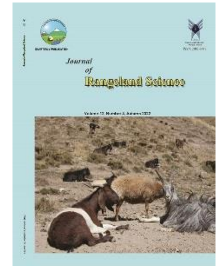
Contents available at ISC and SID
Journal homepage: www.rangeland.ir



Research and Full Length Article:

Ensemble Modeling Approach to Predict the Potential Distribution of *Artemisia sieberi* in Desert Rangelands of Yazd Province, Central Iran

Mohammad Ali Zare Chahouki [A], Peyman Karami[B], Hossein Piri Sahragard [C*]
[A] Prof., Department Rehabilitation of Arid and Mountainous Regions, University of Tehran, Karaj, Iran
[B] Researcher, Department Environmental Sciences, Faculty of Natural Resources and the Environment Sciences, Malayer University, Malayer, Iran
[C] Assistant Prof., Department Rangeland and Watershed Management, Faculty of Water and Soil, University of Zabol, Zabol, Iran *(Corresponding author), Email: hopiry@uoz.ac.ir

Received on: 09/07/2021
Accepted on: 03/10/2021
DOI: 10.30495/RS.2022.685569

Abstract. The object of this study was to compare predictive accuracy of some individual modeling methods versus ensemble modeling approach in estimating the spatial potential distribution and identifying ecological requirements of *Artemisia sieberi* in desert rangelands of Yazd province, Central Iran. For this purpose, the species presence data were collected using the random systematic sampling method in 2019. Individual modeling of the species distribution was performed using Random Forest, Classification and Regression Tree and Generalized Additive Model after preparing environmental variable maps using GIS and geostatistics. Predictive performance of individual models was evaluated using Area Under Curve and Root Mean Square Error statistics. Furthermore, the Ensemble model was used based on the weighted average AUC. The appropriate threshold limit value was calculated based on True Skill Statistic for conversion of continuous maps to binary ones of habitat suitability. Comparison of the performance of individual models showed that the RF model had a more accurate prediction compared to the other models (AUC= 0.971 and RMSE= 0.256). Evaluation of the models implemented using threshold-dependent metrics such as Sensitivity, Specificity, and Kappa index also confirmed this finding. The overall comparison of the results from the three models versus the Ensemble model also indicates the high performance of this model compared to the individual models. Based on Ensemble model results, 45.38% of the study area had a high suitability for the establishment of *A. sieberi*. Based on the analysis of the importance of variables in the RF model, elevation (42%), Clay (40.02%) and pH (38.97%) in 0-30 cm soil depth had the highest effect on the presence of species. In general, Ensemble modeling can reduce the uncertainty and provide more reliable results by combining the results of the different algorithms of individual modeling.

Key words: Species distribution modeling, Weighted average AUC, Ensemble model, Soil properties, Iran's desert rangelands

Introduction

Species distribution models are important tools in proper management of rangelands. In other words, predictive distribution models have an important role in supporting the decisions related to the conservation of plant habitats by providing information on the species ecology. So, the use of these models is increasing dramatically (Sor *et al.*, 2017; Mi *et al.*, 2016). Due to the complex nature of the ecological systems, the selection of key predictive environmental variables is one of the challenges ahead for using these models and can restrict the creation of accurate models. In the meantime, some modeling methods such as machine learning methods can estimate the habitat distribution of plant species accurately by providing a framework to identify these variables (Franklin, 2010; Evans *et al.*, 2011; Piri Sahragard and Zare Chahouki, 2015). Selecting the optimal modeling method can lead to precise predictions of plant species distribution range, based on the ecological requirement, as well as estimating the species distribution where despite the suitability of habitat conditions, the species presence is limited or unknown (Hare *et al.*, 2007). From a practical point of view, knowing which model or group of models produces more accurate predictions at different time and spatial scales is the core of resource management (Stohlgren *et al.*, 2010).

A variety of methods such as regression models and machine learning methods were used in modeling the distribution of plant species. Among the machine learning methods, Random Forest (RF) are usually multiple models created on various sets of data, and the use of these methods is recommended when the goal is to identify the important and influential factors on the performance of the model (Guo *et al.*, 2015). On the other hand, the Generalized Additive Model (GAM) is useful to recognize the shape of species response to environmental factors due to the flexibility in determining the type of the relationship between the response and predictor variables, and is more flexible compared to the generalized linear method (Leathwick *et al.*, 2005). In other words, the GAM model has better performance than GLM model in plant species distribution modeling (Jafarian and Kargar, 2017). Although in the case of the hierarchical effects between predictor variables, decision tree based models such as Classification and Regression Tree are suitable and the results of these methods enjoy more ecological interpretability compared to other machine learning methods (Death and Fabricius, 2000).

In general, the accuracy of modeling habitat suitability depends on several factors such as available information, variables used in modeling, and the theoretical foundations of each modeling method (Austin, 2007). The performance of individual species distribution models varies widely between species and methods and always has some degree of error and uncertainty (Poulos *et al.*, 2012). Measurement error, subjective judgment, and model uncertainty are three sources of uncertainty in modeling, and each of the mentioned items contains several parameters (Ray and Burgman, 2006). Using Ensemble models that integrates the results of individual models and provides a strong model of probability of distribution, one can overcome somewhat of the uncertainties in the modeling results (Marmion *et al.*, 2009).

Chen *et al.* (2015) used various types of predictive models to assess the relative importance of biogeographical variables and predict the distribution of invasive species in China, they reported that RF and MART methods had higher accuracy. Breiner *et al.* (2015) modeled the potential habitat of 105 rare plant species using maximum entropy (MaxEnt) methods, GLM and Generalized Boosted Models (GBM). The results indicated that the performance of Ensemble models, as one of the ways to reduce over-fit and reduce the efficiency of modeling methods, is much higher than that of individual ones. The comparison of predictive accuracy of Tree net, RF, CART and MaxEnt methods in predicting the distribution of rare species showed that RF was the most accurate method to estimate the potential distribution of the species in non-sampling areas (Mi *et al.*, 2017).

Species Distribution Models (SDMs) have different performances; therefore, choosing a model with the highest accuracy in achieving the modeling goals with low habitat variables resulting in the lowest cost is challenging. This is achieved only by recognizing the capabilities and deficiencies of different methods in estimating the distribution range of plant species with different ecological niches and using comparative studies.

Artemisia sieberi, as suitable forage in the steppes of Iran, has an important role in water and soil conservation in desert areas. In recent years, its distribution has faced serious constraints due to the overexploitation (Zare Chahouki and Piri Sahragard, 2016). Therefore, modeling the distribution of this species provides not only the possibility of recognizing potential habitat distribution, but also provides sustainable exploitation and conservation of its habitats. Accordingly, the present study was carried out to compare predictive accuracy of individual modeling methods including GAM, CART and RF versus an ensemble modeling approach in estimating the potential distribution of *A. sieberi*, as well as identifying its ecological requirements in Poshtkouh rangelands of Yazd province in central Iran.

Materials and Methods

The study area

The Poshtkouh rangelands of Yazd province with an area of 170,000 ha, are located between the northern latitude of 31º 04' 27" - 31º 33' 11", and the eastern longitude of 53º 40' 06"-54º 15' 19" (Fig. 1). The maximum and minimum elevation of the study area is 3958 m and 1389 m above sea level. The average precipitation ranges between 270 mm at Shirkouh elevations to 45 mm in the margin of Chah Beiki kavir (Zare Chahouki and Piri Sahragard, 2016).

Data collection and pre- processing

Vegetation sampling was carried out within the pure vegetation types of *A. sieberi* in the homogeneous units using a random-systematic method along 3 to 5 transects of 300 to 500 m. Depending on the topography and changes in vegetation, the distance between transects was 300 and 500 m in mountainous and plain regions, respectively.

The quadrate size, depending on the species type and density, was calculated to be 2 m^2 using the minimal area method. The sample size was determined according to the changes in vegetation and calculated to be 50 plots using the statistical method. Vegetation sampling was performed after determining these parameters. To measure the soil properties, soil samples were taken along each transect from two depths of 0-30 and 30-80 cm (A total of 110 soil samples). Since soil profiles must have a suitable distribution to prepare a soil properties map in addition to sampling in the habitat of the study species, soil sampling was performed in the entire study area at similar depths (Carter *et al.*, 2006). The points harvested in the habitats of other species were considered as absence points for *A. sieberi*.

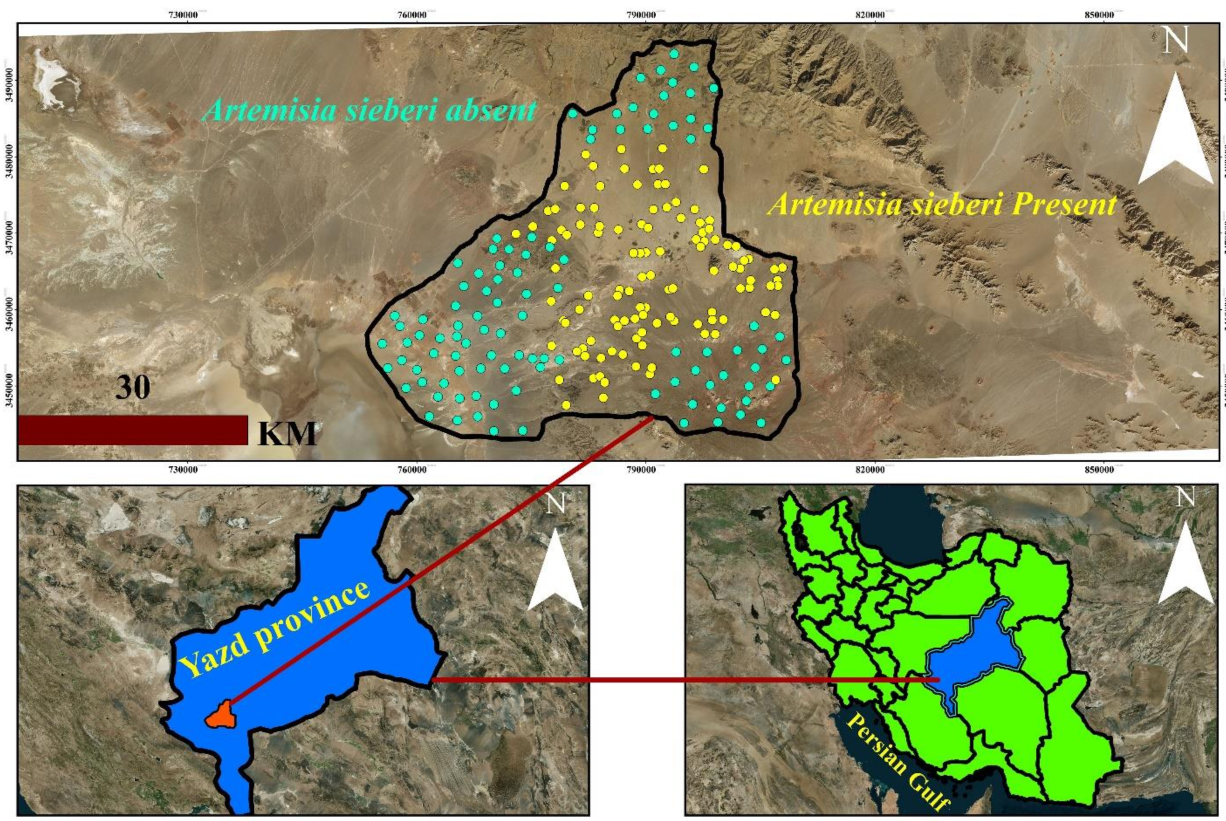


Fig. 1. Map of location of the study area in Yazd province, Iran along with the presence and absence of *A. sieberi*

In the laboratory, the soil properties including gravel, texture, saturated moisture, available moisture, lime, gypsum, organic matter, acidity, electrical conductivity, and soluble salts were measured. The soil properties map was prepared in software GS^+ (version 5.1) Preparation of maps of habitat variables with the same spatial resolution (30 x 30 m pixel size) was performed in the ArcGIS 10.2 software environment. Fig. 2 shows the environment variables used in the modeling process. In general, 19 environmental variables including topography and its related variables and soil variables were used for modeling (Fig 2). In addition to topographic variables such as elevation, gradient percent and roughness of the earth, parameters related to the Heat Load Index and Compound Topographic Index (CTI) were used. Compound Topographic Index (CTI) shows the ratio between slopes in the basin and is calculated by the following equation,

$$CTI = \ln\left(\frac{A_s}{\tan \beta}\right) \text{ (Equation 1)}$$

Where:

A_s is the basin area and

β is the slope gradient in degrees (Regmi *et al.*, 2010).

In order to obtain the heat load of the geographic direction is converted to heat load using equation 2:

$$[1 - \cos(\theta - 45)]_{/2} \text{ (Equation 2)}$$

Where:

The value of θ is Azimuth of direction in degree. The value of the heat load index is between 0 and 1. Heat load in the northeast is zero (the coolest direction) and in the southwest is 1 (the warmest direction) (McCune and Keon, 2002).

In addition, the Roughness Index was used as a benchmark for calculating topographic heterogeneity in a given range using a digital elevation model. The soil properties map was prepared in software GS^+ (version 5.1). Preparation of maps of habitat variables with a similar spatial resolution (30 x 30 m pixel size) was performed in the ArcGIS 10.2 software environment. Fig. 2 shows the environment variables used in the modeling process.

Based on the multicolinearity analysis, organic matter (om) and electrical conductivity (ec) in 0-30 cm and 30-80 soil depths had high variance inflation factor and therefore did not enter into modeling. A total of 112 presence points were recorded for the species. Since the used methods also require absence data, the absence points were obtained by considering the presence of other plant species. In other words, areas where other plant species existed and the presence of this species was not definite were considered absence points for the studied species. A total of 112 presence points and 101 absence points were used for modeling. In order to run the models, all presence / absence data were selected in the ratio of 70 to 30 for training and testing. 72 presence points and 69 absence points for model training and 40 presence points and 32 absence points for model testing were used. Random points were selected using the Random Point command in QGIS software.

Individual Modeling

The RF method is the expanded version of tree regression model, which includes a large number of classification and regression trees (Breiman, 2001). This model shows the importance of the variables using the Mean Decreased Accuracy (MDA) and Mean Decreased Gini (MDG) coefficient approach (Pahlavan-Rad *et al.*, 2014). The RF model was implemented using the Random Package in the R.3.5.2 software. To implement the model, its assumptions were used. First, the model was run with 100 trees, then 500, and eventually 1000 trees to increase the accuracy and reduce the misclassification error.

The CART method is considered as a powerful tool in determining the most important independent variables, solving classifying problems and prediction. This method creates its branches in binary form, and only based independent variables. The best branch creation occurs when the resulting branches are such that in each branch, a class overcomes (overlaps with) the other classes and reduces the variety in the classes as a criterion for evaluating the branches (Pakgohar, 2016). In order to implement the method, the Rpart package was used in R.3.5.2 software.

The use of the GAM modeling technique has been proposed due to some bugs such as unrealistic response curve shape using GLM (Heegaard, 2002). In a generalized model, it is assumed that the Y response variable has a distribution of exponential families with mean μ= E (Y|X1,..., Xp), which is connected to the predictor variable via the link function (g). Generic collective models are expressed as (Eq. 3):

$$g(\mu) = a + \quad \text{(Equation 3)}$$

Where:

fj is the smoothing functions that are estimated using the data and advanced smoothing techniques of distribution chart (Hastie and Tibshirani, 1990).

In this study, the Akaike Information Criterion (AIC) was used to select the appropriate model based on goodness of fit. The overall performance of species-niche model run was compared to other runs using the (AIC) and variable parameter significance p-values. The recession method, Logit activator function and the spline smoother were used in R 3.5.2 software.

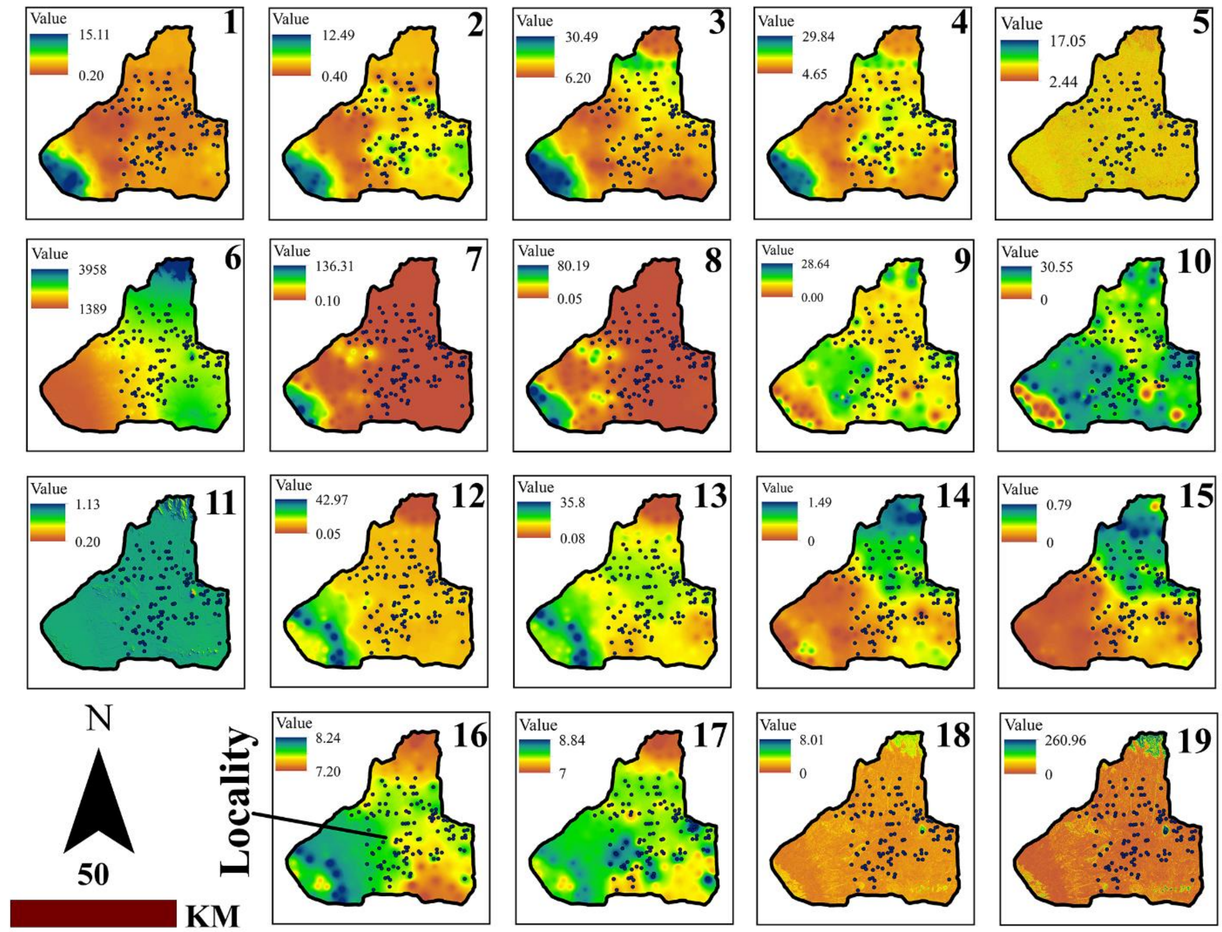


Fig. 2. Environmental variables used in the distribution modeling of *A. sieberi* in study area. The numbers in each map indicate the name of the map. (Maps 1 and 2: Available moisture content of the first and second depth of soil, Maps 3 and 4: Clay content of the first and second soil depths, Map 5: Index of humidity (CTI), Map 6: Digital elevation map, maps 7 and 8: Electrical conductivity of first and second depth, Maps 9 and 10: Gravel size of First and Second depths, Map 11: Heat Load Index (HLI), Maps 12 and 13: Lime Levels in first and second depths, Maps 14 and 15: organic matter of first and second soil depths, maps 16 and 17: pH of the first and second depths of the soil, map 18: roughness of the surface of the earth and map 19: the slope map)

Ensemble modeling approach

Ensembles approach is an appropriate solution to reduce prediction uncertainty in individual modeling. There are various ways to achieve an Ensembles map that is used in various studies. In the present study, based on the nature of the methods used and the available data for evaluating binary outputs, the weighted average based on AUC method from the Ensemble method was used to develop a consensus model. In this method, the habitat suitability i^{th} grid cell (HSi) in the integration model is calculated by the equation 4 (Vignoli *et al.*, 2009).

$$HS_i = \frac{\sum_j (AUC_{mj} \times mj_i)}{\sum_j AUC_{mj}} \text{ (Equation 4)}$$

Where:

mji is the probability of the presence of species in the cell i developed by the j model.

The calculations for the operation mentioned were performed in ArcGIS 10.2. The output map was evaluated using presence / absence points.

Evaluation of model performance

The evaluation of the models used in this study was conducted for continuous and discrete outputs separately. The AUC and RMSE statistics were used to evaluate the continuous maps of habitat suitability. After preparing the prediction map, the optimal threshold of species presence was determined using the Youden Index or True Skill Statistic (TSS) (Allouche *et al.*, 2006; Wilks, 2011). After identifying the correct cutoff value (threshold limit value), binary map of habitat suitability was evaluated using Accuracy, Sensitivity, Specificity and Cohen's Kappa indices in R 3.5.2 software.

Results

a) Evaluation of classification accuracy and overall accuracy of prediction models

Based on the values of the AUC and RMSE statistics, the RF model had less error than the other individual models, and as a result, this model has been able to predict the probability presence of the *A. sieberi* with a higher accuracy (Table 1). The continuous map of habitat desirability of *A. sieberi* derived from the individual used models and the binary map of the habitat suitability after applying the Youdon's threshold limit on the continuous maps is presented in Fig. 3. Statistics of the accuracy evaluation of predictive models after applying the threshold limit and classification of binary maps of for all the used models indicate the superiority of the Ensemble approach compared to individual models (Table 2). The Youdon's index value for the Ensemble map was 0.763. Fig. 4 shows continuous and binary maps of the habitat desirability in the Ensemble model.

The AUC metric value for the Ensemble model in this study was 0.957, which is lower than the RF model and higher than the other single models used in this study. If the predictive power of the model is based on the Sensitivity metrics and Specificity of the model, the Ensemble model has the highest Sensitivity among the models used. The Sensitivity value of this model was 0.917, which shows that the Ensemble model has been able to correctly predict about 91% of the presence points of the species in the study area.

Table 1. Evaluation of the efficiency of individual models used in *A. sieberi* species distribution modeling process

Model	Performance Summary	
	AUC	RMSE
Random Forest	0.971	0.256
CART	0.885	0.387
GAM	0.913	0.356

AUC: Area Under Curve; RMSE: Root Mean Square Error
CART: Classification and Regression Tree; GAM: Generalized Additive Model

Table 2. Evaluation of the efficiency of threshold limit applied in identifying of habitat suitability of *A. sieberi* in the used models

Model	Index of Performance of Cut off					
	Accuracy	Sensitivity	Specificity	Cohen's kappa	Area (ha)	Area (%)
CART	0.861	0.875	0.844	0.719	92717	52.84
GAM	0.889	0.90	0.875	0.775	74832	42.64
Random Forest	0.929	0.897	0.968	0.857	73661	41.98
Ensemble	0.957	0.917	1.00	0.915	79628	45.38

CART: Classification and Regression Tree; GAM: Generalized Additive Model

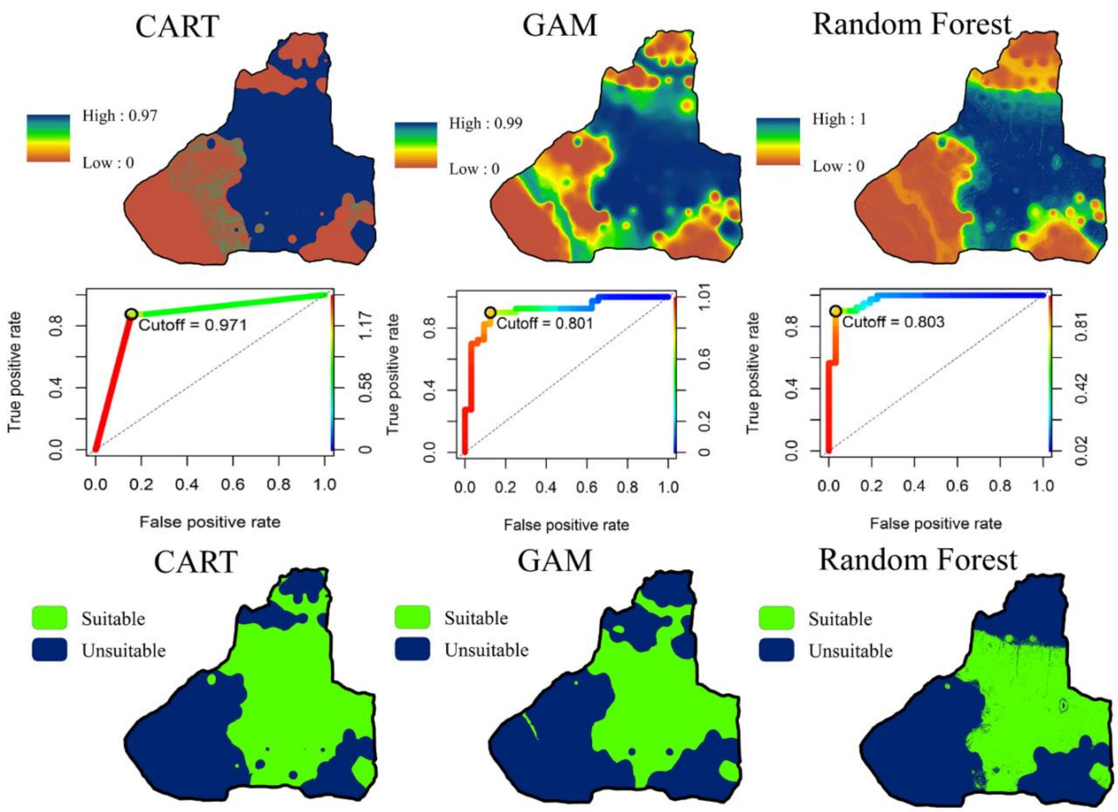


Fig. 3. Prediction probability maps, threshold identified for each map, and binary prediction map of individual models

b) Analysis of the importance of environmental variables

Based on the result of analysis derived from the RF model, elevation has remarkable effect on the presence of the *A. sieberi* species (42%) (Table 3). In addition, the clay content of the second depth (Clay2) and pH of the first depth of soil (pH1) also had the similar effect on the presence and absence of species. On the contrary, moisture index and heat load index had the least effect on the occurrence of the *A. sieberi*. Result shows that by increasing the value of clay 2 up to 16 percent, the presence probability of the species increases and with the increase of clay to more than 16%, the habitat desirability decreases. So, in areas with 25-30% clay 2, the suitability of habitat is zero. Similarly, the response of the species to the pH of the first depth of soil (pH1) was similar to that of the other variables. There was maximum habitat suitability in pH of 7.7 to 7.8, but with increasing this amount, the suitability of habitat decreased and the probability of absence of species increased.

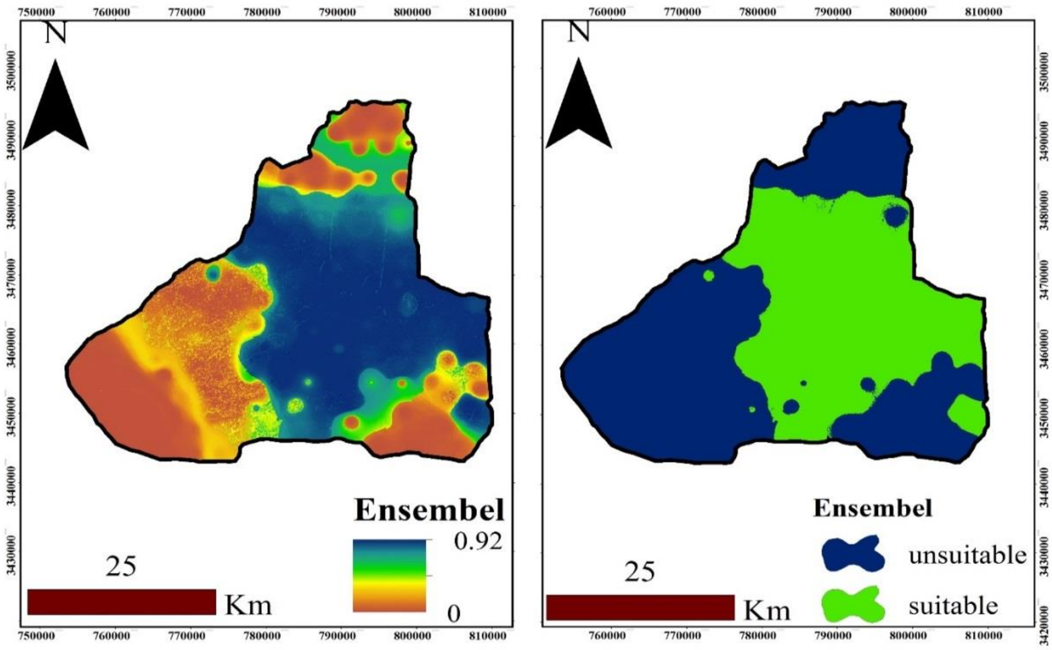


Fig. 4. Continuous (left) and Binary maps (right) of the habitat of *A. sieberi* derived from the Ensemble model

The results of the CART analysis indicate that the variables Clay in 0-30 cm and 30-80 soil depths and Digital elevation Model (DEM) have the most effect on the *A. sieberi* occurance (Fig. 5). In fact, the pruned tree shows that based on the characteristics of clay2, clay1, roughness and pH1, the final decision can be made in relation to presence / absence of species. Accordingly, if the clay content of clay 2 is more than 9.6%, roughness value ≥1.2 and pH1 is greater than 7.6, the presence of the species will occur.

Table 3. Analysis of the variable importance in A. *sieberi* species occurrence derived from RF method

Environmental Variable	Absence	Presence	MDA	MDG
Digital Elevation Model	35.84	38.52	42.00	7.709
Clay2	33.35	38.28	40.20	6.995
pH1	31.31	36.37	38.97	6.555
Clay1	31.42	35.84	38.70	6.714
Available water 1	24.72	30.05	30.80	3.637
Lime1	22.14	24.51	26.07	3.049
Available water 2	19.82	19.75	23.73	2.058
Lime2	16.19	16.34	20.11	1.605
pH2	14.69	14.80	18.05	1.374
Gravel1	9.69	13.63	15.55	1.622
Slope	6.05	9.02	10.62	0.894
Roughness	3.27	10.84	10.12	0.819
Gravel2	5.66	6.30	7.99	0.893
Compound Topographic Index (CTI)	4.20	3.70	5.60	0.443
Heat Load Index (HLI)	-2.87	7.52	4.81	0.429

MDA: Mean Decreased Accuracy; MDG: Mean Decreased Gini

Code 1 and 2 show 0-30 and 30-80 cm soil depths, respectively.

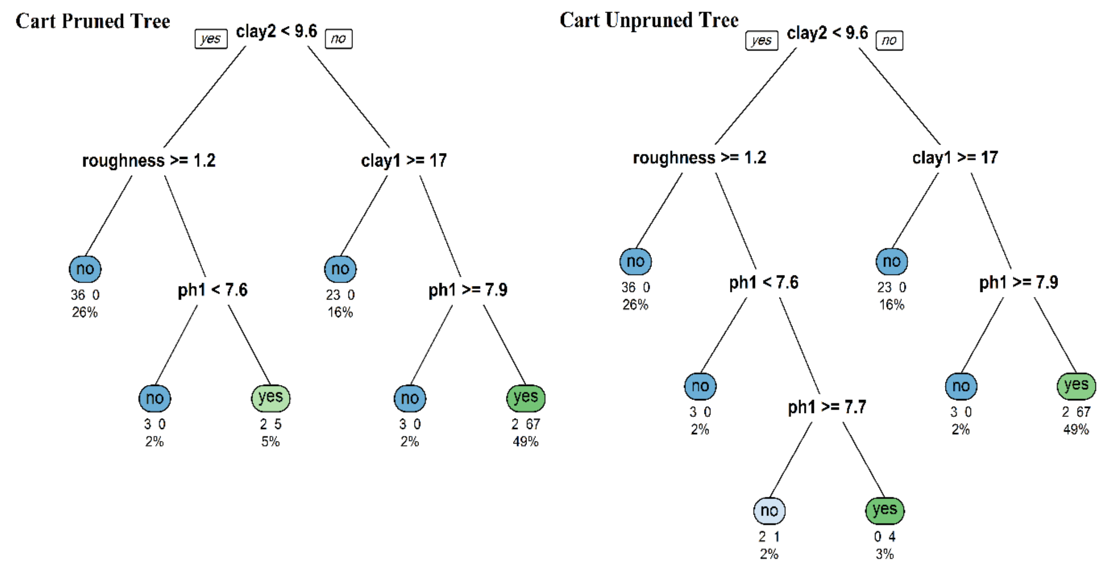


Fig. 5. Pruned and unpruned decision trees derived from the CART method

Based on the results of full model implementation, clay1 and clay2 with the AIC criterion of 93.76 had the highest effect on the presence and absence of *A.sieberi* (Table 4). So, in the reduced model, the value of the AIC statistics for this run with these two variables was 63.68 (Table 5). Accordingly, the effect of clay1 and clay2 variables on the probability of presence is more than other variables in this study. The results show that for clay1 variable, by increasing this variable up to 13%, the habitat suitability was increased, but by increasing the amount of clay1 to more than this amount, the habitat suitability is reduced. This trend is different for clay2 so that by increasing its value to more than 15%, the habitat suitability and probability of habitat of species presence increase (Fig. 6).

Table 4. Predictive variables and degree of freedom derived from running full model in the GAM method

Variable	Df	Standard Error	P-value	GAM Coef.	AIC
pH1	3.139	2.108	0.212	-11.471	
Digital Elevation Model	3.344	0.001	0.390	0.006	
Compound Topographic Index	3.067	0.188	0.998	-0.602	
Lime2	3.062	0.122	0.788	-0.299	
Clay2	2.697	0.195	0.042	-0.257	
Lime1	2.880	0.118	0.812	0.788	
Clay1	2.866	0.157	0.010	0.072	
Heat Load Index	2.994	4.547	0.985	-1.275	93.76
Available water 2	3.029	0.222	0.637	-0.565	
Gravel2	2.984	0.072	0.994	-0.243	
Slope	3.372	0.104	0.994	0.052	
Available water 1	3.194	0.282	0.684	1.088	
Gravel1	2.965	0.095	0.974	0.383	
Roughness	3.301	1.781	0.982	-0.733	
pH2	2.905	1.105	0.837	4.626	

Code 1 and 2 show 0-30 and 30-80 cm soil depths, respectively.

Table 5. Predictive variables and degree of freedom derived from the reduced model in the GAM method

Variable	Df	Standard Error	P-value	GAM Coef.	AIC
Clay2	2.780	0.121	0.023	-0.537	63.68
Clay1	3.079	0.122	0.000	0.881	

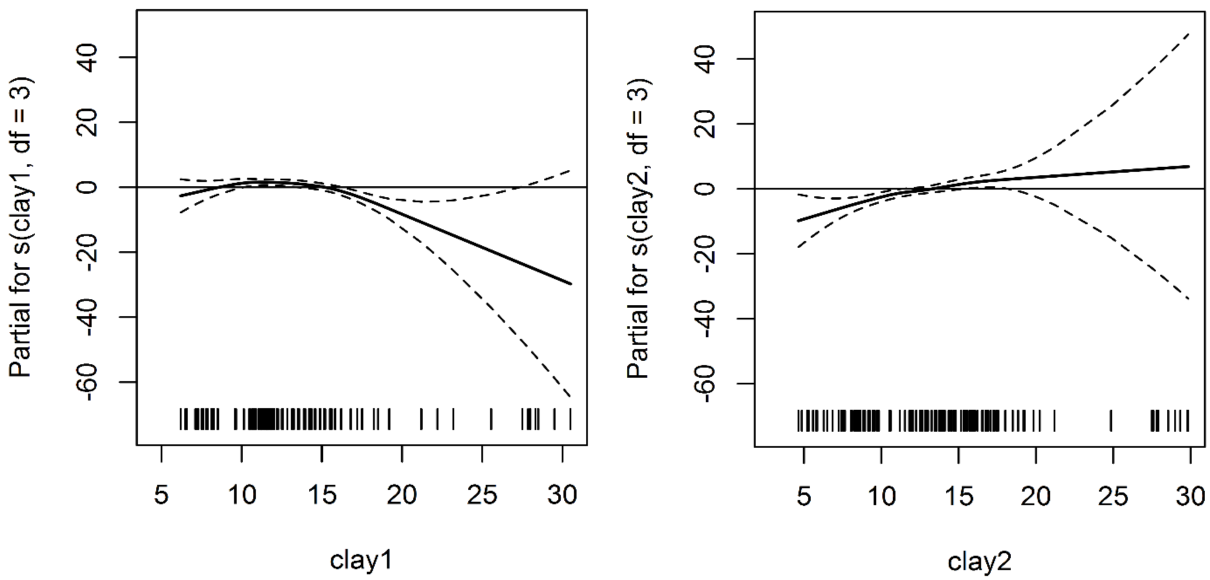


Fig. 6. Relationship between significant environmental variables in the GAM model with the presence of *A.sieberi*

Discussion

Complex ecological analyses require robust and flexible analytical methods that can consider nonlinear relationships, interactions, and missed data. Comparison of the results of continuous maps of habitat desirability prediction showed that based on the individual model performance evaluation, the RF model has better performance than two other models in identifying the suitable habitat of the species (AUC = 0.971; RMSE = 0.256). In other words, the RF model has a higher efficiency in modeling the distribution of species and estimating the potential distribution of species using presence / absence data. The high accuracy of the RF model compared to other models has been reported in similar studies (Benito Garzon *et al.*, 2006; Evans and Cushman, 2009; Guo *et al.*, 2015). The GAM method showed a moderate performance in terms of prediction accuracy, and was ranked among these two methods (AUC= 0.913; RMSE=0.356). GAM is a data-driven method that by discovering nonlinear and non-uniform relationships between the response variable and the predictor variable set, can maximize the predictive efficiency of the response variable (Jafarian and Kargar, 2017). The results of the accuracy evaluation of the models studied also show that the CART model has a lower accuracy in estimating the distribution range of *A. sieberi* species compared to the other two methods (AUC= 0.885; RMSE= 0.387).

On the other hand, the results from the application of the Ensemble model as the resultant of all the individual models showed that the Ensemble models could be more efficient than individual modeling methods. In other words, due to the use of outputs of various models, the Ensemble model in addition to reducing uncertainty will also reduce the weaknesses in individual models (Capinha and Anastácio, 2011). In line with the findings, it has been reported that using Ensemble models as a resultant of all the models can be more efficient than other individual models (Latif *et al.*, 2013; Poulos *et al.* 2012). According to the mentioned points, it should be noted that when models are combined based on the AUC values and the Ensemble model is formed, the performance of the model increases considerably. So, the values for the accuracy and Kappa index for this model would increase to 0.957 and 0.915, respectively, which itself suggests a higher efficiency of the Ensemble model compared to the individual models.

Soil properties are important factors in the distribution of plant species, especially in arid and semi-arid regions (Hosseini *et al.*, 2013; Piri Sahragard *et al.*, 2019). According to the analysis of the importance of influential variables in predicting models, soil properties (including physical properties such as soil texture and chemical properties such as pH and $CaCO_3$) play a crucial role in habitat suitability for the establishment of this species as the elevation does in species distribution. The effect of elevation on the distribution of suitable habitats of this species has been reported in other studies (Hosseini *et al.*, 2013; Kargar-Chigani *et al.*, 2017; Amiri *et al.*, 2019). Contrary to this finding, it has been reported that altitude did not affect the *A. sieberi* distribution, while the average annual temperature affected its distribution (Mousaei Sanjerehei and Rundel, 2017).

Numerous researchers have emphasized the role of different soil properties in the distribution of this species. Accordingly, the direct effect of available soil moisture on species distribution (Zare Chahouki *et al.*, 2012); the positive impact of increasing the amount of lime and soil organic matter on the presence of the studied species in the habitat (Hosseini *et al.*, 2013); the effect of increasing lime, pH, and soil silt content on habitat suitability for this species (Piri Sahragard and Zare Chahouki, 2016), and the effect of the appropriate amount of $CaCO_3$ on species growth (Kargar-Chigani *et al.*, 2017) have also been reported. The positive effect of appropriate amounts of $CaCO_3$ on the growth and distribution of the species can be due to the modification of soil acidity, increasing nutrient uptake, and thus, habitat suitability for the species under study (Hosseini *et al.*, 2013; Piri Sahragard *et al.*, 2019). Consistent with the findings of this study, it has been reported that an adequate amount of $CaCO_3$, especially in soils of arid and semi-arid regions can be beneficial for plant growth, especially this species. The positive correlation between the amount of canopy cover of *A. sieberi* and the amount of $CaCO_3$ confirms this (Sanjerehei, 2012). The amount of soil clay is another feature affecting the habitat suitability of *A. sieberi*. The amount of soil clay as one of the factors affecting soil texture has caused soil texture to be recognized as an important factor in soil moisture, soil formation and aeration, and ultimately affecting the distribution of plant species (Piri Sahragard *et al.*, 2019).

Conclusion

Studying the relationships between environmental factors and plant species distribution and providing reliable scientific resources at the local scale are essential for habitat conservation and restoration programs. Based on the results obtained in the present study, the results of the Ensemble modeling approach are more suitable for this purpose compared to other single modeling methods due to the reduction of error rates of individual modeling algorithms. Thus, the Ensemble-based model could be considered a part of a management support system to adopt the right management decisions in the regeneration of native species such as *A. sieberi* in the rangelands. It leads to the preservation of native species in natural habitats and sustainable use of rangeland vegetation. Physical (soil texture) and chemical properties of soil (pH and $CaCO_3$) and elevation are the variables affecting the distribution of species. Low-elevation areas where soil contains good levels of $CaCO_3$ from an ecological point of view are more suitable for the distribution of this species. In general, besides elevation, soil-related properties such as clay content of the first and second depths have a decisive role in the occurrence of *A.sieberi* species. Accordingly, the central and eastern regions of Poshtkouh rangelands in Yazd province are potentially suitable for species establishment. Therefore, it is suggested that these areas are prioritized in the species habitat development program in these rangelands.

Acknowledgment
Authors would like to express their gratitude to Vice-chancellery for Research and Technology, University of Zabol for funding this project. This work was financially supported by the University of Zabol, Iran (Grant number: IR-UOZ-GR-8694).

References
Allouche, O., Tsoar, A. and Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43(6): 1223-1232.

Amiri, M., Tarkesh, M. and Jafari, R., 2019. Predicting the distribution of *Artemisia Sieberi* Besser under climate change in the steppe and semi-steppe of Iran-Touranian region. Desert Management, 7(13): 29-48.

Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modelling, 200:1-19.

Benito Garzón, M.B., Blazek, R., Neteler, M., deDios, R.S., Ollero, H.S. and Furlanello, C., 2006. Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestr*is L. in the Iberian Peninsula. Ecological Modelling, 197: 383-393.

Breiman, L., 2001. Random forests. Machine Learning, 45(1): 5-32.

Breiner, F.T., Guisan, A., Bergamini, A. and Nobis, M.P., 2015. Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution, 6(10):1210-1218.

Capinha, C. and Anastácio, P., 2011. Assessing the environmental requirements of invaders using ensembles of distribution models. Diversity and Distributions, 17: 13-24.

Carter, G.M., Stolen, E.D. and Breininger, D.R. 2006. A rapid approach to modeling species–habitat relationships. Biological conservation, 127: 237-244.

Chen, L., Peng, S. and Yang, B.,2015. Predicting alien herb invasion with machine learning models: biogeographical and life-history traits both matter. Biological Invasions, 17: 2187-2198.

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

Evans, J.S. and Cushman, S.A. 2009. Gradient modeling of conifer species using random forests. Landscape Ecology, 24: 673-683.

Evans, J., Murphy, M.A., Holden, Z.A. and Cushman, S.A. 2011. Modeling species distribution and change using Random Forest. In: Drew CA, Wiersma YF, Huettmann F (eds.) predictive species and habitat modeling in Landscape Ecology. Springer New York, 139-159.

Franklin, J., 2010. Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press.

Guo, C., Lek, S., Ye, S., Li, W., Liu, J. and Li, Z., 2015. Uncertainty in ensemble modelling of large-scale species distribution: effects from species characteristics and model techniques. Ecological Modelling, 306: 67-75.

Hare, K.M., Hoare, J.M. and Hitchmough, R.A., 2007. Investigating natural population dynamics of Naultinus manukanus to inform conservation management of New Zealand’s cryptic diurnal geckos. Journal of Herpetology, 41: 81-93.

Hastie, T. and Tibshirani, R., 1990. Nonparametric logistic and proportional odds regression. Journal of Applied Statistics, 260-276.

Heegaard, E. 2002. The outer border and central border for species environmental relationships estimated by non-parametric generalised additive models. Ecological Modelling, 157: 2. 131-139.

Hosseini, S.Z., Kappas, M., Zare Chahouki, M.A., Gerold, G., Erasmi, S. and 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.

Jafarian, Z. and Kargar, M., 2017. Comparison of Random Forest (RF) and Boosting Regression Tree (BRT) for prediction of dominant plant species presence in Polour Rangelands, Mazandaran Province, Iranian Journal of Applied Ecology, 6 (1): 41-55.

Kargar-Chigani, H., Javadi, S.A., Zahedi-Amiri, G., Khajeddin, S. J. and Jafari, M. 2017. Vegetation composition differentiation and species-environment relationships in the northern part of Isfahan Province, Iran. Journal of Arid Land, 9**:** 161-175.

Latif, Q. S., Saab, V.A., Dudley, J. G. and Hollenbeck, J.P., 2013. Ensemble modeling to predict habitat suitability for a large-scale disturbance specialist. Ecology and Evolution, 3(13): 4348-4364.

Leathwick, J., Rowe, D., Richardson, J., Elith, J. and Hastie, T., 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshwater Biology, 50: 2034-2052.

Marmion, M., Parvainien, M., Luoto, M., Heikkenen, R.K. and Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modeling. Diversity and Distributions, 15: 59-69.

McCune, B. and Keon, D. 2002. Equations for potential annual direct incident radiation and heat load index. Journal of Vegetation Science, 13: 603-606.

Mi, C., Huettmann, F., Guo, Y., Han, X. and Wen, L., 2017. Why choose Random Forest to predict rare species distribution with few samples in large under sampled areas? Three Asian crane species models provide supporting evidence. PeerJ 5:e2849; DOI 10.7717/peerj.2849.

Mousaei Sanjerehei, M. and Rundel, P. W., 2017. The impact of climate change on habitat suitability for *Artemisia sieberi* and *Artemisia aucheri* (Asteraceae) -a Modeling Approach, Polish Journal of Ecology, 65(1): 97-109.

Pahlavan- Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B. and Bogaert, P. 2014. Updating soil survey maps using random forest and conditioned latin hypercube sampling in the loess derived soils of northern. Iran. Geoderma, 232-234: 97-106.

Pakgohar, A. 2016. Performance comparison of Logistic Regression and Classification Regression tree models for binary dependent variable. Statistical Sciences Extension, 1(2): 7-14.

Piri Sahragard, H. and Zare Chahouki, M.A. 2015. An evaluation of predictive habitat models performance of plant species in Hoze soltan rangelands of Qom province. Ecological Modelling, 309-310: 64-71.

Piri Sahragard, H. and Zare Chahouki M.A. 2016. Comparison of logistic regression and machine learning techniques in prediction of habitat distribution of plant species. Range management and Agroforestry, 37 (1): 21-26.

Piri Sahragard, H., Keshtegar, B., Zare Chahouki, M.A. and Yaseen, Z.M. 2019. Modeling spatial distribution of plant species using autoregressive logistic regression method-based conjugate search direction. Plant Ecology, 220 (2): 267–278.

Poulos, H.M., Chernoff, B. and Fuller, P.L. 2012. Mapping the potential distribution of the invasive red shiner, Cyprinella lutrensis (Teleostei: Cyprinidae) across waterways of the conterminous United States. Aquat Invasions, 7(3): 377-385.

Ray, N. and Burgman, M.A., 2006. Subjective uncertainties in habitat suitability maps. Ecological modelling, 195(3-4), pp.172-186.

Regmi, N.R., J.R., Giardino and Vitek, J. D. 2010. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology, PP.1285-1293.

Sanjerehei, M.M., 2012. Soil-vegetation relationships in arid rangelands (case study: Nodushan rangelands of Yazd, Iran). World Academy of Science, Engineering and Technology, 6(7): 999-1004.

Sor, R., Park, Y.S., Boets, P., Peter, L.M., Goethals, P.L.M. and Lek, S. 2017. Effects of species prevalence on the performance of predictive models. Ecological Modelling, 354:11-19.

Stohlgren, T.J., Ma, P., Kumar, S., Rocca, M., Morisette, J.T., Jarnevich, C.S. and Benson, N. 2010. Ensemble habitat mapping of invasive plant species. Risk Analysis, 30: 224-235.

Vignoli, L., Bologna, M., Bombi, P. and Salvi, D. 2009. Modelling Bedriaga's rock lizard distribution in Sardinia: an ensemble approach. Amphibia-Reptilia, 30(3): 413-424.

Wilks, D.S., 2011. Statistical methods in the atmospheric sciences. New York: Academic press.

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

Zare Chahouki, M.A. and Piri Sahragard, H., 2016. Maxent modelling for distribution of plant species habitats of rangelands (Iran). Polish Journal of Ecology, 64 (4): 453-467.

رویکرد مدل‌سازی اجماعی برای پیش‌بینی پراکنش بالقوه گونه *Artemisia sieberi* در مراتع بیابانی استان یزد، مرکز ایران

محمد علی زارع چاهوکی[الف]، پیمان کرمی[ب]، حسین پیری صحراگرد[ج*]

[الف] استاد گروه احیا مناطق خشک و کوهستانی، دانشکده منابع طبیعی دانشگاه تهران، کرج، ایران

[ب] محقق، گروه علوم محیطی، دانشکده منابع طبیعی و علوم محیطی ملایر، دانشگاه ملایر، ملایر، ایران

[ج] استادیار گروه مرتع و آبخیزداری، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران

*(نگارنده مسئول)، پست الکترونیک: hopiry@uoz.ac.ir

چکیده. پژوهش حاضر با هدف مقایسه دقّت پیش‌بینی برخی از روش‌های مدل‌سازی انفرادی در مقابل رویکرد مدل‌سازی گروهی در برآورد پراکنش بالقوه مکانی و شناسایی نیازهای بوم‌شناختی گونه *Artemisia sieberi* در مراتع بیابانی استان یزد در مرکز ایران انجام شد برای این منظور، داده‌های حضور گونه به روش نمونه‌برداری تصادفی- سیستماتیک در سال ۱۳۹۸ جمع‌آوری شد. بعد از آماده‌سازی نقشه‌های مربوط به متغیرهای محیطی با استفاده از سیستم اطلاعات جغرافیایی و زمین آمار، مدل‌سازی انفرادی پراکنش گونه گیاهی با استفاده از روش‌های RF، CART و GAM انجام شد. عملکرد پیش‌بینی مدل‌های انفرادی با استفاده از آماره‌های AUC و RMSE بررسی شد. علاوه بر این، مدل اجماعی بر اساس براساس میانگین وزنی AUC ایجاد شد. حد آستانه مناسب برای تبدیل نقشه‌های پیوسته به نقشه‌های مطلوبیت رویشگاه، بر اساس شاخص TSS محاسبه شد. مقایسه عملکرد مدل‌های انفرادی مطلوبیت رویشگاه بیانگر آن است که مدل RF در مقایسه با دو مدل دیگر دارای دقّت بیشتری در پیش‌بینی پراکنش گونه مورد مطالعه است (AUC= 0.971 و RMSE= 0.256). بررسی معیارهای توافق نقشه‌های حضور و عدم‌حضور نیز بیانگر برتری این مدل نسبت به دو مدل دیگر بود. مقایسه کلی نتایج حاصل از سه مدل مورد استفاده در مقابل مدل اجماعی نیز بیانگر کارایی بالای این مدل در مقایسه با مدل‌های انفرادی مورد استفاده بود. بر اساس نتایج مدل اجماعی، ۴۵/۳۸ درصد از محدوده مورد مطالعه دارای مطلوبیت بالایی برای استقرار گونه *A. sieberi* است. تحلیل اهمیت متغیرها در مدل جنگل تصادفی(RF) نشان داد که متغیرهای ارتفاع (۴۲٪)، Clay1 (۴۰/۰۲٪) و pH1 (۳۸/۹۷٪) بیشترین تأثیر را بر حضور گونه مورد مطالعه دارند. به‌طور کلی، مدل‌سازی اجماعی می‌تواند با تلفیق صحیح نتایج حاصل از الگوریتم‌های مختلف مدل‌سازی انفرادی، باعث کاهش عدم قطعیت و ارائه نتایج قابل اعتمادتر شود.

کلمات کلیدی: مدل‌سازی پراکنش گونه، میانگین وزنیAUC، مدل اجماعی، خصوصیات خاک، مراتع بیابانی ایران