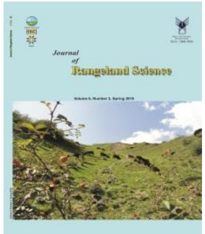
Contents available at ISC and SID

Journal homepage: www.rangeland.ir



Research and Full Length Article:

Modeling of *Artemisia sieberi* Besser Habitat Distribution Using Maximum Entropy Method in Desert Rangelands

Hossein Piri Sahragard[A], Mohammad Ali Zare Chahouki[B]

[A]Assistance Professor, Range and Watershed Department, University of Zabol, Iran (Corresponding Author), Email: hpirys@uoz.ac.ir
[B]Professor, Department of Rehabilitation of Arid and Mountainous Regions, Natural Resources Faculty, University of Tehran, Iran

Received on: 30/12/2014
Accepted on: 19/12/2015

Abstract. Predictive modeling of habitat distribution of range plant species and identification of their potential habitats play important roles in the restoration of disturbed rangelands. This study aimed to predict the geographical distribution of *Artemisia sieberi* and find the influential variables in the distribution of *A. sieberi* in the desert rangelands of central Iran. Maps of environmental variables were generated by GIS software (version 9.3). Predictive distribution maps of *A. sieberi* were produced with Maximum Entropy Method (MaxEnt) and the existing data regarding this species. The agreement of predictive map with the actual map was checked by calculating Kappa coefficient value. Accuracy of predictive models was evaluated using the Area Under the Curve (AUC). Results showed that soil pH and lime content in the surface layer (0-30 cm) and silt percent in both surface and sub-surface soil depths (0-30 and 30-60) had the greatest impacts on the distribution of *A. sieberi* in the study area. Correspondence of actual map with the predictive one was assessed at a satisfactory level (Kappa coefficient = 0.70). MaxEnt is widely used as compared with the other standard methods since it only requires the presence data of a specific plant species to draw the distribution map of its habitat. Additionally, MaxEnt is a generative method and its output can be easily understood by the field practitioners.

Key words: MaxEnt, Geostatistics, Potential habitat, *Artemisia sieberi*, AUC, Kappa coefficient

Introduction

In recent years, modeling the range plants' habitat distribution with GIS facility and statistical methods has progressed very much. On the other hand, distribution patterns and ecological niche of many plant species in the rangeland ecosystems have experienced some considerable changes by the means of anthropogenic activities and natural factors (Zare Chahouki, 2011; Hosseini *et al.*, 2013; Piri Sahragard and Zare Chahouki, 2015). Understanding the changes in the distribution of plant species and determination of potential habitats of major plant species can provide valuable knowledge for the rangeland managers in order to fulfill the restoration practices in the desired degraded rangelands. Use of predictive modeling for mapping the plant habitat distribution can be useful in the restoration operations (Elith *et al.*, 2006). Predictive modeling in a specific area aims to locate the places where the ecological requirements of plant species could be provided or the portion of species potential distribution could be estimated (Anderson and Martinez-Meyer, 2004; Khalasi Ahvazi *et al.,* 2012). Various methods are used to draw the distribution map of plant species habitats while each of these methods has its own advantages and disadvantages. The input data type is one of the challenges in the selection of suitable method. Statistical methods such as regression techniques can be used for predicting suitable habitats for plant species in the places where data are available in terms of species presence and/or absence (Guisan and Zimmermann, 2000; Elith *et al*., 2006; Wisz *et al.,* 2008). Generally, presence data of species are more available than the absence ones. In the case of availability of data, their values in various locations are still questionable. Consequently, the modeling methods that use only presence data as the inputs are of importance (Graham *et al*., 2004). The Maximum entropy method is one of those methods (Phillips *et al*., 2006). MaxEnt is the maximum entropy on the basis of machine learning programs that predict the probability distribution of an individual species occurrence with regard to environmental limitations. The basic idea of MaxEnt is to estimate the unknown probability distribution of a specific species (Phillips *et al*., 2006). MaxEnt is an approach for predicting the species distribution only using presence data and environmental variable (continuous or categorical) layers of a given area. Firstly, the model assesses the environmental layers based on the training data location and secondly, it selects the occurrence probability of a species in the area (Buehler and Ungar, 2001). Basically, when a pixel in an area has equal environmental conditions of the training data, higher values are assigned to this pixel. On the other hand, lower values are attributed to the pixels with different environmental conditions (Negga, 2007). It has been reported that the MaxEnt method works well amongst many other modeling methods (Elith *et al*., 2006) and may remain effective despite its small sample sizes (Hernandez *et al*., 2006; Wisz *et al*., 2008).

Artemisia sieberi habitats have been largely extended across the steppe rangelands of central Iran. It is a desirable forage plant, highly adapted to physical conditions of arid and semi-arid and tolerates the over-grazing while playing a major role in soil and water conservation in the desert rangeland (Zare Chahouki *et al.,* 2010). The objectives of this study were: 1) to assess the geographical distribution of *A. sieberi* in the study area and 2) to find significant variables in the habitat distribution of *A. sieberi***.**

Materials and Methods

Study area

The study site with the area of 3,000 ha is located in the central part of Qom province about 50 km of Qom city, central Iran. The geographic coordinates of the area are 50'50º30" to 50º54'30" E and 34º59'30" to 35'03º30" N (Fig. 1). The study area is in a plain area. Minimum and maximum altitudes in the study area are 796 and 1100 meters above sea level, respectively. This site was chosen because of apparent changes in vegetation cover in relation to soil changes, clarity in vegetation cover variations and possibility of separation in plant communities.

Data collection

Homogeneous units were firstly delineated using basic maps of the study area (digital elevation, aspect, slope and geology maps in the scale of 1:25000). Vegetation sampling was carried out using a randomized-systematic method and taking the physical conditions of the unit into consideration. Four transects with the length of 200-1000 m in each unit were considered for vegetation sampling. Quadrat size was determined by Minimal Area method (Westhoff and Maarel, 1978) which varied from 2 to 25 m^2 depending on the plant species. The sample size was calculated with a statistical method (Cochran, 1977) in each unit. Thus, vegetation samples were taken from 60 quadrats with respect to vegetation cover variations. Vegetation sampling was conducted in the key area with homogeneous conditions. Plant species list, canopy cover percent, habitat boundary, slope, aspect and altitude were recorded in each unit. Eight soil profiles were dug in each unit and soil samples were taken from 0-30 and 30-80 depths. Since most of the root activities happen within the soil depth of 0-50 cm (Bednarek *et al.*, 2005), those depths were selected as the first and second layers for the soil study. Soil characteristics including gravel percent, texture, saturation moisture, available water, lime, gypsum, organic matter, Acidity (pH), Electrical Conductivity (EC) and soluble solute (Na^+, Ca^{2+}, Mg^{2+}, K^+, Cl^-, $Co3^{2-}$, Hco^{3-} and $So4^{2-}$) were measured using standard methods. Soil digital layers were prepared in GIS environment using geo-statistical and kriging interpolation method with the same spatial resolution (pixel size of 30 x 30 m).

Arc GIS 9.3 (available at http://www.esri.com/software/arcgis) and GS^+ 5.0 (http://gamma-design-software-llc.software.informer.com/) were used for mapping the soil properties. Digital elevation map of the area in the scale of 1:25000 was used for drawing the maps of slope, aspect and altitude.

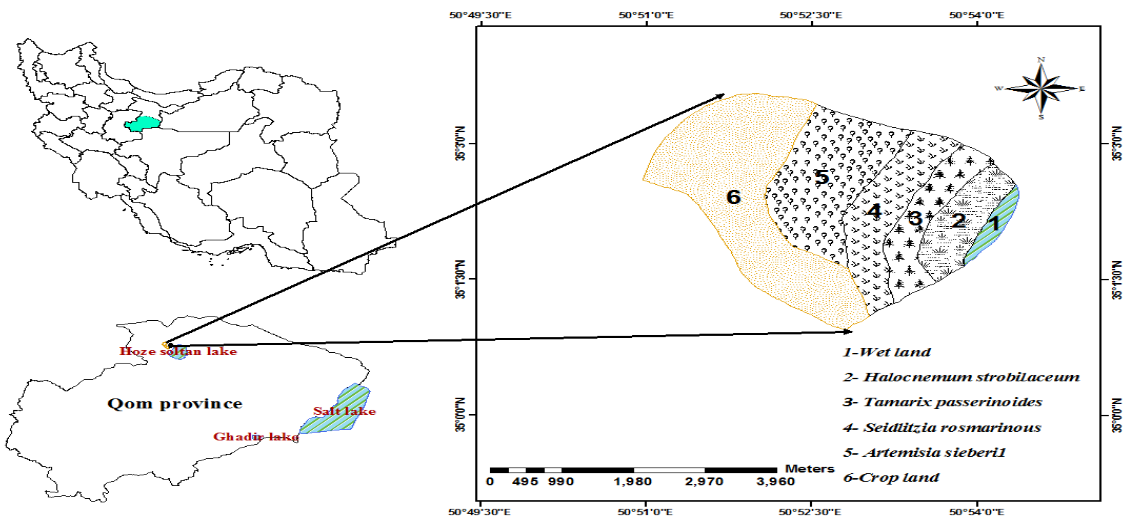


Fig. 1. Location of study area along with vegetation map; the study area is situated in the western Hoz-e-Soltan lake of Qom province, central Iran. The vegetation around the lake is distributed in a strip pattern

Model development

Due to large number of measured variables, principal component analysis (PCA) was used for data reduction in order to find the most important variables while reducing the number of input variables into the MaxEnt model. Matrix of environmental variables for each plant species and PC-ORD 5.0 software were used to conduct PCA analysis. Environmental variables involved saturation of surface soil layer (0-30 cm), silt, gravel percent and gypsum percent in sub-surface soil layer (30-60 cm), lime, clay, pH and EC of both depths. Modeling was performed after preparing the environmental variable maps and their entry in the maximum entropy software. The Area Under the Curve (AUC) of receiver operating characteristic function was used for the evaluation of discrimination ability (Fielding and Bell, 1997). The AUC ranges from 0.5 for an uninformative model to 1 for perfect discrimination. Jackknife analysis was also used to determine the importance of variables. We used MaxEnt 3.1 software (available at http://www.cs.princeton. edu/~ schapire/maxent/) which estimates the probability of species presence ranging from 0 to 1 in which 0 and 1 stand for the lowest and highest probability rates, respectively. Because of 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; Negga, 2007). After determining the optimal threshold using equal sensitivity and specificity method, species presence or absence maps were generated and their coincidence with the actual maps were investigated through calculating the kappa coefficient in the IDRISI 32 release two software.

Results

The results of MaxEnt modeling consist of species distribution maps, importance of predictor variables, and model evaluation with Receiver Operating Characteristics (ROC) curves.

Species distribution maps

After determining the optimal threshold, the agreement of these species distribution maps of two desired species which were derived using the layers of environmental variables at each habitat was checked with the actual maps and then, determined by measuring the Kappa coefficient using IDRISI software. Based on the obtained Kappa coefficients, predictive maps of *Artemisia sieberi* habitat had a very good agreement with the actual ones (Table 1 and Fig. 2).

Table 1. Presence optimal threshold and maps agreement between predictive and actual maps of habitat in the study area

Number	Habitat	Optimal Threshold	Kappa Value	Level of Agreement
1	*Artemisia sieberi*	0.3	0.7	Very good

Importance of Predictor Variable

Based on the Jackknife operation results (Fig. 3), lime, pH1 and silt 2 were identified as the most important variables when they were used individually in the *A. sieberi* habitat. Therefore, these variables have useful information for *A. sieberi* and can provide valuable information about habitat distribution of this species. These results indicated that when other variables such as EC in *A. sieberi* habitat were individually used, they were likely to have little importance in the model gain and when the model was conducted only with these variables, the model gain would not be achieved. This indicates that these variables are not useful for estimating the distribution of species individually.

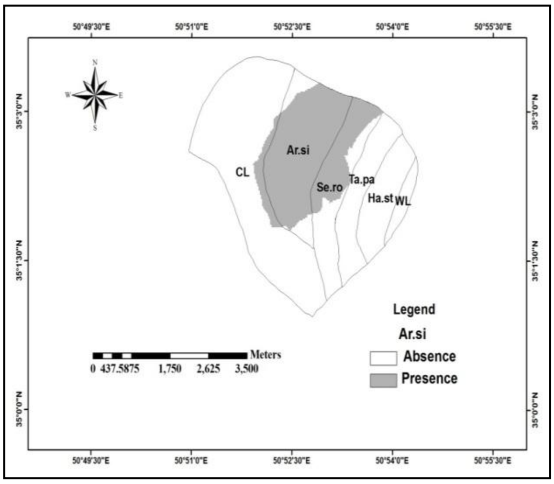


Fig. 2. Predictive and actual species distribution maps for *A. sieberi (*Predictive maps shown darker*)*

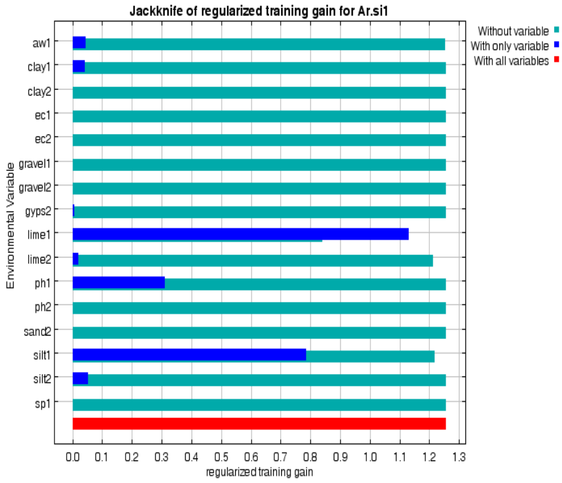


Fig. 3. Jackknife results of variable importance

Response curves

Response curves represent the relationships of environmental variables and suitable habitat distribution of plant species (Fig. 4). This curve can provide useful information about the required environmental threshold for optimal growth of plant species. Response curve analysis of the most important variables of *Artemisia sieberi* showed that the increased lime percent in surface soil depth and the increased silt amount in both soil depths can reduce the presence probability of this species, but the increase in soil pH to 8.5 can increase the presence probability of this species. Therefore, it can be concluded that the habitat with high levels of lime in surface soil (14-16%), silt in both depths (approximately 5 to 10%) and high pH values (8.4-8.8) can provide a suitable habitat for *Artemisia sieberi*.

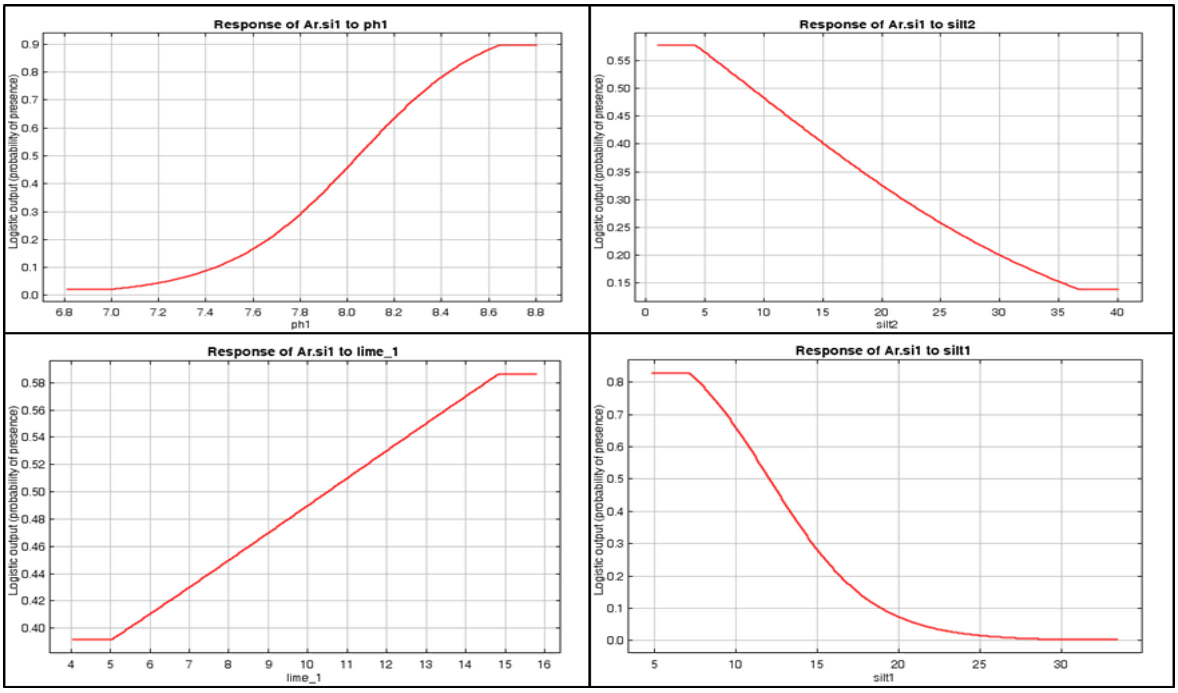


Fig. 4. Response curves of the most influential predictors for *A. sieberi*

Receiver Operating Characteristics (ROC) Curves

In addition to measuring the kappa coefficient for the agreement of desired maps, classification accuracy of predictive habitat models was evaluated using area under ROC curves. According to the obtained AUC values and AUC classification (Sweet, 1988), model predictive accuracy of *Artemisia sieberi* habitat was assessed as an acceptable level (AUC=0.88) (Fig. 5). This is due to the adaptability of *Artemisia sieberi* to diverse habitat conditions. Hence, *A. sieberi* habitat could not be separated with high accuracy by the MaxEnt model.

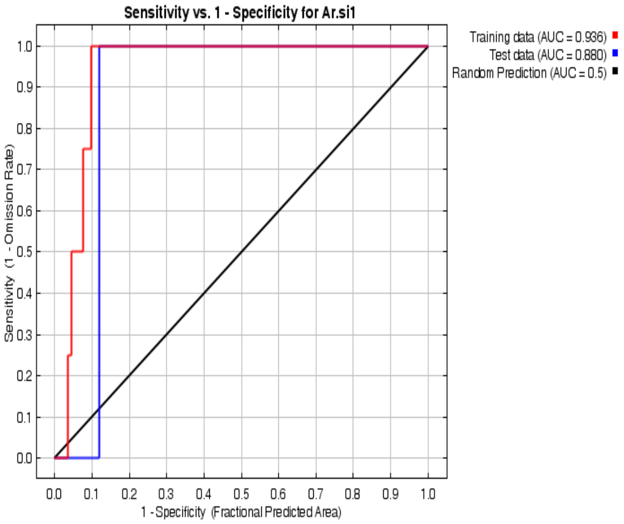


Fig. 5. ROC curves of sensitivity vs. specificity

Discussion

Assessment of accuracy level of model output showed that MaxEnt can successfully predict the occurrence of *Artemisia sieberi* (AUC= 0.88). MaxEnt model indicated that habitat distribution of *A. sieberi* is mainly related to soil characteristics. Soil EC in the Hoze Sultan lake margin is very high, but with distance from the center of playa, soil EC decreased. On the other hand, soil texture was coarse and favorable conditions have been seen for the species which are less resistant to soil salinity. These results indicate that vegetation changes in the study area are largely influenced by soil property gradients. According to the results while getting away from playa center, reducing the soil salinity and increasing the gravel and limestone amounts in the soils, *A. sieberi* has been well established. The MaxEnt model's internal Jackknife test of variable importance showed that soil lime was the most important predictor of *A. sieberi* habitat distribution. These variables presented higher gain (that contained the most information) as compared to the other variables (Fig. 3). Studying the habitat requirements of *A. sieberi* indicated that the highest presence probability of *A. sieberi* is occurred in the soils with high lime content (14-16%). Therefore, *Artemisia sieberi* is calcareous and salt intolerant plant as its density has an inverse relationship with soil salinity (Zare Chahouki, 2012a; Hosseini *et al.*, 2013). In many studies, soil lime content effects on plant growth have been emphasized through its effects on soil pH and the reduction of micronutrients availability such as Zn and Mn (Kourori and Khoshnavis, 2002; Zare Chahouki *et al.*, 2010, 2012b). It can be concluded that vegetation cover variations in the arid regions are chiefly affected by the gradient of soil properties. Soil salinity content was significantly reduced and soil lime content increased by the distance from the center of playa; thus, it will provided favorable conditions for the occurrence of *Artemisia sieberi*.

Caution should be taken when the results of Maximum entropy show that the variable importance is determined based on the MaxEnt algorithm in the Maximum entropy method, which is different from the other applied methods (Phillips *et al.*, 2006). Contrary to GLM and GAM which are diagnostic methods, Maximum entropy method is a generative method and can provide better predictions when the training data are limited (Ng and Jordan, 2001). In this research, the generated predictive maps had a high agreement with the actual distribution maps of desired species. These results indicate that maximum entropy method can be used as an effective method in the prediction of plant species distribution. On the other hand, MaxEnt method does not have many complexities related to the methods that use both presence and absence data whereas it uses only presence data to study the distributions of plant species. In addition, MaxEnt results provide key information about the species tolerance against the effective environmental variables which could be used in the conservation of vulnerable and endangered habitats. This method is useful for avoiding the invasive species infestation in the habitat. Managers can also use the results of this method for the optimal management of resources and the implementation of restoration actions in the relevant areas.

Conclusion

This study reports that the habitat distribution pattern of *A. sieberi* can be modelled with low number of occurrence data and environmental variables using MaxEnt. The MaxEnt model predicted the potential favorable habitat for *A. sieberi* with high success rates (Kappa = 0.7). Thus, the methodology presented here can be used for quantifying the habitat distribution patterns for plant species in the other areas and may contribute to the field surveys and allocation of conservation and restoration efforts.

Literature Cited

Anderson, R. P., Mart´ınez-Meyer, E., 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (Heteromys) of Ecuador. *Biol. Conser.* 116: 167–179.

Buehler, E. C., and Ungar, L. H., 2001. Maximum Entropy Methods for Biological Sequence Modeling. BIOKDD, 60-64.

Bednarek, R., Dziadowiec, H., Pokojska, U., and Prusinkiewicz, Z., 2005. Badania ekologiczno-gleboznawcze (Soil–Ecological Research). PWN, Warszawa. 105p.

Cochran, W. G., 1977. Sampling Techniques. John Wiley and Sons, New York.

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

Fielding, A. H., Bell, J. F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Env. Conserv*. 24: 38–49.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* 19(9): 497–503.

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

Hernandez, P. A., Graham, C. H., Master, L. L., Albert, D. L., 2006. The Effect of Sample Size and Species Characteristics on Performance of Different Species Distribution Modeling Methods. *Ecography* 29: 773-785.

Hosseini, S. Z., Kappas, M., Zare Chahouki, M. A., 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, *Ecological Informatics*, 18: 61-68.

Khalasi Ahvazi, L., Zare Chahouki, M. A., Ghorbannezhad, F., 2012. Comparing Discriminant Analysis, Ecological Niche Factor Analysis and Logistic Regression Methods for Geographic Distribution Modeling of *Eurotia ceratoides* (L.) C. A. Mey. *Jour. Rangeland Sci*., 3(1): 45-57.

Korrouri, S., and Khoshnevis, M., 2000. Ecological and environmental studies of Iranian Juniperus sites, *Research Institute of Forests and Rangelands Press*, 208p.

Negga, H. E., 2007. Predictive Modelling of Amphibian Distribution Using Ecological Survey Data: a case study of Central Portugal, Master thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Ng, A. Y., and Jordan, M. I., 2001. On discriminative versus generative classifiers: a comparison of logistic regression and naive Bayes. *Adv. Neural Inform. Process. Syst*. 14: 605–610.

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

Piri sahragard, H., and Zare Chahouki, M. A., 2015. An evaluation of predictive habitat models performance of plant species in Hoz-e-Soltan rangelands of Qom province. *Ecological Modelling*, 309-310, 64-71.

Sweet, J. A., 1988. Measuring the accuracy of a diagnostic systems. *Science*. 240: 1285-1293.

Verbyla, D. L., and Litvaitis, J. A., 1989. Resampling methods for evaluation of classification accuracy of wildlife habitat models. *Environmental Management*. 13:783-787.

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

Westhoff, V., and Maarel, E. V. D., 1978. The Braun-Blanquette approach. 2nd ed. In: R. H. Whittaker (ed). Classification of plant communities, 287-399.

Zare Chahouki, M. A., Azarnivand, H., Jafari, M., Tavili, A., 2010. Multivariate Statistical Methods as a Tool for Model Based Prediction of Vegetation Types. *Russian Jour. Ecology*, 41(1): 84–94.

Zare Chahouki, M. A., 2011. Multivariate Analysis Techniques in Environmental Science, *Erath and Environmental Sciences*, 1(1): 68-75.

Zare Chahouki, M. A., Khalasi Ahvazi, L., 2012a. 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(2): 68-82.

Zare Chahouki, M. A., Khalasi Ahvazi, L., Azarnivand, H., 2012b. Comparison of three modeling approaches for predicting plant species distribution in mountainous scrub vegetation (Semnan rangelands, Iran), *Polish Jour. Ecology*, 60(2): 105-117.

Zare Chahouki, M. A., and Esfanjani, J., 2015. Predicting potential distribution of plant species by modeling techniques in southern rangelands of Golestan, Iran, *Range Management and Agroforestry Jour.*, 36(1): 66-71.

مدلسازی پراکنش رویشگاه *Artemisia sieberi* Besser در مراتع استان قم با استفاده از روش آنتروپی حداکثر

حسین پیری صحراگرد[الف]، محمد علی زارع چاهوکی [ب]

[الف] استادیار گروه مرتع و آبخیزداری، دانشگاه زابل (نگارنده مسئول)، پست الکترونیک: hpirys@uoz.ac.ir
[ب] استاد گروه احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی دانشگاه تهران

تاریخ دریافت: ۱۳۹۳/۱۰/۰۹
تاریخ پذیرش: ۱۳۹۴/۰۹/۲۸

چکیده. مدلسازی پیش‌بینی پراکنش جغرافیایی و شناسایی رویشگاه‌های بالقوه گونه‌های مختلف گیاهی نقش مهمی در حفاظت و احیای مراتع دارد. پژوهش حاضر با هدف برآورد پراکنش جفرافیایی گونه ***Artemisia sieberi*** و یافتن مهمترین متغیرها در پراکنش این گونه، در مراتع بیابانی ایران انجام شد. نقشه مربوط به متغیرهای محیطی با استفاده از سیستم اطلاعات جغرافیایی (نسخه ۹/۳) و زمین‌آمار ساخته شد. سپس نقشه پیش‌بینی مربوط به پراکنش گونه با استفاده از روش مدلسازی آنتروپی حداکثر و داده‌های مربوط به حضور گونه ساخته شد. میزان تطابق نقشه‌های پیش‌بینی با نقشه‌های واقعی با استفاده از ضریب کاپا مورد ارزیابی قرار گرفت. همچنین دقت مدلهای پیش‌بینی با استفاده از سطح زیر منحنی (AUC) مورد ارزیابی قرار گرفت. بر اساس نتایج حاصل، متغیرهای آهک عمق اول، سیلت عمق اول و دوم و اسیدیته عمق اول به عنوان مهم‌ترین متغیرها در پراکنش گونه ***Artemisia sieberi*** شناخته شدند. میزان تطابق نقشه‌های واقعی و پیش‌بینی نیز در سطح خیلی‌خوب (ضریب کاپای ۰/۷۰) ارزیابی شد. به‌دلیل اینکه روش آنتروپی‌حداکثر برای تهیه نقشه پیش‌بینی پراکنش گونه‌های گیاهی فقط به داده‌های حضور گونه‌ها نیاز دارد، در مقایسه با دیگر روش‌های استاندارد به‌طور گسترده‌ای مورد استفاده قرار گرفته است. علاوه بر این، روش آنتروپی حداکثر یک روش زایا است و نتایج حاصل از این روش می‌تواند به راحتی توسط مدیران اجرایی مورد استفاده قرار گیرد.

کلمات کلیدی: آنتروپی حداکثر، زمین آمار، رویشگاه بالقوه، ***Artemisia sieberi***، AUC، ضریب کاپا