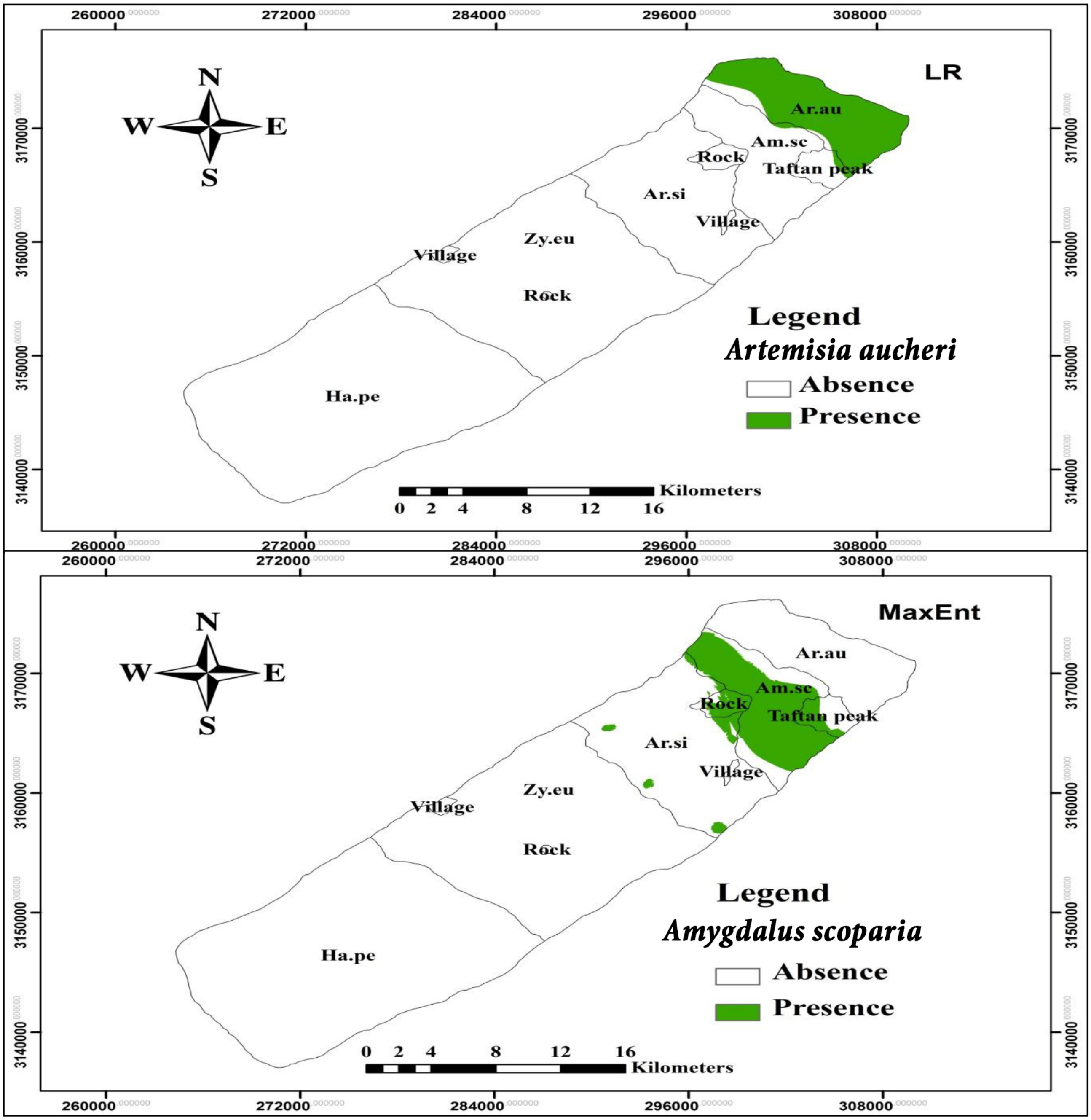

Figure 2. The most accurate predicted and documented maps of *Artemisia aucheri* and *Amygdalus scoparia* habitats resulting from the logistic regression and maximum entropy method, respectively (predicted maps are shown in green).

4. Discussion

Our results indicated that the logistic regression and the maximum entropy models have roughly similar efficiencies in the prediction of studied plant species distribution in the rangelands of western Taftan. Both models were able to predict the distribution of plant species with a small ecological niche (e.g., *Amygdalus scoparia* and *Artemisia aucheri*) more precisely than species with a vast ecological niche (*Artemisia sieberi*). Consequently, both models showed low efficiency in the prediction of distribution of plant species with vast ecological niches in the rangelands of western Taftan. Previous studies reported that the vastness of species ecological niches can negatively affect the accuracy of models generated by logistic regression (Guisan and Zimmermann, 2000; Piri Sahragard and Zare Chahouki, 2015). On the other hand, it can be stated

Table 4. Input variables of predictive models generated by logistic regression, along with their coefficients in the studied plant habitats of western Taftan.

Plant habitat	Intercept	Predictive variables	Coefficient
Haloxylon persicum	2.88	Sand (0–30 cm depth)	0.58
		Altitude	–0.33
		Organic matter (0–30 cm depth)	0.28
Zygophyllum eurypterum	6.57	Gypsum (30–60 cm depth)	1.52
		Silt (0–30 cm depth)	0.43
Artemisia sieberi	–14.64	Lime (0–30 cm depth)	0.64
		pH (0–60 cm depth)	18.23
Amygdalus scoparia	–4.058	Geology structure	0.68
		Sand (0–30 cm depth)	0.44
		Slope	0.32
Artemisia aucheri	11.26	Altitude	4.34
		Lime (0–30 cm depth)	–0.36

Table 5. The most important variables and their contribution in the maximum entropy model for the studied plant habitats of western Taftan.

Plant habitat	Environmental variable	Contribution percent	Desirable range
Haloxylon persicum	Soil available moisture content (0–30 cm)	69.5	14%–15%
	Sand (0–30 cm)	30.3	72%–78%
Zygophyllum eurypterum	Gypsum (30–60 cm)	75.4	35%–50%
	Lime (30–60 cm)	23.8	2%–4%
	Gravel 0–30 cm)	7.2	5%–15%
Artemisia sieberi	Soil available moisture content (0–30 cm)	46.7	3%–6%
	Gypsum (30–60 cm)	53.3	1%–5%
Amygdalus scoparia	Gravel (0–30 cm)	67.0	35%–50%
	Lime (30–60 cm)	30.5	12%–16%
Artemisia aucheri	Altitude	0.96	3000–3600 m
	Lime (0–30 cm)	2.5	2%–4%

that logistic regression is a suitable model for distribution modelling of plant species because of the nonlinear relationship between species and environmental variables and sigmoid curves of the logistic regression functions. The ecological niche of each species can be accurately predicted by using the logistic regression model (Guisan et al., 1999). It is worth mentioning that it is not necessary for all input variables of the model to be statistically significant in logistic regression. The important thing is that combination of all variables can minimize prediction error. Consequently, a predictive model can be formed with a set of significant and insignificant variables (Rossiter and Loza, 2010). It is clear that insignificant variables can improve the prediction accuracy of the model.

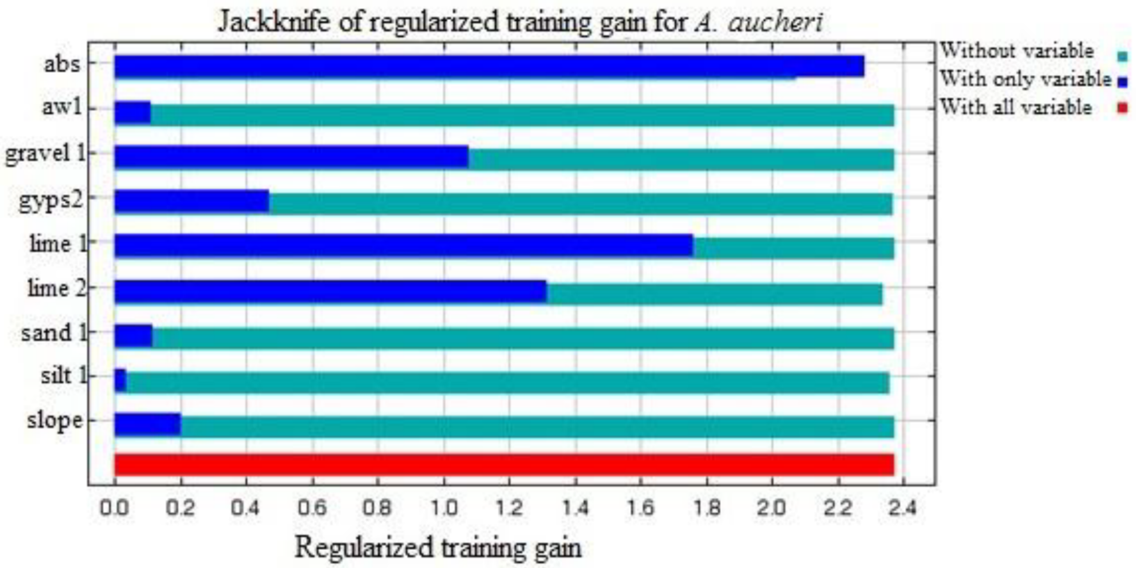

Figure 3. Output of the jackknife test for determination of environmental variables importance value in *Artemisia aucheri* habitat in the rangelands of western Taftan.

In the maximum entropy model, the maximum and minimum values of AUC statistic were 0.96 and 0.84 for *Artemisia aucheri* and *Artemisia sieberi*, respectively. The value of AUC statistic as an index of accuracy of a model in prediction for a species with a vast ecological niche is lower than that for a species with a small niche. Therefore, prediction efficiency of the maximum entropy model is affected by species distribution range or vastness of niche. This method could predict the distribution of species with small ecological niches more accurately in the study. It has been reported that vastness of ecological niche can affect the response curve of a species to predictive variables, distribution range, and efficiency of model (Guisan and Zimmermann, 2000; Luoto and Hjort, 2005; Evangelista et al., 2008; Yang et al., 2013; Piri Sahragard and Zare Chahouki, 2016b). In contrast, Piri Sahragard and Zare Chahouki (2015) reported that maximum entropy was more suitable for distribution modelling of plant species with vast niches. Contrary to logistic regression, maximum entropy does not consider the absence of species. This method predicts the occurrence probability of each species based on the relationship between species distribution and environmental variables (Buehler and Ungar, 2001). Moreover, it is possible to check the effect of each variable on the efficiency of the model separately through AUC statistic. Therefore, variables with low importance can be removed from the model in order to improve its accuracy (Zare Chahouki and Piri Sahragard, 2016).

In general, environmental heterogeneity can generate heterogeneous habitats with a mosaic of conditions in which some areas are favorable for establishment of a specific species, whereas other areas are unfavorable for that species. According to the results of logistic regression and maximum entropy, a large number of environmental variables, including edaphic variables (soil texture, soil lime content, soil moisture content, gypsum percent, gravel percent, and geologic structure) and physiographic variables (slope degree and altitude) contributed to the distribution of studied plant habitats in western Taftan. In other words, a set of soil physicochemical properties and topography of the study area contain invaluable information about the distribution of plant species habitats. Some variables, such as soil EC and pH, made a small contribution to distribution of habitats compared with other environmental variables. The importance of soil properties in distribution of plant habitats has been emphasized by many studies (Zare Chahouki et al., 2010; Abd El-Ghani et al., 2011; Tatian et al., 2011; Hosseini et al., 2013; Piri Sahragard and Zare Chahouki, 2015). In addition, in some geographical conditions the role of physiographic factors such as altitude and slope has been emphasized in occurrence of plant species (Maltez-Mouro et al., 2005; Abdel Khalik et al., 2013; Hosseini et al., 2013; Piri Saharagard and Zare Chahouki, 2016a). Scrutinizing the relationship between environmental variables and species distribution indicates that in flat parts of the studied area where *Haloxylon persicum*, *Zygophyllum eurypterum*, and *Artemisia sieberi* habitats were located the role of soil variables in plant distribution was greater. The results showed that soil variables including texture, organic matter, lime, moisture content, and pH were the input variables of the models for *Haloxylon persicum*, *Zygophyllum eurypterum*, and *Artemisia sieberi*. In other words, the distribution of plant habitats in flat parts of the studied area were mainly affected by soil physicochemical characteristics. However, in the highlands of the studied area, physiographic factors including slope and altitude were the effective variables in distribution of plant habitat, besides soil variables and geologic structure.

It can be concluded that the logistic regression model is basically a linear model that establishes a direct relationship between environmental variables and species presence probability. For this reason, the optimum efficiency of this model is in studies with the objective to investigate a linear relationship between species occurrence probability and environmental variables (for example, relationship between presence probability of a species in a small area in which the environmental gradient is not too long). Meanwhile, maximum entropy explains the relationships between presence probabilities of species with environmental variables in a nonlinear form by providing response curves. In cases where there is a nonlinear relationship between presence probabilities of species with environmental variables, the maximum entropy model will be able to generate a more accurate predictive model when compared with logistic regression. On the other hand, the maximum entropy model compared with logistic regression requires a smaller number of variables to generate an accurate model. This is an important note in practice.

Acknowledgment
This work was supported by the University of Zabol funds (Project No: 40–94). The authors would like to express their gratitude to the Vice-chancellery for Research and Technology, University of Zabol, for funding this project.

References

Abd El-Ghani M, Abo El-Kheir M, Abdel-Dayem M, Abd El-Hamid, M (2011). Vegetation analysis and soil characteristics of five common desert climbing plants in Egypt. Turk J Bot 35: 561-580.

Abdel Khalik K, El-Sheikh M, El-Aidarous A (2013). Floristic diversity and vegetation analysis of Wadi Al-Noman, Holy Mecca, Saudi Arabia. Turk J Bot 37: 894-907.

Araujo MB, Guisan A (2006). Five (or so) challenges for species distribution modeling. J Biogeogr 33: 1677-1688.

Ardestani EG, Tarkesh M, Bassiri M, Vahabi MR (2015). Potential habitat modeling for reintroduction of three native plant species in central Iran. Journal of Arid Land 7: 381-390.

Baldwin RA (2009). Use of maximum entropy modeling in wild life research. Entropy 11: 854-866.

Buehler EC, Ungar LH (2001). Maximum entropy methods for biological sequence modeling.

Proceedings of the 1st International Conference on Data Mining in Bioinformatics, BIOKDD 01: 60-64.

Ebrahimi M, Masoodipour A, Rigi M (2015). Role of soil and topographic features in distribution of plant species (Case study: Sanib Taftan watershed). Ecopersia 3: 917-932.

Elith J, Graham CH (2009). Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography 32: 66-77.

Evangelista P, Kumar H, Stohlgren TJ (2008). Modelling invasion for a habitat generalist and a specialist plant species. Divers Distrib 14: 808-817.

Guisan A, Weiss SB, Weiss AD (1999). GLM versus CCA spatial modeling of plant species distribution. Plant Ecol 143: 107-122.

Guisan A, Zimmermann NE (2000). Predictive habitat distribution models in ecology. Ecol Model 135: 147-186.

Hernandez PA, Franke I, Herzog SK, Pacheco V, Paniagua L, Quintana HL, Soto A, Swenson JJ, Tovar C, Valqui TH et al (2008). Predicting species distributions in poorly-studied landscapes. Biodivers Conserv 17: 1353-1366.

Hosmer DW, Lemeshow S (2000). Applied Logistic Regression. 3rd ed. New York, NY, USA: Wiley.

Hosseini SZ, Kappas M, Zare Chahouki MA, Gerold G, Erasmi S, Rafiei Emam A (2013). Modelling potential habitats for *Artemisia sieberi* and *Artemisia aucheri* in Poshtkouh area, central Iran using the maximum entropy model and geostatistics. Ecol Inform 18: 61-68.

Keating KA, Cherry S (2004). Use and interpretation of logistic regression in habitat-selection studies. J Wildlife Manage 68: 774-789.

Kumar S, Stohlgren TJ (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. Journal of Ecology and Natural Environment 1: 94-98.

Luoto M, Hjort J (2005). Downscaling of coarse-grained geomorphological data. Earth Surf Proc Land 33: 75-89.

Maltez-Mouro S, Garcia LV, Maranon T, Freitas H (2005). The combined role of topography and over story tree composition in promoting edaphic and floristic variation in a Mediterranean forest. Ecol Res 20: 668-677.

Miller J, Franklin J (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecol Model 157: 227-247.

Pearce J, Ferrier S (2000). Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133: 225-245.

Phillips SJ, Dudik M (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161-175.

Phillips SJ, Anderson RP, Schapire RE (2006). Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231-259.

Piri Sahragard H, Zare Chahouki MA (2015). An evaluation of predictive habitat models performance of plant species in Hoze-soltan rangelands of Qom province. Ecol Model 309-310: 64-71.

Piri Sahragard H, Zare Chahouki MA (2016a). Modeling of *Artemisia sieberi* Besser habitat distribution using maximum entropy method in desert rangelands. Journal of Rangeland Science 6: 93-101.

Piri Sahragard H, Zare Chahouki MA (2016b). Comparison of logistic regression and machine learning techniques in prediction of habitat distribution of plant species. Range Manag Agrofor 37: 21-26.

Piri Sahragard H, Ajorlo M (2016). Predicting the distribution of plant species habitats using maximum entropy model (A case study in rangelands of western Taftan, southeastern Iran). Soil Environm 35: 224-234.

Rossiter DG, Loza A (2010). Analyzing land cover change with logistic regression in R [online]. Website http://www.css.cornell.edu/faculty/dgr2/teach/R/R_lcc.pdf [accessed 20 September 2016].

Rushton SP, Ormerod SJ, Kerby G (2004). New paradigms for modelling species distributions? J Appl Ecol 41: 193-200.

Swets JA (1988). Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

Tarkesh M, Jetshcke G (2012). Comparison of six correlative models in predictive vegetation mapping on a local scale. Environ Ecol Stat 19: 437-457.

Tatian MR, Zabihi A, Tamartash R, Shabani M (2011). Determination of indicator species of some soil characteristics by ordination method in Kooh-e-Namak rangelands, Qom. Journal of Environmental Studies 37: 34-52.

Yang XQ, Kushwaha SPS, Saran S (2013). Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. Ecol Eng 51: 83-87.

Zare Chahouki MA (2010). Data analysis in natural resources research using SPSS software. 1st ed. Tehran, Iran: Jahat Press.

Zare Chahouki MA, Azarnivand H, Jafari M, Tavili A (2010). Multivariate statistical methods as a tool for model based prediction of vegetation types. Russ J Ecol+ 41: 84-94.

Zare Chahouki MA, Khalasi Ahvazi L (2012). Predicting potential distributions of *Zygophyllum eurypterum* by three modeling techniques (ENFA, ANN and logistic) in North East of Semnan, Iran. Range Manag Agrofor 2: 68-82.

Zare Chahouki MA, Piri Sahragard H (2016). Maxent modelling for distribution of plant species habitats of rangelands (Iran). Pol J Ecol 64: 453-467.